

UC Berkeley

Green Manufacturing and Sustainable Manufacturing Partnership

Title

Life-Cycle Energy Demand of Computational Logic: From High-Performance 32nm CPU to Ultra-Low-Power 130nm MCU

Permalink

<https://escholarship.org/uc/item/2d47b7dq>

Authors

Bol, David
Boyd, Sarah
Dornfeld, David

Publication Date

2011-04-01

Peer reviewed

Life-Cycle Energy Demand of Computational Logic: From High-Performance 32 nm CPU to Ultra-Low-Power 130 nm MCU

David Bol, *Member, IEEE*, Sarah Boyd, and David Dornfeld

Abstract—Given the exponential growth of the semiconductor industry, it is critical to assess the life-cycle energy demand of its products for appropriate eco-design in nanoelectronics. For computational logic applications, life-cycle energy demand is highly application dependent. In this paper, we study the life cycle of CMOS logic chips for five computational logic applications: from high-performance 32 nm CPUs for servers and laptops to low-power 45 nm processors for set-top boxes and smart phones to ultra-low-power 130 nm MCUs for RFID tags and sensors. For each chip, we model the energy demand for the CMOS processing step of integrated circuit manufacturing as well as for their use phase including both active and stand-by modes. While use-phase energy in active mode is almost two orders of magnitude higher than CMOS processing energy for high-performance CPUs, the energy demand for ultra-low-power MCUs is completely dominated by CMOS processing energy.

I. INTRODUCTION

The semiconductor industry has grown consistently in the last decades thanks in part to its continuous technical advancement and device scaling as described by Moore's law. Today, production volumes of CMOS chips are huge especially in the consumer market. Annual sales for cell phones have exceeded 1 billion units since 2007, according to Gartner. Although CMOS products contain tiny material quantities, they feature significant environmental impacts for two reasons. First, their embodied energy is high due to their complexity and the high purity of the materials, water, and process environment required in manufacturing [1], [2]. Second, the increasing computing capabilities of successive product generations have driven up the use phase power consumption of computational logic in many applications [3], [4]. Even though the energy to process a bit has decreased, device power per area has increased up to its practical limit for advanced applications.

As the application range of CMOS logic is wide, including many types of products, from sensors to advanced CPU, life-cycle impacts can vary considerably. Depending on the application, the life-cycle phase with the highest environmental impact may be either manufacturing or use. In this paper, we perform for the first time an application-aware life-cycle energy evaluation of computational logic chips with the explicit target of identifying the life-cycle phase with the highest

energy demand. The results could further be used to focus the eco-design efforts for reducing the environmental impact of computational logic on the most energy-demanding life-cycle phase. Given the broad application range of computational logic, there is a clear trade-off between the complexity of a fine-grained analysis covering separately numerous applications and the lack of precision of coarse-grained analysis focusing on a single application such as [1], [2], [3]. In this paper, we select five CMOS chips representative of five typical application categories distributed over the wide application range: from high-performance servers to mainstream laptops to set-top boxes to smart phones to ultra-low-power/cost RFIDs. This medium application granularity leverages a balanced trade-off between simplicity of the analysis and precision of the results. Nevertheless, several products and usage scenarios coexist within each application category. The smart phone category for instance also includes ultra-mobile PCs (netbooks, tablet PCs), which may be used in a different fashion. For this reason, there is considerable variation among the use phase scenarios associated with each category. Thus, in addition to data uncertainty, we address model uncertainty by introducing best, typical and worst cases throughout the whole analysis to ensure that results represent the full range of products included within each of the five application categories.

The typical life-cycle of computational applications starts with bare silicon wafer production from silica. CMOS processing is then performed on these bare wafers to create CMOS transistors and interconnects. Processed wafers are tested and diced into Si dies, which are sorted for functionality and performance. Good dies are assembled with a chip package providing the interface to the macro world. Systems such as servers, laptops or cell phones are finally assembled with other CMOS chips (I/O, memory, power management) and other types of components. Use phase and end of life finish the cycle. In modern CMOS technology generations, the energy demand of bare wafer production and die assembly have been shown to be lower than the energy demand for CMOS processing [1], [3]. Moreover, the energy demand for transportation of computational logic chips is very low due to their small weight [3]. Therefore, we restrict the scope of this paper to CMOS processing and use phases.

This paper is organized as follows. In Section II, we present the selected application categories we chose for this analysis with their associated CMOS chips. The modeling of the energy demand for CMOS processing and use phases is then detailed in Section III and the results are discussed in Section IV.

This work was supported by the National Foundation for Scientific Research of Belgium (FNRS).

Dr. Bol is with ICTEAM institute from Université catholique de Louvain, Louvain-la-Neuve, Belgium (e-mail: david.bol@uclouvain.be). Dr. Boyd and Pr. Dornfeld are with the Laboratory for Manufacturing and Sustainability from University of California Berkeley (e-mail: sboyd@me.berkeley.edu, dornfeld@berkeley.edu).

TABLE I
CONSIDERED APPLICATION CATEGORIES OF COMPUTATIONAL LOGIC WITH CORRESPONDING CMOS CHIPS

Typical applications	High-end servers	Mainstream laptops	Set-top boxes and digital TVs	Smart phones and ultra-mobile PCs	RFIDs, eHealth and industrial sensors
Logic circuit	Six-core CPU	Dual-core CPU	Multimedia processor	Application processor	Ultra-low-power MCU
CMOS process	32 nm HP	32 nm HP	45/40 nm LP	45/40 nm LP	130 nm LP + Flash
Die area	240mm ² ±30%	80mm ² ±30%	70mm ² ±30%	50mm ² ±30%	10mm ² ±30%
Lifespan	2 years ±30%	3 years ±30%	4 years ±30%	2 years ±30%	7 years ±30%

II. SELECTED APPLICATION CATEGORIES OF COMPUTATIONAL LOGIC AND TYPICAL CMOS CHIPS

We selected five mainstream application categories of computational logic with the corresponding CMOS chips, typical of ramp-up production in 2010. As presented in Table I, the die area and CMOS process technology of these chips vary significantly to meet the functionality, performance and cost requirements of each application.

As shown in Table I on the high-performance side, high-end professional servers use multi-core microprocessors or CPUs (central processing units) requiring large silicon dies on the most advanced CMOS process generation i.e. 32 nm HP (high performance) in 2010 [6]. Considered life span of servers is 2 years [7]. Lighter dual-core CPUs are used for laptop computers with lower power consumption and cost at the expense of computing performance [6]. A 3-year lifespan is considered for laptops.

On the cost-efficient side, we find two consumer application categories with important similarities regarding their CMOS chips: multimedia processors for set-top boxes and digital TVs, and application processors for smart phones, mobile internet devices, tablet PCs and netbook computers (ultra-mobile PCs) [8], [9], [10]. The die size of these chips is usually smaller than dual-core CPUs for cost effectiveness. They are manufactured on a low-power (LP) CMOS process technology to extend battery life (smart phones and ultra-mobile PCs), to meet energy-efficiency labels regarding stand-by power or to reduce package cost thanks to lower self heating. Moreover, the considered CMOS process node typically lags one generation behind high-performance applications using an optical shrink for further die area savings [10]: 45/40 nm LP in 2010. A 3-year life span of smart phones and ultra-mobile PCs is considered as for laptop computers because these are consumer products, whereas 4 years are considered for set-top boxes because these devices are usually provided by the cable operators who update their hardware more rarely.

Finally, an ultra-low-cost application category featuring active RFID tags, eHealth devices and industrial/habitat sensors require low-performance yet ultra-low-power microcontroller units (MCUs) with a non-volatile memory [11], [12]. Although advanced CMOS technologies are interesting for MCUs as demonstrated in recent research works [13], [14], CMOS process for commercial MCUs lag far behind: a 130 nm LP is typically used in 2010 mainly for mask and fab equipment cost concern but also for low leakage power as well as for the availability at low cost of embedded non-volatile memories such as NAND Flash. Small dies drastically limit the production costs of these devices. As a result from the industrial/professional

context of these applications, we consider a significantly longer life span of 7 years for this CMOS chip.

In this study, we deal with uncertainty by a best/typical/worst-case approach. We consider a ±30% uncertainty on die size as it varies from one company to another. Moreover, we consider an equivalent uncertainty for life span as it is very user/operator specific.

III. LIFE-CYCLE MODEL OF ENERGY DEMAND

In this section, we examine the modeling of life-cycle energy demand. Let us recall here that we restrict the discussion in this paper to the two phases with the highest energy demand potential: CMOS wafer processing and use phases [3].

A. CMOS processing

The 3 CMOS process technologies are detailed in Table II. They share most of the baseline CMOS process [2], [3], [4] with additional features and corresponding process steps at 45/40 nm and 32 nm to face the challenges associated with transistor and interconnect scaling. For example, the standard Si oxide and poly-Si electrode materials of the transistor gate still in use in 45/40 nm LP process [15], [16], [17] have been replaced by high-κ Hf-based oxide and dual-metal atomic-layer-deposited (ALD) electrode in 45 and 32 nm for HP process [18]. Performance-enhancing strain is implemented in the 45/40 nm LP and 32 nm HP CMOS processes using nitride capping and SiGe. The Al bonding pad capping material has also been replaced by Cu in 32 nm HP and the typical number of interconnect layers is also increased [18]. The 130 nm LP CMOS process for industrial sensors features numerous additional process steps due to the need for embedded Flash memory.

Based on these process descriptions, we create representative life-cycle inventories (LCI) by updating and adapting the LCI model for CMOS logic production from [3], [4]. We enhance the previous model by taking into account multi-V_t technology [19] for logic and embedded SRAM blocks with power management, which has been mainstream since the 90nm CMOS generation [9]. We also add the independent process steps for manufacturing analog I/O devices [15]. For Flash memory embedded in the 130 nm LP CMOS logic for industrial sensors, we add generation -specific front-end process steps from the LCI model for NAND Flash presented in [21] to the baseline 130 nm LP logic process. We do not add back-end steps as it is assumed that the interconnect steps are shared between the logic and Flash processes.

Raw results are given in Fig. 1 for equipment electricity (process LCI data), facility electricity (process LCI data) and

TABLE II
MAIN FEATURES AND PROCESS STEPS OF THE CONSIDERED CMOS TECHNOLOGIES

CMOS process	32 nm HP	45/40 nm LP	130 nm LP + Flash
Substrate	STI formation + well implant [†]	STI formation + well implant [†]	STI formation + well implant [†]
Channel	3× logic + 1× SRAM V_t adjusts [†]	3× logic + 1× SRAM V_t adjusts [†]	2× logic + 1× SRAM V_t adjusts [†]
Gate oxide	High- κ Hf-based oxide	Std- κ SiNO oxide	Std- κ SiO ₂ oxide
Gate electrode	Dual-metal ALD + poly-Si capping	Poly-Si RTO [†] + Ni silicide	Poly-Si RTO [†] + Co silicide
Source/drain	Extension [†] , spacer, junction [†] formation	Extension [†] , spacer, junction [†] formation	Extension [†] , spacer, junction [†] formation
Contacts	Ni silicide and W plug damascene	Ni silicide and W plug damascene	Co silicide and W plug damascene
Strain	Dual Ni stress liner capping + eSiGe EPI	Dual Ni stress liner capping + eSiGe EPI	-
Interconnect layers	10 dual-damascene Cu layers	8 dual-damascene Cu layers	6 dual-damascene Cu layers
Interlayer dielectric	Ultra-low- κ SiCOH oxide	Ultra-low- κ SiCOH oxide	Std- κ SiO ₂ oxide
Bonding pads	Cu capping	Al capping	Al capping
Analog I/O transistors	Separate steps for well [†] and V_t implants [†] , SiO ₂ gate oxide and source/drain extension [†] formation - no strain, no halo		
Additional steps	-	-	Flash: additional front-end process [21]

[†] The doping phases of these steps are repeated twice for separate NMOS and PMOS fabrication.

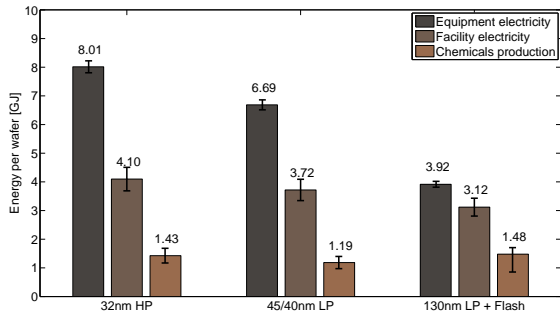


Fig. 1. CMOS processing energy demand per wafer for CMOS technologies used in the considered applications

upstream chemicals production (process and economic input-output LCI). Despite the Flash process addition representing 27% of the energy per wafer, the 130 nm LP CMOS process requires much less energy than the other processes. The 32 nm HP process requires 20% more equipment electricity per wafer than the 45/40 nm LP mainly due to the high- κ metal gate addition and the additional interconnect layers. When compared to [3], the total energy for 45/40 nm LP is higher by 10%, due to the fact that we take multi- V_t technology and analog I/O devices into account in this updated model.

The yield of CMOS processing is considered in 2 steps: wafer (line) yield η_{wafer} and die yield η_{die} . The wafer yield is assumed to be independent of the CMOS process as the productivity target is stronger than the technological challenges for mature high-volume production. We consider a wafer yield of 87% for typical case, including 2% scraps and 10% test/monitor wafers [3]. We further add $\pm 5\%$ uncertainty. The die yield depends on its area [20] and we use an industrial model for yield prediction, which gives typical die yields between 75% for the six-core CPU to 87% for the multimedia processor to 99% for the ultra-low-power MCU. An uncertainty of $\pm(1 - \eta_{die}) \times 0.05$ is added. The resulting annual energy demand per die is given in Section IV.

B. Use-phase power consumption model

The range of computational logic applications considered in this study represents a wide diversity in computation and

data processing capabilities. Consequently, use-phase electrical power consumption varies over several orders of magnitude, from a few mW s to 100W. Moreover, the applications also directly feature very different power budgets for their logic circuits. These targets come from product specifications associated to:

- maximum heat dissipation to meet cooling system capability and Si/package temperature range for servers, laptops, set-top boxes and digital TVs;
- maximum heat dissipation to avoid health hazards for smart phones;
- energy efficiency labels for servers, laptops, set-top boxes and digital TVs;
- battery life for laptops, ultra-mobile PCs, smart phones, biomedical devices and active RFIDs/sensors;

This wide range of power consumption levels requires a generic yet realistic model to get representative data for all the considered applications. The model we consider is based on two distinct contributions to power consumption coming from two modes of operations: active and stand-by. Although most advanced logic circuits often feature several active and stand-by modes for power management concerns [8], [9], [30], their usage highly depends on the operating system and the application scenario. As intermediate modes are not implemented for some of the applications under consideration (set-top boxes, sensors), we consider only active and a single stand-by mode in our model but use best (BC), typical (TC) and worst (WC) cases within an application category to track and bound uncertainty.

We model the average power in active mode P_{act} with a function of the average activity factor in active mode α_F :

$$\begin{aligned}
 P_{act} &= (P_{max} - P_{idle}) \times \alpha_F^{1.5} + P_{idle} \\
 &= P_{max} \times (\alpha_F^{1.5} (1 - 1/\beta_{idle}) + 1/\beta_{idle}), \quad (1)
 \end{aligned}$$

with P_{max} the maximum power at full processing/computing load i.e. when $\alpha_F = 100\%$, P_{idle} the idle power in active mode i.e. when $\alpha_F = 0\%$ and β_{idle} the idle power reduction factor. Indeed, computational-logic applications are rarely operating at full load because logic circuits are designed to support the worst-case load of the application associated to real-time computing/processing demand. Servers for example are typically

TABLE III
USE-PHASE POWER CONSUMPTION OF LOGIC CIRCUITS AND USAGE SCENARIOS IN THE CONSIDERED APPLICATIONS

Typical applications	High-end servers	Mainstream laptops	Set-top boxes and digital TVs	Smart phones and ultra-mobile PCs	RFIDs, eHealth and industrial sensors
Logic circuit	Six-core CPU	Dual-core CPU	Multimedia processor	Application processor	Ultra-low-power MCU
P_{max}	95 W \pm 30%	35 W \pm 30%	3 W \pm 30%	1 W \pm 30%	10 mW \pm 30%
β_{idle}	5 3 2	5 3 2	5 3 2	5 3 2	5 3 2
β_{stb}	20 10 5	20 10 5	20 10 5	10000 100 20	10000 1000 100
α_F [%]	10 30 50	5 20 50	33 66 99	10 30 50	10 33 100
T_{on} [%]	90 95 99	10 23 99	90 95 99	11 71 99	0.03 7 99
T_{act} [%]	90 95 99	5 16 23	6 10 20	6 8 17	0.03 0.07 1

The three values represent best, typical and worst cases regarding life-cycle energy demand, respectively.

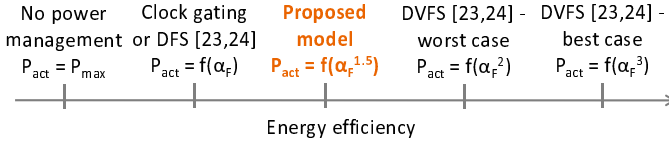


Fig. 2. Energy efficiency models of active-mode power management techniques for logic circuits.

operated at loads between 10 and 50% [22]. Considering P_{max} as the average active-mode power would thus overestimate use-phase energy demand. The considered values for P_{max} and α_F are given in Table III and explained in Section III-C.

Power management techniques such as clock gating [23], dynamic frequency scaling (DFS) or dynamic frequency/voltage scaling (DVFS) [24] are commonly used to save power during low-activity periods. In Eq. (1), we select a superlinear relationship between the P_{act} and α_F with 1.5 exponentiation. This model is a realistic assumption, standing in the middle of the energy efficiency range of power management techniques, as represented in Fig. 2. A fixed term P_{idle} is added to represent the power when the circuit is idle, which comes from leakage currents often close to 1/3 of P_{max} [19], [25] and power driven by always-on peripherals. P_{idle} is modeled as P_{max} divided by a factor β_{idle} corresponding to power management techniques at circuit design and software programming levels to save active-mode power. As shown in Table III, the considered value for β_{idle} is 3 in the typical case, which corresponds to ideal clock gating or dynamic frequency scaling [23] leaving only leakage power ($= P_{max}/3$) when idle (no always-on peripherals). This is consistent with power benchmarks for commercial CPUs in scientific literature [26], [27] and recent web-published studies [28], [29]. Uncertainty is considered through BC and WC β_{idle} values of 5 with active-mode leakage reduction techniques and 2 with no leakage reduction technique and several always-on peripherals, respectively.

Similarly, standby-mode power P_{stb} is modeled with a power reduction factor β_{stb} :

$$P_{stb} = P_{max}/\beta_{stb}. \quad (2)$$

P_{stb} is often orders-of-magnitude lower than P_{idle} because more aggressive power-management techniques such as power gating are used in stand-by mode. Indeed, wake-up time from stand-by mode can be longer and the state of some on-chip memory parts can be lost. As shown in Table III, identical β_{stb} values are considered for six-/dual-core CPUs and the multimedia processor. The best-case β_{stb} of 20 corresponds

to a full power gating of the circuit which totally cancels dynamic power and strongly reduce leakage power [25]. The typical- and worst-cases values represent more realistic and pessimistic scenarios, respectively, with partial power gating. The application processors for ultra-mobile PCs and smart phones feature stronger requirements on stand-by power management. Best- and typical-case β_{stb} values of 1000 [9], [30] and 100 [8] are observed. A value of 20 is considered as the worst case. Finally, ultra-low-power MCUs are intentionally designed in an older 130 nm LP CMOS technology, which features reduced leakage currents. The addition of strong power management techniques in stand-by mode enable β_{stb} values as low as 1000 and 100 [11], [12] for TC and BC, respectively.

C. Usage scenario

The use-phase energy demand per year in active and stand-by modes can simply be expressed by integrating P_{act} and P_{stb} over the annual time spent in these modes, respectively:

$$E_{act} = P_{act} \times T_{act} \times 3.15 \cdot 10^7 \quad (3)$$

$$\begin{aligned} E_{stb} &= P_{stb} \times T_{stb} \times 3.15 \cdot 10^7, \\ &= P_{stb} \times (T_{on} - T_{act}) \times 3.15 \cdot 10^7, \end{aligned} \quad (4)$$

with T_{act} and T_{stb} the fraction of time that the circuit is in active and stand-by modes, respectively, $T_{on} = T_{act} + T_{stb}$ is the fraction of time the circuit is turned on and $3.15 \cdot 10^7$ is the number of seconds in a year.

As the use-phase energy demand model has been thoroughly presented in Section III-B, let us now focus on the considered data for the following parameters P_{max} , α_F , T_{on} and T_{act} for each application. Notice that the uncertainty for P_{max} is considered as bounded to 30% for a given application because the competition between logic circuit design companies with their common power consumption specifications result in a narrow spread of P_{max} .

1) *Six-core CPU*: at full computing/processing load, CPUs for high-end servers consume about 100W as does the 32 nm six-core from [6]. However, their typical α_F in data centers is 30% with 10 and 50% best and worst cases, with almost no stand-by periods ($T_{act} \sim 100\%$) [22]. Servers in data centers are intended to be kept turned on and active most of the time, the down times are only for maintenance purpose. We thus consider T_{on} values of 90% (BC), 95% (TC) and 99% (WC).

2) *Dual-core CPU*: The full load in a 32 nm dual-core CPU for mainstream laptops consumes about 35 W [6], which might correspond to computing-intensive gaming for example. TC

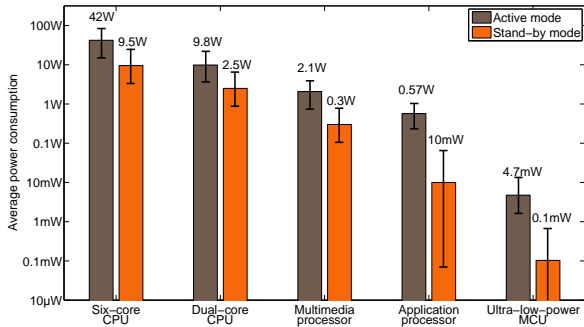


Fig. 3. Average electrical power consumption in active and stand-by modes

represents an office usage 8 hours a day, 5 days a week, 50 weeks a year ($T_{on} = 25\%$) with 70% active time ($T_{act} = 0.7 \times T_{on} = 23\%$) [3]. For the WC, the CPU is assumed to be always on with an 100% active time during office hours. As BC, we select a home computer with 2.5 hours a day of operation and 50% active time. Given the single simultaneous user, α_F values are lower than for the server case.

3) *Multimedia processor*: A full load consuming about 3W [31] may be reached when decoding a high-definition video on one channel while simultaneously recording videos on two other channels (set-top box) or performing simultaneous tasks such as image recognition and database search (digital TV). The multimedia processors in set-top boxes and digital TVs typically remain turned on most of the time to support electronic program guide updating, automated program downloads, rights management, and firmware updates [anonymous]. Their T_{on} is thus similar to six-core CPUs. However, they spend more time in stand-by mode with thus a lower T_{act} and an increased α_F . Average α_F values of 33%, 66%, 100% are considered for BC, TC and WC corresponding to single-channel video decoding, dual-channel decoding/recording and triple-channel decoding/recording, respectively. The 6%, 10%, 20% T_{act} corresponds to 2, 3 and 5 hours a day during 5, 6 and 7 days a week for BC, TC and WC, respectively and 50 weeks a year.

4) *Application processor*: The 1W full load power [8], [9], [30] may be driven when performing simultaneous standard-definition video decoding/encoding for a video conference. As, ultra-mobile PCs and smart phones typically enters stand-by mode very quickly, we consider α_F values slightly higher than for the laptop dual-core CPU. As TC, we consider a smart phone with mid-level usage which is turned off at night: T_{act} is 2 hours a day and T_{on} is 17 hours a day during 365 days a year. WC represents a smart phone with high usage which is kept on at night: T_{act} is 99% and T_{on} is 4 hours a day, whereas BC is typical of an ultra-mobile PC with light usage: T_{act} is 2 hours a day and T_{on} is 4 hours a day, both during 5 days a week and 50 weeks a year.

5) *Ultra-low-power MCU*: The full-load operation is typical for sensor/RFID MCUs but clock frequency tuning at design time can be used to adjust the full load level to the application processing requirements. This can thus be handled by the same model with α_F reflecting the full load adjustment. Typical full-load power consumption is 10mW [11], [12]

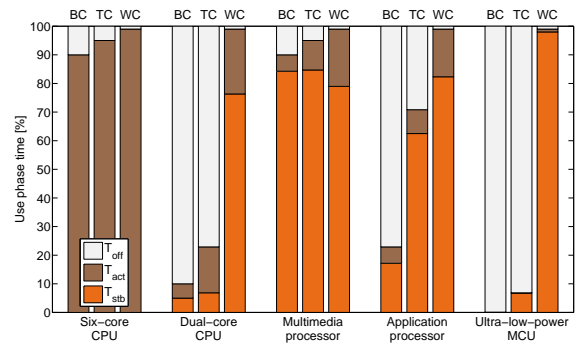


Fig. 4. Use-phase time repartition between modes for best- (BC), typical- (TC) and worst-case (WC) scenarios.

at a baseline 25-MIPS performance level. Clock frequency scaling for performance adjustment is modeled by 10%, 33% and 100% α_F values for BC, TC and WC, respectively. TC represents an active (battery powered) RFIDs for e.g. car keys with 2 hours of operation a day, 6 days a week. WC represents an always-on industrial monitoring sensor system and both TC and WC feature a duty cycle of 1%, which is typical of these applications [32]. As BC, we select a smart card reader with 2 minutes of operation a day, 5 days a week.

Resulting average power consumptions in both modes are depicted in Fig. 3 with time breakdown in Fig. 4.

IV. RESULTS AND DISCUSSION

Fig. 5 summarizes the annual life-cycle energy demand per die for the considered application categories. It is clear that both the magnitude and the breakdown between life-cycle stages of the energy demand highly varies from an application category to another. Energy demand of six-core CPU for servers, 3.8 GJ in the typical case, is 2000 \times higher than the energy of an MCU for sensors, 180kJ in the typical case. The use-phase energy consumed in active mode is dominant for high-performance applications: server's and laptop's CPUs, as confirmed by [3]. However, computational logic chips for low-power consumer applications - multimedia processor for set-top boxes and application processor for smart phones, have a more balanced life-cycle energy breakdown. Notice that annual energy demand for CMOS processing is higher for the application processor despite its smaller die area due to the shorter life span of its applications i.e. smart phones and ultra-mobile PCs. Finally, the life-cycle energy demand of computational logic chips for ultra-low-power applications (MCU for sensors/RFIDs) is totally dominated by the CMOS processing stage.

As shown in Fig. 5, the uncertainty on energy demand is the highest in stand-by mode for the application processor and the ultra-low-power MCU. This is due to the fact the application categories for these logic chips are the broadest with different usage scenarios.

Finally, the world-wide energy demand can be approximated by multiplying the annual energy demand per die by the average life span of the application and the annual sales. For example, when considering only the smart phone market with 1 billion annual sales, the world-wide energy demand for application processors is about 10⁷GJ.

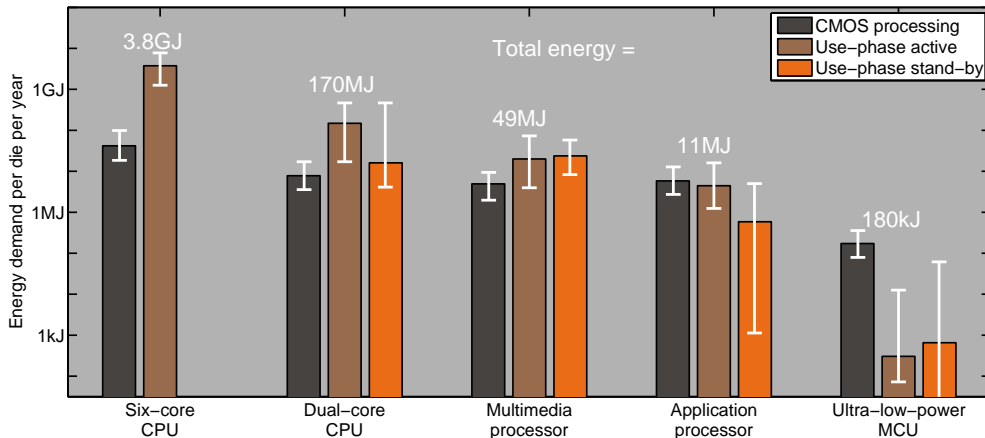


Fig. 5. Breakdown of the annual energy demand per die by life-cycle phases for the considered applications.

V. CONCLUSIONS

In this paper, we performed a life-cycle evaluation of the energy demand of computational logic chips typical of 5 standard applications from high-performance servers and laptops to low-power set-top boxes and smart phones to ultra-low-power sensors and RFID tags. We specifically evaluated the energy demand associated to CMOS processing of bare silicon wafers and to use-phase power consumption in both active and stand-by modes. This evaluation shows that the annual energy demand per die spans over more than 4 orders of magnitude: from 3.8GJ for server's CPUs to 180kJ for sensor's MCUs. Moreover, the dominant life-cycle phase is highly dependent on the application: use-phase energy in active mode dominates for high-performance chips whereas CMOS processing energy dominates for ultra-low-power applications.

REFERENCES

- [1] E. Williams *et al.*: "The 1.7 kilogram microchip: energy and material use in the production of semiconductor devices", in *Environmental Science Technologies*, vol. 36, pp. 5504-5510, 2002.
- [2] N. Krishnan *et al.*: "A hybrid life cycle inventory of nano-scale semiconductor manufacturing", in *Environmental Science Technologies*, vol. 42, pp. 3069-3075, 2008.
- [3] S. Boyd *et al.*: "Life-cycle energy demand and global warming potential of computational logic", in *Environmental Science Technologies*, vol. 43, pp. 7303-7309, 2009.
- [4] S. Boyd *et al.*: "Life-cycle assessment of computational logic produced from 1995 through 2010", in *Environmental Research Lett.*, vol. 5, 8 p., 2010.
- [5] S. Boyd *et al.*: "Life-cycle assessment of semiconductors", Ph.D dissertation, *University of California Berkeley*, 135 p., 2009.
- [6] N. Kurd *et al.*: "Westmere: a family of 32nm IA processors", in *Proc. IEEE Int. Solid-State Circ. Conf.*, 2010, pp. 96-97.
- [7] J. Oliver *et al.*: "Life cycle aware computing: reusing silicon technology", in *IEEE Computers*, vol. 40, pp. 56-61, 2007.
- [8] G. Gerosa *et al.*: "A sub-2 W low power IA processor for mobile internet devices in 32 nm high-k metal gate CMOS", in *IEEE J. Solid-State Circ.*, vol. 44, pp. 71-82, 2009.
- [9] G. Gammie *et al.*: "SmartReflex power and performance management technologies for 90nm, 65nm, and 32nm mobile application processors", in *Proc. IEEE*, vol. 98, pp. 144-159, 2010.
- [10] Y. Kikuchi *et al.*: "A 222mW H.264 full-HD decoding application processor with x512b stacked DRAM in 40nm", in *Proc. IEEE Int. Solid-State Circ. Conf.*, 2010, pp. 326-327.
- [11] Texas Instruments: "MSP430F543xA, MSP430F541xA Mixed-signal microcontroller", datasheets, available at <http://focus.ti.com/docs/prod/folders/print/msp430f5418a.html>, 2010.
- [12] NXP Semiconductors: "LPC110 Cortex-M0 based microcontrollers with industry-leading power and efficiency", product brief, available at <http://ics.nxp.com/literature/leaflets/microcontrollers/pdf/lpc11c00.pdf>, 2010.
- [13] J. Kwong *et al.*: "A 65nm sub- V_t microcontroller with integrated SRAM and switched-capacitor DC-DC converter", in *IEEE JSSC*, vol. 44, pp. 115-126, 2009.
- [14] D. Bol *et al.*: "Interests and limitations of technology scaling for subthreshold logic", in *IEEE Trans. VLSI Syst.*, vol. 17, pp. 1508-1519, 2009.
- [15] F. Boeuf *et al.*: "A conventional 45 nm CMOS node low-cost platform for general purpose and low power applications", in *Proc. IEEE Int. Electron Device Meeting*, 2004, pp. 425-428.
- [16] S. Ekbote *et al.*: "45nm low-power CMOS SoC technology with aggressive reduction of random variation for SRAM and analog transistors", in *Proc. Symp. VLSI Tech.*, 2008, pp. 160-161.
- [17] R. Watanabe *et al.*: "A low power 40nm CMOS technology featuring extremely high density of logic (2100kGate/mm²) and SRAM (0.195 μ m²) for wide range of mobile applications with wireless system", in *Proc. IEEE Int. Electron Device Meeting*, 2008, pp. 1-4.
- [18] R. Watanabe *et al.*: "High performance 32nm logic technology featuring 2nd generation high-k + metal gate transistors", in *Proc. IEEE Int. Electron Device Meeting*, 2009, pp. 659-662.
- [19] K. Roy *et al.*: "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits", in *Proc. IEEE*, vol. 91, pp. 305-327, 2003.
- [20] I. Koren and Z. Koren: "Defect tolerance in VLSI circuits: techniques and yield analysis", in *Proc. IEEE*, vol. 86, pp. 1819-1836, 1998.
- [21] S. Boyd *et al.*: "Life-cycle assessment of NAND Flash", in *IEEE Trans. Semiconductor Manufacturing*, vol. 24(1), 2010.
- [22] L. A. Barroso and U. Hözl: "The case for energy-proportional computing", in *IEEE Computers*, vol. 40, pp. 30-37, 2007.
- [23] A. Chandrakasan *et al.*: "Low-power CMOS digital design", in *IEEE J. Solid-State Circ.*, vol. 27, pp. 473-484, 1992.
- [24] M. Chang *et al.*: "Transistor- and circuit-design optimization for low-power CMOS", in *IEEE Trans. Electron Devices*, vol. 55, pp. 84-95, 2008.
- [25] B. Nikolic: "Design in the power-limited scaling regime", in *IEEE Trans. Electron Devices*, vol. 55, pp. 71-83, 2008.
- [26] A. Mahesri and V. Vasdhan: "Power Consumption Breakdown on a Modern Laptop", in *IEEE Computers*, PACS 2004, LNCS 3471, pp. 165-180, 2005.
- [27] S. Jarp *et al.*: "Evaluation of the Intel Westmere-EP server processor" and "Evaluation of the Intel Nehalem-EX server processor", white papers available at openlab-mu-internal.web.cern.ch, 2010.
- [28] D. Atlavila: "Intel Arrandale Core i5 and Core i3 mobile unveiled", available at hothardware.com/Reviews/Intel-Arrandale-Core-i5-and-Core-i3-Mobile-Unveiled, 2010.
- [29] C. Angelini: "Intels mobile Core i5 and Core i3: Arrandale is for the rest of us", available at www.tomshardware.com/reviews/mobile-core-i5-arrandale,2522.html, 2010.
- [30] R. Islam *et al.*: "Power reduction schemes in next generation Intel ATOM processor based SoC for handheld applications", in *Proc. Symp. VLSI Circ.*, 2010, pp. 173-174.
- [31] Y. Yuyama *et al.*: "A 45nm 37.3GOPS/W Heterogeneous Multi-Core SoC", in *Proc. IEEE Int. Solid-State Circ. Conf.*, 2010, pp. 978-979.
- [32] L. Nazhadili *et al.*: "SenseBench: Toward an accurate evaluation of sensor network processors", in *Proc. Workload Characterization Symposium*, 2005.