electronics

MDPI

Article

Mitigating Self-Heating in Solid State Drives for Industrial Internet-of-Things Edge Gateways

Cristian Zambelli [1,*,†], Lorenzo Zuolo [2,†], Luca Crippa [2,†], Rino Micheloni [2,†] and Piero Olivo [1,†]

[1] Dipartimento di Ingegneria (DE), Università degli Studi di Ferrara, 44122 Ferrara, Italy; piero.olivo@unife.it
[2] Flash Signal Processing Labs, Microchip Technology Inc. (MCHP), 20871 Vimercate, Italy; lorenzo.zuolo@microchip.com (L.Z.); luca.crippa@microchip.com (L.C.); rino.micheloni@microchip.com (R.M.)
* Correspondence: cristian.zambelli@unife.it; Tel.: +39-0532-974993
† These authors contributed equally to this work.

Received: 22 May 2020; Accepted: 16 July 2020; Published: 20 July 2020

Abstract: Data storage in the Industrial Internet-of-Things scenario presents critical aspects related to the necessity of bringing storage devices closer to the point where data are captured. Concerns on storage temperature are to be considered especially when Solid State Drives (SSD) based on 3D NAND Flash technology are part of edge gateway architectures. Indeed, self-heating effects caused by oppressive storage demands combined with harsh environmental conditions call for proper handling at multiple abstraction levels to minimize severe performance slow downs and reliability threats. In this work, with the help of a SSD co-simulation environment that is stimulated within a realistic Industrial Internet-of-Things (IIoT) workload, we explore a methodology orthogonal to performance throttling that can be applied in synergy with the operating system of the host. Results evidenced that by leveraging on the SSD micro-architectural parameters of the queuing system it is possible to reduce the Input/Output operations Per Second (IOPS) penalty due to temperature protection mechanisms with minimum effort by the system. The methodology presented in this work opens further optimization tasks and algorithmic refinements for SSD and system designers not only in the IIoT market segment, but generally in all areas where storage power consumption is a concern.

Keywords: solid state drives; performance; throttling; power consumption; IoT; IIoT; 3D NAND Flash

1. Introduction

The new industrial revolution, mainly fueled by the Internet-of-Things (IoT) paradigm has forced many factories to deal with the issue of data storage. Indeed, an avalanche of bytes coming from sensors, robots, and cameras dispatched in several places of a factory needs collection for real-time data analytics delivery [1]. Cloud storage (either on-premise or remote) is not the prime choice for this operation since it is imperative to guarantee low latency and fast responsiveness to take decisions in the manufacturing process [2]. The best place to do this is close to the data source, colloquially defined as *the Edge*. Gateways hardware at the edge aggregates data and store them for local processing before sending data to the cloud [3]. In this context, Solid State Drives (SSDs) are the primary storage backbone of gateway platforms as they possess most of the sought features in the Industrial IoT world (IIoT), namely high bandwidth and low latency [4,5].

However, the necessity of bringing storage closer to the point where data are generated poses several challenges from a reliability standpoint. Although SSDs are known to outclass traditional magnetic storage like Hard Disk Drives in terms of metrics like Annualized Failure Rate (AFR) and Mean Time Between Failures (MTBF) [6,7], they still could suffer in harsh environments (like those in IIoT) especially when it comes to elevated data storage temperatures or sharp operating temperature

gradients. This behavior is related to the physical working principles of the storage medium inside SSDs, namely the NAND Flash technology [8]. High storage temperatures endanger the non-volatility of data, giving rise to the infamous retention-loss problem [9]. Although it may seem straightforward to accurately control the ambient temperature of the gateway where the SSD is integrated, it is not the only factor to consider. To increase read and write bandwidth, SSD architectures rely on parallel communication channels each associated to a different NAND Flash chip [5]. The higher the parallelism degree, the higher the amount of input/output operations sustainable by the drive thanks to multiple-accessed NAND Flash memories, with a consequent drive's self-heating effect [10]. Such a temperature increase could be detrimental for data retention.

Such heavy temperature sensitivity has been modeled and characterized throughout the years, so that dedicated algorithms can be applied on a firmware level to recover corrupted data with the help of powerful Error Correction Codes (ECCs) and eventually enhance a drive's reliability by retarding uncorrectable failure events [8,9,11]. Nevertheless the continuous race for higher storage density that involves the IIoT world as well has further exacerbated temperature-related issues caused either by the adoption of multi-bits per cell paradigm or by the transition from planar (2D) NAND Flash to vertically integrated 3D NAND Flash technology [12]. This has exposed how acting only at a physical or algorithmic abstraction level could be insufficient, thus calling for an exploration of solutions that span from the SSD micro-architectural parameters to the characterization of 3D NAND Flash technology peculiarities [13]. In addition, we must also point out that reliability is not a standalone concern in SSDs, but it is tightly coupled with performance figures of merit of the drive like response latency, bandwidth, and most of all Quality of Service (QoS) [14].

In this work, we address this challenge by exploring a proper thermal-aware methodology that mitigates the impact of the self-heating effect in SSDs for IIoT edge gateway architectures. Such a method considers the power and performance figures of a Triple Level Cell 3D NAND Flash technology [12] integrated in the drive and exploits the synergy with the operating system of the host to dynamically vary the parameters of a SSD's queuing system in order to minimize performance drop-outs caused by throttling events.

The contributions of this paper can be summarized as follows:

1. We evaluate the impact of a SSD's micro-architectural parameters in its internal queuing system on the performance of the drive. To the best of our knowledge, we carry this activity for the first time considering the IIoT scenario peculiarities;
2. We provide a methodology orthogonal to the state-of-the-art throttling to guard-band self-heating of the drive assuming a monitoring of its internal temperature. This can be achieved by a proper tailoring of the internal drive command queues to achieve the desired power throttling level;
3. We base all our explorations on a co-simulation framework that processes Trace-Driven Benchmarks (TPC-IoT [15]) being able to calculate SSD quality metrics (e.g., IOPS, latency, etc.) according to the measurements results of a 3D NAND Flash chip. The adaptation of the benchmark results to the IIoT scenario conferring solidness to our assumptions.

2. Related Works

Storing the data for IoT and IIoT edge applications through SSDs is a widely adopted strategy for *at-source* analytics and decision making [1,2,16]. Despite those scenarios appearing to be equivalent, there is a clear distinction between the industrial requirements with respect to pure IoT [17]. In harsh environments, reliability is the prime concern. In a storage context this feature is degraded by the temperature. There is a general consensus reported in many large-scale studies [7,8,18,19] that high temperatures may negatively affect the reliability of SSDs by aging the storage media integrated within. Therefore the JEDEC standard for SSDs testing [20] exploits this factor to accelerate failures for reliability assessments. The state-of-the-art methodology to guard-band the temperature increase, and therefore degradation, in the drive is to devise performance throttling supervised by the SSD controller through on-board temperature sensors, as shown in the studies from [10,21].

Such straightforward side-effects are a severe increase on the drive's latency and a general perceived QoS loss by the system. These studies are focused on the characterization and on the modeling of the thermal management of the SSD at a product-level, although considering typical consumer applications that are far from IIoT temperature requirements.

Stand-alone 3D NAND Flash memories have specific testing procedures as well to characterize their temperature sensitivity prior to integration in SSDs [22]. An important issue is in regard of the data retention. Indeed, 3D NAND Flash technology suffers from multiple sources of retention loss caused by temperature-activated charge loss mechanisms [23–25]. Either it is vertical or lateral charge loss through the structure of the memory architecture [12], the outcome is always the same: Temperature corrupts the content of the memory cells to a point where stored data are unrecoverable. Several works in literature discuss how to optimize the NAND Flash characteristics in the drive in order to improve the overall reliability of the system [26,27]. Most of these optimizations are at a system-level abstraction, either considering additional firmware routines to be implemented in the SSD controller, dedicated hardware in the SSD, or through external accelerators directly attached to the interface fabric exploited by the drive to communicate with the host [28,29].

Another important topic to consider is about the simulation/emulation environments that allow an exploration of SSD micro-architectural parameters and storage medium characteristics for thermal management solutions development in IIoT. In [30], a disk emulation strategy dealing with real SSD through a virtual platform was proposed. Fast performance estimations are enabled by a highly abstracted description of the components in the SSD, although losing accuracy under certain workload conditions. SSD trace-driven simulation tools were proposed in [31,32] allowing SSD performance and power consumption evaluation. However, reliability is marginally considered and they still lack the possibility to evaluate micro-architectural effects on the SSD performance like commands pipelining or uncommon queuing mechanisms.

The related research topics presented so far are resumed in Table 1.

Table 1. Related research topics.

Topic	References
SSDs Storage in IoT and IIoT	[1,2,16]
SSD temperature failures	[7,8,18,19]
Thermal management in SSD	[10,21]
3D NAND Flash technology retention temperature sensitivity and mitigation	[23–27,29]
SSD reliability/performance simulators	[30–32]

All the studies presented so far are focused on peculiar topics without considering the picture as a whole. Most of the literature discusses with a limited extent the intertwined relationship between SSD architecture, 3D NAND Flash memory peculiarities, and workload environmental conditions. To the best of our knowledge, this is the first work investigating the effect of temperature on the reliability and performance of SSD embodied in edge gateways for IIoT applications. Moreover, in this work we propose general design methodologies and algorithms that are orthogonal to the state-of-the-art and can be easily included by firmware designers in existing products.

3. Background and Methods Applied

3.1. Edge Gateway and SSD Architectures

In an IIoT edge gateway system architecture (we consider the one presented from Dell in [33] without lack of generality), the SSD plays an important role. As shown in Figure 1, the data captured from a pool of sensors in an industrial environment are connected to the gateway through peripheral ports like RS422/485 or CANbus. The received data packets are temporarily stored in a volatile Double Data Rate Dynamic RAM (DDR DRAM) and then processed by the Central Processing Unit (CPU).

This step is supervised by the Operating System (OS) which orchestrates the data transfers and protocol management. Once the data have been interpreted, they must be transferred through an interconnect fabric (e.g., SATA, PCIe, etc.) on a storage element for future availability. SSD is therefore the gateway component entrusted to this extent. Applications dedicated to data mining like machine learning frameworks or any other data visualization tool rely on big stored datasets to help the manufacturing process. This could be in the form of a simple process report or by proactively altering the production steps through the remote control of actuators like robots connected to the gateway.

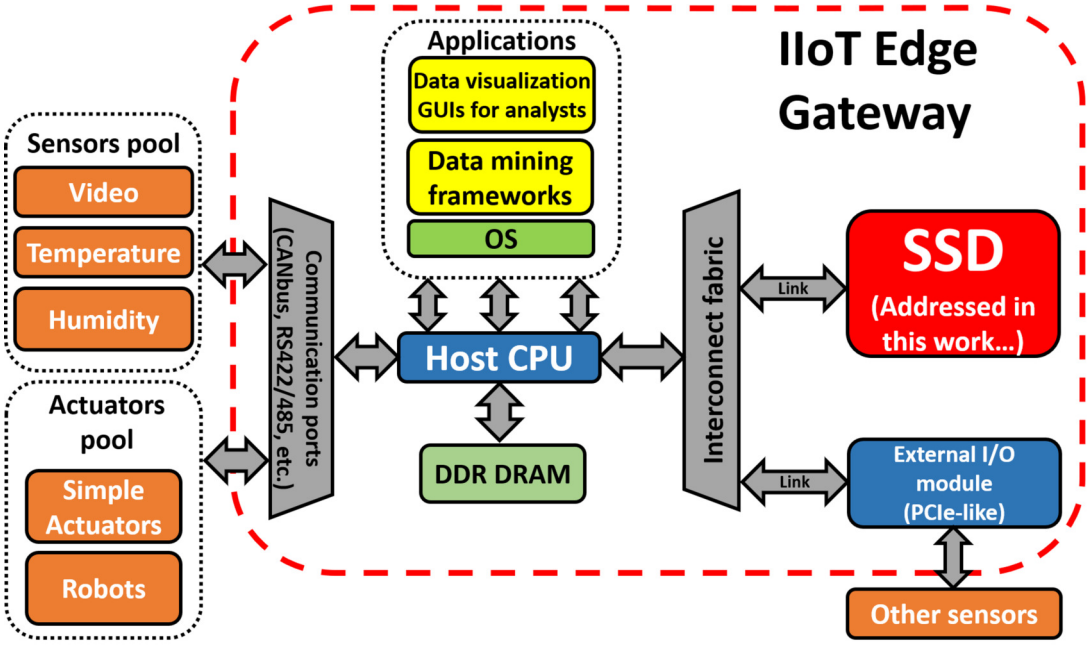

Figure 1. An example of an IIoT edge gateway architecture based on the hardware described in [33]. The SSD component addressed in this work is highlighted in red.

A SSD is a complex electronic system composed by many elements whose layout is presented in Figure 2a. The data arrive or are retrieved to/from the SSD through a host interface sharing the same interconnection fabric of the host (e.g., SATA [34] or PCIe [35]). Internally, the data movement is handled by the *smartness* of the drive, the SSD controller [36], that manages all the reliability and performance firmware routines sometimes with the help of an optional DRAM buffer [8]. This is where Flash Translation Layer (FTL) routines like wear leveling, garbage collection, and block management routines are executed [5]. Other Integrated Circuits (ICs) like temperature sensors or voltage detectors are connected to the SSD controller to help those algorithms fine tune the drive's reliability/performance characteristics. A significant portion of the SSD board is occupied by the storage media (i.e., 3D NAND Flash), which interfaces to the SSD controller through a dedicated memory interface protocol (e.g., ONFI [37] or proprietary). From an architectural standpoint it appears that a SSD is a highly hierarchical piece of architecture as shown in Figure 2b. Besides the SSD controller that integrates a multi-core CPU to run parallel FTL tasks, it is worth mentioning the presence of a channel controller. This hardware block handles the data organization of the memories (organized in N_c parallel communication channels) while at the same time providing a link with a multi-channel Error Correction Code (ECC) engine. The latter block is the one determining the ability of the SSD to handle data corruption and needs careful design to avoid performance flaws during the entire lifetime of the drive. A generic Low-Density Parity-Check ECC designed for SSD application can correct hundreds corrupted bits per 1 KBytes codeword [38].

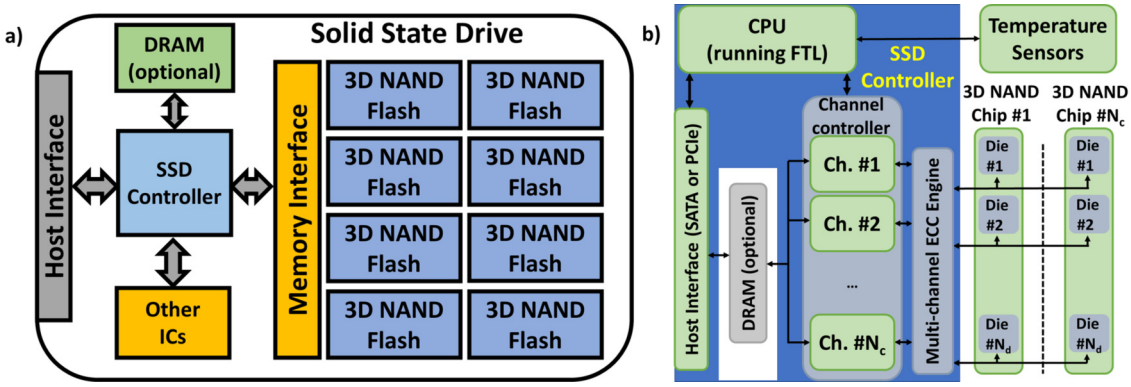

Figure 2. Layout of the components in a SSD (**a**) and an overview of its internal architecture (**b**).

3.2. Characterization and Simulation Tools

To explore the reliability, performance, and power consumption features of SSD architecture in IIoT edge gateways, we exploit the SSDExplorer co-simulator [39]. This Computer Aided Design (CAD) tool allows a fine-grained design space exploration of a drive by allowing modifications of its micro-architectural parameters like command queues, interaction mechanisms with the host system, error recovery policies, and so on. All the simulations performed with this tool consider both the electrical and timing characteristics of the storage medium. Such information comprises of reliability metrics as well (i.e., through the bit error rate or fail bits count), which have been extracted from 3D NAND Flash memory samples with the test equipment and procedures presented in [40]. SSDExplorer has been conceived and designed as a tool for virtual platforms so it can be easily plugged in virtualization environments like QEMU [41] to simulate SSDs in a fully functional machine with a working OS. In this work we setup the machine characteristics (e.g., host DRAM and number of processor cores) to be close to a representative IIoT gateway [33]. Table 2 summarizes the parameters of the host system.

To be consistent with the scenario under investigation, we exploited the TPC-IoT [15,42] benchmark in our findings. This workload mimics the typical data ingestion and query procedures in IoT gateways and can be considered also for IIoT scenarios. It consists of a large dataset representing data from sensors coming from several electric power stations. The single records in the dataset pack identification tag of the sensor, timestamp of the reading, a readout value, etc. for a storage size of 1 Kbytes per record. The workload emulates data injection in the SSD of the gateway on which a real-time data analytics platform can run queries in the background. For all the details about the structure of the benchmarking system and its configuration we would like to refer to the guidelines provided in [15].

Table 2. Configuration of the simulated IIoT edge gateway.

Component	Value
CPU	Dual-core @ 1 GHz
DRAM	2 GB
Interconnection fabric	PCIe gen2 X4
OS	Linux Ubuntu
SSD	from 64 GB up to 512 GB

Concerning the configuration of the SSDs considered in this work, we resume the assumed architectural parameters in Table 3. We considered different SSD sizes to evaluate the impact of parallel channels on the power consumption of the drive and its impact on reliability. Moreover, we speculate the integration of the Triple Level Cell (TLC) 3D NAND Flash technology in such storage platforms to

also project this study in future applications where larger amounts of data would require denser and more complex memory structures.

Table 3. SSD architectural parameters considered in the simulations.

Parameter	Value
Capacity	64–512 GB
Host interface	PCIe gen2 X4
Queue Depth (QD)	Variable from 1 to 64
Channels	1–8
Targets (number of Flash die per chip)	8
Storage medium technology	TLC 3D NAND Flash
3D NAND Flash page	16 KB+parity (4 sectors of 4320 B)
DRAM cache	64–512 MB
ECC	LDPC up to 220 bits
Advanced ECC protection	1 bit soft decoding
Over-Provisioning	30%
Write Amplification Factor	2.4

4. Exploring How to Mitigate the Self-Heating Issue

4.1. Characterizing the Power Consumption in SSDs

When a SSD is continuously accessed at full performance by a host system with data read and write requests as in our study case there is an increase in temperature. This is because a drive's temperature depends not only on the ambient temperature, but also on the different power sources in the SSD architecture that translate in multiple heat sources. As shown in [21,43], it can be observed how the temperature of a SSD increases by several tens of degrees Celsius passing from an idle state to a full performance state. Moreover, since components in the SSD architecture thermally react differently (due to different heat transfer/radiation and thermal dissipation features) there could be up to a ± 15 °C difference from chip to chip on the storage system. This is critical for 3D NAND Flash memories since their average reliability features heavily depends on temperature. Figure 3 shows the number of bit errors, normalized with respect to the ECC correction capability, retrieved in several memory chips as a function of the elapsed time when the storage temperature varies from 55 to 70 °C. Differences are reported up to 5.6 times in the errors number. A case like the one just described is to be avoided since different wear dynamics are experienced by the memories (i.e., one 3D NAND Flash chip could fail precociously) with a burden on SSD reliability.

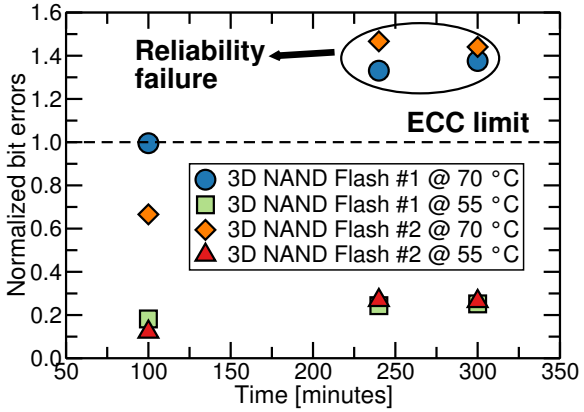

Figure 3. Data retention characteristics as a function of elapsed time measured on two 3D NAND Flash chips storing data either at 55 or 70 °C. Results are normalized with respect to the ECC correction capability provided by the SSD.

Assessing SSD power consumption is therefore a mandatory task to develop strategies that could limit the insurgence of thermal issues. With such a consideration, we breakdown the power contributors in a drive as follows:

1. The SSD controller that is an Application Specific Integrated Circuit (ASIC) whose power consumption linearly increase with time according to the amount of data to process and manage;
2. The 3D NAND Flash memory sub-system whose power contribution depends on the amount of parallel accessed channels and on the operation performed (i.e., data read, write/program or erase);
3. The DRAM buffer used as a cache or as a temporary storage for FTL-metadata structures;
4. Other ICs and passive components for power supply and temperature control.

To investigate the contribution of the sole 3D NAND Flash memory modules on the overall power consumption, we adopted the experimental setup provided in [44]. Such a characterization system extracts the current drawn from the memory power-supply therefore providing the actual power consumption figures for the different operations of the memory. Figure 4a,b shows an example of two extracted power traces from a 3D NAND Flash chip during program and read operations. It is worth mentioning that since we consider a TLC memory in this study, we collected different power traces for each page type (i.e., lower, central, and upper pages). Similar behavior, although with different timings and peak values are found. When multiple memory chips are accessed in parallel on a SSD the Kirchoff's Current Law (KCL) holds on the power supply of the drive so that the memory sub-system power consumption is the sum of single power contributions, as shown in Figure 4c.

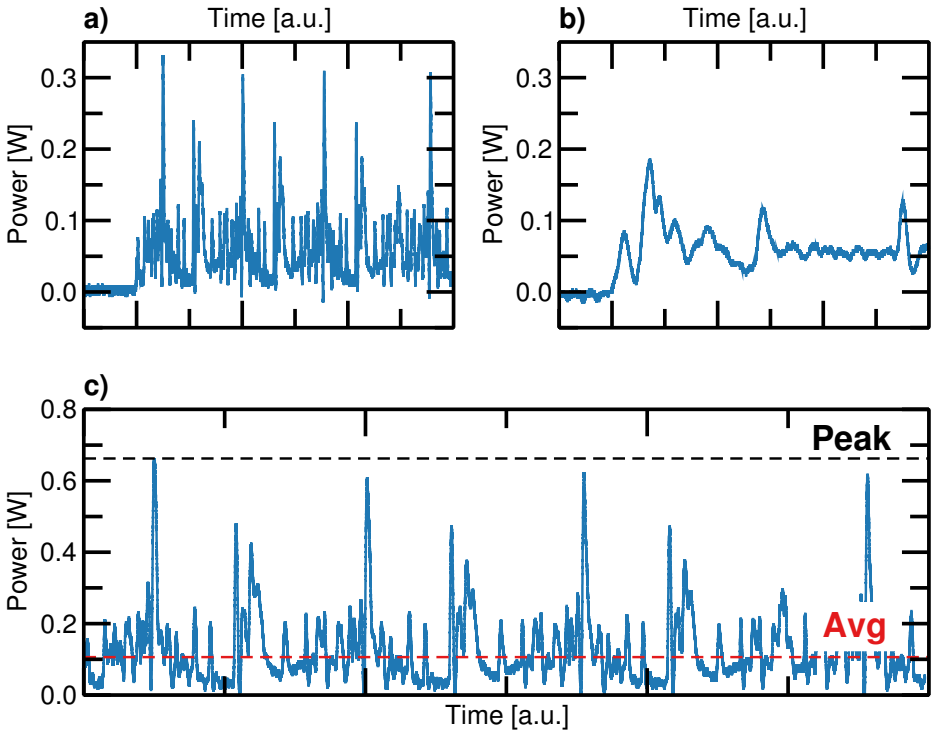

Figure 4. Power figures measured from a 3D NAND Flash chip. The power consumption of a single write operation (**a**), a single read operation (**b**), and a write operation on two parallel accessed chips (**c**) are presented. In the latter, power and average peak values are evidenced.

We then performed a simulation of a 64 GB SSD with a single channel (i.e., single 3D NAND Flash package) and a number of outstanding commands to serve equal to 8. We exploited the 3D NAND Flash power traces presented before. The TPC-IoT benchmark workload is considered in the investigation. Please note that the SSDExplorer can extract the power consumption only for

the memories in the drive (both Flash and DRAM). To assess the total drive power consumption, we considered the peak values reported in [21] for the SSD controller and for the other ICs. We set the former equal to 2 W and the latter equal to 400 mW, respectively. We repeated the benchmark considering also a 512 GB drive with eight parallel memory channels and an increase in the SSD controller power up to 6 W due to a higher amount of channels needing to serve. As observed in Figure 5, the 3D NAND Flash memories contribution on the SSD power consumption weighs from 19.4% to 41% on total, scaling with the channel parallelism. As a general rule, the higher the amount of 3D NAND Flash chips on the SSD, the stronger its contribution will be. This strongly motivates us to find solutions that limit the 3D NAND Flash power consumption to keep the SSD temperature constrained. As a benefit, a minimal degradation of the drive's performance will be ensured while providing high reliability in environments like IIoT where these requirements on storage are enforced day by day.

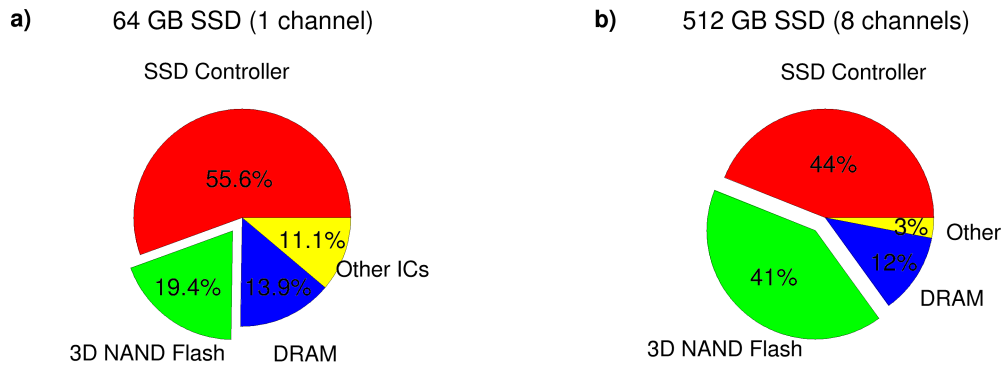

Figure 5. SSD power breakdown for a single channel 64 GB drive **(a)** and for a eight channel 512 GB drive **(b)**.

4.2. The Role of SSD Micro-Architecture on Power Consumption

The data flow in a SSD, both from the read and write perspective is regulated by the SSD controller that allows for the servicing of a number of outstanding operations that depends on the parallelism degree of the storage medium. To maximize the throughput, the host interface of the SSD implements a queue to store a set of commands to be serviced internally by the drive. The maximum number of commands that fills such a queue is defined as *Queue Depth* (QD). This command/operation queuing concept stems from the past of HDDs largely exploiting the SATA communication protocol [34]. Even if communication protocols evolved, SSDs are not different from such paradigm except for the micro-architecture of the queuing system. Besides QD, SSDs have additional queue entities as shown in Figure 6. The highly hierarchical architecture of a drive with different components accessed in parallel calls for multiple queues in the channel controller (see Figure 2b) to take advantage of the storage medium features [5]. In fact, each 3D NAND Flash chip connected to the channel controller features multiple dies (up to eight in this work) and each one basically retains its own queue called *Target Command Queue (TCQ)*. In this case, the drive can sustain commands on a die already busy by another operation. The target command queue is a fixed parameter that depends on the architecture of the firmware run by the SSD controller. However, to provide enough flexibility there is an additional entity stored in the DRAM of the SSD, defined as *frame-buffer*, which collects the maximum amount of transactions processable by all the TCQs.

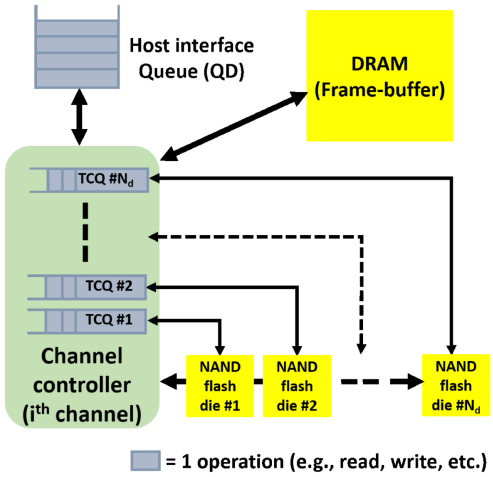

Figure 6. Queuing system in a SSD with single micro-architectural blocks evidenced.

At first glance, we evaluated the impact of QD on the number of input/output operations (measured in IOPS) sustained by the drive and on the latency of a 64 GB SSD-IIoT platform. To enrich this study, we compared the results from the TPC-IoT benchmark execution with those of a 4 kB mixed (i.e., 50%–50%) synthetic read/write workload run on the same gateway architecture. The choice of the latter workload is dictated by the requirements of nowadays file systems exploited by the OS, which tend to align the file size to 4 kB to improve the data throughput [45]. As evidenced in Figure 7, the average drive's IOPS and latency scale with QD as expected ranging from 9 kIOPS at a minimum QD of up to 48 kIOPS. Indeed, the higher the number of commands in the queue, the higher the amount of data to process and so it is for the throughput. Latency straightforwardly increases because the higher the number of commands in the queue, the longer the service time. It is interesting to note that the synthetic 4 kB workload saturates the IOPS sustained by the drive starting from a QD value equal to 8. This is ascribed to the maximum number of parallel addressable 3D NAND Flash targets in the SSD channel, as defined in Table 3. In general, the TPC-IoT produces a higher IOPS amount since at the same time, units write/read more transactions to the SSD. Of cours this will not hold for bandwidth concerns. It is also worth noticing that a too high QD value may collide with the fast responsiveness requests in an IIoT scenario that usually should target few milliseconds [46]. A similar trend from the results are obtained by simulating a 512 GB SSD-IIoT platform except that the bandwidth is eight times higher than in the 64 GB counterpart thanks to a higher number of parallel 3D NAND Flash chips accessed by the SSD channel controller.

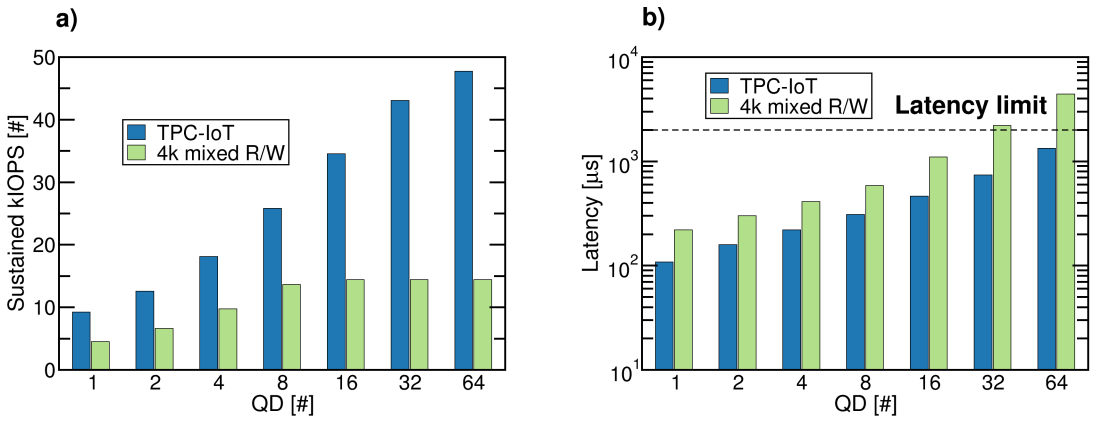

Figure 7. SSD's kIOPS (**a**) and latency (**b**) metrics for a 64 GB storage platform as a function of QD. Both the Trace-Driven Benchmarks (TPC)-IoT and the 4 kB synthetic workloads are considered.

From these simulations we extracted the power consumption of the 3D NAND Flash sub-system in the drive. Two types of power figures are of interest in SSD studies [44]: The average consumption during the entire workload and the peak consumption from the 3D NAND Flash sub-system, as previously highlighted in Figure 4. Figure 8 shows both metrics as a function of QD. Results demonstrate that the average power consumption of the drive, when TPC-IoT or synthetic workload are concerned, correlated with the sustained IOPS since it depended on the number of active 3D NAND Flash targets on the SSD during the workload. Considering the TPC-IoT workload, it was observed that there was an increase of the average power consumption from 100 mW at low QD up to 550 mW at maximum QD. Peak power consumption instead, was a function of the probability in having multiple targets and channels in the SSD active on their peaks of the power figures (see Figure 4). For low QD values (below 8), the TPC-IoT workload generated a low probability of peak overlapping for the different targets mainly due to the small data transfer sizes involved. Each workload saturated the peak power consumption at 704 mW. Once again, the same considerations could be derived for a larger drive like the 512 GB one considered in this study except that its average and peak power magnitude would be eight times higher.

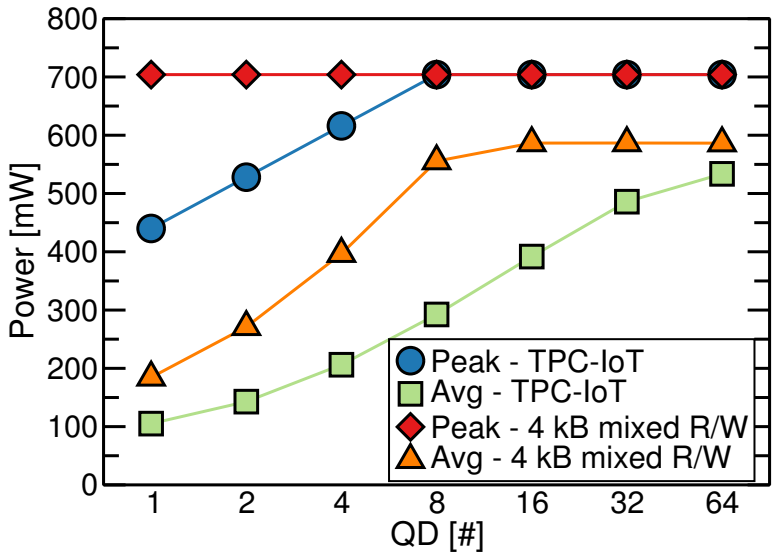

Figure 8. Average and peak power consumption of the 3D NAND Flash memory sub-system of a 64 GB SSD stressed with TPC-IoT and synthetic 4 kB workloads as a function of QD.

Another SSD micro-architectural parameter that impacts power consumption is the frame-buffer size. Its role is twofold since it constrains the amount of data moving in the 3D NAND Flash sub-system and also affects the amount of on-board DRAM allocated for the TCQ collection. Here we consider only simulation results on TPC-IoT since a 4 kB synthetic workload would expose similar trends and would not add significance to the discussion. To better expose the relationship between the frame-buffer size and QD, we performed the simulations with the latter parameter set to 4, 8, and 16. In Figure 9, we reported the kIOPS of the drive and the power consumption of the 3D NAND Flash memory sub-system as a function of the frame-buffer size and for different QD values. Low frame-buffer sizes are associated with a low power consumption (below 250 mW on average) since the managed TCQs by the DRAM were smaller in depth and so the number of parallel 3D NAND Flash active targets. DRAM power consumption decreased as well since the allocated amount of data for TCQ was lower, although this had a minor impact on the overall SSD power consumption and goes beyond the scope of this work. It is worth pointing out that a very low frame-buffer size could increase the probability of Head-of-Line (HoL) blocking events [5] regardless of the QD, with adverse effects on the drive's responsiveness (i.e., latency) since the SSD spends most of the time in a stall condition. In Figure 10, we demonstrate this by showing the number of HoL events detected in the drive during a snippet

of 10^6 transactions during the TPC-IoT execution and the corresponding SSD latency. For this latter metric in particular, we observe an indirect dependency (from 850 µs down to 300 µs) compared with the QD-related results therefore care must be taken in using a frame-buffer as a parameter for power reduction.

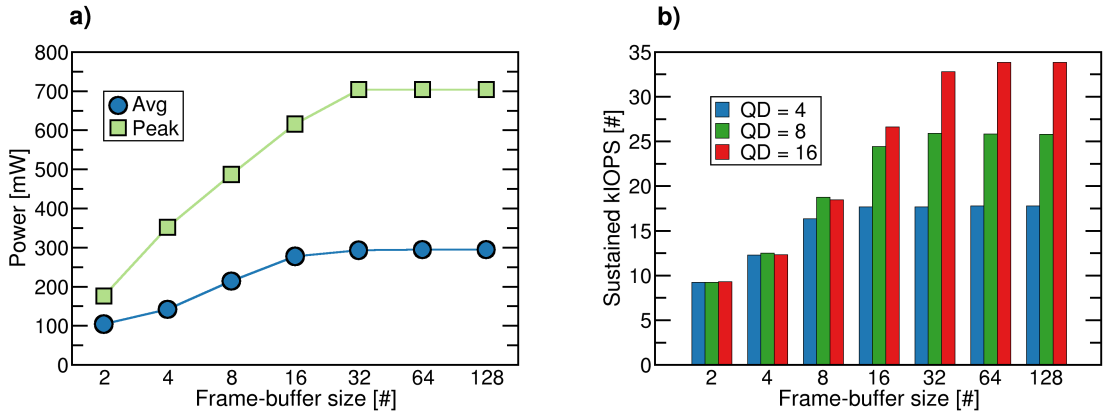

Figure 9. Average and peak power consumption (**a**) of the 3D NAND Flash memory sub-system and SSD kIOPS (**b**) of a 64 GB drive as a function of the frame-buffer size considering QD = [4, 8, 16].

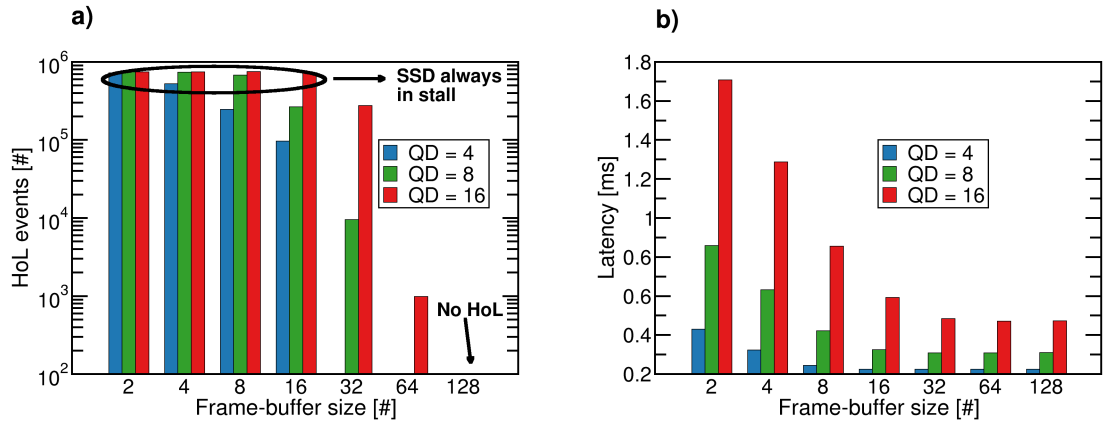

Figure 10. Head-of-Line (HoL) events count (**a**) and SSD latency (**b**) of a 64 GB drive as a function of the frame-buffer size considering QD = [4, 8, 16].

4.3. A Benchmark with Command Submission Time-Based Throttling

In SSDs, the throttling procedure is a common solution to manage thermal issues [10,21]. To assess the drive's temperature, a set of temperature sensors was exploited to measure the temperature of the most critical components in the drive, namely the 3D NAND Flash chips. When the temperature of the worst 3D NAND Flash chip (i.e., the one working with the highest temperature) exceeded a threshold, the performance of the SSD reduced to a significant extent. This would provide a time-window in which the temperature could decrease and as soon as it went below that threshold the performance was brought to a fully operative state. Depending on the scenario and on the severity of the application there could be multiple throttling stages [47]. The easiest way to achieve throttling was by leveraging the OS through the Command Submission Time (CST). This parameter was largely exploited in tuning IoT system-QoS and a general response time in virtual environments [48]. When the temperature sensors on board of the SSD reported to the host OS an alert situation (e.g., it could be through S.M.A.R.T. indicators [49]), the OS could augment the actual time taken to transfer a data read/write command from the OS to the drive. Having less commands to process turned into a lower utilization of the SSD controller resources and of the 3D NAND Flash sub-system, yielding to a lower power

consumption and temperature. We simulated the impact of throttling varying the CST of the host OS in the IIoT edge gateway by analyzing the performance of the larger 512 GB drive. This would magnify the sustained IOPS slow-down and ease the understanding of such issue. Nevertheless, the simulation results still reflected what happens in a smaller 64 GB SSD although on a different performance scale due to the lower number of parallel 3D NAND Flash channel. Figure 11 shows that a highly aggressive throttling with 20 ms CST could slow-down the SSD's sustained IOPS up to 4.93 times. We also evaluated the QoS considering different levels (i.e., different nines in SSDs jargon [5]) of the Cumulative Distribution Function (CDF) of the drive's latency. Increasing the CST would paradoxically improve the QoS since the probability to fill the queuing system in the SSD was lowered due to a higher time interval between commands. However, a drive's responsiveness was heavily traded with IOPS which could be detrimental for applications that requires a high amount of data processed per second like in an IIoT gateway executing real-time data analytics.

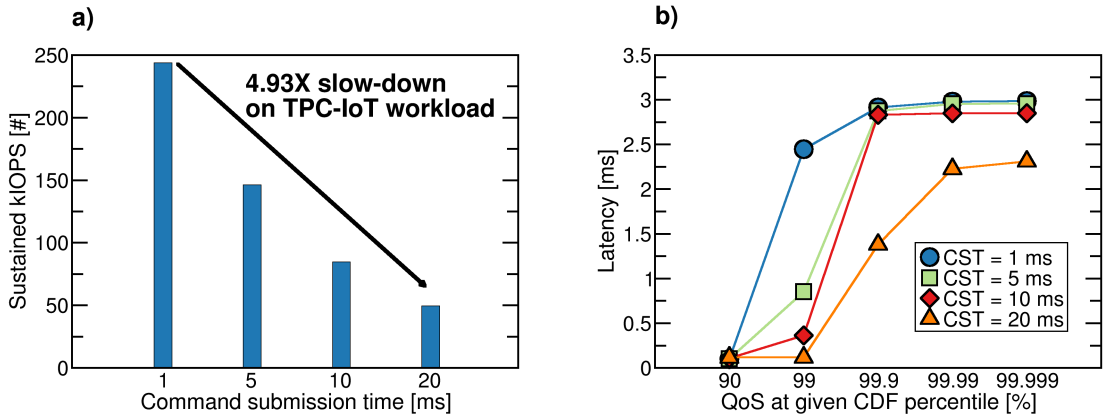

Figure 11. Sustained kIOPS (**a**) and Quality of Service (QoS) at different levels of the latency Cumulative Distribution Function (CDF) (**b**) of a 512 GB drive as a function of the command submission time (throttling aggressiveness).

We previously experienced that both QD and frame-buffer are parameters exploitable for power consumption reduction purposes and for thermal management. Since both are user-definable parameters generally through the OS resources that communicates with the SSD (i.e., through OS drivers), a rule-of-thumb solution in power consumption reduction and therefore on drive's self-heating could be an algorithm orthogonal to the throttling one, as depicted in Figure 12.

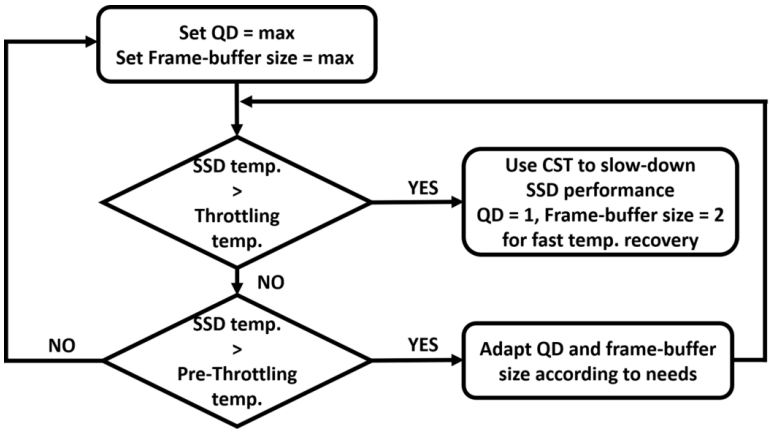

Figure 12. Flow-chart representing our proposed algorithm orthogonal to Command Submission Time (CST)-based throttling.

The drive starts with the maximum QD achievable by the SSD to grant maximum sustained IOPS and then, when a certain temperature threshold level (i.e., pre-throttling temperature) sufficiently below throttling temperature is reached, it progressively reduces it before the drive actually enters in a throttling state. For a fine-grained tuning of the power consumption at a given QD, the frame-buffer size can be varied as well though carefully monitoring possible HoL events that would affect latency. Of course throttling events cannot be completely avoided by only tuning QD and frame-buffer size especially if the SSD sustains a heavy workload for long time frames, but their occurrence can be delayed with the potential benefit on the average sustained IOPS during the workload since the time window in which the SSD operates outside throttling (i.e., the time spent between safe temperature and throttling temperature) is widened. We must remind that with our methodology it is possible to reduce the throttling time and have a fast temperature recovery since we can set the QD and frame-buffer size to minimum values, which is currently not feasible with state-of-the-art throttling algorithm. In Figure 13, we benchmark the state-of-the-art throttling with our proposed methodology. Our SSDExplorer simulator could infer the temperature of the 3D NAND Flash devices integrated in the drive and then simulate throttling from the power traces generated during the submission of the workload using the thermal model and the simulation strategy provided in [50]. Currently, we do not support the modeling of the SSD controller temperature and of the DRAM module. The simulation results show that the sole throttling approach achieved on a generic time-window of the TPC-IoT workload a 161.98 kIOPS average performance, whereas our proposed optimized algorithm materialized in a 184.4 kIOPS average performance. The option devised to manage the fallback from the throttling in our algorithm was to set the QD and frame-buffer size one step below the value achieved before entering the throttling stage. This is only one of the available options that should be explored in future to identify the strategy leading to the best gain. We also expect that in case of an extremely heavy workload sustained by the SSD (i.e., more than 10 or 100 throttling events in the same time window considered in our analysis) our methodology should converge to the state-of-the-art in terms of sustained IOPS. However, the degree of flexibility provided could open unprecedented optimization that leverage as an example with applications, OS, drivers, etc.

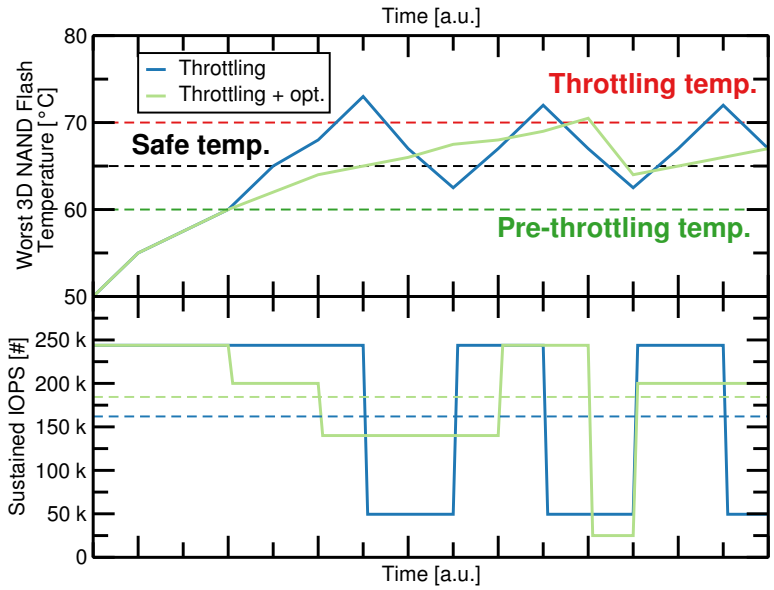

Figure 13. Benchmarking the CST-based throttling with respect to our optimized algorithm (i.e., throttling + opt.). A 512 GB SSD is considered in the study with QD = 8 as a starting point for both approaches.

Eventually, our proposed solution would work well under the assumption that the ambient temperature is sufficiently far from the throttling temperature of the SSD and given the possibility to

smart control the QD, frame-buffer size, temperature, and many other OS-related parameters through proper system drivers.

5. Conclusions

In this work, we have analyzed the self-heating effect in SSDs for IIoT edge gateway applications through the study of the power consumption of the storage medium sub-system. We considered 3D NAND Flash technology in wake of the augmented storage demands for this application scenario. By characterizing the power requirement of the write and read operations through electrical measurements we studied, with the help of a co-simulation environment, the impact on overall consumption (up to 41%) in SSD architectures.

Furthermore, we explored methods of reducing the self-heating effect by acting on the SSD micro-architectural parameters of its queuing system, namely the queue depth and the frame-buffer size. Their role was thoroughly investigated by monitoring SSD sustained IOPS, latency, and power consumption.

Finally, we proposed a methodology orthogonal to command submission time-based throttling that could be implemented by the host operating system and that contributes to reduce performance slowdowns when temperature crosses the throttling temperature. Up to 20 kIOPS on average could be gained. This methodology could be exploited by SSD and system designers for future refinements in multiple scenarios besides IIoT.

Author Contributions: Conceptualization, C.Z., R.M. and P.O.; methodology, C.Z. and R.M.; software, L.Z.; validation, C.Z., L.Z. and R.M.; electrical measurements and resources, L.C.; investigation, C.Z. and L.Z.; data curation, C.Z. and R.M.; writing—original draft preparation, C.Z.; writing—review and editing, L.Z., R.M. and P.O.; visualization, C.Z.; supervision, R.M. and P.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially funded by the UniFE-CCIAA 2018 grant - Bando 2018 per progetti di ricerca finanziati con il contributo della Camera di Commercio, Industria, Artigianato e Agricoltura.

Acknowledgments: The authors would like to thank Gianluca Torsoli and Giovanni Canella for being of valuable help during the experiments and simulations.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
SSD	Solid State Drive
IIoT	Industrial Internet of Things
AFR	Annualized Failure Rate
MTBF	Mean Time Between Failures
ECC	Error Correction Code
QoS	Quality of Service
OS	Operating System
FTL	Flash Translation Layer
ASIC	Application Specific Integrated Circuit
QD	Queue Depth
TCQ	Target Command Queue
CST	Command Submission Time

References

1. Fu, J.; Liu, Y.; Chao, H.; Bhargava, B.K.; Zhang, Z. Secure Data Storage and Searching for Industrial IoT by Integrating Fog Computing and Cloud Computing. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4519–4528. [CrossRef]
2. Karthikeyan, P.; Thangavel, M. (Eds.) *Applications of Security, Mobile, Analytic and Cloud (SMAC) Technologies for Effective Information Processing and Management*; IGI Global: Hershey, PA, USA, 2018; chapter Processing IoT Data: From Cloud to Fog-It's Time to Be Down to Earth, pp. 124–148. [CrossRef]
3. Chen, H.; Jia, X.; Li, H. A brief introduction to IoT gateway. In Proceedings of the IET International Conference on Communication Technology and Application (ICCTA 2011), Beijing, China, 14–16 October 2011; pp. 610–613. [CrossRef]
4. Micheloni, R.; Marelli, A.; Eshghi, K. (Eds.) *Inside Solid State Drives (SSDs)*; Springer: Berlin/Heidelberg, Germany, 2012; chapter SSD Market Overview; pp. 1–17.
5. Zuolo, L.; Zambelli, C.; Micheloni, R.; Olivo, P. Solid-State Drives: Memory Driven Design Methodologies for Optimal Performance. *Proc. IEEE* **2017**, *105*, 1589–1608. [CrossRef]
6. Marquart, T. Solid-State-Drive qualification and reliability strategy. In Proceedings of the IEEE International Integrated Reliability Workshop (IIRW), South Lake Tahoe, CA, USA, 11–15 October 2015; pp. 3–6. [CrossRef]
7. Schroeder, B.; Merchant, A.; Lagisetty, R. Reliability of nand-Based SSDs: What Field Studies Tell Us. *Proc. IEEE* **2017**, *105*, 1751–1769. [CrossRef]
8. Mielke, N.R.; Frickey, R.E.; Kalastirsky, I.; Quan, M.; Ustinov, D.; Vasudevan, V.J. Reliability of Solid-State Drives Based on NAND Flash Memory. *Proc. IEEE* **2017**, *105*, 1725–1750. [CrossRef]
9. Cai, Y.; Luo, Y.; Haratsch, E.F.; Mai, K.; Mutlu, O. Data retention in MLC NAND flash memory: Characterization, optimization, and recovery. In Proceedings of the 2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA), Burlingame, CA, USA, 7–11 February 2015; pp. 551–563. [CrossRef]
10. Zhang, J.; Shihab, M.; Jung, M. Power, Energy, and Thermal Considerations in SSD-Based I/O Acceleration. In Proceedings of the 6th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 14), Philadelphia, PA, USA, 17–18 June 2014.
11. Takahashi, T.; Yamazaki, S.; Takeuchi, K. Data-retention time prediction of long-term archive SSD with flexible-nLC NAND flash. In Proceedings of the 2016 IEEE International Reliability Physics Symposium (IRPS), Pasadena, CA, USA, 17–21 April 2016; pp. 6C-5-1–6C-5-6. [CrossRef]
12. Micheloni, R.; Aritome, S.; Crippa, L. Array Architectures for 3-D NAND Flash Memories. *Proc. IEEE* **2017**, *105*, 1634–1649. [CrossRef]
13. Zambelli, C.; Micheloni, R.; Olivo, P. Reliability challenges in 3D NAND Flash memories. In Proceedings of the 2019 IEEE 11th International Memory Workshop (IMW), Monterey, CA, USA, 12–15 May 2019; pp. 1–4. [CrossRef]
14. Grossi, A.; Zuolo, L.; Restuccia, F.; Zambelli, C.; Olivo, P. Quality-of-Service Implications of Enhanced Program Algorithms for Charge-Trapping NAND in Future Solid-State Drives. *IEEE Trans. Device Mater. Reliab.* **2015**, *15*, 363–369. [CrossRef]
15. Transaction Processing Performance Council (TPC). *(TPCx-IoT) Standard Specification Version 1.0.5*; TPC: San Francisco, CA, USA, 2020.
16. ATP Inc. IoT and IIoT: Flash Storage, Sensors and Actuators in Cloud/Edge. 2018. Available online: https://www.atpinc.com/blog/What-is-iiot-vs-iot-actuators-edge-cloud-storage-sensor-data (accessed on 5 May 2020).
17. Schada, J. The Striking Contrast Between IoT and IIoT SSDs. 2016. Available online: https://www.electronicdesign.com/technologies/iot/article/21802116/the-striking-contrast-between-iot-and-iiot-ssds (accessed on 5 May 2020).
18. Meza, J.; Wu, Q.; Kumar, S.; Mutlu, O. A Large-Scale Study of Flash Memory Failures in the Field. In Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Portland, OR, USA, 15–19 June 2015; pp. 177–190. [CrossRef]
19. Wang, Y.; Dong, X.; Zhang, X.; Wang, L. Measurement and Analysis of SSD Reliability Data Based on Accelerated Endurance Test. *Electronics* **2019**, *8*, 1357. [CrossRef]

20. JEDEC. *JESD218B Solid-State Drive (SSD) Requirements and Endurance Test Method*; JEDEC: Arlington, VA, USA, 2016.
21. Zhang, H.; Thompson, E.; Ye, N.; Nissim, D.; Chi, S.; Takiar, H. SSD Thermal Throttling Prediction using Improved Fast Prediction Model. In Proceedings of the 2019 18th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), Las Vegas, NV, USA, 28–31 May 2019; pp. 1016–1019. [CrossRef]
22. JEDEC. *JESD22-A117 Electrically Erasable Programmable ROM (EEPROM) Program / Erase Endurance and Data Retention Stress Test*; JEDEC: Arlington, VA, USA, 2018.
23. Mizoguchi, K.; Takahashi, T.; Aritome, S.; Takeuchi, K. Data-Retention Characteristics Comparison of 2D and 3D TLC NAND Flash Memories. In Proceedings of the 2017 IEEE International Memory Workshop (IMW), Monterey, CA, USA, 14–17 May 2017; pp. 1–4. [CrossRef]
24. Park, J.; Shin, H. Modeling of Lateral Migration Mechanism of Holes in 3D NAND Flash Memory Charge Trap Layer during Retention Operation. In Proceedings of the 2019 Silicon Nanoelectronics Workshop (SNW), Kyoto, Japan, 9–10 June 2019; pp. 1–2. [CrossRef]
25. Luo, Y.; Ghose, S.; Cai, Y.; Haratsch, E.F.; Mutlu, O. Improving 3D NAND Flash Memory Lifetime by Tolerating Early Retention Loss and Process Variation. In Proceedings of the Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems, Irvine, CA, USA, 18–22 June 2018; p. 106. [CrossRef]
26. Deguchi, Y.; Takeuchi, K. 3D-NAND Flash Solid-State Drive (SSD) for Deep Neural Network Weight Storage of IoT Edge Devices with 700x Data-Retention Lifetime Extention. In Proceedings of the 2018 IEEE International Memory Workshop (IMW), Kyoto, Japan, 13–16 May 2018; pp. 1–4. [CrossRef]
27. Mizushina, K.; Nakamura, T.; Deguchi, Y.; Takeuchi, K. Layer-by-layer Adaptively Optimized ECC of NAND flash-based SSD Storing Convolutional Neural Network Weight for Scene Recognition. In Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018; pp. 1–5. [CrossRef]
28. Karlay Inc.. The KalRay Multi-Purpose-Processing-Array (MPPA). 2016. Available online: http://www.kalrayinc.com/kalray/products/#processors (accessed on 5 May 2020).
29. Zambelli, C.; Bertaggia, R.; Zuolo, L.; Micheloni, R.; Olivo, P. Enabling Computational Storage Through FPGA Neural Network Accelerator for Enterprise SSD. *IEEE Trans. Circuits Syst. II Express Briefs* **2019**, *66*, 1738–1742. [CrossRef]
30. Yoo, J.; Won, Y.; Hwang, J.; Kang, S.; Choi, J.; Yoon, S.; Cha, J. VSSIM: Virtual machine based SSD simulator. In Proceedings of the IEEE Symposium on Mass Storage Systems and Technologies (MSST), Long Beach, CA, USA, 6–10 May 2013; pp. 1–14. [CrossRef]
31. Lee, J.; Byun, E.; Park, H.; Choi, J.; Lee, D.; Noh, S.H. CPS-SIM: Configurable and Accurate Clock Precision Solid State Drive Simulator. In Proceedings of the 2009 ACM Symposium on Applied Computing, Honolulu, HI, USA, 9–12 March 2009; pp. 318–325. [CrossRef]
32. Jung, H.; Jung, S.; Song, Y.H. Architecture exploration of flash memory storage controller through a cycle accurate profiling. *IEEE Trans. Consum. Electron.* **2011**, *57*, 1756–1764. [CrossRef]
33. Dell Inc. Dell Edge Gateway - 5000 Series - Installation and Operation Manual. 2019. Available online: https://topics-cdn.dell.com/pdf/dell-edge-gateway-5000_users-guide_en-us.pdf (accessed on 5 May 2020).
34. Serial ATA International Organization. SATA Revision 3.4 Specifications. 2020. Available online: www.sata-io.org (accessed on 5 May 2020).
35. PCI-SIG. PCI Express Base 3.1 Specification. 2015. Available online: http://www.pcisig.com/specifications/pciexpress/base3/ (accessed on 5 May 2020).
36. Microsemi Inc. (A Microchip company). Microsemi PM8609 NVMe2032 Flashtec NVMe Controller. 2019. Available online: https://www.microsemi.com/product-directory/storage-ics/3687-flashtec-nvme-controllers (accessed on 14 June 2019).
37. Open Nand Flash Interface (ONFI). Open NAND Flash Interface Specification - Revision 4.2. 2020. Available online: http://www.onfi.org (accessed on 5 May 2020).
38. Li, M.; Chou, H.; Ueng, Y.; Chen, Y. A low-complexity LDPC decoder for NAND flash applications. In Proceedings of the 2014 IEEE International Symposium on Circuits and Systems (ISCAS), Melbourne, Australia, 1–5 June 2014; pp. 213–216. [CrossRef]

39. Zuolo, L.; Zambelli, C.; Micheloni, R.; Indaco, M.; Carlo, S.D.; Prinetto, P.; Bertozzi, D.; Olivo, P. SSDExplorer: A Virtual Platform for Performance/Reliability-Oriented Fine-Grained Design Space Exploration of Solid State Drives. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **2015**, *34*, 1627–1638. [CrossRef]
40. Zambelli, C.; King, P.; Olivo, P.; Crippa, L.; Micheloni, R. Power-supply impact on the reliability of mid-1X TLC NAND flash memories. In Proceedings of the IEEE International Reliability Physics Symposium (IRPS), Pasadena, CA, USA, 17–21 April 2016; pp. 2B-3-1–2B-3-6. [CrossRef]
41. QEMU: The FAST! Processor Emulator. 2020. Available online: https://www.qemu.org (accessed on 5 May 2020).
42. Poess, M.; Nambiar, R.; Kulkarni, K.; Narasimhadevara, C.; Rabl, T.; Jacobsen, H. Analysis of TPCx-IoT: The First Industry Standard Benchmark for IoT Gateway Systems. In Proceedings of the 2018 IEEE 34th International Conference on Data Engineering (ICDE), Paris, France, 16–19 April 2018; pp. 1519–1530. [CrossRef]
43. Murakami, K.; Nagai, K.; Tanimoto, A. Single-Package SSD and Ultra-Small SSD Module Utilizing PCI Express Interface. *Toshiba Rev. Glob. Ed.* **2015**, *1*, 24–27.
44. Zambelli, C.; Micheloni, R.; Crippa, L.; Zuolo, L.; Olivo, P. Impact of the NAND Flash Power Supply on Solid State Drives Reliability and Performance. *IEEE Trans. Device Mater. Reliab.* **2018**, *18*, 247–255. [CrossRef]
45. Intel Corp. Partition Alignment of Intel® SSDs for Achieving Maximum Performance and Endurance. 2014. Available online: https://www.intel.com/content/dam/www/public/us/en/documents/technology-briefs/ssd-partition-alignment-tech-brief.pdf (accessed on 5 May 2020).
46. Yu, W.; Liang, F.; He, X.; Hatcher, W.G.; Lu, C.; Lin, J.; Yang, X. A Survey on the Edge Computing for the Internet of Things. *IEEE Access* **2018**, *6*, 6900–6919. [CrossRef]
47. Apacer Technology Inc. Thermal Throttling. Available online: https://industrial.apacer.com/en-ww/Technology/Thermal-Throttling- (accessed on 5 May 2020).
48. Ferreira, A.P. SMARTER: A Smarter-Device-Manager for Kubernetes on the Edge. 2020. Available online: https://community.arm.com/developer/research/b/articles/posts/a-smarter-device-manager-for-kubernetes-on-the-edge (accessed on 5 May 2020).
49. Intel Corp. Intel® Data Center SSDs: Important SMART Attribute Indicators. Available online: https://www.intel.com/content/www/us/en/support/articles/000055367/memory-and-storage/data-center-ssds.html (accessed on 5 May 2020).
50. Wu, Q.; Dong, G.; Zhang, T. A First Study on Self-Healing Solid-State Drives. In Proceedings of the 2011 3rd IEEE International Memory Workshop (IMW), Monterey, CA, USA, 22–25 May 2011; pp. 1–4. [CrossRef]