

Strategies for Reducing Power Consumption and Increasing Reliability in IoT

Ricardo Reis¹

¹ Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91501-970 – Porto Alegre – RS – Brazil
reis@inf.ufrgs.br

Abstract. The Internet of Things (IoT) demands new challenges in the design of computing and electronics components. One of the challenges is the power reduction of this expanding network of connected devices, where the majority is permanently connected. In a large set of applications, another significant issue is reliability, especially on critical areas as health and transport. This paper shows an overview of design strategies that we have developed to reduce power consumption and to increase reliability in circuits that are components of the IoT, as the reduction of the number of transistors in IoT devices, using optimisation techniques and the physical design of circuits tolerant to radiation effects.

Keywords: Internet-of-things, Optimization, Physical Design, Fault Tolerance, Radiation Effects, Nanoelectronics

1 Introduction

The growing number of connected devices in the Internet of Things (IoT) is one of the reasons for the ever increasing increase in the number of transistors produced annually in the world. Figure 1, based on [SIA 2005], shows the number of transistors manufactured annually in the world, year by year. This impressive growth is due to 3 main factors: the increasing number of transistors integrated into a chip, the growing number of products that include embedded chips and the increasing number of manufactured copies of each product. The manufacturing cost of a transistor is relatively cheap. In [The Economist 2010] a comparison is presented between the cost of a grain of rice and the cost of a transistor. The price of a rice grain can be equivalent to the manufacturing cost of more than 125,000 transistors. This would indicate that there is no need to optimise the number of transistors in a design, since the cost of them is relatively small. But the cost of energy required for the operation of a transistor is increasing a lot. We also have to consider that a high-power consumption can reduce the lifetime of a system, as well as increase the effects of variability that can cause an integrated system to malfunction and/or also reduce its useful life. With the increasing connection of electronic and computational devices on the Internet, that is, in the Internet of Things, power consumption problems tend to get worse, and a lot. How much Power Plants we will need to cope with the IoT/IoE (Internet of Everything) world? This is a major issue.

So, an essential keyword on the Internet of Things is **optimisation**, especially the optimisation of power consumption, which must be addressed at all levels of abstraction in the design flow of a computer or electronic system. The total power optimisation is a summation of the optimisation done at each level of design abstraction. So, sustainable computing requires optimisation at all design levels of a computer or electronic system design.

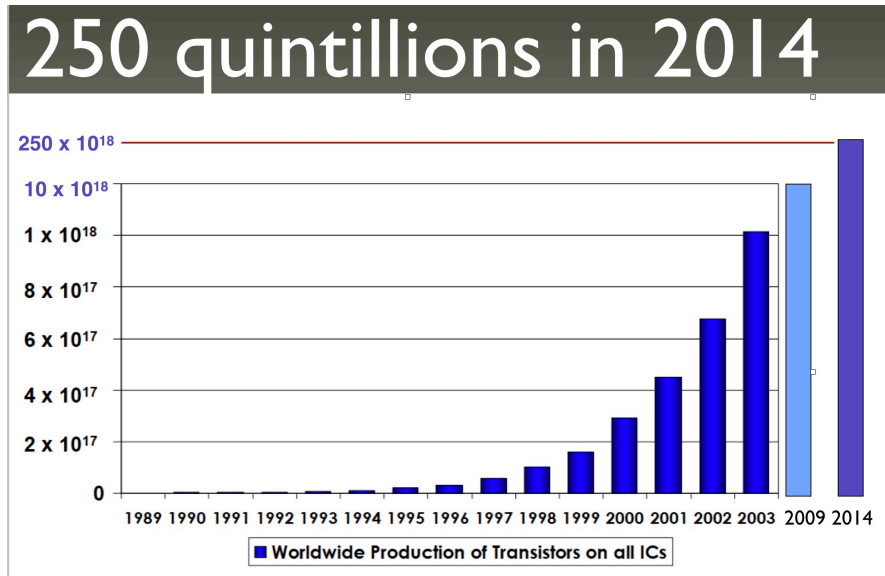


Fig. 1. Number of transistors produced annually in the world [adapted from SIA 2005]

2 Internet of Things

The term Internet of Things has already given rise to several other associated terms, such as the Internet of Health (IoH), Internet of People (IoP) and the Internet of Everything (IoE). In fact, the latter term becomes the most comprehensive, but each one of the others has some specific characteristics. When talking about the Internet of Health, which includes real-time monitoring of a person's clinical conditions, as well as chips injected in a person, the issue of reliability is a key one. And reliability is also related to power consumption in most cases. High power consumption can reduce the lifetime of a system. When it comes to the Internet of People, the issue of people's security and privacy is of great relevance. But in all cases, the importance of optimising energy consumption is growing more and more.

When considering optimisation, it means that integrated systems must increasingly be dedicated to the intended application to optimise the number of components, that means the number of transistors. Another important strategy for optimisation is the

hardware and software codesign, where one can manage the compromise between performance, consumption, and reliability.

Devices connected to the Internet of Things (or the Internet of Everything), can have very different complexities. If it is analysed the complexity considering the number of components, we can find small devices with few transistors and large devices with billions of transistors. Of course, large devices will consume much more power, but we have to consider that most devices on the Internet of Things are devices with a low number of transistors. But, because they are found in large quantities, they can represent a total consumption more important than the consumption of the so-called large devices that are present in a lower number. Therefore, consumption optimisation must be performed on both large and small devices that are present in large quantities. Another aspect to consider is that some devices require the application of reliability techniques (such as those related to transport or health systems), which can increase the number of components, while other devices are not critical, such as a camera or video, where an error in viewing a pixel of an image does not cause significant problems.

Also, we can expect that many systems connected to the Internet of Everything will be Cyber Physical Systems (CPS), that are systems composed by different classes of components like electronic elements, mechanical elements, optical elements, physical sensors, chemical sensors, organic components, and many others. So, it is needed to obtain EDA tools to cope with the design of CPS composed of all these classes of devices.

Figure 2 [The Connectivist 2014] shows an estimate of the number of devices connected to the Internet since 1992 when they were about 1 million devices. By 2020 when it is estimated that there will be more than 50 billion devices connected in the network, and there are currently around 35 billion connected devices. In [Ihsmarkit 2018] the number of devices connected to the network in 2018 is shown by industrial and commercial sectors, where almost half is in the area of communication. The significant growth in the number of connected devices to the Internet has naturally led to a considerable increase in the energy consumed in the Internet of Things. For how long will we have the energy to meet this growing demand? Therefore, it is necessary to use techniques to minimise the energy consumption of each connected device in the Internet of Things.

The Internet of Health (IoH) is a significant way to increase the life of human beings but also to improve life quality. Some of the examples of devices to be connected to the IoH are: Glasses that can advise eye correction; Toothbrush that can find cavities and breath issues; Razor that identify acne; Pacemakers that broadcast data to cardiologist; Underwearables that can provide early detection of cancer and other anomalies; Combs that can scan for fungus and hair loss; Earphones that does measurement of hearing, analysis of emotional level; Watches able to measure parameters like blood pressure, heart rate and others.

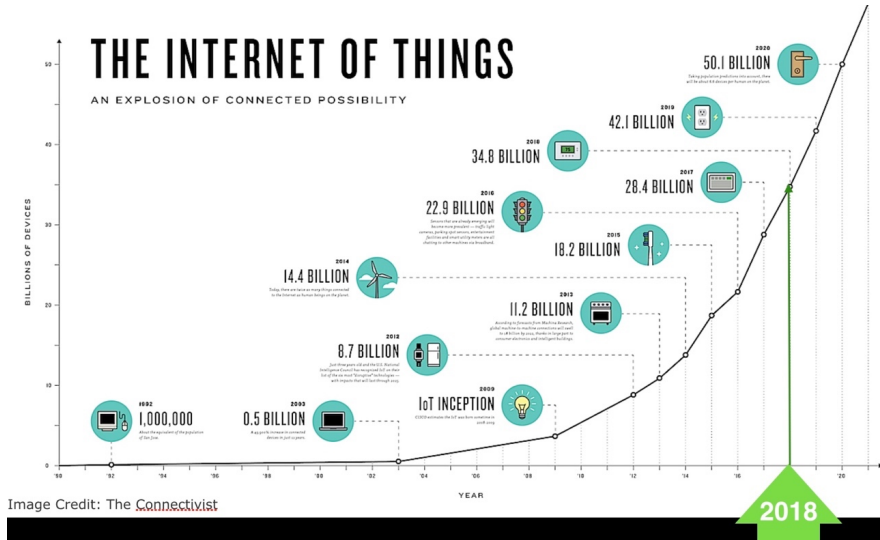


Fig. 2. Number of devices connected on the Internet [adapted from The Connectivist 2014]

In critical areas such as the design of implanted devices (chips) in humans (Figure 3), the reliability of the implanted systems is obviously critical. Some of the techniques used are based on the triplication of circuits and the temporal analysis of the propagation of a signal. Previously, the design of fault-tolerant circuits, to cope with radiation effects, was mainly in circuits that were sent to space. With the reduction of the value of the supply voltage of integrated circuits, nowadays the integrated circuits for use at ground level are also sensitive to errors caused by the radiation incident on the earth. Therefore, in critical areas such as implanted chips in humans, it is necessary to implement radiation effects tolerance techniques [Velazco 2007]. Also, critical systems used on the Internet of Health should be tolerant to any kind of noise (internal or external to the human body). They also must have a larger lifetime as possible, for obvious reasons and also should cope with environmental variability.

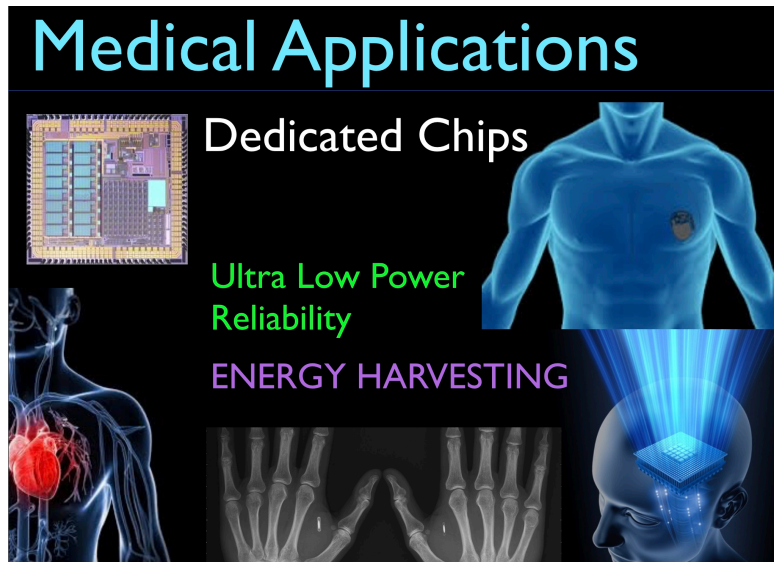


Fig. 3. The implantation of Chip Systems in humans demands reliability and ultra-low consumption

Also, there is the effect of "ageing", that is, the ageing of the circuit, which is more eminent in nanometric technologies [Vasquez et al 2012]. One of the most important effects is known as NBTI (Negative Bias Temperature Instability) that alters the threshold voltage of the PMOS transistors, degrading the operation of the transistor. Another effect that causes failures in circuits throughout their life is the effect of electromigration, which can cause short circuits or rupture of connections (Figure 4). In order to increase the lifetime of the chips, it is necessary to use physical design techniques that reduce the probability of electromigration [Posser 2017]

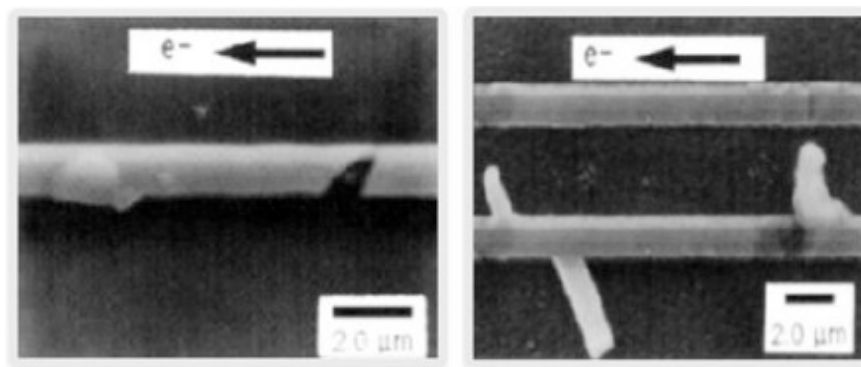


Fig. 4. Exemplo of a Void (open circuit) and hillock (short circuit) (GEDEN, 2011).

3 Electronic Design Automation (EDA) Tools

The use of EDA tools is essential for optimising energy consumption and increasing reliability, as the design flow has a large set of steps as well the number of components of a cheap can reach billions off transistors. In Figure 5 we can see the floorplan of an integrated circuit, where the hotter colours show regions (hot spots) with higher energy consumption, indicating that in some points there is a significant concentration of power consumption. One way to deal with the problem is to modify the placement of the logic cells in the circuit to distribute the cells with the highest energy consumption over the entire circuit area. But this must be done without compromising the area, wire-length and operating frequency specifications (much depends on the routing). Another way is to decrease the number of transistors, since the static consumption is related to the number of transistors [Reis 2011A].

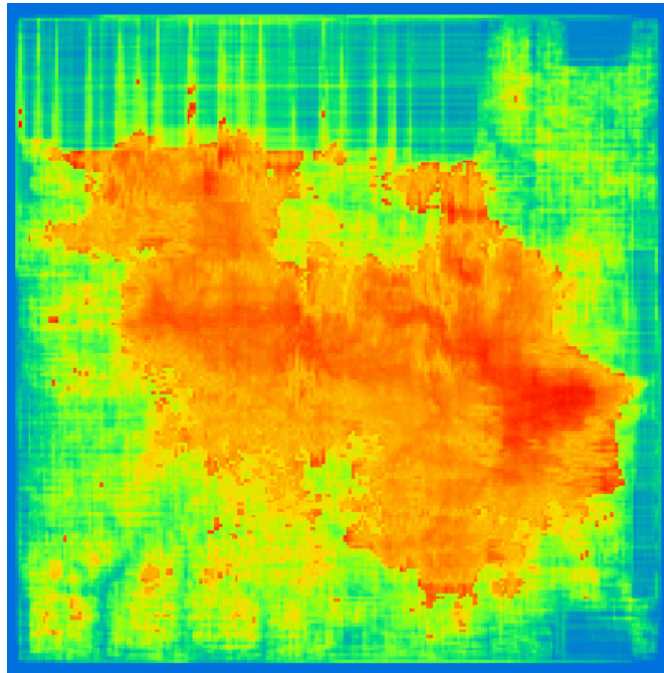


Fig. 5. View of the consumption density on a chip

4 Power Consumption Reduction by Reducing the Number of Transistors

The reduction of the power consumption of a System on a Chip (SoC) is a function of a sum of techniques and strategies of design applied in different levels of abstraction in the design flow of an integrated system [Reis 2010]. The summation of the gains is that

it will set the total gain in power reduction. When we deal with the physical synthesis of a system on a chip, one technique is the optimisation of the number of components, that is, the number of transistors. In Figure 6 [Reis 2011A] we can observe two solutions for the implementation of the same equation. The first solution makes use of 4 basic logic gates (3 NOR 2-input ports and one CMOS inverter), using a total of 14 transistors. The second solution makes use of only one logic gate, which performs the same function but with only 8 transistors. That is, the second solution, having a reduction in the number of transistors, will also have a proportionally smaller static power consumption. Furthermore, in the example of Figure 6, we can see that the first solution also has 3 connections between the basic gates (and therefore even vias and contacts) that are eliminated in the second option with only one logic gate.

This elimination of connections is increasingly important because it decreases the number of connections to be implemented using the different metal layers. The decrease in the number of connections decreases the density of connections and, therefore, increases the routability of the circuit and also contributes to reduce the average length of the connections, which implies in a reduction of the delay. In modern technologies, the delay in connections is so or more significant than the delay in the switching of logic gates. A greater spacing between the connections also contributes to an increase of reliability, due, for example, to the reduction of the possibility of electromigration, as already mentioned above.

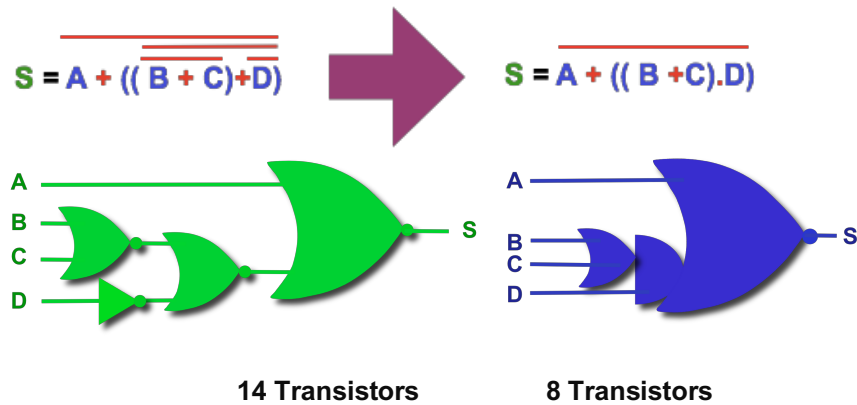


Fig. 6. Two options for the implementation of the same function [2011A Reis]

The reduction of the number of transistors depends on the use of efficient Electronic Design Automation (EDA) tools that transform the logical equations of a system so that in addition to mapping equations in CMOS gates, make optimum use of complex logic gates. In [Conceição 2016] we present a tool to reduce the number of transistors in a circuit through the fusion of networks of transistors that present fanout equal to 1. Also, it is fundamental the use of an automatic synthesis tool that can perform the automatic layout of any logical function. There is no use to achieve a logical optimisation if it is necessary to map (transform) the equations according to the logic gates available in a traditional cell library [which have few functions, in general, no more than 100 functions], as is still done when using traditional EDA systems. This mapping step is called technology mapping, and it represents a step of deoptimization. With this aim, we have developed automatic layout synthesis tools such as ASTRAN (Ziesemer 2015) (Figure 7), which allows automatic generation of the layout of any network of transistors [Reis 2011].

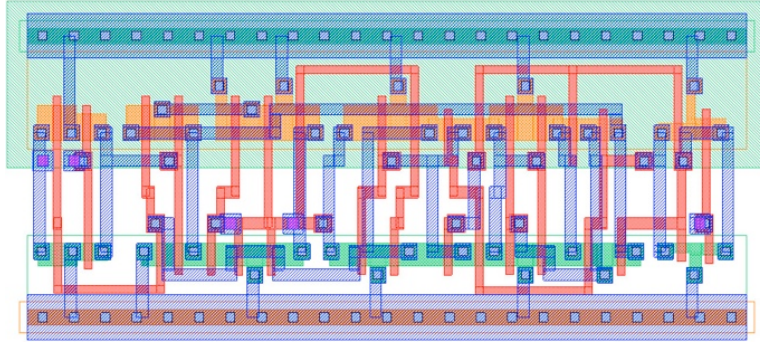


Fig. 7. Transistor Network Layout Generated Automatically [Ziesemer 2015]

Another technique to reduce consumption is through the sizing of the transistors. Modern integrated circuit manufacturing technologies show a significant increase in static power consumption that is often greater than dynamic power consumption. One way to mitigate power consumption, especially the static one, is to carry out a sizing of transistors to optimise power consumption. In [Reimann 2016] significant decreases in consumption are obtained through the use of automatic transistor sizing tools. This is also called cell selection, where the cells are selected from a cell library. In this case, cell selection means the selection of cells with a specific size and V_{th} (threshold voltage). In traditional cell libraries, one function has in general 3 sizings (one for less area, one for less power, and one for less delay) and 3 V_{th} (threshold voltage).

5 Reliability

As in the reduction of power consumption, in the design of critical systems, it is needed to use techniques to increase reliability at different levels of design abstraction. At the

architectural level, a very applied method is the redundancy of modules, especially triple module redundancy (TMR) [Kastensmidt 2006]. Another is the temporal redundancy [Nicolaidis 1999] where a signal traverses two paths, one with higher delay and another one with less delay. The difference of delay must be longer than the duration of a transient. Comparing the signal after traversing the two paths indicates whether there has been a transient propagation or not. At the physical level, we can apply different techniques to reduce or avoid problems such as electromigration [Posser 2015]. In the example of Figure 8, the position of the output pin in the centre (point 4) increases the lifetime of the circuit because it allows reducing the maximum density of current in the segments of the metal layer.

In [Velazco 2007] it is presented a series of works aimed at mitigating the effects of radiation on integrated circuits. In [Kastensmidt 2006] [Neuberger 2014] [Gennaro 2017] [Aguar 2016] [Lazzari 2011] we present some of the results that our research group has obtained in the development of techniques aiming the design tolerant to faults due to transients, as the effects due to radiation.

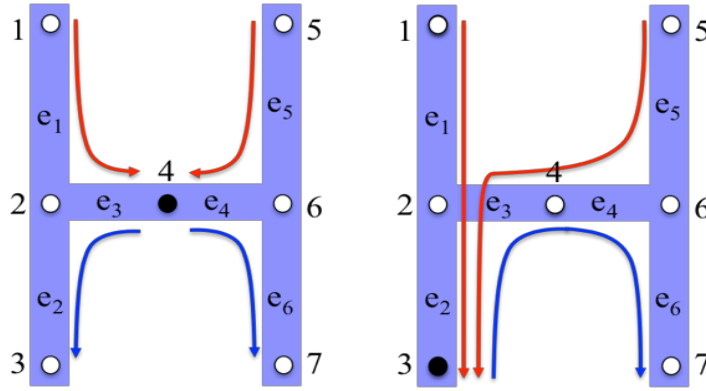


Fig. 8. Changing current density with the change of position of the output pin [Posser 2015]

6 Hardware Accelerators

The evolution of computer architectures, that today means, the evolution of microprocessor architectures has been very significant. In the 1970s, one marketing argument from microprocessor producers was the number of instructions that the microprocessor could execute as well as the clock frequency of the microprocessor. In the last decades, there has been a change of paradigm, discontinuing the race for the increase of the clock frequency, because the increment of the clock means an increase of the dynamic consumption. Instead, there was an increase in the number of cores (CPUs) aiming at increasing performance. Initially with homogeneous cores and later with heterogeneous cores.

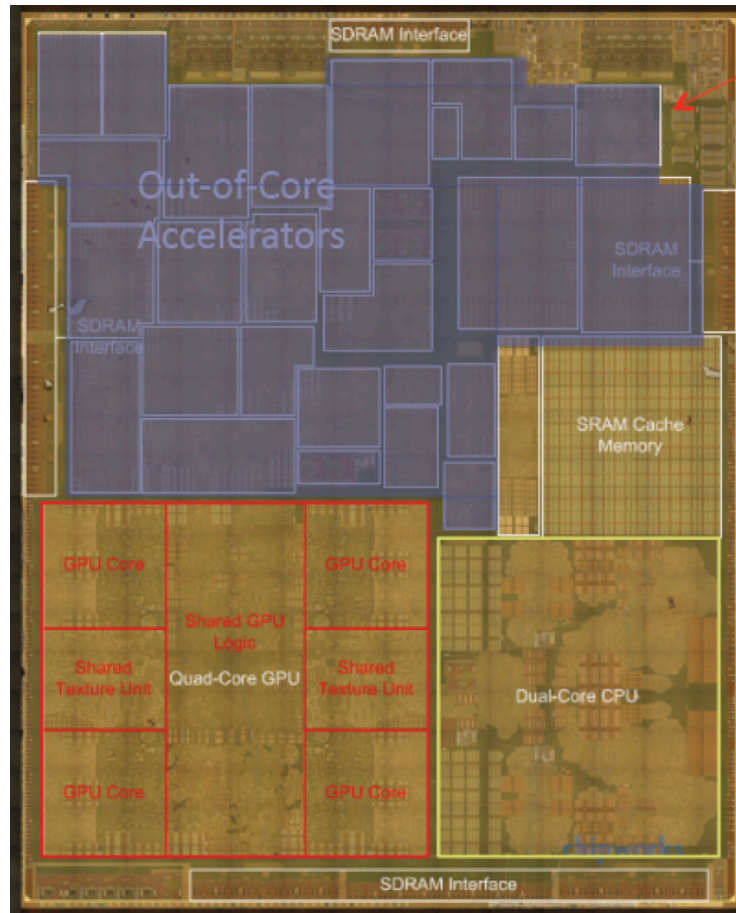


Fig. 9. Apple 8 Floorplan with 29 Hardware Accelerators [Anatech, 2014], [SHAO 2016]

Currently, we can find chips with multiple CPUs and several GPUs (as can be seen in Figure 9 [Shao 2016] showing the floorplan of the A8 microprocessor (from Apple). In this same figure, it can be observed that about half of the area is occupied with hardware accelerators, which are modules dedicated to the execution of a specific function. For example, an encryption module placed next to the output/input pins and which will encode the output data and decode the received data. So, the execution of this function will be faster, because it is done by a dedicated module (that means smaller) and with only the needed number of components to perform that function. It also will consume less power.

A more important fact is that the use of hardware accelerators leads to greater energy efficiency (allowing more sustainable computing), mainly due to the reduction in the number of components used to perform a function. At any given time, only the hardware accelerators in use at that time are being powered. So, the hardware accelerators that are not in use are disconnected from the power supply. This strategy is also known as "Dark Silicon". We can even predict architectures consisting essentially of hardware accelerators, with only one or two small CPUs to manage these hardware accelerators.

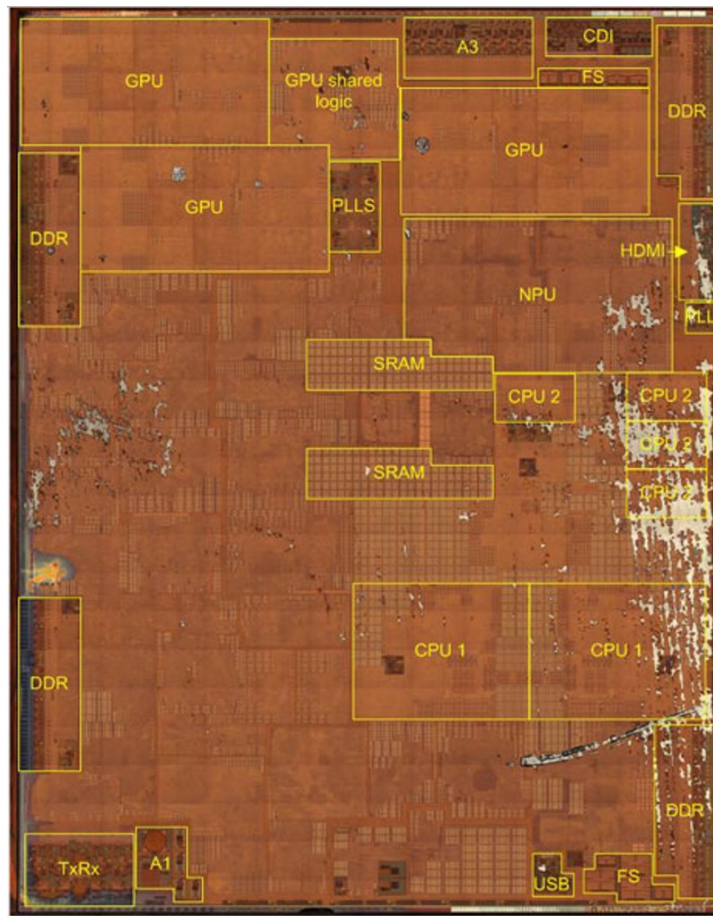


Fig. 10. Apple 11 Floorplan with an NPU [Techinsights, 2017]

The introduction of an NPU in A11 is another element characterising the heterogeneity of the SoC (chip system). And we can expect increasingly heterogeneous architectures, with dedicated modules for different operations to be performed by a SoC. In Figure 10 [Techinsights, 2017] the floorplan of the Apple A11 microprocessor is presented, where one of the modules is an NPU (Neural Processing Unit). NPU is mostly dedicated to

facial recognition [Techinsights, 2017], processing machine learning tasks more efficiently, consuming less energy than CPUs do. The CPUs occupy about 15% of the area of the chip and 6 GPUs occupy about 20% of the area. Most of the area is filled with the hardware accelerators. That is, it is growing in the architecture of Apple microprocessors the use of hardware accelerators.

7 Conclusions

To have sustainable computing, when the number of connected devices in the Internet of Things is fast increasing, it is fundamental the design of devices optimised regarding energy consumption. Most of the chips produced today use much more transistors than necessary to perform a function. So, there is a significant space for the optimisation of the number of components. In many devices related to critical applications, the application of techniques for fault tolerance is also fundamental, as nowadays circuits at ground level can have faults due to radiation effects. The reduction of power consumption must be treated at all design abstraction levels in a synthesis flow of integrated systems, from the specification of them in high-level languages to the physical synthesis. It was presented several works that were developed to reduce the power consumption and increase the reliability of integrated systems on a chip, and more details are shown in the mentioned references. The keyword in the age of the Internet of Things is **optimisation**.

8 Acknowledgements

We thank CNPq, FINEP, Fapergs, and CAPES for financial support for the development of our team's work, as well as the master's and doctoral students of PGMICRO and PPGC and students of Scientific Initiation who have contributed to the research works that served as the basis for this paper.

References

1. AGUIAR, Y., ZIMPECK, A., MEINHARDT, C., REIS, R. (2016), "Permanent and Single Event Transient Faults Reliability Evaluation EDA Tool", *Microelectronics Reliability*, Volume 64, September 2016, Pages 63-67, published by Elsevier B.V., 2016. ISSN: 0026-2714.
2. ANANTECH (2014), <https://www.anandtech.com/show/8562/chipworks-a8>
3. CONCEIÇÃO, C., MOURA, G., PISONI, F., REIS, R. (2017), "A Cell Clustering Technique to Reduce Transistor Count", 24th IEEE International Conference on Electronics, Circuits and Systems – ICECS2017, Batumi, Georgia, December 5 - 8, 2017, p. 186-189, DOI 10.1109/ICECS.2017.8291996
4. GENNARO, R., ROSA, F., OLIVEIRA, A., KASTENSMIDT, F., OST, L., REIS, R. (2017), "Analyzing the Impact of Fault Tolerance Methods in ARM Processors under Soft Errors Running Linux and Parallelization APIs", *IEEE Transactions on Nuclear Science*, Volume: 64, Issue: 8, August 2017, ISSN: 1558-1578, DOI: 10.1109/TNS.2017.2706519

5. GEDEN, B., "Understand and avoid electromigration (EM) & IR-drop in custom IP blocks, Synopsys, 2011.
6. LAZZARI, C., WIRTH, G., KASTENSMIDT, F., ANGHEL, L., REIS, R. (2011), "Asymmetric Transistor Sizing Targeting Radiation-Hardened Circuits", *Journal on Electrical Engineering*, Springer, DOI10.1007/s00202-011-0212-8, June 2011.
7. KASTENSMIDT, F., CARRO, L.; REIS, R. (2006), "Fault-Tolerance Techniques for SRAM-Based FPGA", Springer. April 2006, 183 p., ISBN 0-387-31068-1
8. NEUBERGER, G., WIRTH, G., REIS, R., (2014) "Protecting Chips Against Hold Time Violations Due to Variability", Springer, 107 p., 2014. ISBN 978-94-007-2426-6. DOI 10.1007/978-94-007-2427-3
9. NICOLAIDIS, M. (1999), "Time redundancy based soft-error tolerance to rescue nanometer technologies". In: IEEE VLSI TEST SYMPOSIUM, 17., 1999. Proceedings... IEEE Computer Society, 1999. p. 86-94.
10. POSSER, G., FLACH, G., WILKE, G., REIS, R. (2011), "Gate Sizing Minimizing Delay and Area", ISVLSI2011. IEEE Computer Society Annual Symposium on VLSI, Chennai, India, July 4-6, 2011. p. 315-316, ISBN 978-0-7695-4447-2. DOI 10.1109/ISVLSI.2011.92
11. REIMANN, T., SZE, C., REIS, R. (2016), "Challenges of Cell Selection Algorithms in Industrial High Performance Microprocessor Designs", *Integration*, Elsevier B. V., Vol. 52, January 2016, Pages 347-354, ISSN: 0167-9260, doi:10.1016/j.vlsi.2015.09.001
12. REIS, R., (2010) "Redução de Consumo pela Otimização de Componentes", SEMISH 2010, Anais do 37º Seminário Integrado de Software e Hardware, Belo Horizonte, 21 a 22 de julho de 2010, p. 371-379, ISSN: 2175-2761.
13. REIS, R. (2011A), "Design Automation of Transistor Networks, a New Challenge". IEEE International Symposium on Circuits and Systems, ISCAS2011, Rio de Janeiro, Brasil, May 15-19, 2011. IEEE Press. p. 2485-2488, ISBN: 978-1-4244-9472-9. DOI 10.1109/ISCAS.2011.5938108
14. REIS, R. (2011B), "Power Consumption & Reliability in NanoCMOS", IEEE NANO, 11th International Conference on Nanotechnology, Portland, USA, August 15-19, 2011 (**invited talk**), p.711-714. ISBN 978-1-4577-1515-0, DOI:10.1109/NANO.2011.6144656
15. The Connectivist (2014), available at: <http://ow.ly/i/5vph6/original>
16. The Economist (2010), September 6, 2010.
17. SIA (2015), Semiconductor Industry Association, Rebooting the IT Revolution, available at <http://www.semiconductors.org/clientuploads/Resources/RITR%20WEB%20version%20FINAL.pdf>
18. IHSMARKIT (2018), IoT Trend Watch 2018, available at: https://ihsmarkit.com/forms/thankyou.html?efid=t+m2jEyFYkQYyoP3YvuHA==&&gasc_id=862037098&&gasc_label=scrXCLnM7m0Q6siGmwM
19. Techinsights (2017), <http://techinsights.com/about-techinsights/overview/blog/apple-iphone-8-teardown/>
20. VAZQUEZ, J., CHAMPAC, V., ZIESEMER, A., REIS, R., TEIXEIRA, I., SANTOS, M. e TEIXEIRA, P. (2012), "Delay Sensing for Long-Term Variations and Defects Monitoring in Safety-Critical Applications", IN: *Analog Integrated Circuits and Signal Processing*, Volume 70, Number 2, 249-263, February 2012, Springer, ISSN 0925-1030, DOI: 10.1007/s10470-011-9789-0.
21. VELAZCO, R., FOUILLAT, P, REIS, R. (2007), "Radiation Effects on Embedded Systems", Springer, June 2007. ISBN 978-1-4020-5645-1

22. Yakun Sophia Shao (2016), "Design and Modeling of Specialized Architectures, PhD Thesis, Harvard", May 2016. Available at: <https://ysshao.github.io/papers/shao2016-dissertation.pdf>
23. ZIESEMER, A., REIS, R. (2015), "Physical Design Automation of Transistors Network", *Microelectronics Engineering*, V. 148, p. 122-128, December 2015, Elsevier B.V., ISSN: 0167-9317, doi:10.1016/j.mee.2015.10.018