# State of the Art on Technology and Practices for Improving the Energy Efficiency of Data Storage

Marcos Dias de Assunção[a], Laurent Lefèvre[b]

[a]*IBM Research Brazil*
*Rua Tutóia, 1157*
*04007-900 - São Paulo, SP, Brazil*
`marcosda@br.ibm.com`
[b]*ENS de Lyon - INRIA - LIP*
*46 allée d'Italie 69364 Lyon Cedex 07 - France*
`laurent.lefevre@inria.fr`

## Abstract

Information is at the core of any business, but storing and making available all the information required to run today's businesses have become real challenges. While large enterprises currently face difficulties in providing sufficient power and cooling capacity for their data centres, midsize companies are challenged with finding enough floor space for their storage systems. Data storage being responsible for a large part of the energy consumed by data centres, it is essential to make storage systems more energy efficient and to choose solutions appropriately when deploying infrastructure. This chapter presents the state of the art on technologies and best practices to improve the energy efficiency of data storage infrastructure of enterprises and data centres. It describes techniques available for individual storage components – such as hard disks and tapes – and for composite storage solutions – such as those based on disk arrays and storage area networks.

*Keywords:* data storage, energy efficiency, best practices, storage taxonomy

## 1. Introduction

Information is at the core of any business, but storing and making available all the information required to run today's businesses have become real challenges. With the storage needs of organisations expected to grow by a factor of 44 between 2010 and 2020 [1], efficiency has never been so popular. The constant fall in the price per GB of storage led to a scenario where it is simpler and less costly to add extra capacity than to look for alternatives to avoid data duplicates and minimise other inefficiencies.

As the cost of powering and cooling storage resources becomes an issue, inefficiencies are no longer accepted. Studies show that large enterprises are currently faced with the difficult task of providing sufficient power and cooling capacity, while midsize companies are challenged with finding enough floor space for their storage systems. As data storage accounts for a large part of the energy consumed by data centres, it is crucial to make storage systems more energy efficient and to choose the appropriate solutions when deploying storage infrastructure.

This chapter discusses technologies that improve the energy efficiency of data storage solutions. Moreover, it describes best practices that – in addition to the use of the discussed technologies – can improve the energy efficiency of storage infrastructure in enterprises and data centres.

## 2. Taxonomy of Data Storage Solutions

With the goal of providing reproducible and standardised assessment of the energy efficiency of storage solutions, the Storage Networking Industry Association (SNIA) has created the SNIA Emerald Power Efficiency Measurement specification [2]. As part of the specification, SNIA has proposed a taxonomy for storage products to ease the evaluation of energy efficiency of different storage equipments and allow comparisons among devices produced by different manufacturers. This taxonomy, which has been adapted by the ENERGY STAR program [3], classifies storage products in terms of operational profile and features, and has the following main categories:

- **Online**: defines features and functionalities for online, random-access storage products. The products in this category must have a Maximum Time To First Data (MaxTTFD) smaller than 80 ms.

- **Near Online**: category that defines features and functionalities for near-online, random-access storage products, which may employ Massive Arrays of Idle Disks (MAIDs) or Fixed Content Aware Storage (FCAS) architectures and can have a MaxTTFD greater than 80 ms.

- **Removable Media Library**: defines characteristics of storage products that rely on manual or automated media loaders, such as tape archive systems [4]. Data access is sequential and the MaxTTFD is between 80 ms and 5 minutes.

- **Virtual Media Library**: category for sequential-access storage products that rely on optical or disk-based storage media such as optical jukeboxes [5]. The MaxTTFD must be bellow 80 ms.

- **Adjunct Product**: storage appliances that support a Storage Area Network (SAN) and provide advanced management capabilities. The user accessible data is prohibited and the MaxTTFD must be smaller than 80 ms.

- **Interconnect Element**: defines features and functionalities of managed inter-connect elements within a SAN with a MaxTTFD under 80 ms.

Each product category defines a set of attributes that are common to products within the category as well as ranges to certain attributes (*e.g.* MaxTTFD between 80ms and 5 minutes). Each category is further divided into smaller sub-categories that take into account several factors such as connectivity, no single point of failure and service-ability. The measurement specification [6] also defines metrics and a methodology to evaluate the power efficiency of a certain range of products within some of the proposed categories. The metrics and benchmarks are discussed in more detail in Section 6.

Storage solutions such as disk arrays are composed of drives that provide the raw storage capability and additional components that allow the interface to the raw storage and improve the reliability of the storage solution. In addition, a tape library often comprises several tape loaders. Hereafter we adopt SNIAs terminology and refer to the individual components that compose the raw-storage capability of storage solutions as **storage devices** (*e.g.* tape loaders, hard disk drives and solid state-drives), whereas a composite storage solution such as a network attached product is termed as a **storage element**. When discussing schemes for improving the energy efficiency of storage solutions, these are mainly the two levels at which most techniques apply. Therefore, we first describe energy efficient concepts for individual devices, and then analyse how these techniques are currently used to improve the energy efficiency of storage solutions or elements.

The use of different tiers of storage depends on the requirements of the applications that will run on the infrastructure. For example, applications that rely on services delivered to customers via the Web require the use of web servers and benefit from fast response time. It is not uncommon to use technologies that rely on high-performance disk drives or solid-state drives. The levels of Redundant Array of Independent Disks (RAID) and replication depend on how critical the services are and require a careful analysis when designing the storage infrastructure. Organisations that are required to store data for long periods due to legal or business requirements, such as government offices, can benefit from carrying out backups on tape. A detailed analysis of the requirements of applications and the features provided by storage solutions at different tiers are crucial for planning the deployment of storage solutions on data centres.

## 3. Device-level Solutions

This subsection describes energy efficient solutions that operate at the device level. We also describe tape based systems as a device-level approach, though they are often solution aggregates as tapes appear as an alternative to technology that relies either on hard-disk drives of solid-state drives.

### *3.1. Tape Based Systems*

Tapes are often mentioned as one of the most cost-efficient types of media for long-term data storage. Although in recent years tapes have been viewed as outdated, analyses have showed that [7, 8]:

- Under given long-term storage scenarios, such as backup and archival in mid-sized data centres, hard disk drives can be on average 23 times more expensive than tape solutions and cost 290 times more than tapes to power and cool [7]. Although the costs of disk subsystems have decreased and their capacity has increased, especially for Serial Advanced Technology Attachment (SATA) drives, tape continues to be the most economical solution for long-term storage requirements [7, 8].

- Data consolidation using tape-based archival systems can considerably decrease the operational cost of storage centres [8]. Tape libraries with large storage capacity can replace islands of data via consolidation of backup operations, hence reducing costs with infrastructure and possibly increasing its energy efficiency.

With archival life of thirty years and large storage capacity, tapes make always an appealing solution for data centres with large long-term backup and archival requirements. Hence, for an environment with multiple tiers of storage, tape-based systems are still the most power-efficient solutions when considering long-term archival and low retrieval rate of archived files. There are disk library solutions that attempt to minimise the impact of the energy consumption of disk drives by using techniques such disk spin-down – discussed in the next section. For example, EMC's Disk Library 5200 uses 2GB SATA drives that can be put into idle mode when the data they store is not accessed. Although disk libraries tend to be more energy consuming than tape systems, they commonly present better performance when considering throughput and data access time. The next sections discuss some techniques for improving their energy efficiency of hard disk drives (*e.g.* disk spin-down and variable disk speed).

### *3.2. Hard Disk Drives*

Hard Disk Drivess (HDDs) have long been the preferred media for non-volatile data storage that offers fast write and retrieval times. A hard disk comprises one or more rotating rigid platters on a motor-driven spindle placed within a metal case, also know as disk enclosure. Data is recorded/read by heads that float above the platters. An actuator arm is responsible for moving the heads across the platters, allowing each head to access almost the entire surface of the platter as it spins.

The presence of moving parts such as motors and actuator arms are often mentioned as responsible for most of the power consumed by hard drives. In order to improve the data throughput, manufacturers increase the speed at which platters spin, which in turn increases the power consumed by hard disks. Platters spinning at speeds of 15K-rpm are common for current high-throughput hard disk drives.

Several techniques have been proposed to improve the energy efficiency of hard disks. There are schemes to store data in certain regions of the platters possibly reducing seek time and requiring less mechanical effort by the actuators when retrieving the required data. Controlling the speed at which platters spin is also a technique that can save energy. Attempts have also been made to reduce the power consumption during idle periods. Techniques in this context include spinning platters down and parking the heads at the secure zone after a factory-set period of inactivity; approach commonly termed as disk spin-down [9]. Energy efficient drives proposed by manufacturers generally spin at lower speeds when compared with their high-performance counterparts. Some 5.4K-rpm SATA drives are argued to reduce power consumption by up to 30% with less than 10% degradation in random I/O performance over traditional 7.2K-rpm SATA drives[1]. Moreover, instead of stopping platters completely, some manufactures offer the feature of spinning the platters at variable speed, adjusted according to the workload.

In response to the Energy Star Program that requires PC manufacturers to equip their PCs with an automatic power-saving mode during non-operation, HDD manufacturers have established and implemented idle and standby states for HDDs. During these states, techniques such as disk spin-down and variable speed, described above, are used. Figure 1 illustrates the power saving modes of PCs, where hard disks often implement idle and standby states. If a PC reaches sleep state, then virtually all HDD operations cease.

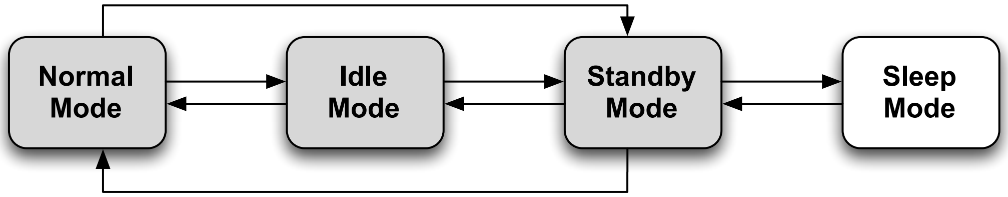

Figure 1: Power saving modes (shaded boxes are states generally reached by HDDs).

The implementation of idle states varies across solutions, where the number of disabled components typically increases as a drive reaches certain idleness thresholds. Seagate's PowerChoice[2] technology [10], for example, implements three distinct idle states. The specific energy-saving steps implemented by each PowerChoice state for a 7.2K-rpm drive are as follows:

- **Idle_A**: Disables most of the servo system, reduces processor and channel power consumption and platters continue rotating at 7.2k-rpm.

- **Idle_B**: Disables most of the servo system, reduces processor and channel power consumption, heads are parked, and platters continue rotating at 7.2k-rpm.

[1] Technical specification of Dell PowerVault MD1000 Direct Attached Storage Arrays

[2] PowerChoice is a trademark of Seagate.

- **Idle_C**: Disables most of the servo system, reduces processor and channel power consumption, heads are parked, and platters have their speed reduced.

- **Standby_Z**: Heads are parked, driver motor is spun down, and drive responds only to non-media access commands.

The intermediate idle states have recovery times that are generally shorter than restarting a disk that has been spun-down completely. Comparing two high-end HDDs, Table 1 shows that the consumption at standby mode is generally close to 80% less than the idle consumption. It is argued that these approaches can lead to substantial savings on RAID systems [11] and MAIDs [12].

Table 1: Power saving using disk spin-down in standby modes.

| **Drive Description** | **Power Consumption** | | | **Power Saving (%)*** |
|---|---|---|---|---|
| | **Read/Write** | **Idle** | **Standby** | |
| Western Digital RE4 1TB 7.2k-rpm | 7.9 | 5.9 | 0.7 | 88.13 |
| Seagate Constellation ES 1TB 7.2k-rpm | 10.8/Read<br>9.6/Write | 6.0 | 1.3 | 78.33 |

* Savings comparing the standby and idle power consumptions.

Although spinning disks down can compromise performance, manufacturers explore additional techniques such as larger caches and read/write command queuing to minimise its impact. Furthermore, schemes have been proposed at the operating system and application levels to increase the length of periods of disk inactivity and hence benefit from techniques such as spin-down and variable spinning speed. Some of these approaches consist of rescheduling data-access requests by modifying the application code or data layouts. There are also less intrusive techniques that provide compiler customisations that re-schedule the data access requests during compilation without the need of modifying application source code. Although these techniques can reduce power consumption, it is a common belief that constant on-off cycles can reduce the life time of HDDs.

As motors and actuators are responsible for most power consumed by hard disk drives, a tendency for making drives more energy efficient is to use

Small Form Factors (SFFs), which are 70 percent smaller than 3.5-inch enclosures. A chassis designed with enough volume for 16 3.5-inch drives might be redesigned to hold up to 48 2.5-inch hard-disk drives without increasing the overall volume [13].

Packing high-performance hard drives into 2.5-inch enclosures reduces their power consumption by making motors and actuators smaller and allowing drives to emit less heat. Manufacturers claim that for Tier-1 2.5-inch HDDs, their Input/Output Operations Per Second per Watt (IOPS/W) can be up to 2.5x better than comparable 3.5-inch Tier-1 drives [14]. In addition, less power is required for cooling due to smaller heat output and reduced floor space requirements.

Table 2: Power consumption of two of Seagate's high throughput HDDs.

| **Specifications** | **Cheetah 15K.7 300GB*** | **Savvio 15K.2 146GB*** | **Difference** |
|---|---|---|---|
| Form Factor | 3.5" | 2.5" | – |
| Capacity | 300GB | 146GB | – |
| Interface | SAS 6Gb/s | SAS 6Gb/s | – |
| Spindle Speed (rpm) | 15K | 15K | – |
| Power Idle (W) | 8.74 | 4.1 | 53% less |
| Power Active (W) | 12.92 | 6.95 | 46.2% less |

* Obtained from data sheets available at the manufacturer's website.

Table 2 shows the approximate power consumed by two models of high performance hard disk drives produced by Seagate. It is evident that the smaller form factor takes substantially less power. When active, it consumes approximately 46% less power than its 3.5-inch counterpart, whereas this difference can reach 53% when the disk is idle. If one compares the power consumed by one disk drive alone, it might not look substantial. However, when we multiply the consumption by a large number of drives and hours, the difference starts to become considerable. Considering the cost to power 24 drives over a year, taking the active power consumption as an example and a price of 0.11 Euros per kWh, 24 3.5-inch drives would cost approximately 298 Euros to power whereas 24 2.5-inch HDDs would cost 160 Euros per year. The savings are around 138 Euros per year with only 24 drives. In data centres with storage systems with hundreds or thousands of disks, the savings can easily reach figures in the thousands of Euros.

### *3.3. Solid State Drives*

Solid State Drives (SSDs) are equipped, among other components, with flash memory packages and a controller responsible for various tasks (Figure 2) [15]. Unlike HDDs, SSDs have no mechanical parts such as motors and moving heads. Currently available SSDs rely on NAND-based flash memory, and employ two types of memory cells according to the number of bits a cell can store. Single-Level Cell (SLC) flash can store one bit per cell and Multi-Level Cell (MLC) memories can often store 2 or 4 bits per cell. Most affordable flash memories and SSDs rely on MLC while high-end devices are often based on SLC. NAND-based memory cells have a limited number of writes, generally between 10,000 and 100,000, which at first makes one question the reliability of SSDs.

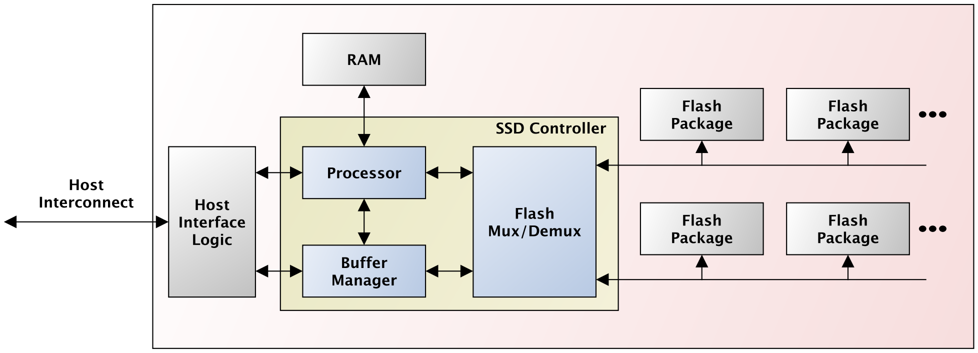

Figure 2: SSD main components [15].

Mean Time Between Failure (MTBF) of SSDs is generally improved by packaging additional memory cells in the SSDs, transparently replacing defective cells, and applying "wear levelling" algorithms that insure uniform wear of the flash memory. In addition, hard-disk drive failures are generally catastrophic, leading to complete drive malfunction or serious performance degradation, whereas SSDs can continue to operate normally even if cells fail. The defective blocks can be easily isolated and no longer used by the SSD controller.

The memory in SSD is organised in pages whose size varies from 512 to 4096 bytes, and all read and write operations take place at page granularity.

Pages are combined in blocks of 128, 256 or 512KB. Due to design issues and the limited number of writes allowed by memory cells, a write operation requires that cells be erased before the new content is written, and erase operations are block-wise. Therefore, a page can be modified (*i.e.* written) only after the whole block to which it belongs is erased, which makes write operations significantly more costly than reads in terms of performance and energy consumption [16]. Manufacturers such as Intel aim to improve the write performance via several techniques such as Native Command Queuing (NCQ). Intel's recent 510 series of SSDs [17] present read and write latencies of $65\mu$s and $80\mu$s respectively, which is much lower than the latency of 2.5-inch Serial attached SCSI (SAS) 15K-rpm HDDs . In addition, the implementation of **TRIM**[3] can improve the write performance by allowing the operating system to notify the SSD drive about data blocks that have been released due to the deletion of a file, for example. This allows the SSD controller to make optimisations of erase commands that further improve the performance of write operations. The erase operations can be executed in background before further requests to write the page contents. DeVetter and Buchholz [18] summarise some of the advantages of SSDs over HDDs for mobile environments (Table 3). Although the requirements of enterprises differ from those of mobile users, some characteristics of SSDs are also advantageous to data centres, such as their improved performance, reliability and reduced power consumption.

In spite of its write limitations, SSDs have considerably better read-performance than hard-disk drives [16]. Customer applications with mostly random data access requirements see the greatest benefit from SSDs over hard disk drives [19]. Due to the lack of mechanical parts, SSDs create less heat and can be packed into smaller enclosures, thus decreasing the floor space and cooling requirements. Table 4 presents a simple comparison between a Seagate's Pulsar enterprise SSD and a high performance SAS 15k-rpm HDD. The SSD consumes approximately 87% less power than the 15k-rpm HDD in active mode, and around 82% less in idle mode. In practice, however, the energy savings will depend on how the storage solutions use the SSDs

[3] TRIM is a command that allows an operating system to inform an SSD which blocks of data are no longer in use and can be erased internally. As various file systems often update structures for handling information of free blocks without actually updating the media, TRIM enables the SSD to perform garbage collection by erasing blocks before future write operations take place.

Table 3: Hard-disk drives versus solid-state drives.

| **Hard-Disk Drives** | **Solid-State Drives** |
|---|---|
| More fragile due to moving parts such as rotating platters and mechanical arms. | Stronger because there are no moving parts. |
| Requires more power and emits more heat. | Equipments can run cooler and more efficiently. |
| Decreased performance as file fragmentation increases. | Consistent performance because frag-mentation is not an issue. |
| Greater risk of data loss and hard disk failure when transported. | More resistant to bumps and drops. |
| Slower responsiveness and performance due to time required by disk spin up and mechanical movements. | Faster responsiveness and performance due to no drive spin up time and no mechanical arm movement. |

and HDDs, and the characteristics of the workload applied to the storage equipments.

When considering the cost of MB per dollar, SSDs frequently lag behind hard disk drives. The scenario is however different when considering the cost per Input/Output Operations Per Second (IOPS). Table 5 presents a comparison between the IOPS cost of a few IBM enterprise HDDs and SSDs [20]. Furthermore, as the price per GB of flash memory declines at a faster rate than the increase in capacity of hard-drives, SSDs become a very complementary technology to balance performance, availability, capacity and energy across different application tiers [21]. Although purely-SSD based storage solutions are available, their use is often recommended as a means to complement the performance of systems based on other storage medias. Later sections discuss advantages of disk-arrays with mixed storage (*i.e.* mixing hard-disk drives and SSDs).

### *3.4. Hybrid Hard Drives*

Some hard-disk drives have been equipped with large buffers made of non-volatile flash memories that aim to minimise data writes or reads on the platters. These disks are usually called Hybrid Hard Drives (HHDs). Several algorithms have been proposed over the years for utilising the buffer offered

Table 4: Comparison of a high throughput HDD and an SSD counterpart.

| Specifications | Savvio 15K.2 73GB* | Pulsar SSD 50GB* | Difference |
|---|---|---|---|
| Form Factor | 2.5” | 2.5” | – |
| Capacity | 73GB | 50GB | – |
| Interface | SAS 3Gb/s<br>SAS 6Gb/s | SATA 3Gb/s | – |
| Spindle Speed (rpm) | 15K | – | – |
| NAND Flash Type | – | SLC | – |
| Power Idle (W) | 3.7 | 0.65 | 82.4% less |
| Power Active (W) | 6.18 | 0.8 | 87% less |

* Obtained from data sheets available at the manufacturer’s website.

Table 5: IOPS and cost for HDDs and SSDs.

| Metrics | HDD (3.5” 15K) | HDD (2.5” 15K) | SLC SSD | MLC SSD |
|---|---|---|---|---|
| Write IOPS | 300 | 250 | 1600 | 3000 |
| Read IOPS | 390 | 300 | 4000 | 20000 |
| Cost per IOPS | 0.52<br>(146 GB) | 0.83<br>(146 GB) | 0.09<br>(50 GB) | 0.04<br>(50 GB) |

by this type of drive [22]. By using this large buffer, the platters of the hard drive are at rest almost at all times, instead of constantly spinning as in HDDs. This additional flash memory can minimise the power consumed by storage solutions by reducing the power consumed by the motors and mechanical arms. These drives can present potentially lower power requirements when compared to hard-disk drives, but the offerings by manufacturers are, as of writing, very limited. The Seagate Momentus XT hybrid drive is an example of this technology.[4]

## 4. Solutions for Storage Elements

As discussed earlier, we adopt SNIA's terminology to discuss and assess storage solutions. In earlier sections, we analysed the existing solutions for improving the energy efficiency of individual storage components (*i.e.* storage devices) such as hard-disk drives and solid-state drives. The next sections assess how these device-level techniques are used and combined to improve the energy efficiency of composite storage solutions such as disk arrays, direct attached storage and networked storage.

When choosing networked storage solutions and designing storage area networks, it is essential to know the application that will use the storage resources. An application that creates several small blocks of data at random might require SAS or Fibre Channel (FC), Fibre Channel over Ethernet (FCoE) [23] or iSCSI connectivity. Applications that create large data blocks sequentially, such as video servers, streaming media and high performance computing, might benefit from SATA and FC connectivity. Table 6 presents a list of different types of applications and the recommended drive type and network connectivity required to maximise performance [24].

A storage element or storage solution deployed on a data centre generally comprises several components, including disk arrays, controllers, network switches, hard-disk drives, solid-state drives, power supplies, fans and Power Distribution Units (PDUs). Moreover, a storage solution can be composed of software systems used to, among other features, manage different storage tiers and backup. Disks tend to be the components that consume most power in a storage solution, and hence we start our discussion on disk arrays and techniques used to improve their energy efficiency.

[4] Technical specifications of the Seagate Momentus XT series of HHDs, available at the manufacturers' website at: http://www.seagate.com.

Table 6: Applications' performance, drive and connectivity requirements [24].

| Application | Performance Requirement | Best Drive Type / Best Connectivity |
|---|---|---|
| Email – Microsoft Exchange | IOPS intensive | FC or SAS disks, SAS or FC connectivity |
| File serving | MB/s intensive | SAS or SATA disks and Ethernet option for iSCSI, CIFS or NFS* |
| Sensor Data Collection | MB/s intensive | SAS or FC (SATA option) disks, SAS or FC connectivity |
| Database – OLTP | IOPS intensive | SAS or FC disks, SAS or FC connectivity |
| Data warehouse | MB/s and IOPS intensive | SAS or FC disks, SAS, InfiniBand or FC connectivity |
| D2D** backup – VTL+ | MB/s | MAID$^t$, FC connectivity |
| Data analysis | MB/s or IOPS | FC or InfiniBand |
| Active archives | MB/s | MAID, FC |

* NFS and CIFS are the primary file systems used in network-attached storage.

** Disk-to-Disk.

+ Virtual Tape Library.

*t* Massive Array of Idle Disks, explained in later sections.

### *4.1. Disk Arrays and MAIDs*

A disk array is a storage system that contains multiple disk drives. It can be Just a Bunch of Disks (JBODs), in which case the controller is an external module that interfaces with the array. Several of current storage arrays use Switched Bunch of Diskss (SBODs) or Extended Bunch of Diskss (EBODs), which give better response times. Hence, an array solution generally comprises controllers, which make arrays differ from disk enclosures by having cache memory and advanced features such as RAID. Common components of a disk array include:

- **Array controllers**: devices that manage the physical disk drives and present them to the servers as logical units. Usually a controller contains additional disk cache and implements hardware level RAID.

- **Cache memories**: as described above, an array can contain additional cache memories for improving the performance of read and write operations.

- **Disk enclosures**: an array contains a number of disk drives, such as HDDs and SSDs. It can contain a mix of different drive types. The size of the disk enclosures depends on the used form factor (*e.g.* 2.5-inch or 3.5-inch hard-disk drives).

- **Power supplies**: a disk array can contain multiple power supplies in order to increase its reliability in case one of the supplies fails.

Although disk arrays can be directly attached to servers through a series of interfaces, they are often part of a more sophisticated storage system such as network attached storage or storage area network; described later.

As mentioned earlier, in order to improve their reliability and fault tolerance, disk arrays are commonly equipped with multiple power supplies. It is important that these supplies be power efficient and have a minimum power factor. Furthermore, the disk drives are the most power consuming elements in the array. Thus, it is crucial to choose drives that are efficient and provide features that can minimise power consumption under the expected workload. For example, data archives can be more energy efficient by using disks with large storage capacity, while this is often not the case of high I/O applications. The RAID level also affects the energy efficiency of a storage system, since drives used for protection are not used to retrieve data, but consume energy like the other drives. As an example, Table 7 shows different RAID levels and their storage efficiency [25].

Table 7: RAID types and efficiency [25].

| **RAID Level** | **Storage Efficiency*** |
|---|---|
| RAID 1 | 50% |
| RAID 5 (3+1) | 75% |
| RAID 6 (6+2) | 75% |
| RAID 5 (7+1) | 87.5% |
| RAID 6 (14+2) | 87.5% |

* Storage efficiency here means the percentage of the disks capacity that is made available for actual data storage.

As discussed earlier, it is important that the power supplies of storage arrays be power efficient. Properly sized power supplies benefit systems in both idle and active modes. Furthermore, it is relevant to work closely with the provider of storage equipments to choose solutions suitable to the expected workload and that have been designed with energy efficiency in mind. Disk arrays that utilise (i) disks with variable speeds, (ii) disks with spin-down features and (iii) mixed storage, can help minimise the energy consumed by the storage subsystem and reduce costs.

The efficiency of several power saving features often depend on the workload; hence, the importance of working closely with providers of data storage solutions. For example, as described in the next section, current technology on MAIDs can lead to savings of up to 70% [26]. The energy savings can generally be substantial when MAID technology is applied to near-line storage where the storage resources can remain idle for large periods of time.

### *4.1.1. Options for improvement of energy and cost efficiency*

MAID is a technology that uses a combination of cache memory and idle disks to service requests, only spinning up disks as required [12]. Stopping spindle rotation on less frequently accessed disk drives can reduce power consumption (see Figure 3). Manufacturers such as Fujitsu allow customers to specify schedules with periods during which the drives should be spun down (or powered off) according to the workload or backup policies. Fujitsu also employs a technique in which drives are not spun up at the same time to minimise peak usage scenarios. These techniques come at hand for solutions targeted at back-up and archival as the drives can be spun down when the backup operations are not taking place.

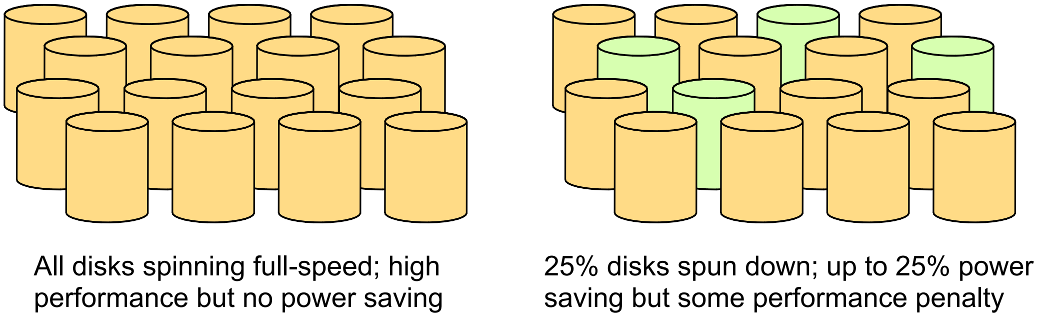
All disks spinning full-speed; high performance but no power saving
25% disks spun down; up to 25% power saving but some performance penalty

Figure 3: Pictorial view of MAIDs [27].

How much power MAID features can save depends on the application that uses the disks and how often the application accesses the disks. As discussed earlier, EMC reports savings of up to 30% in power usage in a fully loaded CLARiiON CX4-960 environment if more than 50% of the data is infrequently accessed [28]. The criteria used to decide when drives are spun down (or put into standby mode) or spun up, also have an impact on energy savings as well as in performance. As an example of standby criteria, in EMC's FLARE system [28], hard disk drives of a RAID group enter standby mode when both storage and processors report that the drives have not been used for 30 minutes. Similar threshold is used by Fujitsu's ECO mode, where by default the ECO mode starts after 30 minutes of no disk access. ECO mode also allows the administrator to specify operation periods during which the motors of hard disk drives should not stop.

When initially conceived, MAID techniques enabled HDDs to be either on or off, which could incur considerable application performance penalties if data on a spun-down drive was required and the disk had to be spun back up. MAID techniques are said to have reached their second generation, where they implement Intelligent Power Management (IPM) with different power saving modes and performance [29]. An example of MAID 2.0 is Nexsans Assureon, SATABoy and SATABeast solutions that implement intelligent power management with its AutoMAID[5] technology. AutoMAID has multiple power saving modes that align power consumption to different quality of

[5] AutoMAID is a trademark of Nexsan Corporation.

service needs. The user can configure the trade-off between response times and power savings. Nexsan claims that by enforcing the appropriate policies to determine the required level of access speed and MAID levels, a reduction of up to 70% in power requirements can be achieved [30]. The typical MAID-level configuration settings of AutoMAID are as follows:

- **Level 0**:
  - Normal operation, drives at 7.2K-rpm, heads loaded.
- **Level 1**:
  - Hard-disk drive heads are unloaded.
  - Sub-second recovery time.
- **Level 2**:
  - Hard-disk drive heads are unloaded.
  - Platters slow to 4k-rpms.
  - 15-second recovery time.
- **Level 3**:
  - Hard-disk drives stop spinning (sleep mode; powered on).
  - 30 to 45 second recovery time.

Other power conservation techniques for disk arrays have been proposed, such as the Popular Data Concentration (PDC) [31] and file allocation mechanisms [9]. The rationale is to perform consolidation by migrating frequently accessed data to a subset of the disks. By skewing the load towards fewer disks, others can be transitioned to low-power consumption modes. It was found that it is possible to conserve a substantial amount of energy during periods of light load on the servers as long as two-speed (or variable speed) disks are used.

Another important issue refers to scalability. When choosing storage solutions, a recommended practice is to employ systems that allow for further storage bays to be added as the storage demand grows [25]. Hence, it is important to design the system to the intended workload and then scale using small storage bays to reduce eventual inefficiencies.

### *4.2. Direct Attached Storage*

Direct-Attached Storage (DAS) consists of a data storage system attached to a host without a network in between. It typically comprises drive enclosures such as disk arrays connected to a host bus adapter. The main protocols for interconnecting DAS and hosts are SATA, eSATA, SCSI and SAS.

DAS solutions benefit from the energy-efficiency improvements achieved by the equipments described in the previous sections, such as hard-disk drives, SSDs and disk arrays. Manufacturers of DAS have been pursuing a few additional solutions that, along with carefully designed data-management policies, can improve the energy efficiency of DAS systems. These solutions include:

- Improvements of power supply units. As DAS solutions usually have multiple power supplies for reliability purposes, it is important to choose supplies whose efficiency is certified (*e.g.* 80PLUS Certified power supplies[6]).

- Use of large capacity hard disk drives for certain applications. For applications that do not demand high-performance storage, it is usually more energy efficient to use drives with larger capacity. Typical SATA disk drives consume up to 50% less power per terabyte of storage than Fibre Channel drives [32].

- Co-existence of mixed drives in the same disk enclosure to enable vertical storage tiering. Existing storage solutions can maintain different types of media and can take advantages of these differences according to the data access patterns. High-capacity, low-power disk drives with medium to high-performance disk drives in tiered storage subsystems, and disk drive spin-down features can reduce power and cooling requirements [33].

- Introduction of small-form-factor enclosures that save floor space in data centres and can decrease the energy footprint by using more power-efficient 2.5-inch HDDs [34, 13]. As discussed beforehand, 2.5-inch disks can generally consume up to 50% less power and 70% less space than 3.5-inch drives.[7]

[6] http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx

[7] Technical specification of Dell PowerVault MD1220 storage solution

- Use of more energy-efficient RAID levels and mechanisms. As demonstrated in Table 7, different RAID levels present different storage efficiencies. When considering data protection some RAID levels, such as RAID 6, present a significant amount of overhead processing. In addition, high performance RAID 6 implementations can provide the same performance as RAID 5 and up to 48% reduction in disk capacity requirements compared with RAID 10 [35].

- Variable and temperature controlled fans designed to deliver optimal performance and energy efficiency. In EMC CLARiiON CX4, the adaptive cooling functionality intelligently monitors airflow and temperature within the storage processor chassis and adjusts blower and fan speeds based on system activity, constantly adapting to changing environmental needs.[8]

### *4.3. Storage Area Networks and Network Attached Storage*

To avoid the creation of information islands, often mentioned as a drawback of DAS systems, SANs attempt to consolidate the data by enabling storage equipments to be accessed by servers via network generally on a per-block manner using protocols such as iSCSI, Fibre Channel Protocol (FCP) and FCoE. The main components or layers of a SAN include [23]:

- **SAN Connectivity or Fabric**: it is the actual network part of a SAN. The connectivity of storage and server components generally uses FC technology. SANs can interconnect the storage equipments together into several network configurations. Some of the components employed at this layer are hubs, switches, gateways and routers.

- **SAN Storage**: it is the layer where the storage equipments, and consequently the data, reside. It contains all the disk drives, tape drives and other storage devices. Storage equipments are attached directly to the network, so that storage can be distributed across the organisation, or be centralised in order to foster consolidation, ease management, and reduce cost.

- **SAN Servers**: server infrastructure is the main reason for using a SAN solution. The server infrastructure can comprise a range of server

[8] Brochure on EMC CLARiiON CX4: the Best Energy Efficiency in Midrange Storage

platforms, such as Unix, Linux and Windows. This layer also includes Host Bus Adapters (HBAs) and the software running on servers, which allows HBAs to communicate with the SAN fabric.

Several applications can benefit from SAN solutions: high-performance applications can use a SAN for storing data and check-pointing; via thin provisioning, some applications can allocate storage from a SAN on demand; database applications that require fast access time to data can benefit from the low-latency block-level data access offered by SANs; backup operations across the enterprise can be centralised; and server virtualisation can make heavy use of a SAN to store virtual machine images, snapshots, and enable virtual machine migration.

As a SAN may not require an IP address, costly operations such as converting data blocks into IP packets can often be avoided. However, iSCSI is sometimes used by SAN solutions with the goal of minimising cost and reusing existing Ethernet technology. As it can transfer SCSI commands over IP networks, iSCSI can facilitate data transfer across Wide Area Network (WAN) and the Internet. A SAN environment differs from network-attached storage solutions in the sense that it generally does not offer tools to expose storage devices to servers as file-level services.

Network Attached Storage (NAS), on the other hand, is a specialised server with its own IP address that is made available to multiple clients or servers over a network. Standard protocols such as iSCSI and Fibre Channel are used to communicate with NAS systems, thus allowing for heterogeneous environments where different operating systems can read and write data on NAS servers. Unlike SANs that use block-level protocols, at the communication level NAS solutions frequently utilise file-level protocols such as Network File System (NFS) and Common Internet File System (CIFS). SAN and NAS can be combined in ways that consolidate networked storage. A NAS gateway can connect to disk arrays or tape systems on a storage area network.

Manufacturers of SAN and NAS solutions often attempt to curb the power consumption of their systems by applying some of the DAS concepts described beforehand, and by reducing the power consumption of the network equipments, such as Fibre Channel and iSCSI switches, and HBAs. Hence, many of the techniques for improving the energy efficiency of storage equipments described above for other solutions are also applicable to SAN and NAS. We list below a few other techniques that can be utilised.

### *4.3.1. Combining Server and Storage Virtualisation*

By combining server virtualisation with storage virtualisation it is possible to create disk pools and virtual volumes whose capacity can be increased on demand according to the applications' needs. Typical storage efficiency of traditional storage arrays is in the 30-40% range. Storage virtualisation can increase the efficiency to 70% or higher according to certain reports [35], which results in less storage requirements and energy savings.

Storage virtualisation technologies can be classified in the following categories [24]:

- **Block-level virtualisation**: this technique consists in creating a storage pool with resources from multiple network devices and making them available as a single central storage resource. This technique, used in many SANs, simplifies the management and reduces cost.

- **Storage tier virtualisation**: this virtualisation technique is generally termed as Hierarchical Storage Management (HSM) and allows data to be migrated automatically between different types of storage without users being aware. Software systems for automated tiering are used for carrying out such data migration activities. This approach reduces cost and power consumption because it allows only data that is frequently accessed to be stored on high-performance storage, while data less frequently accessed can be placed on less expensive and more power efficient equipments that use techniques such as MAID and data de-duplication.

- **Virtualisation across time to create active archives**: this type of storage virtualisation, also known as active archiving, extends the notion of virtualisation and enables online access to data that would be otherwise offline. Tier virtualisation software systems are used to dynamically identify the data that should be archived on disk-to-disk backup or tape libraries or brought back to active storage.

Storage virtualisation is a technology that complements other solutions such as server virtualisation by enabling the quick creation of snapshots and facilitating virtual machine migration. It also allows for thin provisioning where actual storage capacity is allocated to virtual machines when they need to write data rather than allocated in advance.

### *4.3.2. Thin Provisioning*

Thin provisioning, a technology that generally complements storage virtualisation, aims to maximise storage utilisation and eliminate pre-allocated but unused capacity. With thin provisioning, storage space is provisioned when data is written. Reserve capacity is not defined by the maximum storage required by applications; it is generally set to zero. Volumes are expanded online and capacity is added on the fly to accommodate changes without disruption. For example, NetApp's FlexVol technology is a storage virtualisation technology that allows storage managers to virtually allocate capacity to users without physically allocating it. Storage is physically allocated when it is actually used [36]. Fujitsu's ETERNUS works with the notion of threshold alarms, which when triggered allow the system to allocate more physical storage capacity to virtual volumes in order to improve performance. Thin provisioning can lead to energy savings because it reduces the need for over provisioning storage capacity to applications.

### *4.3.3. Horizontal Storage Tiering*

For efficient use of storage infrastructure, it is important to design and enforce sound data management policies that use different tiers of storage according to: how often the data is accessed, whether it is reused and for how long it has to be maintained (for business or regulatory purposes). For deciding on archival and backup policies, Chistofferson illustrates the use of different storage technologies according to the probability of data reuse and time over which the data must be stored [37].

Manufacturers of data storage solutions have proposed software systems that allow for seamless and automatic tiering by moving data to the appropriate tier based on ongoing performance monitoring; for example, EMC2s Fully Automated Storage Tiering, IBMs System Storage Easy Tier, Compellent's Data Progression and SGIs Data Migration Facility.

### *4.3.4. Vertical Storage Tiering*

Techniques for providing storage tiering at the level of arrays and storage elements can help improve performance and reduce power consumption. For example, employing a solution that uses both SSDs and HDDs can improve the application's performance by moving data frequently accessed to SSDs and benefit from the larger storage capacity of HDDs for storing less frequently accessed data. Finding a good mix of different types of drives aiming to reduce the energy footprint of the storage systems is hence possible via vertical tiering.

*4.3.5. Consolidation at the Storage and Fabric Layers*

Consolidation of both data storage and networking equipments can lead to substantial savings in floor space requirements and energy consumption. Some manufacturers argue that by providing multi-protocol network equipments, the network fabric can be consolidated on fewer resources, hence reducing floor space, power consumption and cooling requirements.[9] In addition, the increasing use of blade servers and migration of virtual machines encourage the use of networked storage, which then allows for improvements in storage efficiency by means of consolidation [35].

Storage consolidation is not a recent topic. In fact, SANs have been providing some level of storage consolidation and improved efficiency for several years by permitting the sharing of arrays of disks across multiple servers over a local private network, and avoiding islands of data. Hence, moving DAS to networked storage systems offers a range of benefits, which can increase the energy efficiency. These benefits include [35]:

- **Capacity sharing**: Administrators can improve storage utilisation by pooling storage capacity and allocating it to servers as needed. Hence, it helps reducing the storage islands caused by direct attached storage.

- **Storage provisioning**: Storage can be provisioned in a more granular way. Volumes can be provided at any increment, in contrast to allocating physical capacity or entire disks to a particular server. In addition, volumes can be resized as needed without incurring server downtime.

- **Network boot**: This allows administrators to move not only the servers data to the networked storage, but also the server boot images. Boot volumes can be created and accessed at boot time, without the need for local storage at the server.

- **Improved management**: Storage consolidation removes many of the individual tasks for backup, data recovery and software updates. These tasks can be carried out centrally using only one set of tools.

[9] Brochure on Next Generation IBM Blade Center Virtual Fabric.

Manufacturers of storage equipments have provided various consolidated solutions generally under the banner of unified storage. Traditionally, enterprise storage uses different storage systems for each storage function. One solution might be deployed for online network attached storage, another for backup and archival, while yet a third is used for secondary or near-line storage. These equipments can use different technologies and protocols. With the goal of minimising cost by reducing floor space and power requirements, unified-storage solutions usually accommodate multiple protocols and offer transparent and unified access to a storage pool regardless of the storage tier where the data is located [38] (*e.g.* NetApps Data ONTAP, EMCs Celerra Unified Storage Platforms). Software systems are used to migrate data across different storage tiers according to their reuse patterns.

### *4.3.6. Data De-Duplication*

Storage infrastructures generally store multiple copies of the same data. Several levels of data duplication are employed in storage centres, some required to improve the reliability and data throughput, but there is also waste that can be minimised, thus recycling storage capacity. Current SAN solutions employ data de-duplication (de-dupe) techniques with the aim of reducing data duplicates. These techniques work mainly at the data-block and file levels and commonly consist of the following steps:

- Splitting the data into individual chunks (files, blocks or sub-blocks);
- Calculating a hash value for each chunk and keeping the hash in an index; and
- Comparing the hash value of the original data with the hash of new data that needs to be stored, to verify whether the new data can be ignored or not.

In addition to the level of data de-duplication (*e.g.* block or file level), de-dupe techniques also differ on when the data de-duplication is performed: before or after data is stored on disk. Both techniques have advantages and shortcomings. Although it leads to decreases in storage media requirements, performing de-duplication after the data is stored on disk re-quires cache storage that is used for removing duplicates. However, for backup applications, performing de-duplication after storing the data usually leads to shorter backup windows and smaller performance degradation.

Moreover, data de-duplication techniques differ on where data de-dupe is carried out: at the source (client) side, target (server) side, or by a de-duplication appliance connected to the server [39]. When considering data backup, the techniques present advantages and disadvantages as shown in Table 8.

Although data de-duplication is a promising technology for reducing waste and minimising energy consumption, not all applications can benefit from it. For example, performing data de-duplication before the data is stored on disk could lead to serious performance degradation, which would be unacceptable for database applications. Applications and services that retain large volumes of data for long periods are more likely to benefit from data de-duplication. The more data one organisation has and the longer it needs to keep it, the better are the results that data de-duplication technologies yield. In general, data de-duplication works best for data backup, data replication and data retention.

The actual storage savings achieved by data de-duplication solutions vary according to their granularity. Solutions that perform hashing and de-duplication at the file-level tend to be less efficient. However, they pose a smaller overhead. With the block-level techniques, the efficiency is generally inversely proportional to the block size.

As data de-dupe solutions enable organisations to recycle storage capacity and reduce media requirements, they are also considered a common approach to reduce power consumption. By using delta versioning for example, data centres can reduce the amount of data that is transferred across the network or replicated. Incremental and differential backup solutions (*e.g.* IBMs Tivoli Storage Manager [39]) reduce the amount of data an organisation stores on its SAN infrastructure. Some organisations report reductions between 47% and 70% of their data footprint using NetApps data de-duplication solutions [41]. EMC Data Domain de-duplication systems are claimed to reduce the amount of disk storage needed to retain enterprise data by up to 30x.[10]

### *4.3.7. Data Compression*

By efficiently compressing and decompressing data on the fly, capacity can be recycled. Data compression has long been used in data communications to minimise the amount of data transferred over network links. Techniques such

[10] EMS Data Domain, http://www.datadomain.com.

Table 8: Advantages and drawbacks of different de-duplication approaches [39, 40].

| Approach | Advantages | Disadvantages |
|---|---|---|
| **Source-side (client-side)** de-duplication performed at the data source (e.g. by a backup client), before transferring to target location. | • De-dupe before transmission conserves network bandwidth.<br>• Awareness of data usage and format allow more data reduction.<br>• Processing at the source may facilitate scale-out. | • De-duplication consumes CPU cycles on the file/ application server.<br>• Requires software deployment at source (and possibly target) endpoints.<br>• Depending on design, may be subject to security attack via spoofing. |
| **Target-side (server-side)** de-duplication performed at the target (e.g. by backup software or appliance). | • No deployment of client software at endpoints.<br>• Possible use of direct comparison to confirm duplicates. | • De-duplication consumes CPU cycles on the target server or storage device.<br>• Data may be discarded after being transmitted to the target. |
| **Appliance** Appliances can perform WAN data de-duplication or storage-based de-duplication at the target. | • The appliance is a separate component that does not depend on the backup software.<br>• Processor cycles are spent on the appliance. | • Redundant data is sent over the network.<br>• WAN-based de-dupe results in redundant data on storage. If storage-based and WAN-based de-duplication are used together, it is difficult to select what data is de-duplicated.<br>• Not aware of file content; appliance tries to de-duplicate data that should not be de-duplicated. |

as minimising redundant and recurring bit patterns can prove to be efficient to reduce both the amount of data stored and the storage hardware requirements. According to EMC,[11] the block data compression techniques used in CLARiiON solutions can reduce data footprints by up to 50%. IBM Real-time Compression claims to enable clients to keep up to 5 times more data online by compressing up to 80% of data in real-time, without performance degradation.[12]

## 5. Recommendations for Best Practices

This section provides an overview of best practices adopted to reduce the power consumption and improve the energy efficiency of storage resources in enterprises and data centres. The SNIA and NetApp, for example, have released recommendations that describe best practices for data storage in data centres [42, 32, 36]. The best practices for improving energy efficiency frequently revolve around some principles that are described in this section. There are other techniques, however, which were described beforehand such as MAID and hard disk spin-down. It is also important to mention that as new types of equipments are made available, such as SSDs, existing file systems must be adapted since they have long been designed to improve the performance of other types of media.

In addition to the best practice principles described in the recommendations, there are other improvements that are applicable at the data centre level. These improvements or solutions do not relate specifically to storage equipments and include better air-conditioning systems, increase in data centre temperature, use of server virtualisation, more efficient power distribution units and Uninterruptible Power Supply (UPS) technologies.

### *5.1. Improve Storage Reliability*

Current storage architectures have been designed expecting that equipments will fail. If equipments are more reliable and expected to fail less, storage redundancy can be reduced thus decreasing the energy consumed by the overall infrastructure. This aspect is not heavily mentioned in the best practice reports – being touched upon when mentioning how to select

[11] Technical specification on EMC CLARiiON CX4 Series.

[12] IBM Real-Time Compression,
http://www-03.ibm.com/systems/storage/solutions/rtc/index.html

appropriate RAID levels – but should be taken into account. Equipments with a larger MTBF could demand less redundancy and consequently reduce the energy consumption of storage solutions. Hence, if layers of storage in data centres move towards using more reliable hardware, considerable energy savings could be achieved. It is hence important to always keep one eye on recent technologies that increase the MTBF

### *5.2. Efficient Data Management*

One of the main causes of the current data explosion faced by data storage facilities is the number of redundant copies of data that organisations maintain. Email is often mentioned as an example among the villains of data duplication in enterprises [36]. Users sending emails with large attachments are likely to increase the email server's database unnecessarily as most servers will forward a copy of the original file to each email recipient. The situation is worsened by the fact that file formats are getting richer – documents embed videos and audio files – and the users can further copy the original file to the hard drives of their personal computers or to store it in their network area. The duplicate data could be backed up indefinitely.

Therefore, policies for efficient data management, replication, and retention are crucial to reduce an organisations data footprint and maintain the energy costs under control [42]. Technologies that provide features to ease these tasks should be considered over other traditional approaches. It is important not only to use technologies that reduce the number of data duplicates, but also to change the organisation's behaviour. Some of the approaches for designing energy efficient data management policies include [42, 32, 36]:

- Prioritise data in terms of its business value. Some types of data lose their value as time goes by whereas others increase their business value after a few months or years. It is important to identify the business value of the data managed by the organisation in order to devise policies to proactively move the data to the appropriate storage solution.

- Move the data to the appropriate storage class. As discussed above, different types of data have their own business values. Data that is not mission-critical may not require high-performance storage medium. By identifying which data is not required in a timely fashion, it is possible to move data to the appropriate storage class, hence using more energy-efficient solutions – such as tape libraries and MAIDs – to store data that is not mission-critical.

- Structure Service Level Agreements (SLAs) to reward efficient data management. As an example, the data centre provider can apply price discrimination when offering storage solutions to hosted applications and services. Pricing storage according to its performance, and offering discounts to clients who move non-critical data to lower-performance storage areas or volumes can provide incentives to clients to adopt data management policies that take into account the business value of their data.

- Constantly review the information that is essential to business. As mentioned beforehand, data de-duplication and compression solutions are important as they help reduce data duplicates and the data footprint. However, it is important to constantly review what information needs to be stored and what can be simply deleted without affecting the business. Reviewing the information that is essential to business guarantees that useless data is not backed up indefinitely.

- Manage data backup and archiving efficiently. A common problem in organisations is to confuse data archiving with backup [37]. Identifying the time value of data helps manage data more efficiently by defining which data needs to be archived, preventing an organisation for wasting storage capacity by backing up several times, data that should be placed in an archival using more energy efficient media.

Therefore, in addition to the data de-duplication and consolidation approaches presented in previous sections, an important aspect is to have clear and efficient data management policies that – in addition to minimising unnecessary data duplication – classify data according to its importance and define how data must be retained. By establishing the data retainment requirements, it is possible to decide on tiered storage architectures and assign data to layers according to the their relevance, taking the energy consumption of tiers into account. It is also possible to utilise software systems that take advantage of storage tiers automatically. A clear study on how data duplicates can be eliminated, which data must be backed up, and what can be archived, is important to minimise data storage capacity requirements and consequently lower the energy footprint of a data centre.

Another technique related to data management is to employ thin provisioning of storage servers along with server and network virtualisation. To benefit from these approaches, however, it is essential to know the applications and their workloads.

### *5.3. Data De-duplication and Consolidation*

Data de-duplication is very important to eliminate data duplicates and recycle storage capacity. In the email example presented in the previous section, copies of the file sent in the original message could be eliminated, hence preventing storage capacity from being allocated to store useless copies of the same file. As discussed earlier, most data de-duplication techniques work at two levels: files and blocks. File de-duplication is less effective since hashes are computed for files instead of blocks. Hence, even if two files are 99.9% identical, storage capacity will be allocated to store both files completely.

As data de-duplication can be performed at different moments (*i.e.* in band or out-of-band) and for different classes of storage (*e.g.* primary, backup and archival) it is important to choose solutions that strike a balance between performance and storage savings. Regardless the selected solution, it is evident that minimising the storage requirements is likely to reduce the energy footprint of the storage infrastructure.

De-duplication can also be used with other techniques, such as server virtualisation, by preventing duplicated data from being produced in the first place. In server virtualisation, several copies of virtual machine images are commonly created to run the servers required to host application services. By using techniques that create virtual clones of virtual machine images, such as FlexClone from NetApp, it is possible to reduce the storage requirements for storing the images.

In addition to data de-duplication and virtualisation, storage consolidation can be achieved by other means. As discussed earlier, unified storage solutions can reduce the floor-space required by the storage infrastructure. Such techniques allow for example that SANs be consolidated at the fabric level by providing switches and directors that communicate via multiple protocols, such as IP and Fibre Channel.

### *5.4. Tiered Storage and Virtualisation*

The benefits of virtualisation and of automating the migration of data across different tiers of storage have been discussed beforehand. When designing the storage infrastructure of a data centre, it is important to provision the different tiers appropriately and have clear policies for data migration, backup, archival and data retrieval. Factoring power consumption in migration policies is important to achieve a balance of performance and energy savings.

The use of active archiving can provide considerable savings in energy consumption since infrequently used data can be moved to more energy efficient storage solutions. Technologies that facilitate active archiving are hence recommended to improve the energy efficiency of storage infrastructures that store data with various access patterns.

### *5.5. Thin Provisioning*

Storage solutions that enable thin provisioning can avoid that storage capacity be wasted by pre-allocating storage resources that are not actually used by applications. Thin provisioning allows creating virtual volumes that appear to have a given capacity, but the actual physical capacity is allocated as applications demand it. This allows organisations to recycle capacity, use fewer resources and as a consequence minimise the energy consumption.

### *5.6. Use Energy Efficient Drives*

In addition to using technologies such as MAIDs, it is relevant to employ energy efficient drives in disk-array based solutions. Using drives that provide larger IOPS per watt can increase the overall efficiency of a storage solution.

### *5.7. Shift to Solid State Drives*

Although SSDs are still expensive when compared to traditional hard disk drives, they can be considered for applications demanding high performance or for tiered storage architectures. SSDs should, therefore, be considered as storage cache or for applications that demand high-performance storage. The energy savings they can achieve with applications that present random data access patterns is substantial compared to more traditional media such as HDDs.

## 6. Community Efforts and Benchmarks

Manufacturers of storage equipment generally use the power consumption under idle state to indicate that a specific power-efficient solution saves energy when compared to a non-efficient counterpart. Actual energy savings are, however, highly dependent on the application workloads and the data-management policies in place. Some metrics take into account performance factors such as data throughput and the energy footprint of centres that use the equipments. Some metrics often found in the literature are listed as follows:

- **GB per Watt**: this metric takes into account the storage capacity of devices and can favour different equipments according to the manner it is employed. For example, SSDs are considerably less power consuming than HDDs, but they have more modest storage capacities. Several SSDs may be required to achieve the same storage capacity of a high-end hard disk, which in turn can make the energy savings of SSDs look unappealing.

- **MB/s per Watt and IOPS per Watt**: these are metrics that take into account the performance of equipments. The former considers the throughput in MB/s per Watt and the latter the number of operations per second. Although these metrics take performance into account, they may not incentivise manufacturers to put effort in minimising the power consumed by equipments during periods of inactivity. In addition, considering performance metrics such as IOPS without taking into account response time is not meaningful as applications often face problems under long storage response times.

- **Power supply efficiency**: considerable attention is given to the efficiency of power supplies and distribution units as they account to the electricity loss of storage equipments. Metrics that evaluate the ratio of DC output power to AC input power are considered in this scenario.

- $CO_2$ **footprint and total annual energy bill**: these are more exotic metrics often mentioned in product descriptions. Although important, the $CO_2$ footprint is frequently difficult to estimate as it depends on the source of electricity used by the data centre. Moreover, when showing reductions in the annual electricity, companies use workloads and scenarios that may not reflect the reality of most costumers.

Existing work has proposed some variances of the aforementioned metrics to evaluate the performance of different types of storage [16]. There are also attempts such as the ENERGY STAR Program Requirements to stipulate minimum efficiency requirements for power distribution units of data centre storage hardware such as of disk arrays.

### *6.1. Storage Performance Council*

The Storage Performance Council (SPC) has developed a set of benchmarks for evaluating the performance of storage solutions (*i.e.* SPC-1, SPC-1C, SPC-2 and SPC-2C) [43]. These benchmarks provide methodologies to evaluate, validate and publish performance results that enable the comparison of different storage solutions. The family SPC-1* of benchmarks are used to evaluate the performance of storage solutions when processing Online Transaction Processing (OLTP) applications such as DBMS and e-mail servers, whereas SPC-2* benchmarks assess the performance of storage when used for large sequential processing. These benchmarks contain extensions that aim to provide a methodology and metrics to assess the energy efficiency of the storage systems (*i.e.* SPC-1/E, SPC-1C/E, SPC-2/E and SPC-2C/E). A summary of both benchmarks and their metrics if provided by Poess *et al.* [44].

The energy extensions use the metrics defined in their parent benchmarks and are included in the energy results. They provide the basis for comparing performance and energy consumption. In addition, the energy extensions define:

- A measurement methodology for power consumption such as the types of equipments accepted and their accuracy.
- Disclosure requirements concerning the electricity supply, power distribution units, among others.
- The energy efficiency metrics.

SPC-1/E and SPC-1C/E, for example, work with the idea of three profiles, which describe the conditions in environments that impose light, moderate and heavy demands on the system. When applying the energy profiles, the heavy operation is associated with measurements obtained when the System Under Test (SUT) is processing 80% of the IOPS peak rate reported in the performance test; the moderate operation is associated with measurements taken at 50% of the IOPS peak rate of the performance test; and the idle operation uses measurements taken during the idle test phase that precedes the performance test when the energy consumption is evaluated.

The metrics reported when using this benchmark are summarised as follows:

- **Nominal Operating Power**: a weighted average of the power consumption at different load operations, where the weight is the average number of watts observed in each of the profiles.

- **Nominal Traffic (IOPS)**: similar to the metric above, the nominal traffic is a weighted average of the IOPS rates at different load operations, where the weight is the average number of watts observed in each of the profiles.

- **Operating IOPS/Watts**: assesses the efficiency with which the I/O traffic can be sustained. It is the ratio of the Nominal Traffic to the Nominal Operating Power.

- **Annual Energy Use (kWh)**: estimates the average energy use computed across three selected environments, over the course of a year. The Annual Energy Use is given by: 0.365*24*(Nominal Operating Power).

### *6.2. SNIA's Green Storage Initiative*

The SNIA Green Storage Initiative (GSI) aims to advance energy efficiency and conservation in networked storage. The Green Storage Technical Working Group focuses on developing test metrics for measuring and evaluating energy consumption, whereas the GSI targets at publicising best practices for energy efficient networked storage. One of the efforts of these groups is the SNIA Emerald Program [2], which provides a repository of vendor storage system power efficiency measurement and related data.

The Emerald program provides a methodology for measuring and evaluating the energy consumed by equipments that fall in some of the categories of the storage taxonomy that it proposes. Similarly to SPC benchmarks, SNIA's measurement methodology divides tests in different phases where the power consumption of equipments both in ready idle and active states can be assessed. The evaluation starts with a SUT conditioning phase, followed by an active test and completes with a Ready Idle Test. The SNIA considers storage systems and components to be in ready idle state when they are configured, powered up, connected to host systems and capable of satisfying I/O requests from those systems, but no I/O requests are being submitted from the host systems.

Furthermore, the specification defines a set of pass/fail tests to check the presence of Capacity Optimisation Methods (COM). They are intended to check the presence and activation of capacity optimisation techniques. The methodology proposed by SNIA also stipulates minimum duration for test phases and additional requirements such as reporting the temperature and humidity of the data storage room.

## 7. Conclusions

This chapter discussed the state of the art on techniques and best practices for improving the energy efficiency of data storage solutions. Current techniques for improving energy efficiency of storage solutions act mainly at two levels, namely the level of individual devices and at the level of storage elements such as disk arrays and storage area network equipment.

The efficiency of most solutions available for storage elements is highly dependent on the application workloads under which they operate. Hence, there is no fit-all solution. Data centre architects, operators and personnel responsible for equipment procurement should work closely with providers of storage equipment to employ solutions that best fit their performance and energy requirements. On environment with multiple storage layers, sound data management policies are essential for storing data on the most appropriate layer.

## Acknowledgments

The work presented in this chapter has been supported by the PrimeEnergyIT project (an European project financially supported by the Intelligent Energy in Europe program).

## References

[1] J. Gantz, D. Reinsel, The digital universe decade – are you ready?, IDC iVIEW (May 2010).

[2] SNIA Emerald power efficiency measurement specification, SNIA Green Storage Initiative (August 2011).

[3] ENERGY STAR program requirements for data center storage, draft 1, ENERGY STAR PROGRAM (2010).

[4] Consolidate storage infrastructure and create a greener datacenter, White paper, Oracle (April 2010).

[5] Power and cost efficient data storage, Hie Electronics, Inc., http://www.hie-electronics.com (2012).

[6] User guide for the SNIA Emerald power efficiency measurement spec, SNIA Green Storage Initiative (October 2011).

[7] D. Reine, M. Kahn, Disk and tape square off again – tape remains king of the hill with lto-4, Clipper Notes.

[8] Consolidate storage infrastructure and create a greener datacenter, Oracle White Paper (April 2010).

[9] E. Otoo, D. Rotem, S. Tsao, Analysis of trade-off between power saving and response time in disk storage systems, in: IPDPS 2009, 2009, pp. 1–8.

[10] Seagate PowerChoice technology provides unprecedented hard drive power savings and flexibility, Technology Paper (2010).

[11] J. Wang, H. Zhu, D. Li, eRAID: Conserving energy in conventional disk-based RAID system, IEEE Transactions on Computers 57 (3) (2008) 359–374.

[12] D. Colarelli, D. Grunwald, Massive arrays of idle disks for storage archives, in: Supercomputing 2002, Los Alamitos, USA, 2002, pp. 1–11.

[13] Small form factor disk drives  the economic power of lower power consumption, White Paper Fujitsu,
http://www.fujitsu.com/downloads/COMP/fcpa/hdd/sff-sas2_wp.pdf.

[14] Seagate Savvio 15K.2 data sheet, Technical Specification (2010).

[15] N. Agrawal, V. Prabhakaran, T. Wobber, J. D. Davis, M. Manasse, R. Panigrahy, Design tradeoffs for ssd performance, in: USENIX 2008 Annual Technical Conference, Berkeley, USA, 2008, pp. 57–70.

[16] O. Mordvinova, J. M. Kunkel, C. Baun, T. Ludwig, M. Kunze, Usb flash drives as an energy efficient storage alternative, in: E2GC2 2009, Banff, Canada, 2009, pp. 175–182.

[17] Intel solid state drive 510 series: Experience the 6gb/s hard drive alternative, Product Brief, Intel Corporation (2011).

[18] D. DeVetter, D. Buchholz, Improving the mobile experience with solid-state drives, intel white paper, january 2009., Intel Information Technology Whitepaper (January 2009).

[19] Dell solid state disk (ssd) drives  high performance and long product life, Dell Whitepaper (2010).

[20] Enterprise solid state drives for ibm bladecenter and system x servers, IBM Redbooks Product Guide (February 2012).

[21] G. Schulz, Achieving energy efficiency using FLASH SSD, StorageIO (December 2007).

[22] T. Bisson, S. A. Brandt, D. D. E. Long, NVCache: Increasing the effectiveness of disk spin-down algorithms with caching, in: MASCOTS 2006, Washington, USA, 2006, pp. 422–432.

[23] J. Tate, F. Lucchese, R. Moore, Introduction to Storage Area Networks, IBM Redbooks, 2006.

[24] J. Everett, F. Christofferson, S. Sahajpal, Advanced Disk Solutions for Dummies: SGI and LSI Limited Edition, John Wiley & Sons, Ltd., 2011.

[25] Power efficiency and storage arrays: Technology concepts and business considerations, EMC Whitepaper (May 2008).

[26] Nexsan energy efficient AutoMAID technology, Wikibon Green Validation Report, Wikibon Energy Lab (September 2009).

[27] P. Chu, E. Riedel, Green storage ii: Metrics and measurement, Storage Networking Industry Association (SNIA) (2008).

[28] An introduction to EMC CLARiiON CX4 disk-drive spin down technology, EMC Whitepaper (October 2009).

[29] G. Schulz, MAID 2.0: energy savings without performance compromises – energy savings for secondary and near-line storage systems, StorageIO (January 2008).

[30] CalTech relies on nexsan reliability and power efficiency to store two petabytes of critical NASA data, Nexsan 10 Minute Case Study (2009).

[31] E. Pinheiro, R. Bianchini, Energy conservation techniques for disk array-based servers, in: ICS 2004, New York, USA, 2004, pp. 68–78.

[32] L. Freeman, Reducing data center power consumption through efficient storage, NetApp White Paper (Jul. 2009).

[33] The efficient, green data center, EMC Whitepaper (October 2008).

[34] B. Craig, T. McCaffrey, Optimizing nearline storage in a 2.5-inch environment using Seagate Constellation drives, Dell Power Solutions (Jun. 2009).

[35] Storage consolidation for data center efficiency, BLADE Network Technologies White Paper (Jun. 2009).

[36] T. McClure, Driving storage efficiency in san environments, Enterprise Strategy Group (ESG) White Paper (Nov. 2009).

[37] F. Christofferson, Time value of data – creating an active archive strategy to address both archive and backup in the midst of data explosion, SGI White Paper (2010).

[38] P. Feresten, R. Parthasarathy, Unified storage architecture enabling today's dynamic data center, NetApp Whitepaper (October 2008).

[39] Data deduplication in tivoli storage manager v6.2 and v6.1, Product Guide, IBM (2011).

[40] D. Cannon, Data deduplication and tivoli storage manager, Tivoli Storage, IBM Software Group (September 2007).

[41] Polysius reclaims space, extends storage life with NetApp and datalink, NetApp Success Stories (2008).

[42] T. Clark, A. Yoder, Best practices for energy efficient storage operations, SNIA Green Initiative (Oct. 2008).

[43] Spc specifications, http://www.storageperformance.org/specs (2011).

[44] M. Poess, R. O. Nambiar, K. Vaid, J. M. Stephens Jr., K. Huppler, E. Haines, Energy benchmarks: A detailed analysis, in: 1st International Conference on Energy-Efficient Computing and Networking (e-Energy 2010), Passau, Germany, 2010, pp. 131–140.

## Appendix A. List of Acronyms

- CIFS – Common Internet File System
- COM – Capacity Optimisation Methods
- DAS – Direct-Attached Storage
- EBOD – Extended Bunch of Disks
- FC – Fibre Channel
- FCAS – Fixed Content Aware Storage
- FCoE – Fibre Channel over Ethernet
- FCP – Fibre Channel Protocol
- GSI – Green Storage Initiative
- HBA – Host Bus Adapter
- HDD – Hard Disk Drives
- HHD – Hybrid Hard Drive
- HSM – Hierarchical Storage Management
- IOPS – Input/Output Operations Per Second
- IOPS/W – Input/Output Operations Per Second per Watt
- IPM – Intelligent Power Management
- JBOD – Just a Bunch of Disk
- MAID – Massive Arrays of Idle Disk
- MaxTTFD – Maximum Time To First Data
- MLC – Multi-Level Cell
- MTBF – Mean Time Between Failure
- NAS – Network Attached Storage

- NCQ – Native Command Queuing
- NFS – Network File System
- OLTP – Online Transaction Processing
- PDC – Popular Data Concentration
- PDU – Power Distribution Unit
- RAID – Redundant Array of Independent Disks
- SAN – Storage Area Network
- SAS – Serial attached SCSI
- SATA – Serial Advanced Technology Attachment
- SBOD – Switched Bunch of Disks
- SFF – Small Form Factor
- SLA – Service Level Agreement
- SLC – Single-Level Cell
- SNIA – Storage Networking Industry Association
- SPC – Storage Performance Council
- SSD – Solid State Drive
- SUT – System Under Test
- UPS – Uninterruptible Power Supply
- WAN – Wide Area Network