

IEEE Recommended Practice for Developing Energy-Efficient Power-Proportional Digital Architectures

IEEE Communications Society

Developed by the
Green ICT Standards Committee

IEEE Std 1924.1™-2022

STANDARDS

IEEE Recommended Practice for Developing Energy-Efficient Power-Proportional Digital Architectures

Developed by the

Green ICT Standards Committee
of the
IEEE Communications Society

Approved 26 April 2022

IEEE SA Standards Board

Abstract: A set of guidelines is presented in this recommended practice for the development of energy-efficient and power-proportional digital architectures so that energy is only consumed when computations are under way and energy is reduced in the non-operating state. The purpose of this practice is to provide guidelines for the designers and developers of digital architectures for creating power-proportionality at different levels of the system.

Keywords: digital circuits, energy efficiency, energy-efficient architectures, hardware accelerators, IEEE 1924.1™, interconnects, Internet of Things, machine learning assisted design, on-chip memory, on-chip power distribution networks, power-proportional, processors, sensors, servers

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2022 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published 30 November 2022. Printed in the United States of America.

IEEE is a registered trademark in the U.S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

PDF: ISBN 978-1-5044-8773-3 STD25468
Print: ISBN 978-1-5044-8774-0 STDPD25468

IEEE prohibits discrimination, harassment and bullying. For more information, visit <https://www.ieee.org/about/corporate/governance/p9-26.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE Standards documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page (<https://standards.ieee.org/ipr/disclaimers.html>), appear in all standards and may be found under the heading “Important Notices and Disclaimers Concerning IEEE Standards Documents.”

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents are developed within IEEE Societies and subcommittees of IEEE Standards Association (IEEE SA) Board of Governors. IEEE develops its standards through an accredited consensus development process, which brings together volunteers representing varied viewpoints and interests to achieve the final product. IEEE Standards are documents developed by volunteers with scientific, academic, and industry-based expertise in technical working groups. Volunteers are not necessarily members of IEEE or IEEE SA and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE makes no warranties or representations concerning its standards, and expressly disclaims all warranties, express or implied, concerning this standard, including but not limited to the warranties of merchantability, fitness for a particular purpose and non-infringement. In addition, IEEE does not warrant or represent that the use of the material contained in its standards is free from patent infringement. IEEE standards documents are supplied “AS IS” and “WITH ALL FAULTS.”

Use of an IEEE standard is wholly voluntary. The existence of an IEEE Standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity, nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document, should rely upon his or her own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus development process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English version published by IEEE is the approved IEEE standard.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE SA Standards Board Operations Manual shall not be considered or inferred to be the official position of IEEE or any of its committees and shall not be considered to be, nor be relied upon as, a formal position of IEEE. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that the presenter's views should be considered the personal views of that individual rather than the formal position of IEEE, IEEE SA, the Standards Committee, or the Working Group. Statements made by volunteers may not represent the formal position of their employer(s) or affiliation(s).

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE or IEEE SA. However, **IEEE does not provide interpretations, consulting information, or advice pertaining to IEEE Standards documents.**

Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its Societies and subcommittees of the IEEE SA Board of Governors are not able to provide an instant response to comments, or questions except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in evaluating comments or in revisions to an IEEE standard is welcome to join the relevant IEEE working group. You can indicate interest in a working group using the Interests tab in the Manage Profile & Interests area of the [IEEE SA myProject system](#).¹ An IEEE Account is needed to access the application.

Comments on standards should be submitted using the [Contact Us](#) form.²

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not constitute compliance to any applicable regulatory requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

Data privacy

Users of IEEE Standards documents should evaluate the standards for considerations of data privacy and data ownership in the context of assessing and using the standards in compliance with applicable laws and regulations.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under US and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, neither IEEE nor its licensors waive any rights in copyright to the documents.

¹Available at: <https://development.standards.ieee.org/myproject-web/public/view.html>.

²Available at: <https://standards.ieee.org/about/contact/>.

Photocopies

Subject to payment of the appropriate licensing fees, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400; <https://www.copyright.com/>. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every 10 years. When a document is more than 10 years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit [IEEE Xplore](#) or [contact IEEE](#).³ For more information about the IEEE SA or IEEE's standards development process, visit the IEEE SA Website.

Errata

Errata, if any, for all IEEE standards can be accessed on the [IEEE SA Website](#).⁴ Search for standard number and year of approval to access the web page of the published standard. Errata links are located under the Additional Resources Details section. Errata are also available in [IEEE Xplore](#). Users are encouraged to periodically check for errata.

Patents

IEEE Standards are developed in compliance with the [IEEE SA Patent Policy](#).⁵

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE SA Website at <https://standards.ieee.org/about/sasb/patcom/patents/>. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or if any licensing agreements are

³Available at: <https://ieeexplore.ieee.org/browse/standards/collection/ieee>.

⁴Available at: <https://standards.ieee.org/standard/>.

⁵Available at: <https://standards.ieee.org/about/sasb/patcom/materials/>.

reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

IMPORTANT NOTICE

IEEE Standards do not guarantee or ensure safety, security, health, or environmental protection, or ensure against interference with or from other devices or networks. IEEE Standards development activities consider research and information presented to the standards development group in developing any safety recommendations. Other information about safety practices, changes in technology or technology implementation, or impact by peripheral systems also may be pertinent to safety considerations during implementation of the standard. Implementers and users of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, and interference protection practices and all applicable laws and regulations.

Participants

At the time this recommended practice was completed, the EECH Working Group had the following membership:

Sara R. Biyabani, *Chair*
Daniel Wong, *Co-Vice Chair*
Hyeran Jeon, *Co-Vice Chair, Co-Secretary*
Rida Khan, *Co-Vice Chair, Co-Secretary*

Mahtab Alam
Mark Azadpour
Ahmed Biyabani
Azza Eltraify

Tim Godfrey
Taeyoung Kim
M. Sabarimalai Manikandan

Mohammad Musa
Dipti Pal
Suresh Sugumar
Young Jin Yoon

The following members of the individual Standards Association balloting group voted on this recommended practice. Balloters may have voted for approval, disapproval, or abstention.

Mark Azadpour
Sara Biyabani
Koti Reddy Butukuri
Pin Chang
Fabrizio Granelli

Marco Hernandez
Werner Hoelzl
Hyeran Jeon
Piotr Karocki
M. Sabarimalai Manikandan

Venkatesha Prasad
R. K. Rannow
Lisa Ward
Daniel Wong
Rachel Wood

When the IEEE SA Standards Board approved this recommended practice on 26 April 2022, it had the following membership:

David J. Law, *Chair*
Ted Burse, *Vice Chair*
Gary Hoffman, *Past Chair*
Konstantinos Karachalios, *Secretary*

Edward A. Addy
Ramy Ahmed Fathy
J. Travis Griffith
Guido R. Hiertz
Yousef Kimiagar
Joseph L. Koepfinger*
Thomas Koshy
John D. Kulick

Johnny Daozhuang Lin
Kevin Lu
Daleep C. Mohla
Andrew Myles
Damir Novosel
Annette D. Reilly
Robby Robson
Jon Walter Rosdahl

Mark Siira
Dorothy V. Stanley
Lei Wang
F. Keith Waters
Karl Weber
Sha Wei
Philip B. Winston
Daidi Zhong

*Member Emeritus

Introduction

This introduction is not part of IEEE Std 1924.1-2022, IEEE Recommended Practice for Developing Energy-Efficient Power-Proportional Digital Architectures.

This recommended practice specifies a set of guidelines for the development of power-proportional digital architectures. It presents independent recommendations at multiple levels of the system, starting from the lowest level of semiconductor process technology, through digital circuits, system components, modules and interconnects, and up to the system and multi-system level, including the hardware and the operating system as well as power supplies.

The focus of this recommended practice is to present specific recommendations pertinent at different levels of the system so that energy is only consumed when computations are under way and energy is reduced in the non-operating state.

Acknowledgments

The Working Group would like to posthumously thank IEEE Fellow Earl McCune for inspiring the need for this recommended practice.

Figure 4 and Figure 5 reprinted with permission from Jeon, H., “Resource Underutilization Exploitation for Power Efficient and Reliable Throughput Processor,” PhD Thesis, University of Southern California, 2015 [B30].

Figure 6 reprinted from Jeon, H., H. A. Esfeden, N. B. Abu-Ghazaleh, D. Wong, and S. Elango, “Locality-Aware GPU Register File,” *IEEE Computer Architecture Letters*, vol. 18, no. 2, pp. 153–156, 2019 [B29]. © IEEE 2019.

Figure 12 reprinted with permission from International Roadmap for Devices and Systems, 2016 Edition, More Moore White Paper, accessed March 7, 2022, https://irds.ieee.org/images/files/pdf/2016_MM.pdf.

Figure 14, Figure 15, and Figure 16 reprinted from Carrano, R. C., D. Passos, L. C. S. Magalhaes, and C. V. N. Albuquerque, “Survey and Taxonomy of Duty Cycling Mechanisms in Wireless Sensor Networks,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 181–194, First Quarter 2014 [B11]. © IEEE 2014.

Contents

1. Overview	10
1.1 Scope	10
1.2 Purpose	11
1.3 Word usage	11
2. Normative references	11
3. Definitions, acronyms, and abbreviations	11
3.1 Definitions	11
3.2 Acronyms and abbreviations	12
4. Framework for power-proportional energy efficiency recommendations	13
5. Fundamentals of power and energy efficiency	14
5.1 General	14
5.2 Static power	14
5.3 Dynamic power	15
5.4 Server processor power	15
5.5 Power and energy efficiency	16
5.6 Recommended practice: Power-proportional design versus low-power design	17
6. Process and gate-level techniques	17
6.1 General	17
6.2 Static power and leakage	18
7. Circuit-level techniques	21
7.1 General	21
7.2 Adaptive power management techniques	21
8. Component-level techniques	24
8.1 General	24
8.2 Processors and SoCs	24
8.3 Networks	31
9. Platform-level techniques	35
9.1 General	35
9.2 Platforms with high power demand	36
9.3 Platforms with low power demand	38
10. System architecture-level techniques	47
10.1 Techniques at the multiprocessor architecture level	47
10.2 Task migration techniques in clusters of servers	48
10.3 Load redistribution techniques in Mobile Cloud Computing	48
11. Supply-side technique	49
11.1 Background	49
11.2 The problem	50
11.3 The solution	50
11.4 Recommended practice	51
Annex A (informative) Overview of C-states and S-states	54
Annex B (informative) Bibliography	59

IEEE Recommended Practice for Developing Energy-Efficient Power-Proportional Digital Architectures

1. Overview

As the collective energy consumption of all Information and Communications Technology (ICT) products has surpassed that of worldwide air travel, hardware implementations of those products need to include the impact of those implementations on energy efficiency.

The need is to reduce the power consumption of devices when in idle mode and in transition to an “on” state at logic switching speeds in addition to the operating “on” state. Meeting this goal substantially reduces the energy consumption of digital devices, and it leads to power-proportionality of device operation. The cost of implementing and operating a standardized digital device is also generally reduced. This is particularly important for large server farms and cloud computing installations, yet it is equally important for embedded computing devices at the opposite end of the power scale.

Two desirable results directly follow when energy efficiency is improved to the extent that the laws of physics allow. First is that energy consumption is also reduced (i.e., this is the purpose and measurable effect of energy efficiency). Second is that the cost of implementing and operating such a standardized device is also generally reduced.

1.1 Scope

This recommended practice specifies a set of guidelines for the development of power-proportional digital architectures so that energy is only consumed when computations are underway.

Digital architectures could encompass individual devices such as CPUs and systems on chip (SoCs), or specialized computing platforms such as smartphones, smartwatches, and individual servers, or larger systems and networks of distributed compute and data servers. These architectures could consist of components such as processors, specialized accelerators, memory, interconnects, storage, networks, and power supplies.

In power-proportional power supplies, various techniques can be used to conserve and store energy when computation is halted, keeping it available for immediate use upon restart. The objective of such power supplies is to respond to nanosecond variations in the computing load in the presence of complementary metal-oxide semiconductor (CMOS) static power dissipation.

The focus of this work is to create a framework for creating power-proportionality at different levels of the system, starting with the device level and extending to the highest system architecture level, and to present specific recommendations at each level.

Given the vast number of public and proprietary techniques for energy efficiency, this recommended practice is not intended to be an exhaustive list of all such techniques.

1.2 Purpose

The purpose of this recommended practice is to provide a set of guidelines for the designers and developers of digital architectures that help ensure that power is consumed mostly when useful computational work is under way, and is consumed the least when idle.

1.3 Word usage

The word *shall* indicates mandatory requirements strictly to be followed in order to conform to the standard and from which no deviation is permitted (shall equals is required to).^{6,7}

The word *should* indicates that among several possibilities one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required (should equals is recommended that).

The word *may* is used to indicate a course of action permissible within the limits of the standard (may equals is permitted to).

The word *can* is used for statements of possibility and capability, whether material, physical, or causal (can equals is able to).

2. Normative references

The following referenced documents are indispensable for the application of this document (i.e., they must be understood and used, so each referenced document is cited in text and its relationship to this document is explained). For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments or corrigenda) applies.

There are no normative references in this document.

3. Definitions, acronyms, and abbreviations

3.1 Definitions

For the purposes of this document, the following terms and definitions apply. The *IEEE Standards Dictionary Online* should be consulted for terms not defined in this clause.⁸

C-state: Processor idle sleep states.

data sensing: The acquiring of data by a sensor.

⁶The use of the word *must* is deprecated and cannot be used when stating mandatory requirements, *must* is used only to describe unavoidable situations.

⁷The use of *will* is deprecated and cannot be used when stating mandatory requirements, *will* is only used in statements of fact.

⁸*IEEE Standards Dictionary Online* is available at: <https://dictionary.ieee.org/>. An IEEE Account is required for access to the dictionary, and one can be created at no charge on the dictionary sign-in page.

duty cycle: **(A)** In power electronics, the fraction of one period in which a signal or system is active. It is expressed as a ratio or a percentage. **(B)** In IoT device (MAC layer) communications, the ratio of the idle listening duration to the wake-up duration. The wake-up duration is the time spent in the active mode by a node.

energy efficiency: **(A)** At the circuit level, the ratio of total output to total input (total output/total input). **(B)** At the system level, ratio of energy consumed for unit of useful work to total energy input into the system.

heterogenous multiprocessors: Multiprocessors that contain cores or central processing units (CPUs) of different capabilities and characteristics, such as functionality, performance, and power consumption, such that an appropriate core or CPU is matched to the demands of the application or task. This leads to greater energy efficiency. For example, a task demanding highest capabilities is executed on a high performance, high energy-consuming core or CPU, whereas a background task with low compute demands is run on a lower performance more energy-efficient core or CPU.

operating system: System software that manages the hardware and other system resources and provides services to the application program.

power proportional computing: Computing where power is drawn when the device is in the operating ON state and power is minimal (or ideally zero) in the non-operating OFF state. Also, energy is only consumed when computations are under way and energy is reduced in the non-operating state

S-state: System-level sleep idle states [Advanced Configuration and Power Interface (ACPI)]

scheduler: The component of the operating system that controls which compute tasks are run on which specific hardware resource, such as a processor or server or a cluster. The kernel mechanism that realizes scheduling, i.e., task scheduling.

sensor activation: The switch on /off function of a sensor using control logic.

3.2 Acronyms and abbreviations

ACPI	Advanced Configuration and Power Interface
AI	artificial intelligence
AVS	adaptive voltage scaling
BAN	body area network
CMOS	complementary metal-oxide semiconductor
CNN	convolutional neural network
CPU	central processing unit
DVFS	dynamic voltage and frequency scaling
DVS	dynamic voltage scaling
EM	electromigration
GPS	Global Positioning System
GSM	global system for mobile communications
HoL	Head-of-Line
IC	integrated circuit

IoT	Internet of Things
LSTM	long short-term memory
MAC	medium access control
ML	machine learning
MOS	metal-oxide semiconductor
MP	multiple physical
NN	neural network
NoC	Networks-on-Chip
PDN	power-delivery network
PM	power management
RNN	recurrent neural network
SoC	System on Chip
SOI	Silicon on Insulator
TDMA	time division multiple access
VC	virtual channel

4. Framework for power-proportional energy efficiency recommendations

The body of work currently under way to make computing and communications components and systems energy efficient is vast and there is no standardized method to classify the various categories because of different architectures.

This recommended practice specifies a framework consisting of two parts:

- A “demand side,” where energy is consumed to perform a computation or communication transfer
- A “supply side,” where the energy is produced to supply to the demand side where that energy is being consumed

The demand side focuses on power-proportional energy efficiency techniques for:

- Process and gate level
- Digital circuits and VLSI design
- Components at the chip level: processors, SoCs, memory, interconnects, and networks
- Low power devices: Internet of Things (IoT), wearables, body area network (BAN)
- Computer systems and servers

The supply side focuses on power-proportional energy efficiency techniques for power supplies, especially switching power supplies.

Figure 1 shows the layered view of the framework, starting with the silicon at the bottom and the system architecture at the top.

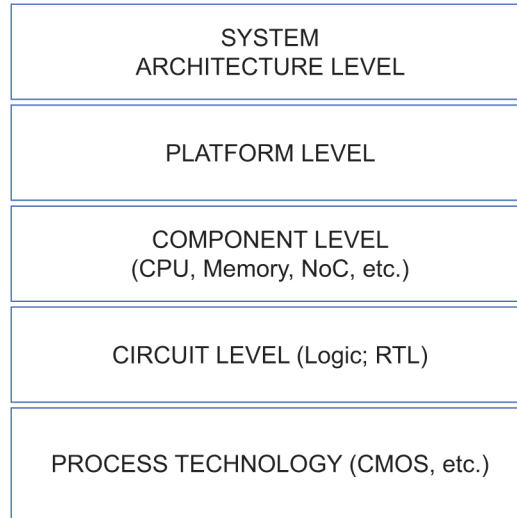


Figure 1—Layered view of the IEEE 1924.1 framework

Demand-side techniques are presented in [Clause 6](#) through [Clause 10](#), and a supply-side technique is presented in [Clause 11](#). For each topic in this recommended practice, the context of the problem is described, followed by the description of the problem(s) that affects power efficiency. Proposed solutions to address the problem are then described, together with any associated tradeoffs. Finally, recommendations for improving power efficiency are provided in the recommendation subclauses.

The fundamentals of power and energy efficiency are discussed in [Clause 5](#).

5. Fundamentals of power and energy efficiency

5.1 General

Power consumption in a digital circuit consists of two components:

- a) Static power
- b) Dynamic power

$$P_{total} = P_{static} + P_{dynamic} \quad (1)$$

NOTE—In this recommended practice, terms with lowercase subscripts denote power terms for simple digital circuits (e.g., $P_{dynamic}$) and power terms with uppercase subscripts denote processor power terms (e.g., $P_{DYNAMIC}$).⁹

5.2 Static power

Static power is consumed when the circuit is not operational and is largely due to leakage from sub-threshold currents. It is somewhat constant in comparison to dynamic power. It can be calculated in terms of the leakage current with [Equation \(2\)](#):

$$P_{static} = N \times K \times V \times I_{leak} \quad (2)$$

⁹Notes in text, tables, and figures are given for information only and do not contain requirements needed to implement the standard.

where

- N is the number of transistors in the circuit
- K is the proportionality constant, which depends on the specific circuit design
- V is the supply voltage
- I_{leak} is the leakage current

5.3 Dynamic power

Dynamic power is consumed when the circuit is operating and can be variable. It is due to the charging and discharging of the transistor intrinsic capacitor when switching state between high and low, and due to short-circuit currents during transitions.

It is calculated with [Equation \(3\)](#):

$$P_{dynamic} = \frac{1}{2} C \times V^2 \times A \times f \quad (3)$$

where

- C is the capacitance
- V is the supply voltage
- A is the activity factor
- f is the clock rate

5.4 Server processor power

Similarly, power dissipated by server processors has a dynamic and a static component:

$$P_{D,Processor} = P_{DYNAMIC}(f, V) + P_{STATIC}(T, V) \quad (4)$$

where the dynamic component is a function of clock frequency f and operating voltage V , and the static power component is a function of V and of the temperature T .

The dynamic power component follows the relationship based on capacitor charge/discharge cycles:

$$P_{DYNAMIC}(f, V) = \alpha N C V^2 f \quad (5)$$

where

- α is the average fraction of circuits that are switching
- N is the number of switching circuits available
- C is the effective switching capacitance for each switching circuit, and is a metric based on process technology

In order to reduce the dynamic power, in [Equation \(5\)](#), one needs a minimum operating voltage, the lowest practical clock frequency, a CMOS process with small C value, and a design with the least amount of circuitry that results in a small N value.

The dynamic voltage and frequency scaling (DVFS) technique is commonly used to manage, as shown in [Equation \(5\)](#), by scaling V and f to operate near a local power minimum for the compute job requirements of

the moment. Within an Advanced Configuration and Power Interface (ACPI) implementation, this control corresponds to the processor states P0 through P16 [B1].¹⁰

CMOS processes in the nanometer range have a significant static power component. However, semiconductor vendors do not routinely publish mathematical models for this static power.

$$P_{STATIC}(T, V) = K \times 1.04^T \times V^3 \quad (6)$$

where

- T is the operating temperature in Kelvin
- V is the core voltage V normalized to its maximum value V_{MAX}
- K is the proportionality constant which depends on the specific circuit design

This static power remains, even when the processor dynamic power in Equation (5) is stopped by setting frequency, $f=0$ (corresponding to the ACPI state C2). In order to entirely stop this static power, Equation (6) shows that the only option is to bring the device voltage V to zero.

5.5 Power and energy efficiency

5.5.1 General

Total power is the sum of static and dynamic power and can be approximated as given in Equation (7):

$$P_{total} \approx I_{leak} V + C V^2 f \quad (7)$$

Since power (in units of watts) is the rate of energy usage (in units of joules), energy is a product of power and time, $E = P \times t$.

Setting time to 1 unit, results in $E = P$, for $t = 1$. Energy efficiency can then be defined in terms of the ratio of power.

Energy efficiency is defined differently at the circuit level versus the system level. The following two subclauses (5.5.2 and 5.5.3) describe the details at the two levels.

5.5.2 Circuit level

Energy efficiency at the circuit level is defined as:

Energy efficiency at the Circuit level = Ratio of total output power per unit time to the total input power per unit time.

Energy efficiency of digital circuits (in units of joules/clock cycle) can be defined in terms of energy consumed (in unit of joules) per clock cycle of circuit operation. For example, “18.8 fJ/cycle (10.5 nW at 560 kHz)” [B56].

In low-power design, such as for batteryless IoT sensing systems that spend large amounts of time in standby mode, low-leakage operation is prioritized and such circuits are run at slow clock speeds in the hertz to kilohertz range to reduce dynamic power consumption. Static power is measured as leakage power (in units of picowatts), and dynamic power as energy per transition [in units of femtojoules (fJ)] [B57].

¹⁰The numbers in brackets correspond to those of the bibliography in Annex B.

5.5.3 System level

Energy efficiency at the system level is defined as:

Energy efficiency at the System level = Ratio of energy consumed for a unit of useful work to total energy input into the system in a unit of time.

Because a system is far more complex than a simple circuit, and is composed of multiple complex and independent circuits with multiple asynchronous clocks, it is difficult to precisely calculate the aggregate power input and output for all asynchronous circuits.

Also, there are multiple circuits and components necessary for the running of the main processor or CPU, such as the management processor(s), which do not directly contribute to the operation of the main processor/CPU where the computational operations of interest are run but are nonetheless required for the system to run correctly.

This leads us to differentiate between “useful work” (done on the main processor or CPU) and “all work” (done by components and circuits required for the running of the overall system). Hence, the energy efficiency at the system level is defined as the ratio of energy consumed for “useful” work to the total energy input into the system in the same duration of time.

5.6 Recommended practice: Power-proportional design versus low-power design

Power-proportional designs and low-power designs optimize different parameters and address different use cases. For example, power-proportional techniques are used in data center servers that consume considerable power and low-power techniques are used in battery-powered devices. The parameters are mainly total power utilized and total power dissipated.

Power-proportional designs should

- Draw power in the ON state proportional to the number of computational operations executed by the workload. Hence, substantial heat dissipation can be tolerated in ON state.
- Reduce power utilized when the system is idle *in the OFF state*. Ideally, the system should draw zero-power in the idle state.
- Reduce power dissipated as heat (*in both ON and OFF states*). However, the only way to eliminate or reduce heat in the OFF state (that is, static power) is to switch off the voltage.

Low-power designs should reduce both:

- Overall power utilized in both the *ON and OFF states*
- Power dissipated as heat *in both ON and OFF states*

6. Process and gate-level techniques

6.1 General

Power consumption in complementary metal-oxide-semiconductor (CMOS) circuits can be divided between static and dynamic dissipation with the dynamic component being predominant. This is because of the work done in switching the circuit to do useful computation. [Equation \(1\)](#) and [Equation \(3\)](#) can be approximated as:

$$P_{total} \sim C \times V^2 \times f \quad (8)$$

where

P_{total} is power
 C is switch capacitance
 V is supply voltage
 f is operating frequency

The switch capacitance in Equation (8) is the capacitance at a given node in the circuit, for example, at any logic gate output. Every low-to-high logic transition in a digital circuit corresponds to transferring charge, hence causing a change of voltage and drawing energy from the power supply.

The design variables in Equation (8) can be varied at the semiconductor process level, circuit design level, or architectural level with the aim of minimizing overall energy dissipation. This is usually a complex process of tradeoffs between multiple competing goals including accuracy, speed, and area.

6.2 Static power and leakage

6.2.1 General

Static energy consumption is the result of the flow of electrons during idle non-computationally useful periods of operation. This can be caused by short circuit currents, bias, and leakage currents. For example, during the transition on the input of a CMOS gate both p and n channel devices may conduct simultaneously, briefly establishing a short from the supply voltage to ground.

There are, however, a few specialized cases where statically biased gates are used, such as programmable logic arrays. Alternate design techniques have become the norm to avoid this situation. Despite the adoption of power efficient circuit strategies, the use of newer lower voltage semiconductor processes has led inevitably to leakage currents becoming the primary component of static energy consumption. Where previously it was a secondary order effect static power consumption doubles with every new modern process node.

6.2.2 Capacitance

Since energy consumption in CMOS circuitry is directly proportional to capacitance minimizing circuit node capacitances is a high priority for lowering power dissipation. This can be achieved at the architectural design level as well as at the logic and process levels.

Connections to off-chip resources, such as auxiliary memory, normally have much greater capacitance than connections to on-chip components. As a result, computer architects have avoided accessing external memory for energy consumption reasons quite apart from the obvious speed considerations. But since on-chip memory can be expensive there is a tradeoff somewhere in the design process by optimizing the overall system using on-chip caches and registers. A desirable side effect of using fewer external components and infrequent switching of external pins is a savings in terms of dynamic power consumption.

On-chip, one of the main limitations on clock frequency is routing capacitance. Circuits that are able to run faster can do so because of a lower routing capacitance. Also, they dissipate less power at a given clock frequency. Optimizing the clock frequency of the high-level system design, possibly with a performance exceeding requirement, can yield a reduction in overall energy consumption.

6.2.3 Background

One of the earliest and most cited references [B13] arose from the motivation to create portable battery-powered applications demanding intensive computation. Techniques discussed include reducing power consumption in CMOS digital circuits while maintaining computational throughput. Other techniques for low-power operation shown include those which use the lowest possible supply voltage. Several optimizations at

the architectural, logic style, circuit, and technology levels are also described. Lastly, a scaling strategy based on architectural considerations is presented. The conclusion is that the optimum operating voltage in terms of reduced power consumption can be much lower than that determined by other scaling factors and is achieved by settling for increased silicon area.

The book *Low-Power CMOS Design* [B12] presents principles of architectural optimization enabling supply voltages at and below 1 V standard CMOS technology. Various other techniques are presented for minimizing the switched capacitance. These techniques include optimizing signal correlations, minimizing spurious transitions, optimizing sequencing of operations and activity-driven power down. Such techniques have since become standard across the industry. There is also discussion of the fundamental limits of power consumption at all levels of the design hierarchy as well as a description of adiabatic switching yielding very efficient energy per computation for a given voltage.

The principles of power-aware design [B45] and power reduction specifically in microprocessor design [B8] are also foundational resources for any would-be designer focusing on “green” computation even in otherwise demanding energy intensive artificial intelligence (AI)-related applications such as real-time video and speech analysis. A computation engine such as a microprocessor provides its services continuously or in short bursts where a framework is presented for an energy-efficient design methodology. This has also become a standard technique. A general overview [B20] provides useful information to many readers.

6.2.4 Standard practices using supply voltage, variable clock frequency

6.2.4.1 General

A straightforward and effective way of reducing power at the technology level is to reduce the supply voltage, because the power consumption scales quadratically with the voltage. The resulting decrease in circuit speed or performance needs to be balanced against the power savings. In order to offset or maintain the same throughput, additional circuitry may be needed with its own attendant costs.

Energy-efficient design or the use of integrated circuits (ICs) in portable applications requiring longer battery life is routinely implemented with variable clock frequency and a voltage controlled by the operating system. The variable voltage and frequency represent a tradeoff between delay and power. Reducing clock frequency does not necessarily reduce power, since the system needs to run longer to do the same work.

As voltage is reduced, the delay of the circuit increases. A common approach to power reduction starts by increasing the performance of the module, perhaps by increasing parallelization. Once an acceptable performance is decided, the voltage is reduced as much as possible while maintaining performance. This design method often results in larger circuit area. The basic tradeoff is between area and power.

Dynamically turning off the clock to unused logic or peripherals is another straightforward way to reduce CMOS power consumption, which is inherently tied to clock frequency. Such control can be exercised at the hardware level or at the operating system level. For example, the clocks to most of the sections are turned off during their idle or sleep modes. Devices may require different amounts of time to wake up from different sleep modes.

6.2.4.2 Asynchronous operation

CMOS is a preferred technology for many low-power applications because gates only dissipate energy when they are switching. However, many gates switch because they are connected to the clock, not because they have new inputs to process. Hence, a synchronous circuit block wastes power when its constituent logic is not utilized. A methodology that helps reduce energy dissipation due to un-needed circuit switching activity is asynchronous design.

The largest single gate on an IC is frequently the clock driver. The clock driver evenly distributes a clock signal, the timing reference, to all parts of a circuit whether or not any given portion requires it during a given