Techniques of Low Power Digital Design: A Survey

DALIA EL-DIB
Dalhousie University
Department of Electrical and Computer Engineering
Halifax, NS
CANADA
dafeldib@alumni.uwaterloo.ca

Abstract: For two decades, low power/energy design has been a major design constraint. The explosion in digital communications and the desire to preserve battery life time, improve system reliability, and reduce cooling costs has pushed for extensive research in low power/energy digital design. In this paper low energy versus low power will be discussed. Then the basics of low power design trends, major techniques and recent challenges will be demonstrated and discussed.

Key–Words: Digital Design, low power, energy efficient

1 Introduction

In predicting the future of integrated electronics, Gordon Moore predicted in 1965 that the number of components per chip will double every year in the period till 1975 [1] reaching 65,000 components on a single quarter-inch semiconductor. In 1975, Moore reduced the rate to a doubling every two years due to integrating more microprocessors which are in general less dense in electronic circuits [2]. In 1995, Moore's stated that his projection is not going to stop soon [3]. In fact, Moore's projection (rule) was considered one of the driving forces of electronics industry. It challenged technologists to deliver annual breakthrough in manufacturing Integrated Circuits (ICs) to comply with Moore's law. The transistor density is defined as the transistor number on one square silicon die that is generally considered the largest manufacturable die [4]. Such a die was able to hold over 7 billion transistors in 2014. Thus, Moore's law worked perfectly and was continuously fulfilled and has caused many of the most important changes in the electronics manufacturing technology.

Since the 1970s, the most dominant electronics manufacturing technologies used were bipolar and nMOS transistors [5]. However, these consume non-negligible power even in static (non-switching) state. Consequently, by 1980s, the power consumption of bipolar designs and its cooling solution costs were considered too high to be sustainable. This caused an inevitable switch to a slower, but lower-power Complementary Metal Oxide Semiconductor (CMOS) technology (Figure 1).

At that time, CMOS transistors consumed lower power largely because static (leakage) power was neg-ligible if compared to dynamic (switching) power. Along with fulfilling Moore's law, the aim is always to increase processing power of electronic circuits. This is achieved by scaling down the technology, increasing the number of components per chip, and increasing the frequency of operation as shown in Figure 2.

Additionally, to satisfy Moores law, die size has been increasing at the rate of 7% per year, the operating frequency has doubled every two years, However to meet the performance goal, the supply voltage scaled down by only 15% every two years, rather than the theoretical 30% causing reduction in power consumption not tracking Moores Law [4]. Therefore, a direct consequence of Moore's law is the exponential increase in power density with every technology generation. In the late 2004 with scaling down the CMOS fabrication technology to 45 nm and downwards, we encountered an exponential increase in leakage power to the extent that it is comparable to dynamic power and can even dominate the overall power dissipation. Also, with integrating more and more components, the power increases dramatically which causes a challenge regarding excessive thermal dissipation. This challenge along with other physical, material, and economical challenges are bringing the CMOS scaling along with Moore's law to an end [7]. Thus, another paradigm shift in computing electronics was inevitable, namely the shift to multi-core computing as shown in Figure 3. This was in the aim to increase performance while keeping the hardware simple, retain acceptable power consumption levels and transfer complexity to higher layers of the system design abstraction, including software layers [5].

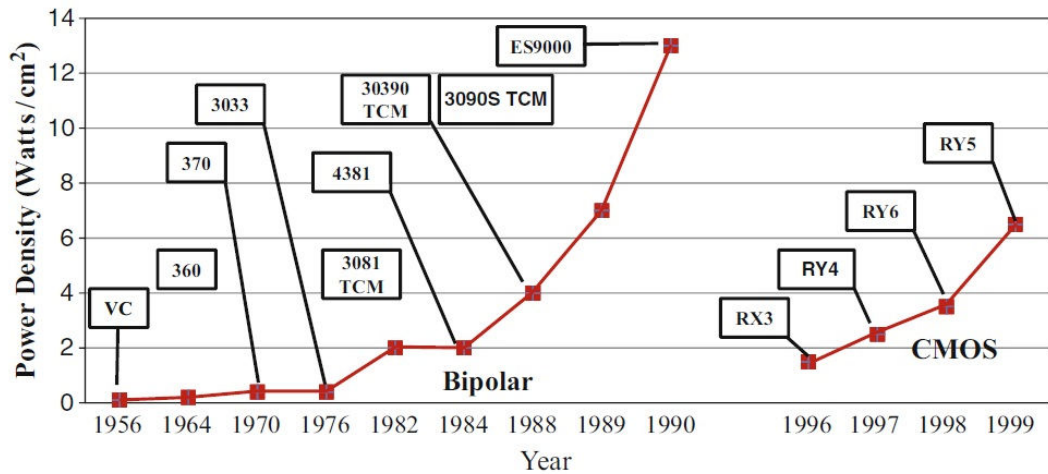This paper is organized as follows: Section 2 out-

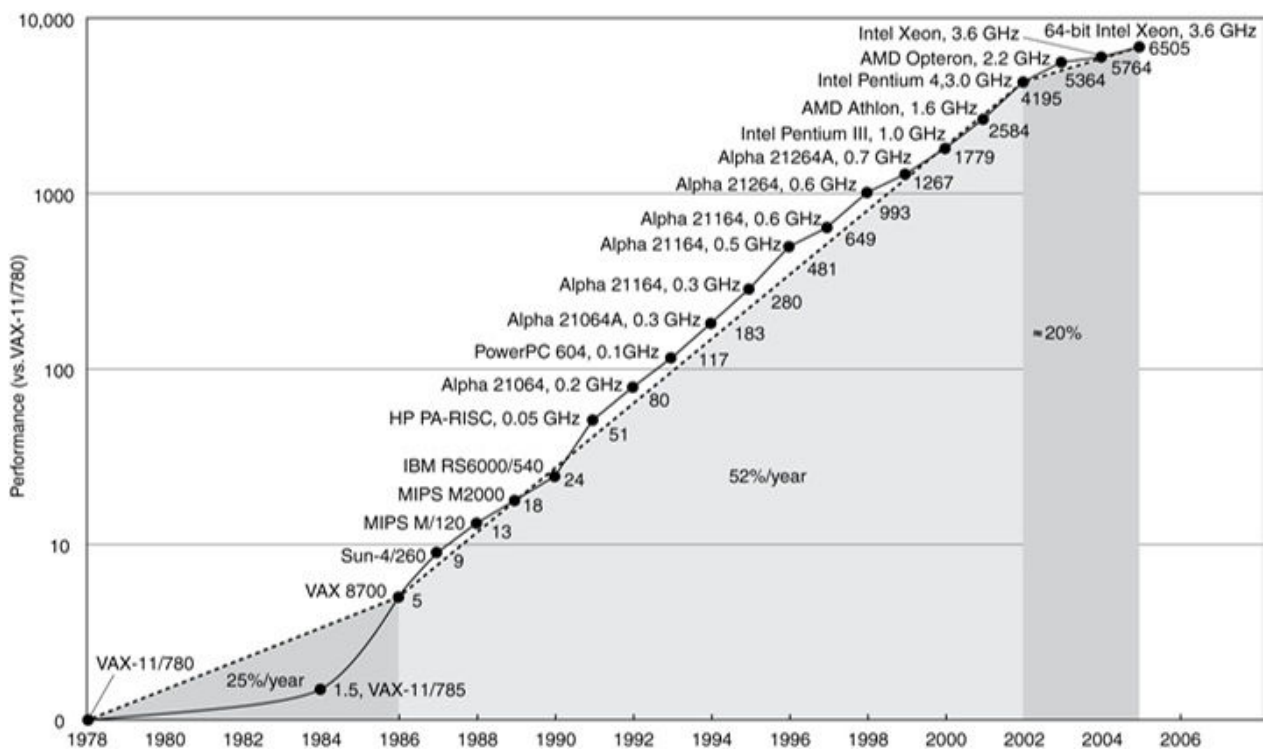Figure 1: Power Density Trend in IBM processors. (see Ref. [5])

Figure 2: Uniprocessor Performance. (see slides of Ref. [6])

lines the difference between the two terms of low power and low energy. Then, section 3 describes the different components of the CMOS power dissipation. Section 4 outlines the different low power digital design techniques at all abstraction level. Finally, section 5 presents the conclusions.

2 Power versus Energy Consumption

Both terms low energy and low power are exchangeably used although energy consumption is different from power consumption. For example, a specific task needs a specified amount of energy E to
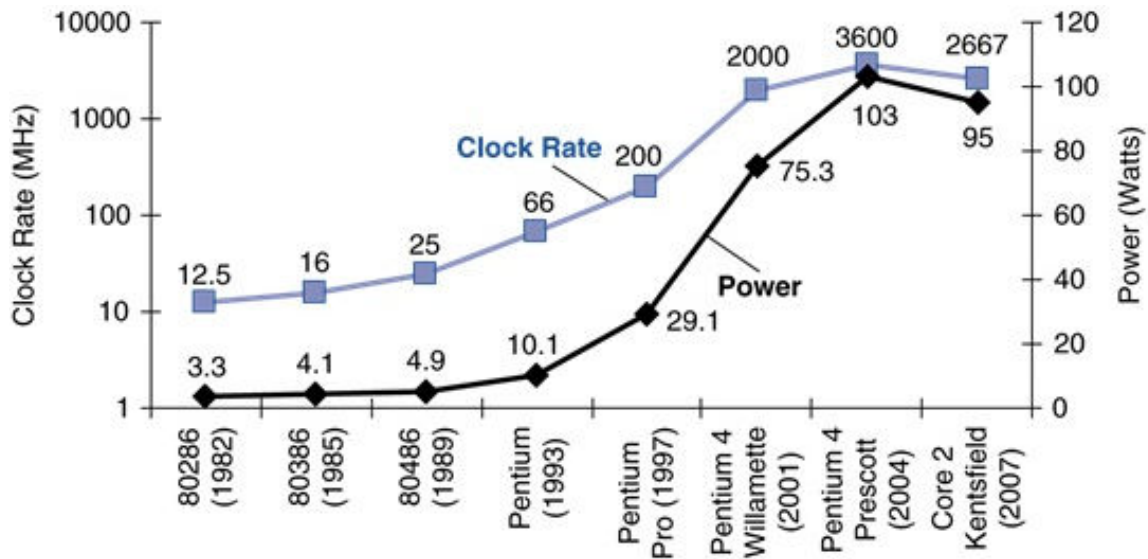
Figure 3: Clock Rate Versus Power. (see slides of Ref. [6])

complete over time T. Its power consumption P is the rate at which energy is consumed (E/T). The time needed to complete the task can be increased by reducing the frequency of operation for example. Whereas, the same amount of energy is still needed to complete the task. Thus, the power consumption is reduced, however the energy consumption (area under the graph) is still the same as shown in Figure 4.
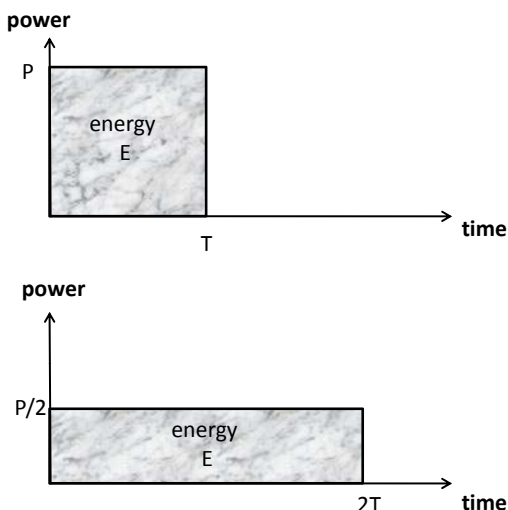

Figure 4: Power versus Energy.

Consequently, low power design is not equivalent to low energy design. Even motivation for both can be different. Yet, motivation for both low power design and low energy design are usually hard to distin-guish and are best combined under one title. In re-search, sometimes low energy is the main motivation for the work, but the research appears under the title low power. So, it is a commonly accepted wrong in-terchanging of terms used in research. For the rest of the paper only the term low power will be used al-though (low energy is sometimes meant) unless stated differently in the research cited.

2.1 Motivation for low power design

Motivations for low power design actually vary ac-cording to the application or device used and can be summarized as follows:

○ Huge data centers are working around the clock to support billions of digital transactions around the world like those of google, facebook, amazon etc.. . Data centers are composed of millions of servers and consume a vast amount of energy in operation and in cooling. It is highly desired to reduce the power/energy cost associated with the explosion in digital information.

○ The increasing demand for high performing and complex portable systems in the area of communica-tion, computation and comsumer electronics aggra-vated the demand for low power design to guarantee longer battery-life time [8]. In fact, prolonging bat-tery life time of battery operated hand held devices is the main concern of most consumers. However, battery capacity doubles only every decade whereas processing capacity doubles every other year [9].

- Even, light and small wearable implantable battery operated systems require long battery operation time to increase time before replacing with a new device [4].

- In battery-less systems where energy is harvested from the environment, like the chips in various sensor wireless networks, energy available is limited. Thus, low power design is needed to be able to operate within the limits of the energy harvested from the environment. These systems will be widely used in the Internet of thing (IoT) such as smart grid [4].

- The limitation on peak power (maximum allowable power dissipation) is a driving motivation for low power. Excessive heat caused by increased power consumption reduces the system reliability and signal integrity.

- The cost associated with packaging and cooling high performing ICs such as microprocessors to get rid of excess heat caused by increased power consumption is becoming too expensive.

3 Power Dissipation Components

Recognizing power saving techniques can be achieved only after identifying the dominant sources of power dissipation and where power is dissipated in digital designs. An IC's power consumption is mainly composed of static power and dynamic power.

$$P_{total} = P_{dynamic} + P_{static} \tag{1}$$

Dynamic power consumption is frequency-dependent and results from one of the following three sources: Switching power, short circuit power and glitching power [10]. Switching power is consumed during the charging and discharging of capacitive nodes (see Figure 5). Short circuit power occurs during the momentary current flow that occurs when two complementary transistors conduct during a logic transition, which arises from long rise or fall times of input signals. (see Figure 6). Glitching power occurs due to the finite delay of the logic gates which cause spurious transitions at different nodes in the circuit. The dominant part of the dynamic power is the switching power which can be represented with the following equation:

$$P_{dynamic} = \alpha C_L V_{dd}^2 f. \tag{2}$$

Where α is the switching activity of the circuit, C_L is the effective capacitance of the circuit, V_{dd} is the supply voltage (or rail to rail voltage) and f is the operating frequency.

Static power typically comes from leakage current and DC current sources. Static power consumption has many components and has many paths as
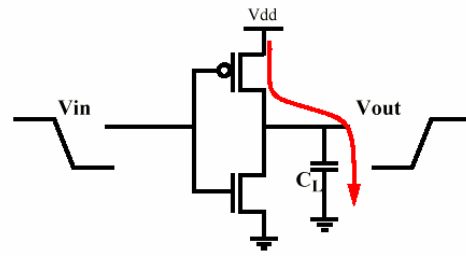

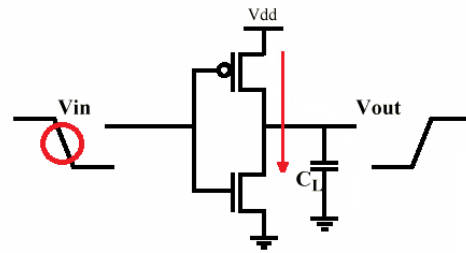
Figure 5: Switching Power

Figure 6: Short Circuit Power

shown in Figure 7. The most important contributor to static power in CMOS is the subthreshold leakage which is exponentially dependant on $(V_{gs} - V_T)$, where V_{gs} is the gate to source voltage and V_T is the threshold voltage. Another part of leakage is caused by reduced gate oxide thickness t_{ox} which increases gate oxide tunneling current. All parts of leakage current are increased excessively due to scaling down of technology which necessitates reducing V_T and t_{ox} to keep up with higher processing requirements.
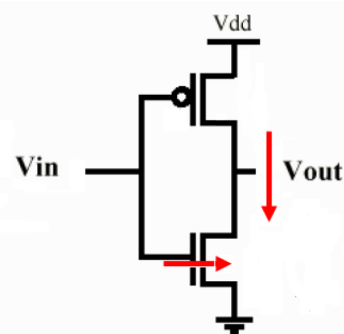

Figure 7: Leakage Power

4 Low Power Techniques

There is no single solution for low power design. Low power design methodologies and techniques must be applied at all design abstraction levels including system, algorithm, architecture, logic, circuit, device and technology levels. A compromised combination

of techniques at many levels can result in order of magnitude of power reduction. However, some of the low power techniques reduce power in exchange of reduced performance, compromised reliability, increased chip area, or a combination of these. Eventually, one has to reach a compromise between power, performance, and cost to satisfy overall design requirements. Any modification at a higher level design abstraction will definitely affect all subsequent design abstraction levels to comply with the changes at that higher level. So any technique suggested at a high level can be included in all subsequent levels. However, to avoid repetition, each technique will be mentioned only at the highest level it can be applied to.

4.1 System-level low power design techniques

4.1.1 Memory Centric Design

The ever increasing performance requirements of digital signal processing systems causes the capacity of memory integrated in systems to become larger; this increases the power consumption of such systems. It is suggested to incorporate the entire memory hierarchy into system optimization. Thus, many domain-specific approaches to improve the system energy efficiency were proposed [11].

For example, the design of motion estimation accelerator under a 3D logic-DRAM integrated heterogeneous multi-core system framework is studied to replace off-chip commodity DRAM access. Also, a joint source-channel coding and channelization for embedded video processing with the abundant Flash memory storage is suggested [11].

4.1.2 Multi-V_{th}

The threshold voltage V_{th} must be scaled down to maintain a high driving current and improve performance. However, with the scaling-down of V_{th}, sub-threshold leakage current increases exponentially. Thus, high V_{th} devices are used on non-critical paths to reduce static leakage power without incurring a delay penalty and low V_{th} are used on timing critical paths [12].

4.1.3 Variable Frequency and Clock-Gating

Higher operating frequencies generally translate to higher power dissipation. According to trade offs of speed versus power, the frequency of the design which meets the power constraint can be chosen. In this case variable frequencies are enabled according to current

needs. Frequencies can even be cut off parts of the design in idle mode using clock-gaters [13].

4.1.4 Dynamic Voltage Scaling (DVS)

In DVS systems the performance level (voltage level and/or frequency of operation) is reduced during periods of low utilization. In this case, the system performs the tasks in a longer time when high performance is not required.Thereby, reducing power consumption if careful limits for the upper and lower limits of voltage/frequency are chosen [14, 15]. However, DVS can cause performance loss due to system overhead [16].

4.1.5 Multi-V_{dd} and voltage island design

Multi-V_{dd} is an effective method to reduce both leakage and dynamic power, by assigning different supply voltages to cells according to their timing criticality [17]. In a multi-V_{dd} design, cells of different supply voltages are often grouped into a small number of voltage islands (each having a single supply voltage), in order to avoid complex power supply system and excessive amount of level shifters [18]

4.1.6 Power Gating

Power-gating turns off V_{dd}/V_{ss} to unused blocks temporarily to mitigate leakage power. It cuts off the devices from their power or ground sources, thus, reducing the leakage current flow by creating a break in the power path to unused portions of the design when possible [19].

4.2 Architecture and RTL level low power design techniques

4.2.1 Parallelism: multi-core processing

Mere parallelism was historically used for increasing performance at the expense of larger area and higher power consumption. However, parallelism can be used to reduce power consumption while pertaining the same throughput. Figure 8 presents a four core multiplier architecture to replace a one multiplier.

Increasing the number of cores to 4 can be associated with reducing the frequency to the quarter. while operating at quarter of the original frequency, the supply voltage can be reduced, thereby, reducing the power consumption dramatically as suggested in Table 1.

The estimated reduction in dynamic power consumption while keeping the same performance is as
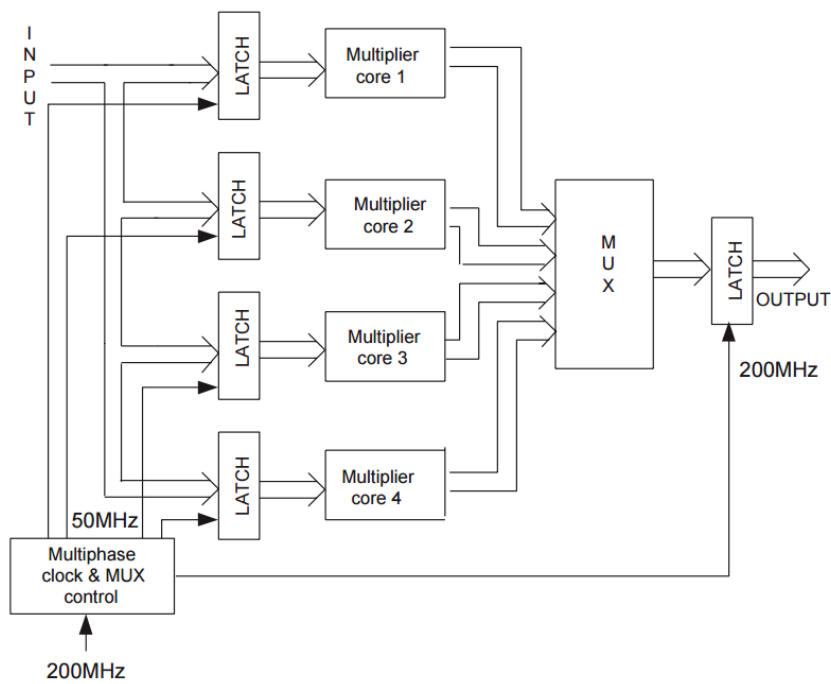
Figure 8: A four core multiplier architecture [10]

Table 1: Power in multicore architectures.

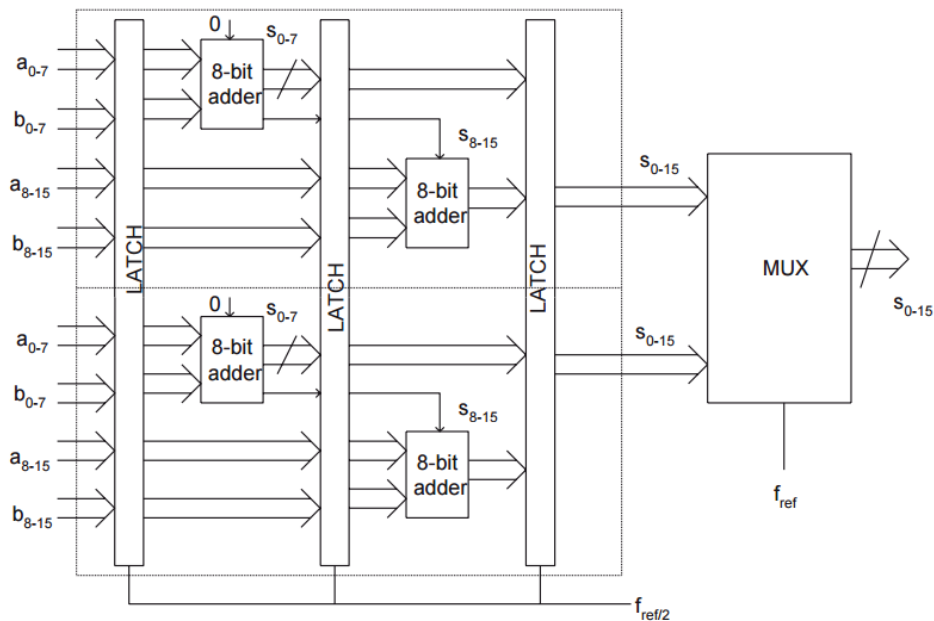| Number of cores	Clock(MHz)	Supply Voltage	Total Power
1	200	5	15
2	100	3.6	8.94
4	50	2.7	5.2
8	25	2.1	4.5

Figure 9: Parallel-pipelined realization of 16-bit adder [10]

follows:

$$P_{dyn\_Parallel} = \alpha(4C_L)(\frac{2.1}{5}V_{dd})^2(\frac{f}{4}) \quad (3)$$
$$= 0.1764\alpha P_{dynamic},$$

where $P_{dyn\_Parallel}$ is the estimated reduced dynamic power after applying parallelism.

4.2.2 Combining Parallelism with Pipe lining

Combining parallelism with pipe lining can reduce dynamic power dramatically if the same performance is required. The architecture of a 16-bit adder by combining both pipe lining and parallelism is shown in Figure 9.

In this case the effective capacitance is increased, however the speed can be reduced along with the supply voltage. The estimated reduction in dynamic power consumption while keeping the same performance is as follows:

$$P_{dyn\_Parallel\_Pipe} = \alpha(2.5C_L)(0.3V_{dd})^2(\frac{f}{2}) \quad (4)$$
$$= 0.1125\alpha P_{dynamic},$$

where $P_{dyn\_Parallel\_Pipe}$ is the estimated reduced dynamic power after applying both parallelism and pipelining techniques to reduce power consumption without degrading performance. Power reduction after applying pipe lining is very close to that after applying parallelism only, but throughput is increased using pipe lining thereby improving performance as well.

4.2.3 Operations Reduction

The number of operations to perform a task is reduced to reduce the associated power consumption. For example, using coding (gray encoding or one-hot encoding) over data transferred can reduce switching activity, using operand decomposition can reduce the power requirements of multiplication [20], modifying the coding algorithm can reduce the number of memory accesses [21], etc...

4.2.4 Asynchronous Design

A clock signal in synchronous designs consumes power even when the circuit is idle, but asynchronous circuits by default move into the idle state and involve no transition in the circuit during that state [22]. In addition, in an active asynchronous system, only the subsystem that is in use dissipates dynamic power. Thus, dynamic power can be reduced by reducing the dependency of the clock signal in the design by opting for asynchronous logic. There exist many asynchronous

design techniques, a performance comparison is presented by Joshi [23].

4.3 Device-level low power design techniques

4.3.1 3D IC Design

To avoid the problems associated with long interconnects including, higher power consumption, larger delays and smaller bandwidths, stacking of dies was suggested. Then, interconnecting the dies is done using vertical VIAs. This would result in much shorter interconnects reducing all disadvantages of long interconnects [24]. Recently four-tier Monolithic 3D ICs wer suggested [25].

4.3.2 FinFets

Multi-gate or tri-gate architectures, also known as FinFET technology (or non-planar CMOS gates) is a promising option to increase performance while maintaining the power consumption or to reduce the power consumption while maintaining the performance. This can help in keeping up with Moore's law [26, 27, 28, 29]. However, it is still facing many design challenges and is a very hot research area.

4.3.3 Nanoelectromechanical-System Switches(NEMS)

A NEMS is a CMOS-compatible mechanical relay with near-infinite OFF-resistance and low ON-resistance. Hybrid NEMS-CMOS technology takes advantage of both near-zero-leakage characteristics of NEMS devices along with high ON current of CMOS transistors [30]. For example, NEMS can be used for power-gating. Here, a NEMS switch completely eliminates OFF-state leakage, yet is compact enough to be contained on die [19]. Many other designs using NEMS to reduce static power consumption of ICs are being suggested [31, 32, 33] and a lot of research is ongoing.

Whether designing digital designs using an Electronic Design Automation (EDA) tool or using a Hardware Description Language (HDL) many of the above techniques can be applied simultaneously. However, a few extra precautions can be taken when designing using a HDL [34]. Some of them are listed below.

4.4 HDL low power design techniques

Improper tool usage or HDL coding can result in unnecessary power consumption, here are a few precautions to follow to attain low power when using a HDL.

○ Reducing data transition on a bus can be attained by assigning a default constant value to the bus instead of transitioning from a value to another when not needed. This might induce some registers, but the associated power reduction can be worth it.

○ Resource sharing can be attained by writing the code in a manner avoiding code redundancy as it might induce redundant hardware.

○ Making sure to include start and stop conditions for counters such that they do not keep counting unnecessarily.

○ Clock Gating can be achieved by using specific available FPGA Hardware and writing the corresponding HDL code [35].

○ Careful placement of logic block on the FPGA chip can create power aware designs [36].

○ Allowing automatic synthesis tool power optimization and carefully analyzing synthesis reports to identify spots where power consumption can be reduced.

5   Conclusion

There are more mobiles than there are humans. Mobiles comprise the dominant part of our Information-Communication-Technologies (ICT). ICT is mainly composed of our computers, smart phones and digital TVs along with its supporting computer server farms which support the cloud. In 2013, it was estimated that our ICT consumes 1,500 tera Watt-hours of energy per year which approaches 10% of the world's electricity with a huge impact on economy and environment [37]. So, the real cost of computation in our exploding ICT world is simply the cost of power consumption. Making this cost cheaper is very important and it seems that all the above low power digital design techniques are still not enough. We need to think creatively to deliver breakthrough techniques of low power digital design.

References:

[1] G. Moore, "Cramming more components onto integrated circuits," *Electronics Maganize*, vol. 38, no. 8, pp. 52–59, 1965.

[2] G. E. Moore *et al.*, "Progress in digital integrated electronics," in *Electron Devices Meeting*, vol. 21, 1975, pp. 11–13.

[3] G. E. Moore, "Lithography and the future of moore's law," in *SPIE's 1995 Symposium on Mi-crolithography*. International Society for Optics and Photonics, 1995, pp. 2–17.

[4] N. N. Tan, D. Li, and Z. Wang, *Ultra-low power integrated circuit design*. Springer, 2014, vol. 1801466741.

[5] P. R. Panda, B. Silpa, A. Shrivastava, and K. Gummidipudi, *Power-efficient system design*. Springer Science & Business Media, 2010.

[6] D. A. Patterson and J. L. Hennessy, *Computer organization and design: the hardware/software interface*. Newnes, 2013.

[7] N. Z. Haron and S. Hamdioui, "Why is cmos scaling coming to an end?" in *2008 3rd International Design and Test Workshop*. IEEE, 2008, pp. 98–103.

[8] J. M. Rabaey and M. Pedram, *Low power design methodologies*. Springer Science & Business Media, 2012, vol. 336.

[9] J. Wu, Y.-L. Shen, K. Reinhardt, H. Szu, and B. Dong, "A nanotechnology enhancement to moore's law," *Applied Computational Intelligence and Soft Computing*, vol. 2013, p. 2, 2013.

[10] A. Pal, *Low-Power VLSI Circuits and Systems*. Springer, 2014.

[11] Y. Li, "Memory-centric low power digital system design," Ph.D. dissertation, Rensselaer Polytechnic Institute, 2012.

[12] M. Anis and M. Elmasry, *Multi-threshold CMOS digital circuits–managing leakage power*. Springer, 2003, vol. 3.

[13] R. Chadha and J. Bhasker, *An ASIC Low Power Primer: Analysis, Techniques and Specification*. Springer Science & Business Media, 2012.

[14] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and practical limits of dynamic voltage scaling," in *Proceedings of the 41st annual Design Automation Conference*. ACM, 2004, pp. 868–873.

[15] J. Kim, F. Gruian, and D. Shin, "Dynamic voltage scaling for low-power hard real-time systems," in *The VLSI handbook*, W.-K. Chen, Ed. CRC press, 2016, ch. 18.

[16] V. Sundriyal and M. Sosonkina, "Runtime power-aware energy-saving scheme for parallel applications," *Computer Science Technical Reports, Iowa State University*, 2015.

[17] M. D. Wong, "Low power design with multi-vdd and voltage islands," in *2007 7th International Conference on ASIC*. IEEE, 2007, pp. 1325–1325.

[18] ——, "A low power design methodology with multi -vdd and voltage islands," Ph.D. dissertation, University of California, 2007.

[19] M. B. Henry, "Emerging power-gating techniques for low power digital circuits," Ph.D. dissertation, Virginia Tech, 2011.

[20] Z. Abid, D. A. El-Dib, and R. Mudassir, "Modified operand decomposition multiplication for high performance parallel multipliers," *Journal of Circuits, Systems and Computers*, p. 1650149, 2016.

[21] D. A. El-Dib and M. I. Elmasry, "Modified register-exchange viterbi decoder for low-power wireless communications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 2, pp. 371–378, 2004.

[22] A. Jayasekar and S. Vimalraj, "Low power digital design using asynchronous logic," Master's thesis, San Jos State University, 2011.

[23] M. Joshi and R. Patel, "Performance comparison of different asynchronous design methodologies," *Digital Signal Processing*, vol. 8, no. 5, pp. 130–134, 2016.

[24] S. K. Lim, *Design for high performance, low power, and reliable 3D integrated circuits*. Springer Science & Business Media, 2012.

[25] K. M. Kim, S. Sinha, B. Cline, G. Yeric, and S. K. Lim, "Four-tier monolithic 3d ics: Tier partitioning methodology and power benefit study," in *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. ACM, 2016, pp. 70–75.

[26] K. Pathak and G. T. Arasu, "Design and characterization of shorted gate finfet for low power circuits," in *Electrical, Electronics, Signals, Communication and Optimization (EESCO), 2015 International Conference on*. IEEE, 2015, pp. 1–4.

[27] S. Ferwani, S. Khandelwal, and R. Shrivastava, "Low power finfet based operational amplifier with improved gain at 45 nm technology regime," *Journal of Nanoelectronics and Optoelectronics*, vol. 11, no. 3, pp. 377–381, 2016.

[28] V. S. Kumar, S. Saravanan, P. Deepa, and S. Priyanka, "Design and implementation of low power finfets using adiabatic logic," *Middle-East Journal of Scientific Research 24 (Techniques and Algorithms in Emerging Technologies)*, 2016.

[29] N. Gupta, A. Makosiej, A. Vladimirescu, A. Amara, and C. Anghel, "Ultra-low-power compact tfet flip-flop design for high-performance low-voltage applications," in *2016 17th International Symposium on Quality Electronic Design (ISQED)*. IEEE, 2016, pp. 107–112.

[30] H. F. Dadgour and K. Banerjee, "Hybrid nems-cmos integrated circuits: A novel strategy for energy-efficient designs," *IET computers & digital techniques*, vol. 3, no. 6, pp. 593–608, 2009.

[31] S. Yazdanshenas, B. Khaleghi, P. Ienne, and H. Asadi, "Designing low power and durable digital blocks using shadow nanoelectromechanical relays," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2016.

[32] J. L. Muñoz-Gamarra, A. Uranga, and N. Barniol, "Cmos-nems copper switches monolithically integrated using a 65 nm cmos technology," *Micromachines*, vol. 7, no. 2, p. 30, 2016.

[33] J. H. Kim, J. Xiang, Z. C.-y. Chen, and S. Kwon, "Nanowire nanoelectromechanical field-effect transistors," Feb. 25 2016, uS Patent 20,160,056,304.

[34] K. Buch, "Hdl design methods for low-power implementation," *EInfochips, Dec*, 2009.

[35] A. Gupta, S. Murgai, A. Gulati, and P. Kumar, "Design and implementation of low power clock gated 64-bit alu on ultra scale fpga," in *ADVANCEMENT IN SCIENCE AND TECHNOLOGY: Proceedings of the 2nd International Conference on Communication Systems (ICCS-2015)*, vol. 1715. AIP Publishing, 2016.

[36] A. A. E. Zarandi, A. S. Molahosseini, L. Sousa, M. Hosseinzadeh, and K. Navi, "Area-delay-power-aware adder placement method for rns reverse converter design," in *2016 IEEE 7th Latin American Symposium on Circuits & Systems (LASCAS)*. IEEE, 2016, pp. 223–226.

[37] M. P. Mills, "The cloud begins with coal," *Digital Power Group. Online*

at: http://www.   tech-pundit.   com/wp-content/uploads/2013/07/Cloud_Begins_With_Coal. pdf, 2013.