International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-2, Issue-7, July 2014

VLSI Power Efficiency, Leakage, Dissipation and Management Techniques: A Survey

Sandeep Kaur, Sarabdeep Singh

Abstract— **Modern processor and controlling systems are using increasingly sized-up on-chip cache memory. With this there has been significant increase in leakage power consumption. This reason, accounts for overall and cache power management research issue in processor design. The System-On-Chip (SoC) revolution challenges both design and test engineers, especially in the area of power dissipation. A circuit or system consumes more power in test mode than in normal mode. This extra power consumption can give rise to severe hazards in circuit reliability or, in some cases, can provoke instant circuit damage. It can create problems such as increased product cost, difficulty in performance verification, reduced autonomy of portable systems, and decrease of overall yield. Technological advances have improved the performance and features of embedded systems development. We also describe sources of power dissipation and leakage in CMOS circuits and there varying degrees of freedom in the low power design. This survey will enable engineers and researchers to get insights into the techniques for improving cache power efficiency, power management techniques, about the available low power testing techniques during testing. and motivate them to invent novel solutions for enabling low-power operation of caches.**

Index Terms— **CMOS, DRAM, HDD, MRU, LRU, LFSR.**

I. INTRODUCTION

As we are entering into an era of green computing, the design of energy efficient IT solutions has become a topic of paramount importance [1]. Recently, the primary objective in chip design has been shifting from achieving highest peak performance to achieving highest performance-energy efficiency. Achieving energy efficiency is important in the design of all range of processors, such as battery-driven portable devices, desktop or server processors to supercomputers. To meet the dual and often conflicting goals of achieving best possible performance and best energy efficiency, several researchers have proposed architectural techniques for different components of the processor, such as processor core, caches, DRAM (dynamic random access memory) etc. For several reasons, managing energy consumption of caches is a crucial issue in modern processor design. With each CMOS (complementary metal oxide semiconductor) technology generation, there is a significant increase in the leakage energy consumption [2], [3]. According to the estimates of International Technology Roadmap for Semiconductors (ITRS); with technology scaling, leakage power consumption will become a major industry crisis, threatening the survival of CMOS technology itself [4]. Further, the number of processor cores on a single chip has greatly increased over years and future chips are expected to have much larger number of cores [5]. Finally, to bridge the gap between the speed of processor and main memory, modern processors are using caches of increasingly larger sizes.

Electronic systems can be viewed as collections of components, which may be heterogeneous in nature. Some components may have mechanical parts, e.g., hard-disk drives (HDD's), or optical parts, e.g., displays. For example, a cellular telephone has a digital very large scale integration (VLSI) component, an analog radio-frequency (RF) component, and a display. Such components may be active at different times, and correspondingly consume different fractions of the telephone power budget. Similarly, main components of portable computers are VLSI chips, HDD, and display. It is often the case that the HDD and the display are the most power-hungry components [6], and thus their effective use is key to achieving long operating times between battery recharges. To be competitive, an electronic design must be able to deliver peak performance when requested. Nevertheless, peak performance is required only during some time intervals. Similarly, system components are not always required to be in the active state. The ability to enable and disable components, as well as of tuning their performance to the workload (e.g., user's requests), is key in achieving energy-efficient designs.

VLSI circuit designers are excited by the prospect of addressing these challenges efficiently, but these challenges are becoming increasingly hard to overcome [7] Test currently ranks among the most expensive and problematic aspects in a circuit design cycle, revealing the ceaseless need for innovative, test-related solutions. As a result, researchers have developed several techniques that enhance a design's testability through DFT modifications and improve the test generation and application processes. Traditionally, test engineers evaluated these techniques according to various parameters: area overhead, fault coverage, test application time, test development effort, and so forth. But now, the recent development of complex, high-performance, low-power devices implemented in deep-submicron technologies creates a new class of more sophisticated electronic products, such as laptops, cellular telephones, audio- and video-based multimedia products, energy efficient desktops, and so forth. This new class of systems makes power management a critical parameter that test engineers cannot ignore during test development. Testing

We believe that this survey will help the researchers and designers in understanding the state-of-the-art in power management of embedded systems and also motivate them to further improve the energy efficiency of embedded systems. In a paper of this length, it is not possible to do justice to the broad range of developments in the field of embedded

Manuscript received July 20, 2014.

Sandeep Kaur, Research Scholar, Chandigarh Engineering College, Landran, Mohali.

Sarabdeep Singh, Asstt. Professor, ECE Department, Chandigarh Engineering College, Landran, Mohali.

systems and hence, we take the following approach to limit the scope of the paper. We include only those research works that propose methods for improving energy efficiency and also evaluate it. Those works which only evaluate performance improvement are not included although they may also lead to better energy efficiency. We review application and architectural level techniques and not circuit-level techniques. Since different techniques have been evaluated using different platforms and methodologies, we only focus on their fundamental research idea and do not present the qualitative results.

A. *Energy and Power Modelling*

Power consumption in CMOS circuits can be static or dynamic. Current used from the power supply causes static power dissipation in the system. Dynamic dissipation occurs during output switching because of short circuit current, and charging and discharging of load capacitance. For existing CMOS technology, dynamic power is the dominant source of power consumption, although this might change for future high-scale integration.

B. *Sources of Power Consumption*

We briefly review the sources of power consumption in embedded systems and refer the reader to previous work [8, 9] for more details. The power consumption of embedded systems can be broadly divided in two categories, namely dynamic power and static power. The dynamic power (P_{dyn}) consumption arises from charging and discharging of the load capacitance, and the short circuit currents. The leakage power (P_{leak}) arises due to leakage currents that flow even when the device is inactive. Thus, we have

$$P_{dyn} = \alpha C V^2 F \qquad (1)$$

$$P_{leak} = I_{leak} V \qquad (2)$$

Here α shows the switching activity, F shows the operating frequency and V shows the operating voltage. I_{leak} shows the leakage current. With CMOS scaling the leakage power is increasing dramatically [7]. DVFS based techniques work by reducing dynamic energy, while the techniques which transition the system to low-power aim to reduce leakage energy. For a given CMOS technology generation, dynamic power consumption can be reduced by adjusting voltage and frequency of operation or by reducing the activity factor. It is clear that, for a given CMOS technology generation, the opportunity of saving leakage energy lies in redesigning the circuit to use low-power cells, reducing the total number of transistors or putting some parts of caches into low (or zero) leakage mode. Based on these essential principles, several architectural techniques have been proposed

C. *Terminology*

Test power is a possible major engineering problem in the future of SoC development. As both the SoC designs and the deep-submicron geometry become prevalent, larger designs, tighter timing constraints, higher operating frequencies, and lower applied voltages all affect the power consumption systems of silicon devices. [4]

D. *Energy*

The total switching activity generated during test application, energy affects the battery lifetime during power up or periodic self-test of battery-operated devices.

E. *Average Power*

Average power is the total distribution of power over a time period. The ratio of energy to test time gives the average power. Elevated average power increases the thermal load that must be vented away from the device under test to prevent structural damage (hot spots) to the silicon, bonding wires, or package.

F. *Instantaneous Power*

Instantaneous power is the value of power consumed at any given instant. Usually, it is defined as the power consumed right after the application of a synchronizing clock signal. Elevated instantaneous power might overload the power distribution systems of the silicon or package, causing brown-out.

G. *Peak Power*

The highest power value at any given instant, peak power determines the component’s thermal and electrical limits and system packaging requirements. If peak power exceeds a certain limit, designers can no longer guarantee that the entire circuit will function correctly. In fact, the time window for defining peak power is related to the chip’s thermal capacity, and forcing this window to one clock period is sometimes just a simplifying assumption. For example, consider a circuit that has peak power consumption during only one cycle but consumes power within the chip’s thermal capacity for all other cycles. In this case, the circuit is not damaged, because the energy consumed which corresponds to the peak power consumption times one cycle will not be enough to elevate the temperature over the chip’s thermal capacity limit (unless the peak power consumption is far higher than normal).

H. *Sources of Power Dissipation*

Power dissipation in digital CMOS circuits is caused by sources such as the leakage current, dependent on the fabrication technology, consists of reverse current in the parasitic diodes between source and drain junction diffusions and the bulk substrate region in a MOS transistor, and sub-threshold current which arises due to inversion charge that exists at the gate voltages which are the threshold voltage, the standby current which is the DC current drawn continuously from Vdd to ground, the short-circuit (rush-through) current which is due to the DC path between the supply rails during output transitions, the capacitance current which flows to charge and discharge capacitive loads during logic changes.

II. POWER MANAGEMENT FOR LIMITED SIZE AND BATTERY

Power management in embedded systems is important for battery-operated mobile embedded system; energy supply is a crucial limitation. Power consumption in systems leads to heating, which should not exist in several domains such as embedded systems. Further, the small size of these systems also limits the amount of heat-dissipation that can be managed. Smaller power consumption enables use of smaller power supplies and reduced heat dissipation overhead, which also reduces the cost, weight and area of embedded systems. Thus power management can lead to easier system design.

III. LEAKAGE ENERGY SAVING APPROACHES

A. An Overview

As explained before, leakage energy saving approaches work by turning off a part of the cache to reduce the leakage energy consumption of the cache. Based on the data retentiveness of turned-off blocks, the leakage energy saving techniques are classified into two broad types, namely state-preserving and state-destroying techniques. The state-preserving techniques turn off a block while preserving its state (e.g., [10], [11]). This means that when the block is reactivated, it does not need to be fetched from next level of memory. The energy saving techniques turn off cache at the granularity (unit) of certain cache space, such as a single way or a single block at a time. Based on this granularity, leakage energy saving techniques can be classified as way-level [12], [13], cache sub-block level [14]. To demonstrate the typical values of the different cache parameters, we take the example of an 8-way set-associative cache of 2MB size with 64B block size and 8-byte sub-block. To achieve high granularity with selective-ways approach requires use of highly-associative caches, which also have high access time and energy. Selective-sets approach can potentially provide large granularity, however, in practice, it is observed that reducing the cache size below 1/8 or 1/16 significantly increases the miss-rate [15], [16]. Since leakage energy varies exponentially with the temperature, an increase in chip temperature increases the leakage energy dissipation in caches, which, in turn, further increases the chip temperature. To take chip temperature into account while modelling and minimizing leakage energy, several techniques have been proposed [17], [18], [19].

For both state-preserving and state-destroying leakage control, architectural techniques make use of some well-known circuit-level mechanisms. Powell et al. [20] propose a circuit design named ‘gated Vdd ’, which facilitates state-destroying leakage control. This technique adds an extra transistor in the supply voltage path or ground path of the SRAM (static random access memory) cell. For reducing the leakage energy of the SRAM cell, this transistor is turned off and by stacking effect of the transistor, the leakage current is reduced by orders of magnitude. For reducing the leakage energy of the SRAM cell, the cache controller switches the operating voltage of the cell to low voltage, thus putting the cell in low-leakage mode. When this line is accessed the next time, the supply voltage is again switched to high, thus the cache-block consumes normal power. Kim et al. [10] propose a “super-drowsy” circuit design and Agarwal et al. [21] propose a gated-ground circuit design, both of which behave similar to the drowsy cache, except that they only require a single voltage supply. Similarly, another state-preserving circuit design, named multi-threshold CMOS (MTCMOS), dynamically changes the threshold voltage of the SRAM cell by modulating the back-gate bias voltage to transition the cell to low-leakage mode.

Mohyuddin et al. [22] propose a technique for saving leakage energy by maintaining different ways of a cache at different state-preserving power saving modes depending on their replacement priorities. Going from the MRU way to the LRU way, cache lines are kept in increasingly aggressive power saving mode which also have increasingly larger penalties of cache line wakeup. To dynamically reconfigure caches using the selective-ways approach, program response for different number of cache ways needs to be estimated. For this purpose, researchers generally utilize utility monitors based on Mattson stack algorithm. Similarly, for utilizing selective-sets approach, researchers generally use set-sampling method and multiple auxiliary tags for getting profiling information.

IV. APPROACHES FOR SAVING BOTH DYNAMIC AND LEAKAGE ENERGY

Several studies present reconfigurable cache architectures which offer flexibility to change one or more parameters of cache. By taking advantage of the flexibility offered by these architectures, both dynamic and leakage energy can be saved. Several researchers have presented techniques for synergistically using both leakage and dynamic energy saving techniques. For example, Giorgi and Bennati [24] demonstrate that using filter cache [23] reduces the number of accesses to L1 cache, which, in turn, enables effectively using leakage energy saving techniques in L1 caches. Similarly, Keramidas et al. [25] propose a way-selection based technique for additionally saving dynamic energy in the caches which use decay-based leakage energy management. Their technique works on the observation that in a cache, using cache-decay mechanism [26] for saving leakage energy, several cache-blocks may be dead. Thus, by making an early determination of these dead blocks, the accesses to these cache blocks can be avoided, which leads to saving of dynamic energy of the cache. Since way-selection mechanism, unlike way-prediction mechanism, gives definite information about a cache miss, it always leads to uniform cache hit latency.

A. Enabling Green Computing

It has been estimated that the ICT (Information and communications technology) contributes nearly 3% in the overall carbon footprint [27]. Thus, power management in embedded systems is also important for achieving the goals of green computing.

B. Using Power Modes

In embedded systems, the hardware typically provides a range of operating modes which can be used to save energy. Different modes consume different amount of power and take different time to return back to the normal mode. In general, the modes with lower energy consumption also take the largest time to return to the normal mode and vice versa. For saving energy while keeping the performance loss bounded, these modes should be judiciously used. Also, while a low-power mode can be used when the system is idle, the system must return to the normal mode for actually servicing a request or performing the task.

Hoeller et al. [28] propose an interface for power management of hardware and software components. They method allows applications to express when certain components are not being used and based on this information, individual components, subsystems or the whole system can be transitioned to low-power modes. This frees the programmer from the task of individually managing the power consumption of each component. Huang et al. [29] propose an energy saving technique which works by adaptively controlling the power mode of the embedded system according to historical arrivals of tasks. Their technique takes decision regarding when to transition the system to low-power from normal-power mode or vice versa, based on the relative time overhead and energy advantage from mode transition and the consideration of meeting the deadlines of the tasks.

Awan et al. [30] Propose an approach for saving energy in embedded systems using multiple low-power modes. Their technique computes the break-even time for each mode using offline analysis. Further, since early completion of high-priority task creates slack, their technique accumulates this task and uses it to save extra leakage energy in lower priority tasks by allowing the device to stay in low-power mode for longer time.

C. *Saving Energy in Specific Components*

Severalresearchers propose micro-architectural techniques for saving energy in specific components of embedded systems. These techniques leverage application properties or variation in workload to dynamically reconfigure the component of the system to save energy. The technique uses software-based RAM compression to increase the effective size of the memory. The memory compression is used only for those applications which may gain benefit in performance or energy from the compression. For such applications, compression of memory data and swapped-out pages is performed in an online manner, thus dynamically adjusting the size of the compressed RAM area.

D. *Problems Induced by Excessive Test Power*

When dealing with high-density systems such as modern ASICs and SoCs, a non-destructive test must satisfy all the power constraints defined in the design phase. In addition to preventing destruction of the CUT, cost, reliability, autonomy, performance-verification, and yield-related issues motivate power consumption minimization during test.[31] The cost constraints of consumer electronic products typically require plastic packages, which impose a tight limitation on power dissipation. Unfortunately, excessive switching activity during test leads to increased current flows in the CUT, making the use of expensive packages for the removal of excessive heat imperative. Moreover, electro migration causes the erosion of conductors and subsequently leads to circuit failure. As the temperature and current density are major factors that determine electro migration rate, the elevated temperature and current density severely decrease CUT reliability. This phenomenon is even more severe in circuits equipped with BIST because such circuits might be tested frequently in, for example, online BIST strategies. Not only the reliability but also the autonomy of battery-powered remote and portable systems suffers from increased activity. Remote system operation occurs mostly in standby mode with almost no power consumption, interrupted by periodic self-tests. Hence, power savings during test mode directly prolong battery lifetime.

V. METHODS FOR POWER TESTING

A. *Low Transition TPGs*

One common technique to reduce test power consumption is the design of low transition TPGs. Most of these techniques modify the design of the LFSR (or other forms of TPGs such as cellular automata) in such a way as to reduce the transitions in the primary inputs of the CUT for test-per-clock BIST or inside the scan-chain for scan-based BIST. An example of the low transition TPG for test-per-clock schemes is the approach presented in [23]. This approach, called DS-LFSR. The proposed design, called low transition random test pattern generator (LT-RTPG), is composed of an LFSR, a k-input AND gate, and a toggle flip-flop T-FF. Some cells of the LFSR are connected with the inputs of the k-input AND gate, the output of the AND gate is connected with the CUT (the T-FF output will not toggle in m-cells will have the same value in most cases. Thus the power while scanning-in a test vector not while scanning out the captured response. Also, in order to get a high fault-coverage, a long test sequence is needed. put of the T-FF, and the output of the T-FF is connected with the scan-chain input Sin). Since the output of the AND gate (input of the T-FF) is 0 in most of the cases, of the clock cycles, and hence the transition probability in the CUT will decrease. The main drawback of this system is that it reduces the average power while scanning-in a test vector not while scanning out the captured response. Also, in order to get a high fault-coverage, a long test sequence is needed.

B. *Test Vectors Reordering*

The test vectors reordering techniques aim to reduce the switching activity by modifying the order in which the test the number of transitions between two consecutive vectors is reduced (i.e. the Hamming distance between two consecutive vectors is minimum), then the WSA will be reduced in the whole CUT [32].

C. *Scan Cells Reordering Techniques*

Another category of techniques used to reduce the power consumption in scan-based BIST is the use of scan-chain cells ordering techniques [33]. Changing the order of the scan cells in each scan-chain can reduce the switching activity, and hence power dissipation, in scan designs. In the case of a deterministic set of test patterns, the best order of cells is the one that gives the best compromise between reducing the transitions in the scan cells both while scanning in test patterns and while scanning out captured responses.

D. *Vector Filtering Techniques*

The test vectors that are generated by TPGs such as LFSRs are pseudorandom vectors. The fault detection capability of these vectors quickly reaches diminishing returns. Hence, after running a sequence of test vectors and detecting many faults, then only a few of the subsequent test vectors can still detect new faults. The vectors that do not detect new faults, but do consume power when applied to the CUT, can be filtered or inhibited from being applied to the CUT [34]. These algorithms, in general, use extra logic (e.g. decoder circuitry). Using prior knowledge of the sequences of test vectors generated by TPGs such as the LFSR, they can prevent some sequences from being transmitted to the CUT by knowing the first and last vectors in this sequence. Thus they reduce the power consumption in the CUT.

E. *Low Power Test Vector Compaction*

In scan-based circuits, in order to reduce the test data volume, compacting techniques are introduced to merge several test cubes. However, compacting test vectors greatly increases the power dissipation (it could be several times higher). Thus, low power test vector compaction techniques have been introduced to minimize the number of test cubes generated by the ATPG tool by merging test cubes that are compatibles in all bit positions under a power constraint [35]. By carefully merging the test cubes in a specific manner, the number of transitions in the scan-chain can be minimized.

F. Scan Architecture Modification

This technique involves modifying the scan architecture by inserting new elements and partitioning the scan-chain into segments. In [36] the scan-chain is partitioned into N segments where only one segment is active at a time. This technique reduces the average power consumption in the CUT, but it will not affect the power will be enabled by using the gated clock trees instead of scan enable signals as was used in the previous technique.

G. Adaptive Shift Power Control technique

To reduce the scan shift power consumption in logic BIST by using highly correlated test stimulus bits among adjacent scan cells, all existing methods only manipulate test stimulus sequences generated by LFSR in various ways and the test responses are ignored completely. Although it has been observed that the Hamming distance between a test stimulus and its captured test response is typically small, the test stimulus of a test pattern is loaded into the scan chains at the same time as the test response of the previous test pattern is unloaded from the scan chains.

VI. INCREASING ENCODING EFFICIENCY OF LFSR RESEEDING-BASED TEST COMPRESSION

Usually, the deterministic test set to be encoded by LFSR reseeding tends to have a biased probability for the logic value 1 or 0 at each primary input. The biased inputs are fixed to the logic value 1 or 0 with some combinational logic, so that the amount of data to be encoded by the LFSR can considerably be reduced. The combinational logic for bit fixing has to set some primary input to the logic value 0 (or 1), if the corresponding probability of the logic value 0 (or 1) is one. Otherwise, the test pattern from the pseudorandom test pattern generator, such as an LFSR is directly applied to the CUT.

VII. CONCLUSION

Driven by continuous innovations in CMOS fabrication technology, recent years have witnessed wide-spread use of multi-core processors and large sized on-chip caches for achieving high performance. However, due to this, total power consumption of processors is rapidly approaching the "power-all" imposed by thermal limitations of cooling solutions and power delivery. Thus, to be able to continue achieving higher performance using technological scaling, managing the power consumption of processors has become a vital necessity. In this paper, we have reviewed several architectural techniques proposed for managing dynamic and leakage power in caches. A qualitative survey on low power testing techniques and its methodology was carried out. While analyzing, all dimensions of power during chip testing was considered as parameters. Low power design requires a rethinking of the conventional design process, where power concerns are often overridden by performance and area considerations. This clearly highlights the need of power management in embedded systems. To cope with these challenges, power management is necessary at all levels, viz. chip-design level, micro architectural level, application level and system level. We believe that our survey will enable researchers and engineers to understand the state-of-the-art in micro architectural techniques for improving cache energy efficiency, motivating them to design novel solutions for addressing the challenges posed by future trends of CMOS fabrication and processor design and in addressing the challenges of power consumption and architecting highly-energy efficient embedded systems of tomorrow.

REFERENCES

[1] S. Murugesan, "Harnessing green IT: Principles and practices," IT professional, vol. 10, no. 1, pp. 24-33, 2008.

[2] S. Borkar, "Design challenges of technology scaling," Micro, IEEE, vol. 19, no. 4, pp. 23 -29, jul. 1999.

[3] G. Gammie, A. Wang, H. Mair, R. Lagerquist, M. Chau, P. Royannez, S. Gururajarao, and U. Ko, "Smartreflex power and performance management technologies for 90 nm, 65 nm, and 45 nm mobile application processors," Proceedings of the IEEE, vol. 98, no. 2, pp.144-159, 2010.

[4] "International technology roadmap for semiconductors (ITRS)," http://www.itrs.net/Links/2011ITRS/2011Chapters/2011ExecSum.pdf, 2011.

[5] S. Borkar, "Thousand core chips: a technology perspective," in 44th annual Design Automation Conference. ACM, 2007, pp. 746-749.

[6] J. Lorch and A. Smith, "Software strategies for portable computer energy management," IEEE Personal Commun., vol. 5, pp. 60-73, June 1998.

[7] S. Wang and S.K. Gupta, "DS-LFSR: A New BISTTPG for Low Heat Dissipation," Proc. Int'l Test Conf. (ITC 97), IEEE Press, Piscataway, N.J.,1997, pp. 848-857.

[8] X. Zhang, K. Roy, and S. Bhawmik, "POWERTEST: A tool for energy conscious weighted random pattern testing ", Proceedings of International Conference on VLSI Design, pp. 416-422, January 1999.

[9] B. Pouya and A. Crouch, "Optimization Trade-offs for Vector Volume and Test Power," Proc. Int'l Test Conf. (ITC 00), IEEE Press, Piscataway, N.J.,2000, pp. 873-881.

[10] N. Kim, K. Flautner, D. Blaauw, and T. Mudge, "Single-VDD and single-VT super-drowsy techniques for low-leakage high-performance instruction caches," in International Symposium on Low power electronics and design (ISLPED), 2004, pp. 54-57.

[11] H. Hanson, M. Hrishikesh, V. Agarwal, S. Keckler, and D. Burger, "Static energy reduction techniques for microprocessor caches," IEEE Transactions on VLSI Systems, vol. 11, no. 3, pp. 303 -313, 2003.

[12] A. Bardine, M. Comparetti, P. Foglia, G. Gabrielli, C. Prete, and P. Stenstr öm, "Leveraging data promotion for low power D-NUCA caches," in 11th EUROMICRO Conference on Digital System Design Architectures, Methods and Tools (DSD). IEEE, 2008, pp. 307-316.

[13] S. Petit, J. Sahuquillo, J. Such, and D. Kaeli, "Exploiting temporal locality in drowsy cache policies," in 2nd conference on Computing frontiers. ACM, 2005, pp. 371-377.

[14] M. A. Z. Alves et al., "Energy savings via dead sub-block prediction," in International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), 2012.

[15] S. Mittal and Z. Zhang, "EnCache: Improving cache energy efficiency using a software-controlled profiling cache," in IEEE International Conference On Electro/Information Technology, Indianapolis, USA, May 2012.

[16] S.-H. Yang, B. Falsafi, M. D. Powell, K. Roy, and T. N. Vijaykumar, "An integrated circuit/architecture approach to reducing leakage in deep-submicron high-performance I-caches," in 7th International Symposium on High-Performance Computer Architecture (HPCA), 2001.

[17] S. Kaxiras, P. Xekalakis, and G. Keramidas, "A simple mechanism to adapt leakage-control policies to temperature," in International Symposium on Low Power Electronics and Design (ISLPED), 2005, pp. 54-59.

[18] L. Yuan, S. Leventhal, J. Gu, and G. Qu, "TALk: A Temperature Aware Leakage Minimization Technique for Real-Time Systems," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 30, no. 10, pp. 1564-1568, 2011.

[19] L. He, W. Liao, and M. Stan, "System level leakage reduction considering the interdependence of temperature and leakage," in Design Automation Conference, 2004, pp. 12-17.

[20] M. Powell, S.-H. Yang, B. Falsafi, K. Roy, and T. Vijaykumar, "Gated Vdd: a circuit technique to reduce leakage in deep-submicron cache

memories," in International Symposium on Low power electronics and design (ISLPED), 2000, pp. 90 - 95.
[21] A. Agarwal, H. Li, and K. Roy, "DRG-cache: a data retention gated-ground cache for low power," in Design Automation Conference, 2002, pp. 473-478.
[22] N. Mohyuddin, R. Bhatti, and M. Dubois, "Controlling leakage power with the replacement policy in slumberous caches," in 2nd conference on Computing frontiers. ACM, 2005, pp. 161-170.
[23] J. Kin, M. Gupta, and W. Mangione-Smith, "The filter cache: an energy efficient memory structure," in 30th International symposium on Microarchitecture (MICRO), 1997, pp. 184-193.
[24] R. Giorgi and P. Bennati, "Reducing leakage in power-saving capable caches for embedded systems by using a filter cache," in workshop on MEmory performance: DEaling with Applications, systems and architecture, 2007, pp. 97-104.
[25] G. Keramidas, P. Xekalakis, and S. Kaxiras, "Applying decay to reduce dynamic power in set-associative caches," High Performance Embedded Architectures and Compilers, pp. 38-53, 2007.
[26] S. Kaxiras, Z. Hu, and M. Martonosi, "Cache decay: exploiting generational behavior to reduce cache leakage power," in 28th international symposium on Computer architecture (ISCA), 2001, pp. 240-251.
[27] L. Smarr, "Project greenlight: Optimizing cyber- infrastructure for a carbon-constrained world," Computer, vol. 43, no. 1, pp. 22-27, 2010.
[28] A. Hoeller, L. Wanner, and A. Fröhlich, "A hierarchical approach for power management on mobile embedded systems," From Model-Driven Design to Resource Management for Distributed Embedded Systems, pp. 265-274, 2006.
[29] K. Huang, L. Santinelli, J.-J. Chen, L. Thiele, and G. C. Buttazzo, "Adaptive power management for realtime event streams," in Asia and South Pacific Design Automation Conference, pp. 7-12, 2010.
[30] M. A. Awan and S. M. Petters, "Enhanced race-to-halt: A leakage-aware energy management approach for dynamic priority systems," in 23rd Euromicro Conference on Real-Time Systems (ECRTS), pp. 92- 101, IEEE, 2011.
[31] X. Zhang, K. Roy, and S. Bhawmik, "POWERTEST: A tool for energy conscious weighted random pattern testing ", Proceedings of International Conference on VLSI Design, pp. 416-422, January 1999.
[32] S. Chakravarty and V. Dabholkar, "Two techniques for minimizing power dissipation in scan circuits during test application", Proceedings of Asian Test Symposium, pp. 324-329, November 1994.
[33] S. Ghosh, S. Basu, and N. Touba, "Jointminimization of power and area in scan testing by scan cell reordering", IEEE Computer Society Annual Symposium on VLSI, pp. 246-249, February 2003.
[34] S. Manich, A. Gabarro, M. Lopez, J. Figueras, P. Girard, L. Guiller, C. Landrault, S. Pravossoudovitch, P. Teixeira, and M. Santos, "Low power BIST by filtering non-detecting vectors", Journal of Electronic Testing: Theory and Application (JETTA), 16(3), pp. 193-202, June 2000
[35] P.-H. Wu, T.-T. Chen, W.-L. Li, and J.-C. Rau, "An efficient test-data compaction for low power VLSI testing", IEEE International Conference on Electro/Information Technology, pp. 237-241, May 2008.
[36] L. Whetsel, "Adapting scan architectures for low power operation ",Proceedings of International Test Conference, pp. 863-872, October 2000.