MIND: A Black-Box Energy Consumption Model for Disk Arrays

Zhuo Liu[1,2], Jian Zhou[1], Weikuan Yu[2], Fei Wu[1*], Xiao Qin[2], and Changsheng Xie[1]

[1]Wuhan National Laboratory for Optoelectronics
[1]Key Laboratory of Data Storage Systems, Ministry of Education, China
[1]School of Computer Sci. and Tech., Huazhong University of Sci. and Tech., Hubei, Wuhan 430074, China
[2]Department of Computer Sci. & Software Engineering, Auburn University, AL 36830, USA
[*]Corresponding Author: wufei@hust.edu.cn

***Abstract*—Energy consumption is becoming a growing concern in data centers. Many energy-conservation techniques have been proposed to address this problem. However, an integrated method is still needed to evaluate energy efficiency of storage systems and various power conservation techniques. Extensive measurements of different workloads on storage systems are often very time-consuming and require expensive equipments. We have analyzed changing characteristics such as power and performance of stand-alone disks and RAID arrays, and then defined MIND as a black box power model for RAID arrays. MIND is devised to quantitatively measure the power consumption of redundant disk arrays running different workloads in a variety of execution modes. In MIND, we define five modes (idle, standby, and several types of access) and four actions, to precisely characterize power states and changes of RAID arrays. In addition, we develop corresponding metrics for each mode and action, and then integrate the model and a measurement algorithm into a popular trace tool – blktrace. With these features, we are able to run different IO traces on large-scale storage systems with power conservation techniques. Accurate energy consumption and performance statistics are then collected to evaluate energy efficiency of storage system designs and power conservation techniques. Our experiments running both synthetic and real-world workloads on enterprise RAID arrays show that MIND can estimate power consumptions of disk arrays with an error rate less than 2%.**

***Index Terms*—Energy Consumption, Disk Arrays, Black-Box**

I. Introduction

With the rapid development of large-scale data centers, reducing power consumptions has become one of the primary goals in building data servers. As storage systems account for a large proportion of total energy [4], many energy-saving techniques for storage systems spring up with the urgent need for reducing storage energy [6][16][22][23][19][8][11][21][17]. In spite of remarkable progress in the design of energy conservation techniques, evaluation approaches lag far behind. A number of benchmarking and simulation tools are developed to evaluate or mimic the performance of storage and file systems, such as TPC-C [7], IOmeter [13], DiskSim [5], Blktrace [2]. However, how to measure or model the power consumption is still a big problem. To measure the power consumptions of data centers with hundreds of storage nodes, thousands of voltage and current meters should be installed, which will not only bring large expenses of instruments, labor and time, but also affect present architecture and performance of storage systems. Therefore, a fine-tuned power modeling for online storage systems is highly expected. Existent simulation models for power consumptions of storage systems are mainly designed for single disk drive, there is still a lack of an integrated model for RAID array based power-aware storage systems [14]. This problem motivates us to implement a simulation tool to estimate power consumption of energy-efficient disk arrays under various workloads. At the heart of the simulation tool, a power consumption model plays an important role to predict energy dissipation.

In this study, we address the need of a power consumption model and develop accordingly. MIND defines a precisely defined power model for RAID disk arrays. A RAID disk array built with controller and parallel disks is considered as a black box with "Four Types of Actions and Five Types of Mode". Four actions include random read, random write, spin-up and spin-down; Five modes include sleep, idle, multi-random access, sequential access, and cached access, reflecting different energy consumption profiles of disk arrays. To make this analytical model practical and compatible with present benchmarking tools, we also develop state-time and action-count analyzers, which can, respectively, record the length of time spent at each power mode, and action counts of RAID arrays. In doing so, we obtain how much power RAID arrays may consume during every power mode and for every type of action, which together amount to the total power consumption. Taken together, we have validated MIND as an accurate and lightweight modeling tool for power-aware mass storage systems users.

II. Related work

In 2003, Zedlewski *et al.* designed and implemented the Dempsey tool to simulate the power consumption of disk drives by computing seeking, rotation, reads, writes and idle periods separately [20]. And The Drive-Thru tool developed by Peek *et al.* can be used to evaluate power management policies of file systems by replaying file traces [15]. Heath *et*

978-1-4577-1221-0/11/$26.00 @2011 IEEE

al. provided a fine-grained software suite to emulate temperature of server clusters, in which disk power consumption is approximated to be linearly proportional to disk utilization [9]. Allalouf *et al.* presented a power model for disk arrays, which estimates power by transforming front-end workloads to back-end disk workloads [1]. In this model, the requirements of uniform workloads and accurate performance metrics from users limit its wide use. In addition, the dynamic power management (e.g., spin-ups and spin-downs of disks) is not considered. Sivathanu *et al.* modeled performance and energy consumptions for disk array, which considers RAID power as a sum of energy consumed by individual disks, while power consumption of controller's processor and memory is omitted [18]. However, in our observations, for enterprise disk arrays with dedicated hardware RAID controllers, the controller's power is not so minimal that should be omitted. Actually, in our measurement for the Huawei S8000 array with over 200 disk drives and in-house disk array with 8 disk drives, the components excluding disks consume 50% to 60% of the arrays' total power consumptions. Moreover, it's not reasonable to simplify the behaviors of a disk array as an adds-up of individual disks. In addition, modern enterprise disk arrays are equipped with redundant components such as double power supplies and multi-controllers for reliability consideration, making the power consumed by components other than disk drives even significant.

Therefore, it should be reasonable to consider the disk array as a whole or a black box with several predefined actions and statuses in modeling their power consumption and performance. With this kind of encapsulation, the model of disk array is supposed to be more portable and much easier to work with in evaluating energy-efficiency of storage systems. As a result, we present the power model of MIND which takes into account not only disk array different workload modes but also different actions including spin-ups and spin-downs of disks. Compared to the existent models mentioned above, our linear-like black box model is relatively simpler; in addition, the inclusion of RAID controller and dynamic power management issues makes the model more applicable, convincing, and able to accurately characterize the energy profiles of various storage systems.

III. MIND METHODOLOGY

A. Black-Box Energy Consumptions Modeling

We performed the following four steps to develop the MIND energy consumption model.

- First, we started this study by comprehensive power and performance measurements on stand-alone disk drives and several RAID arrays. With our in-house evaluation platform for mass storage systems, abundant measurements were proceeded on different disk arrays, standalone hard disks under various workloads.
- Second, we analyzed the impact of I/O load on performance and energy efficiency of disk arrays and developed a power consumption model for disk arrays. In this model, we considered five modes (i.e., idle, sleep, and three access types) and four kinds of state transitions or instant actions which capture the behaviors of the disk arrays used in our empirical studies. An example for transition curve of power consumption is shown in Figure 1.

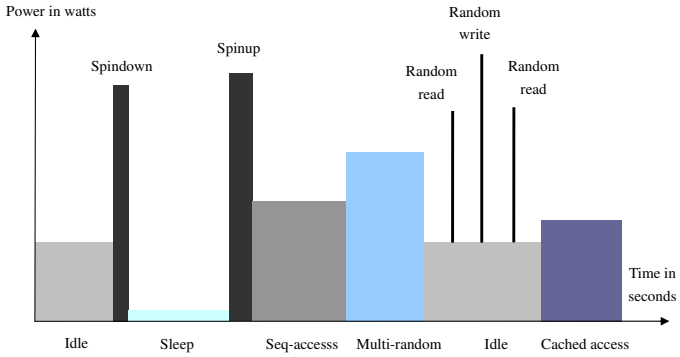

Fig. 1. Power Modeling of Actions and Modes

- Next, with respect to each power mode, we quantified energy consumption of disk arrays as a function of access types and throughput of workloads. Typically, in modes of sequential and multi-random access, the power consumptions of disk arrays demonstrate highly related to their IO throughput in terms of MBPS and IOPS. See more details in next section.
- Finally, we incorporated the model into blktrace - a widely used trace tool. Finally, in light of blktrace coupled with our model, we evaluated energy power consumptions and I/O performance of large-scale storage systems using a wide range of I/O traces thus verifying the correctness of the power consumption model.

By analyzing our power measurement results on standalone disk drives and RAID arrays, we discovered the power changing characteristics with the workloads, and succeeded in constructing a black-box model for a disk drive RAID array, in which the detailed behaviors of individual disks, controllers and other components are transparent to us. Therefore, we can determine the power consumption of RAID based on its state and realtime IO throughput rate. As described below, our power model estimates the energy consumption of a RAID array by recording the time spent on each of the five modes and counts for each of the four actions, and multiplying them by corresponding power values of modes or energy values of actions. The power in each mode is further divided into three parts: one is power consumed by disks (P_{disks}), the second is power consumed by controllers and other components ($P_{controllers}$), the third part is the power that is only related to IO throughput of the whole disk array, which has been taken out from P_{disks} and $P_{controllers}$ and considered independently. In this way, P_{disks}, $P_{controllers}$ are only relevant to the mode that RAID stands and the number of disks on-board; thus, we

get the modeling formulas for the calculation of RAID power consumptions as follows:

$$E_{RAID} = \sum_{i=1}^{5} P_{RAID\,in\,mode\,i} * t_{in\,i} + \sum_{j=1}^{4} E_{action\,j} * counts_{of\,j}$$

$$P_{RAID\,in\,multi-random} = P_{disks\,in\,random} + P_{controllers} + f_r(R)$$

$$P_{RAID\,in\,sequential} = P_{disks\,in\,sequential} + P_{controllers} + f_s(R)$$

$$P_{RAID\,in\,idle} = P_{disks\,in\,idle} + P_{controllers}$$

$$P_{RAID\,in\,cache-acc} = P_{disks\,in\,idle} + P_{controllers\,in\,cache-acc}$$

TABLE I
DESCRIPTIONS FOR THE NOTATIONS IN THE POWER MODEL

Notation	Description
R	Throughput rate (MBPS)
E	Energy consumptions (Joule)
P	Power consumptions (Watt)
t	Time elapse (Second)
P_0	Power consumptions under idle mode
k	Quotient of power changes to throughput rates (Watt/MBPS)

As showed in Table II, four kinds of actions are defined in our power model: random read, random write, spin-down and spin-up. Random read and write are defined to be those sparse and not large IO requests to disk array, typically, average IOPS in that time window is smaller than 1. We consider these statuses to be still idle with additional instantaneous energy jumps. The time durations of random read and write are modeled to be zero. Spin-down and spin-up are actions taken by dynamic power management (DPM) to stop all or part of the disk drives of the array from rotating for energy-conservation, in case there are no IO requests coming for a certain period (e.g., 10 minutes). A spin-down action puts the disk array from idle state to sleep state, and a spin-up action awakes the disk array from sleep state to be active state. The time durations for a disk array spin-down and spin-up range from several to tens of seconds, varying with devices. The basic difference between an action and an status is that the time duration of an action is constant for a given device, while the duration of a status is uncertain. That is why we multiply the power (in watts) of each type of status by time elapse and calculate the energy (in joules) for each type of action by counts.

TABLE II
ENERGY CONSUMPTIONS BY FOUR TYPES OF ACTIONS (IN JOULES)

Name of Action	Energy by controller(in Joules)	Energy of disks (in Joules)
Random read	$E_{c-randomread}$ (Duration=0)	$E_{disks-randomread}$
Random write	$E_{c-radnomwrite}$ (Duration=0)	$E_{disks-radnomwrite}$
Spin down	$E_{c-spindown}$ ($Duration = T_{sipndown}$)	$E_{disks-spindown}$
Spin up	$E_{c-spinup}$ ($Duration = T_{spinup}$)	$E_{disks-spinup}$

As showed in Table III, there are five types of power modes in the MIND model: idle, sleep, sequential access, multi-random access and cached access. Idle mode means the disk array is not serving any request, sleep mode is when the disk array has been spinned-down, and cached access mode indicates that the disk array controller can satisfy the requests from the controller cache without troubling the disks. Sequential access and multi-random access are the two most usual modes when the disk array is working actively. Their power consumptions are functions of the realtime overall throughput - R of disk array in terms of MBPS, so $f_s(R)$ and $f_r(R)$ are added to our model. We differentiate the sequential and multi-random modes by their average request size and MBPS. In each time window, Average request size = Average MBPS / Average IOPS. When the average request size is larger than strip size of the RAID disk array and MBPS is larger than a certain ratio of peak throughput in a time window, the status is defined as a sequential access mode; otherwise, the status is considered as a multi-random access mode. Notice that the throughput is the overall average condition of all concurrent I/O streams with different access patterns.

TABLE III
POWER CONSUMPTIONS IN FIVE TYPES OF MODES(IN WATTS))

Name of Status	Power of controller(in Watts) plus f(R)	Power of disks(in Watts)
Idle	$P_{controller-idle}$	$N * P_{disk-idle}$
Sleep	$P_{controller-sleep}$	$N * P_{disk-sleep}$
Sequential access	$P_{controller-idle} + f_s(R)$	$N * P_{disk-seq}$
Multi-random	$P_{controller-idle} + f_r(R)$	$N * P_{disk-random}$
Cached access	$P_{controller-cache}$	$N * P_{disk-idle}$

With abundant measurements for disk arrays, we observed that the functions of power consumptions to throughput of disk arrays in sequential and multi-random access modes can be approximated as two Piecewise Linear Functions respectively. The reason of piecewise linear could be explained as follows: for each access mode, when IO requests ratio is still relatively low, the power consumption of disk array is linear to the overall throughput with a certain coefficient because more IO requests motivate the disk arrays to do more data transfer and seek operations which consume energy; and when IO requests ratio is relatively high, the disk array has more opportunities to merge IO operations in the request queue which will reduce average disk seek operations for each request, thus lowering the linear coefficient for the second section of the power curve.

k_{r1}, k_{r2}, k_{s1} and k_{s2} are quotients of power changes to throughput rates changes (Watt/MBPS). R_{r0} and R_{s0} are turning points of each piecewise functions, which are to certain ratios of the disk array's peak workloads in terms of MBPS. For example, the peak throughout of our in-house disk array is 100 MBPS, R_{r0} is 10 MBPS, R_{s0} is 30 MBPS. Empirically, k_{r1} is larger than k_{r2} and k_{s1} is larger than k_{s2}. That is to say, for multi-random mode, when the throughout of disk array is relatively small, the power consumption grows fast with

the increasing of throughput; and then after the turning point of R_{r0}, the rake ratio becomes much smaller. For sequential mode, the situation is similar though with different values of rake ratio and turning point.

For multi-random mode,

$$f_r(R) = \begin{cases} k_{r1} * R + c_{r1}, & if\ R < R_{r0}, \\ k_{r2} * R + c_{r2}, & if\ R >= R_{r0}, \end{cases}$$

For sequential mode,

$$f_s(R) = \begin{cases} k_{s1} * R + c_{s1}, & if\ R < R_{s0}, \\ k_{s2} * R + c_{s2}, & if\ R >= R_{s0}, \end{cases}$$

B. Trace-Replay Based Implementation of MIND

To implement the MIND model into real experimental environment, two issues are needed to be solved. One is how to acquire the values of constants and coefficients in the modeling formulas; the other issue is how to acquire the time durations spent in each mode and the IO throughput needed for throughput-aware power calculation, and the counts of every type of action. For the first issue, power consumptions for idle, cached access, sleep modes of disks and controllers, time duration and energy consumptions of spin-down and spin-up actions are obtained by multiple tests and averaging. To acquire the energy consumptions of random read and random write action, we use the way of 2-trace method: $E_{action} = \frac{E_2 - E_1}{n_1 - n_2}$, similar to that used by [20]. Typically, energy consumption of random write depends on the parity type of the RAID, for example, a random write to RAID5 array will cause two disks to proceed seek operations. To acquire the coefficients needed for the $f(R)$ functions, we replay a series of traces that cover various workloads on the disk array and collect their real-time IO throughputs and energy-consumptions. With the acquired results, the approach of least squares adjustment and linear fit is used to work out the coefficients. In practical use, the manufactures of the disk arrays can provide the standard values of these needed coefficients for certain devices, so that users can predict the power consumptions with the MIND model without proceeding the power measurement step.

For the second issue, we integrated the MIND model with a widely used block-level trace tool - blktrace[2]. As displayed in Figure 2, blktrace replays a certain IO trace and generates workloads to disk array. We inserted a module which can collect throughput in terms of IOPS and MBPS in every time window and manage transition event signals with time stamps. All these real-time information are sent to MIND model for further analysis. The size of time window is configurable, and is set to be 5 seconds in our evaluation. The MIND model consists of four major components: judger, throughput accumulator, mode time accumulator and action counter. The throughput accumulator records throughput values of multi-random and sequential access modes in four parts, respectively for coefficients of k_{r1}, k_{r2}, k_{s1}, k_{s2}. The mode time accumulator sums up time for each type of status while the action counter accumulates the counts for each type of action. When a group of information for a time window arrives , the judger will determine which status the array stands and what actions it has done and then update the other three components correspondingly. Therefore, both the real-time power consumption in each time window and the total energy consumption of the disk array can be computed.

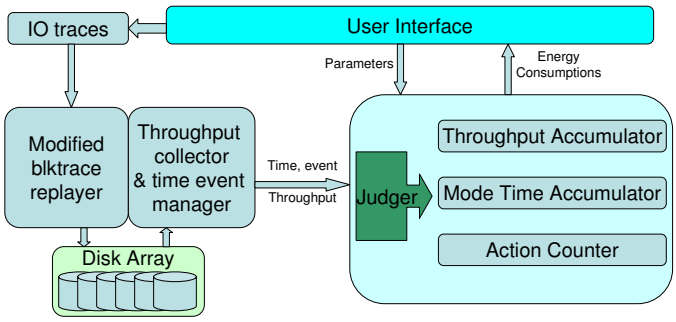

Fig. 2. Implementation of MIND by integration with blktrace

IV. Evaluation

A. Experiment setup

To evaluate our implementation of MIND with blktrace, we setup the prototype on a machine with Intel Dual-Core E5200 2.5GH CPU, 2G Bytes DDR2 memory running RHEL 4 Linux-2.6.18 operating system. The target disk array was our in-house disk array with 6*500G 7200RPM SATA disks on-board. A RAID5 volume with 32KB strip size is setup on the disk array. The disk array and testing platform are connected by 4Gbps FC channel. A ZH-101 power analyzer was used to acquire real-time power consumption by measuring real-time AC current and voltage of the disk array's power line.

In terms of workloads, we replayed 125 traces which were collected on the target disk array when generating synthetic workloads. The 125 traces have different request sizes (512B, 4KB, 16KB, 64KB, 1MB), read/write ratios (0%, 25%, 50%, 75%, 100%) and random/sequential ratios (0%, 25%, 50%, 75%, 100%). The maximum throughput intensity among all the traces on the disk array is about 110 MBPS.

In addition to synthetic traces, we chose a section of cello99 trace file, collected on a HP Unix server [10], in which the read ratio is 58%, average request size is 6 KB. And we also replayed a webserver trace containing web requests for a week on the O4 machine of a web server in the Department of Computer Science, Florida International University[3]. Its average request size is 21.5 KB and data set is 23.31GB.

B. Evaluation Results

For the 125 synthetic traces, we replayed them by adjusting their replay intensity from 10%, 20% to 100% thus acquiring 1250 different workloads [12]. As showed in Figure 3, the over 1200 blue points depict the average power consumptions and throughput of the disk array under each workload by measured. The two lines are the piecewise linear approximations of their power consumptions by MIND modeling, pink line with

squares for sequential access mode and dotted yellow line for multi-random access mode, respectively. As we can see, the two modeling lines fit most of blue power points accurately with the max error rate below 2%. To make the comparison clearer, Figure 4 further shows the fitting result of sequential access mode line to the measuring results of 1MB-0%random-100%read trace. The blue line shows the power consumptions of the disk array when replaying 1MB-0%random-100%read trace under load intensity varying from 10% to 100%.

And Table IV shows the error rate statistics of MIND power estimation for synthetic traces, webserver trace and HP cello99 trace. For the webserver and cello99 trace, the total error rate represents the deviation rate of modeling total energy consumptions (in Joules) to the measured energy consumptions. And the max error rate shows the largest deviation rate of modeling power to the measured power consumptions (in watts) for all time windows in each trace. For the 1250 synthetic tests, because they keep the disk array in an uniform workload during each replaying, the meanings of the statistics are a little different. We use the total error rate to represent the average deviation for all traces, and the max error rate indicates the largest one among them. As we can see, the total error rate of modeling is about 1%, and max error rate is below 2.5% for all situations. This proves that MIND can provide high accuracy for estimating energy consumptions of disk arrays under both synthetic workloads and real-world workloads.

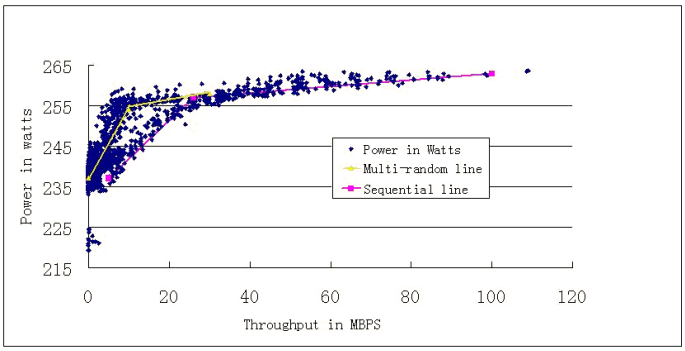

Fig. 3. Modeling lines vs measured points under synthetic workloads

TABLE IV
ERROR RATE STATISTICS OF MIND POWER ESTIMATION

Traces	Total Error Rate	Max Error Rate
Synthetic(in average)	0.85%	2.10%
Webserver	1.12%	2.21%
Cello99	1.30%	2.40%

V. CONCLUSION AND FUTURE WORK

Present disk array modeling techniques are either heavy-weight or hard to implement, we present a power modeling methodology for disk arrays to solve this issue, which refers to the disk array as a black-box object with certain working modes and actions. In addition, we implement this MIND model by integrating a widely used trace tool – blktrace. With this prototype, users who want to evaluate energy efficiency of disk-array based storage systems only need to acquire a limited number of parameters from the disk arrays, and then run trace-based workloads without investing much effort on installing energy measuring equipments. The MIND model can then generate the targeted throughput performance and energy-consumptions. We validate the MIND model by running large amounts of synthetic traces and some real traces. Our experimental results show that MIND can estimate the power consumptions of disk arrays accurately. Therefore, MIND is demonstrated to be helpful for designers and users of power-aware storage systems. Specially, for those who do not have disk arrays in hand, it's pretty easy to run MIND model with simulation tools like DiskSim[5]. Moreover, as energy consumptions for cooling are no longer negligible in storage systems, modeling the relationship of performance, temperature and energy is also of great interest for system practioners and designers.

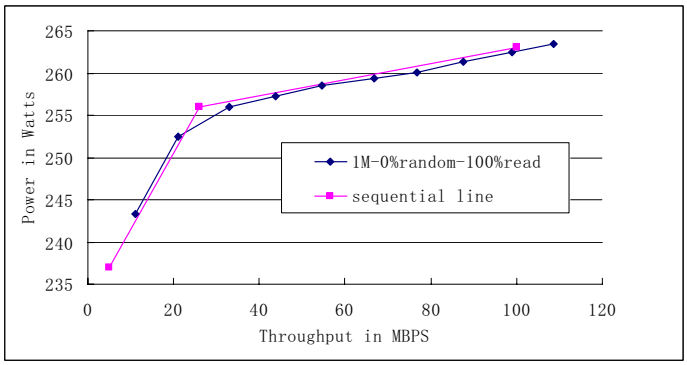

Fig. 4. The sequential modeling line vs measured points under 1MB-0%random-100%read workload

ACKNOWLEDGMENTS

We are very grateful to Lin Cao and Cristi Cira for their helpful comments. This research is sponsored by the National Basic Research 973 Program of China under Grant No. 2011CB302303, the National Natural Science Foundation of China under Grant No. 60933002, and the National 863 Program of China under Grant No. 2009AA01A402, and in part supported by the U.S. National Science Foundation awards CNS-0917137 and CRI-1059376.

REFERENCES

[1] M. Allalouf, Y. Arbitman, M. Factor, R. Kat, K. Meth, and D. Naor. Storage modeling for power estimation. In *Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference*. ACM New York, NY, USA, 2009.

[2] J. Axboe and A. Brunelle. Blktrace User Guide, 2007.

[3] M. Bhadkamkar, J. Guerra, L. Useche, S. Burnett, J. Liptak, R. Rangaswami, and V. Hristidis. BORG: block-reORGanization for self-optimizing storage systems. In *7th USENIX Conference on File and Storage Technologies*, 2009.

[4] R. Bianchini and R. Rajamony. Power and energy management for server systems. *Computer*, 37(11):68–76, 2004.

[5] J. Bucy and G. Ganger. The DiskSim simulation environment version 3.0 reference manual. 2003.

[6] D. Colarelli and D. Grunwald. Massive arrays of idle disks for storage archives. In *Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, pages 1–11. IEEE Computer Society Press Los Alamitos, CA, USA, 2002.

[7] T. P. P. Council. *TPC-C*. http://www.tpc.org/tpcc/, Online.

[8] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke. DRPM: dynamic speed control for power management in server class disks. In *Annual International Symposium on Computer Architecture*, volume 30, pages 169–181. IEEE Computer Society; 1999, 2003.

[9] T. Heath, A. Centeno, P. George, L. Ramos, Y. Jaluria, and R. Bianchini. Mercury and freon: temperature emulation and management for server systems. *ACM SIGARCH Computer Architecture News*, 34(5):106–116, 2006.

[10] H. Labs. *Tools and traces*. http://www.hpl.hp.com/research/ssp/software/, Online, 2007.

[11] D. Li and J. Wang. EERAID: energy efficient redundant and inexpensive disk array. In *Proceedings of the 11th workshop on ACM SIGOPS European workshop*. ACM New York, NY, USA, 2004.

[12] Z. Liu, F. Wu, X. Qin, C. Xie, J. Zhou, and J. Wang. TRACER: A Trace Replay Tool to Evaluate Energy-Efficiency of Mass Storage Systems. In *Proceedings of the IEEE International Conference on Cluster Computing*, 2010.

[13] OSDL. *Iometer project*. http://www.iometer.org/, Online, 2004.

[14] D. Patterson, G. Gibson, and R. Katz. A case for redundant arrays of inexpensive disks (RAID). In *Proceedings of the 1988 ACM SIGMOD international conference on Management of data*, pages 109–116. ACM, 1988.

[15] D. Peek and J. Flinn. Drive-Thru: Fast, Accurate Evaluation of Storage Power Management. In *Proceedings of the Annual USENIX Technical Conference*, 2005.

[16] E. Pinheiro and R. Bianchini. Energy conservation techniques for disk array-based servers. In *Proceedings of the 18th annual International Conference on Supercomputing*, pages 68–78. ACM New York, NY, USA, 2004.

[17] X.-J. Ruan, A. Manzanares, S. Yin, Z.-L. Zong, and X. Qin. Performance Evaluation of Energy-Efficient Parallel I/O Systems with Write Buffer Disks. In *Proc. 38th International Conference on Parallel Processing*, 2009.

[18] S. Sivathanu, L. Liu, and C. Ungureanu. Modeling the performance and energy of storage arrays. In *International Conference on Green Computing*, pages 229–242. IEEE, 2010.

[19] C. Weddle, M. Oldham, J. Qian, A.-I. A. Wang, P. Reiher, and G. Kuenning. Paraid: a gear-shifting power-aware raid. *ACM Transactions on Storage*, 3(3):1–13, 2007.

[20] J. Zedlewski, S. Sobti, N. Garg, F. Zheng, A. Krishnamurthy, and R. Wang. Modeling Hard-Disk Power Consumption. In *2th USENIX Conference on File and Storage Technologies*, 2003.

[21] Q. Zhu, Z. Chen, L. Tan, Y. Zhou, K. Keeton, and J. Wilkes. Hibernator: helping disk arrays sleep through the winter. *ACM SIGOPS Operating Systems Review*, 39(5):177–190, 2005.

[22] Q. Zhu, F. David, C. Devaraj, Z. Li, Y. Zhou, and P. Cao. Reducing energy consumption of disk storage using power-aware cache management. In *Proceedings of the 10th International Symposium on High Performance Computer Architecture*, volume 118. IEEE Computer Society Washington, DC, USA, 2004.

[23] Q. Zhu, A. Shankar, and Y. Zhou. PB-LRU: a self-tuning power aware storage cache replacement algorithm for conserving disk energy. In *Proceedings of the 18th annual international conference on Supercomputing*, pages 79–88. ACM New York, NY, USA, 2004.