# Systems Architectural Challenges for Transitional and Compatible to CMOS Technologies in Giga-Scale Hardware Integration

**Sergio Bampi[1], Altamiro Susin[2], Ricardo Reis[1]**

[1] PGMicro/PPGC e Instituto de Informática – Univ. Federal do Rio Grande do Sul

[2] PGMicro e Depto. de Eng. Elétrica – Univ. Federal do Rio Grande do Sul (UFRGS)

Porto Alegre - Brazil

{bampi,reis}@inf.ufrgs.br, altamiro.susin@ufrgs.br

***Abstract.*** *The transition from silicon technologies to alternative ones has been touted as an inevitable development that will affect the architectural research agenda for computing. Henceforth, the computing scenario beyond silicon-based electronics is grounded as a technical grand challenge of enormous impact on society, not just as a research topic in Computer Science and Engineering. This paper is a contribution to the establishment of a Grand Challenge research agenda that addresses fundamental challenges in the field of nano- and microelectronics that definitely will impact Computing. This work proposes a likely scenario for hardware technology evolution and challenges for the next 20 years, which were not at all dealt with in the document of the Grand Challenges by Brazilian Computer Science, 2006. In that time frame both radically new and evolutionary technologies will emerge, evolve and will be gradually selected. This paper contends that such selection will occur to make those technologies compatible with nano-scaled CMOS in silicon, not to replace it entirely. We propose that transitional technologies will rather co-exist and be built upon a basic CMOS technology platform. Radically new devices at the 1-10 nm scale will most likely be built on a silicon substrate with the same technical requirements (such as cleanness, lithographic resolution, etc) of current CMOS industry. The authors propose a more concrete scenario for architectural challenges, which dismisses the belief that multidisciplinary research on very diverse technologies (from materials to abstract computing systems) will lead the computer engineering into a post-silicon era. Such multidisciplinary research is even more important today, but it is far from leading to a non-silicon scenario. The grand challenges from the Computer Science point of view are refined into a more realistic transition: total replacement of CMOS will not occur for at least 20 years, instead new forms of integration, hierarchically ordered from the micron-level, to sub-micron level (500nm to 100nm), down to nano-scaled transistors in silicon (further down to 10 nm). In this hierarchy, at the bottom, it is highly possible that disruptive molecular-level devices (self-assembled in the scale of 2 to 10 nanometers) will eventually be production-worthy for Giga- and Tera-scale devices integration. Structures like graphene-based carbon tubes or planes are the most viable candidates for molecular devices. There is little assurance that overcoming the major technology barriers will happen in the next 10 years for carbon electronics. In this paper the computer-systems relevant issues of systems power dissipation, hardware design complexity, resilience to systems failures and fraud are dealt with as the overwhelming, challenging computing research topics that will guide future research in computing architectures at the giga-scale integration beyond 2020.*

## 1. Introduction

CMOS (complementary metal-oxide-semiconductor) circuits on silicon substrates of various types (bulk, silicon-on-insulator SOI, strained silicon with silicon-germanium alloys, etc) have been the dominant devices on which ICT (digital information and

communication technologies) hardware and software have been developed for about 30 years. The integrated circuit and discrete silicon devices industries are even older, namely 50 years old in 2009 since Jack Kilby´s patent in 1959 [Kilby 1959]. The estimate is that the worldwide semiconductor industry reached US$261.9 billion revenue in 2008 [Gartner 2008], a 4.4 % decline from 2007, due to the 2008 economic downturn worldwide. In 2007 the revenue topped at US$273 billion per year, a dollar-mark that is forecasted to be surpassed only in 2011, after the economic recovery expected in 2010. Electronics manufacturing worldwide produces roughly US$1.6 Trillion per year, and it has grown about 3 percentage points over the average growth of the rest of the world-manufacturing sector, consistently in the period 1996-2008. IT services enabled by equipment produced by this industry will continue to grow at even higher rates, powered by software as-services and by the increasingly pervasive communication services that will grow above the average world GDP growth rate for the next decades. IT and wireless communications devices are forecasted to move into most objects the humans relate to. Chips with simple processors and communicators will be embedded into buildings and most life-related engines, as ubiquitous as paints or information ducts. The ever-decreasing cost of chips in this micro- and nano-worlds provides the efficiency gains that will make this IT explosion viable. The integration paradigm in semiconductor circuits – reliable and low-cost due to mass-production - will ultimately make it viable to connect $10^{11}$ to $10^{12}$ objects simultaneously to the internet, roughly hundreds of IP-powered systems or nodes per individual, in average.

The grand challenges in computer science will be shaped by this tera-scale number of complex embedded hardware, autonomous systems and communicating devices, all interacting much like general-purpose computers do today over the Internet protocols. What will be the key enablers to this internet-of-things? In our view, towards 2030 the main-stay of the electronics integration technology will continue to be nano-scaled CMOS devices manufactured on silicon wafers. To which disruptive, yet-in-research devices will be added, to be made compatible with silicon ultra-clean in-fab processing. For instance, solid-stage storage devices with 5-10 nm-sized devices for each 1-bit of RAM, integrated in the range of 256Gbits/chip is within reach of mass-production in the next 10 years. Still manufactured in CMOS. Also other types of sensors, which will be compatible with – if not on the same - silicon substrate as the dense CMOS devices, will bring the "sensing-and-computing" hardware integration into a great variety of products and applications. With great benefits resulting thereof for health-care, cheap communications, and intelligent machines. These will sense tens of environmental variables and will act more and more autonomously.

Mature electronic packaging technologies at the 10-100 micron scale will continue to evolve to provide increasing hardware power efficiency. The 3-D integration will make new systems into compact volumes – with silicon dies in them, interconnected by fine wires (10 μm to 40 $\mu$m wide). While this is an evolutionary scenario drawn in this paper, its enormous technical and scientific impacts in terms of building ever-more complex systems have to be modeled, designed and fabricated as computer systems in the giga-scale era. Most importantly, because such systems are of heterogeneous nature, in which sensing physical events is as important as communicating information over RF devices; all with many digital processors and dedicated software embedded at their core. Moreover, one can assert that the impact on Computer Science brought by disruptive device technologies, even at the molecular

level, is most often a change of focus at the system to model, instead of the introduction of a radically new scientific methodology in Computer Science. Hence, Computer Systems research will continue to thrive on this method: system modeling, verify/refine the computing model, actual design, and finally, realize and fabricate those complex systems. Designing them today is an extremely complex task, in which hardware integration capabilities in silicon already surpass the capacity to design complex systems-on-chip timely. Impacts on the ICT (information and telecommunication technologies) due to this heterogeneous integration will be enormous.

While this paper addresses the grand challenges for the global hardware industry with respect to giga- and tera-scale integration, a few points are made with respect to the local ICT industry in Brazil. Current capabilities of this local industry in this part of the world show shortcomings in the engineering expertise to drive into the computing components the intelligence needed in the future ICT products. The rather insignificant number of local innovative companies in nano- and microelectronics certainly affects negatively the competitiveness of the entire ICT industry in Brazil. The leaders in modeling, specification, designing and also making of those nano-circuits are poised to rip the most economic benefits in the world ICT market. Clever and IP-powered electronic devices will also unveil new large IT applications and markets.

This paper is organized as follows: in section 2 we address the Brazilian Computer Society [SBC 2006] grand challenge number 3 and the ITRS roadmap for the world electronics industry, revising the related work for understanding the issues presented to the grand challenge. Section 3 briefly addresses roadblocks to continuing down-scaling (size reduction) of circuits on planar structures and the emerging alternatives for silicon. Section 4 presents the Systems-on-chips challenges, which are more relevant to overcome when dealing with computer systems design. In Section 5 the impacts for the future of the hardware industry will be addressed, while section 6 concludes this paper.

## 2. Related Work

The SBC Society Grand Challenges Workshop in May 2006 has presented five grand challenges for advanced research in CS in Brazil towards 2016 [SBC 2006]. The "impacts on Computer Science research of the transition from silicon to new technologies" was singled out as one of the challenges for the decade following. Our paper contends that this transition is not at all possible in the timeframe of the next 10 years. Moreover, the most likely scenario in the global nano-electronics industry hints for transitional and non-replacing integrated technologies to be introduced with silicon devices – far from full replacement of the latter. For reasons that are dealt with in sections 3 and 4 of this paper, the SBC document referred to a transition that will not occur in 20 years, certainly not by 2016. An incongruence that is not technically based. The Society's document is oblivious with respect to the actual hardware grand challenges, including the design of computer systems on-chip. Which are by far the most relevant computer systems - today and well into the future. This paper proposes a considerable "*aggiornamento*" to the SBC grand challenges document.

The most comprehensive industry experts panel publishes the ITRS (International Technology Roadmap for Semiconductors) roadmap [ITRS 2007]. The roadblocks to future progression in terms of integration into a silicon surface are well

mapped, and several problems with yet unknown technical solutions are pointed out and updated yearly in that roadmap. The significance of those roadblocks is by no means saying that silicon will be replaced as the implementation technology – instead it is pointing to likely events that could be "*showstoppers*" for more integration onto silicon. Once overcoming them, the industry envisions the goal of fabricating memories with 128 to 256 Gbits/chip by 2020, compared to 8G -16G bits today. The industry forecasts these densities using nano-CMOS, still on silicon chips. This will evolve, as long as the physical channel lengths of 5 to 6 nm would be mass-produced by 2020 compared to 23nm of nano-CMOS 45nm technologies currently (2009) in production in selected state-of-the-art nano-electronics production facilities. This minimum physical length evolution is shown in Fig. 2, in a log y-scale, to show the enormous scaling since 1970. This dimension downscaling alone does not provide the full picture. A factor of 4 to 5 in the transistor length reduction, from today to 2020, can lead to a 15X to 20X integration gain, at best, if measured in logic gates per square millimeter of silicon area, considering the logic part of a system-on-chip design. The point is that logic gates are very different than digital cells for storing bits. For this reason, we foresee a large divergence in the semiconductor industry paths for two kinds of products: a) memory chips, b) processing (analog or digital) chips. For both types, the 3-D integration will provide another leap in terms of systems integration. With gains of the order of 10X - 100X in integration density, at a lower cost than planar-only downscaling and integration.

The ITRS industry panel assesses continuously the emerging devices, which are candidates to replace silicon. Figure 1 from ITRS conceptually separates the emerging devices in terms of the state variables used to enable in the physical world the digital state abstraction (charge, as in CMOS devices, phase state of a molecule or atom, and spin or magnetic number state of a given electron level or molecule, respectively). The devices in the red polygon (resonant tunneling device - RTD, single-electron SET switch) are non-conventional devices also fabricated on silicon planar technology. Currently the carbon nanostructures (single plane, carbon nano-tubes - CNTs, multi-walled tubes, etc) are the most likely candidates for the non-silicon transistor. These devices are being pursued at the research laboratories precisely because they can be manufactured by keeping compatibility with conventional planar CMOS silicon fabrication. Low dimensionality structures (2-D monoatomic layers) require an extremely clean surface on which to be built, like those of silicon wafers today. In fact, 2-D mono-layers of atoms on semiconductors were demonstrated in labs since the late 1970s. The aforementioned compatibility is the main reason as to why CNTs are considered the alternative transistor, replacing nanowires of silicon – or most likely sharing the silicon wafer surface with silicon nanowires over insulators. The alternatives in spintronics or molecular devices will not be described here, even though they deserve careful discussion as revolutionary alternative candidates in the future technology roadmaps. These emerging devices are being pursued in pre-competitive research phase exactly because they do not rule out the compatibility with current silicon technologies. These alternatives are in the infancy of materials and devices research, well behind the systems integration capability of current nano-CMOS technologies.

## 3. The "More Moore" Brick-Walls

The technical and physical roadblocks that are forecast for silicon technology evolution down the integration path have different natures and assumptions. They are divided into limitations of different natures, namely: i) economical ii) technical, iii) electrical, and iv) physical/materials. An example of an economical constraint is the capital costs required for system-on-chip components fabrication in mass volumes. The ICT market requires volume production. The R&D labs have working 5nm Si devices and carbon nano-tubes devices with smaller diameter – in proof-of-concept experiments, not production. Many alternatives are being explored in the semiconductor companies, mostly to be compatible not to replace CMOS. The complexity and cost issues become decisive when moving those to densely packed devices in high-volume manufacturing. High fabrication costs today already require larger and larger chip volumes to recover the capital costs incurred on fab infra-structure. Memory chips are one class of components that find very large market demand to justify such investments. The processors+memory (general purpose multi-processors indeed) paradigm are another type of such component. Integrated optical cameras are another class of product in high volumes today. Other products for high volume manufacturing are rare to emerge.

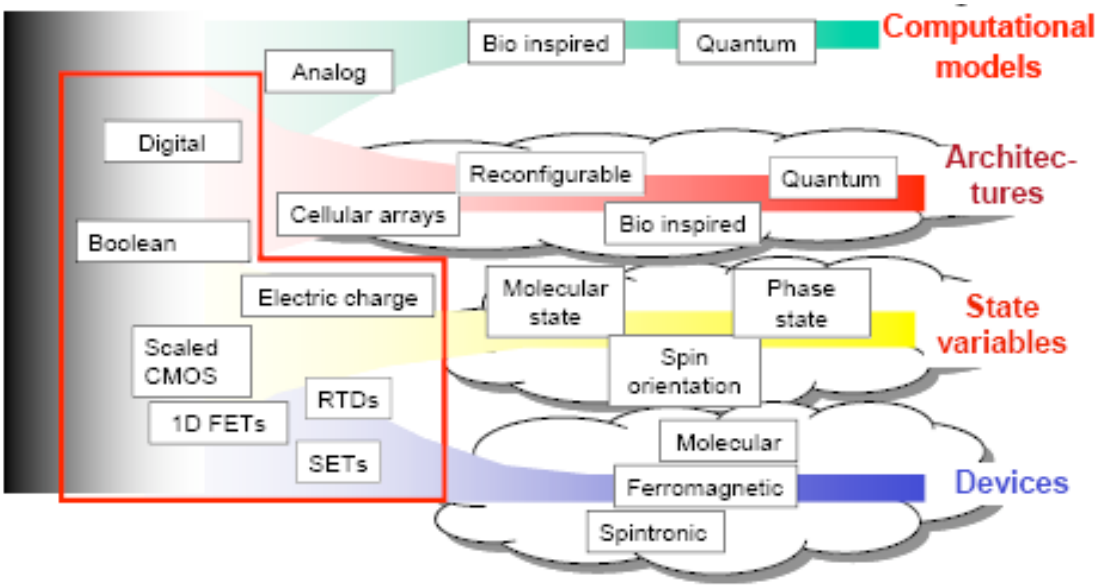


**Figure 1. Conceptual hierarchy of alternative computing devices (ITRS, 2007)**

The limits or "brick-walls" imposed to silicon scaling that are of a physical nature are shown in Figure 2 [Iwai 2007]. The oxide thicknesses in MOSFET transistors today have reached 1 (one) nanometer to 1.5 nm, which is the lower limit for direct metal-to-semiconductor electron tunneling through the insulator. This limit cannot be surpassed, and the solution was to replace silicon-dioxide by another insulator with a higher dielectric constant. This material solution is already in production at leading fabrication of 45nm nano-CMOS today. Device structures at 10nm sizes have quantum-mechanical behavior since the conduction band electron wave-length is typically around 10nm. This limit will be attained by the physical gate lengths (Lg) of production transistors before 2016, as marked in Figure 2. At 0.3 nm, this distance is just the separation between individual nuclei of the atoms in the solid, which is a physical scale

at which the electron orbitals are probabilistically distributed as quantum particles binding the atoms. This limit is marked in Figure 2, while it is clearly physically unattainable.

An example of an electrical limit today for MOS transistors is the ratio of on-state to off-state currents. To operate as a switch and to store bits in capacitors, this ratio in the MOS transistors has to be 100 or more. This in turn requires capacitor voltages around 200mV or more at room temperature, which is an electrical limit to the Vdd supply scaling that is shown in Figure 2 also. Currently high volume devices operate at 0.8V to 1V supplies. Giga-scale circuits have high power densities within their computing cores even at this low-supply. Reducing this energy supply is advantageous for lowering power, however it cannot be done due to the electrical limitations of off-to-on currents.

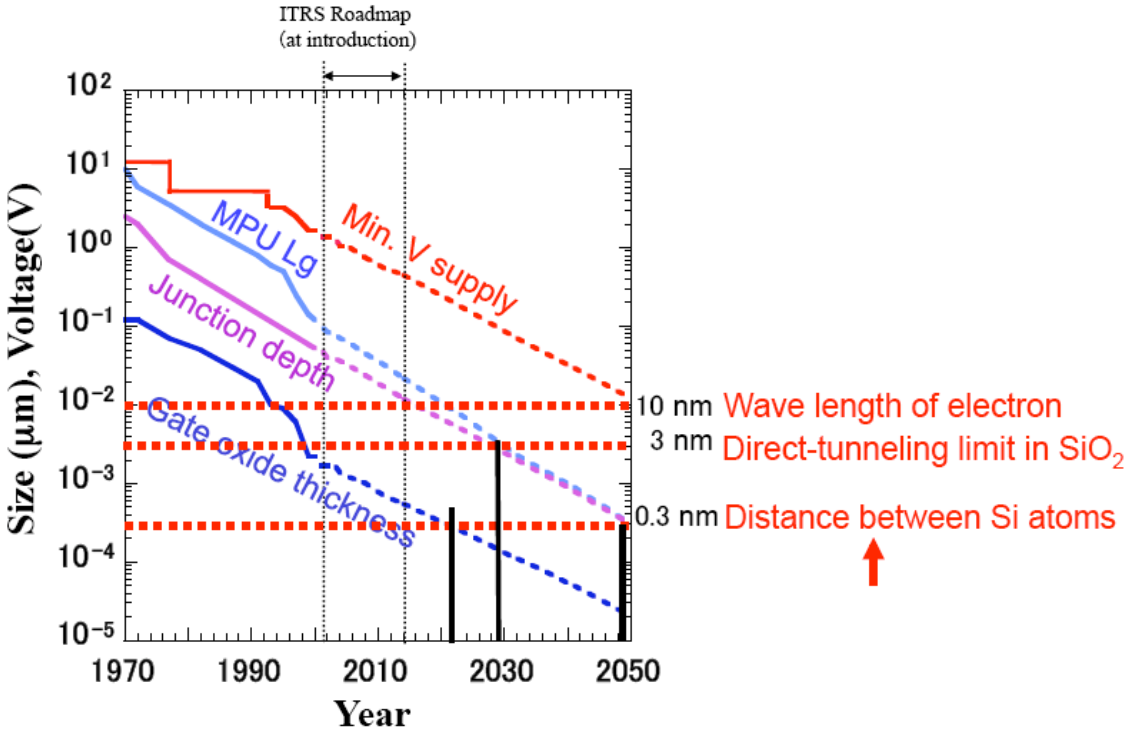


**Figure 2. Evolution of CMOS technology: the minimum supply voltage, oxide thickness and junction depth. Comparison to basic physical limits – electron wavelength in conduction-band and distance between silicon atoms (Iwai, 2007).**

Those are limits to the evolution of current CMOS silicon chips, which combined to the economics of this industry will eventually call for a "show-stopper" in terms of CMOS down-scaling in the Lg (physical gate) curve in Figure 2. The first ITRS roadmaps (ITRS 2007) were introduced to predict the evolution of the CMOS technology in the time period marked in Figure 2 by the top arrow. Some industry experts forecast that at 16nm CMOS node (for which 7 to 9 nm Lg will be the physical transistor length, the same magnitude of the in-band electron wavelength in Figure 2), the circuits will be at the limit of its technical and economical viability, at least for processors. And this could occur before 2020. Memory devices could experience even further down-scaling, hence an increasing divergence between memory and processor

CMOS technologies is forecasted by the authors. It is important that the challenges for CMOS downscaling be understood as a combination of factors of the natures above-mentioned: i) economical ii) technical, iii) electrical, and iv) physical or materials.

While those roadblocks combined may hinder the mainstream use of 16nm CMOS in 2020, they will not drive silicon technology out of the mainstay, or make current silicon technologies less viable. The "more-Moore" brick-walls will not be dealt with in detail here, since they are most out of the scope of architectural challenges. The relevant point is that no other economically and technically viable integration technology other than CMOS is within sight. The economic gains of the hardware systems integration will be moved into the "more than Moore" devices, which are addressed in the next section.

## 4. Grand Challenges for Systems-On-Chips Integration

The most important computing devices in the future will be at the leaf-end of the Web, as always-on, ambient-aware micro-devices. For their impact to be far-reaching, they will have to be compact, resilient to failures and attacks, communicating through RF. They are ubiquitous, in most senses of their operation. This will be the economic driver to sustain the growing of the computing infrastructure into most objects. The main challenges for the design of these new machines are listed below, according to the characteristics of the hardware architectures that will need considerable research and breakthroughs.

### 4.1. Heterogeneous Hardware Integration

Silicon integration technology will not reach the end of affordable fabrication on planar structures on-wafer in the foreseeable future. Going below 10nm CMOS in all film levels is still uncertain for technical reasons aforementioned. The new trend of 3-D, multi-stack, multi-die integration will be an enabling technology that will push the integration level not only a factor of 100 over conventional planar silicon: it will allow heterogeneous integration of systems-in-package that will contain: transducers to electro-optical devices, sensors, actuators, besides the usual processor and memory subsystems. The so-called more-than-Moore path to integration will provide more complex and new functions of high value for the IT devices. The 3-level stack of mass-produced digital cameras is today just an example of a system technology that has just begun. By 2020 3-D integration will make viable powerful massively parallel computers on few cubic centimeters, with 1K to 10K CPUs, for which the power efficiency and dissipation is the single most important design constraint.

An important impact on computing and communication will happen with the opto-electronics integration. Currently the transducers like lasers and photodetectors for serial optical communication links are done in III-V (like InP, InGaAsP) compound semiconductors. This is a niche market that may merge into 3-D packages with silicon substrates. Recent developments on photonic devices with nano-structured silicon show promising perspectives for optic emitters on silicon – an important breakthrough. This will provide more integration, not an optical computer, since these new devices are well suited for serial communication links, not information processing. The photo-electronics

systems in the future could move onto silicon substrates, narrowing the market for III-V semiconductor devices like LEDs and discrete semiconductor lasers.

### 4.2. Nano-power Computing for Autonomic Systems

The research on autonomic systems will strive on a technology that has yet to be discovered: computing systems that can reliably compute in bursts, at very low MIPS rates, and still be able to communicate to the next hierarchy at the network, using average nano-watts over long periods of time (years) and few micro-watts over short communication times. Heterogeneous integration will be key again to provide intelligent energy scavenging from the environment for such devices. This is a well-known research field in CS, which still requires breakthrough-engineering solutions at the circuit level.

### 4.3. Design and Architectural Complexity

The sequential computing paradigm and the ever-increasing complexity of systems-on-chip are considerable challenges to be overcome at the system design tools. Massively parallel systems are viable to integrate for general computing onto single chips. The applications and the parallel programming models are key to demonstrate an efficient use of such parallelism. The restrictions are well known since the mid-80s research on parallel – and room-scaled systems: While few parts could be computing, other parts sit idle wasting power. For this reason, the software stack necessary to exploit effectively the massively parallel systems also needs challenging innovation.

The application-specific complex systems can decompose the applications and map it beforehand to specific silicon cores. Some could be programmable cores, some certainly dedicated cores. These systems are, for instance, multiprocessors SoCs that are specifically designed for certain applications. This is viable for network processing, media processors (audio & video streams), and pattern matching processing. General computing is a different challenge though. Designing in parallel and making effective usage of hundreds of processors simultaneously is a methodology yet to be dealt with effectively in the computer science research community.

### 4.4. Simple IP objects – sense-compute-communicate

The future Net or Internet will have $10^{11}$ to $10^{12}$ objects interconnected – from PDAs to vehicles to mundane appliances. Communication hardware has to be cheap (integrated in silicon, with processors and memory), hence mass-produced. And this ubiquitous computing era will be enabled in silicon chips – not in molecular level devices, for certain. Molecular devices will find applications in dense memories at Tera-scale bits/cm$^2$, but for radiation fields for communication over 1m to 100m the technology requires tens of $\mu$W or even mW power, and this RF systems requires much larger silicon area than Mbits of storage. The first appliances to reach 100% connectivity are the communication and entertainment gadgets (like phones, audio and video systems) that empower individuals to communicate.

The market driver for electronics now and into decades ahead will be at the leaf-cells of this net-of-things, in local wireless personal networks (PANs). And these PANs will have simple objects that have to do simple tasks to empower the Net to a ubiquitous phase: to sense, to compute efficiently and to communicate with a small bandwidth to

the next IP hop. Most connected objects will have one IP address, but locally will interoperate with multiple sensors and computing devices – which are to be cheap and simple to design. For those, the conventional nano-scale CMOS, at the current level of 65nm or below is more than sufficient. In fact the 45nm CMOS technology available today is too expensive for most system needs in ubiquitous communication. Energy transducing at $\mu$W level requires area, not area reduction. Smaller, in this case, does not mean better. With sensors integration on-chip this is often the case also. High performance analog sub-systems, with large signal-to-noise ratio, also do not scale in total silicon area as they are designed in more advanced CMOS.

### 4.5. Electronic Design Automation Tools

The quality of the SoCs design is directly dependent of the quality of the EDA (Electronic Design Automation) tools available. So, the challenge is to find out algorithms that can emulate the skills and strategies of experienced designers and cope with the restriction and features of new fabrication technologies. The variability of recent and future technologies demands new design tools sets, tuned to new design methodologies, which could provide reliable SoCs even using non-reliable components. Also to minimize power and delay there is a strong request to reduce at most the amount of transistors used to implement a function. This request depends on the construction of an all-new physical design approach.

## 5. Impacts on The Global Hardware Industry

The ICT hardware is produced in complex industries for the global IT market. The electronics industry is globally a 1.6 Trillion US dollars industry, in which the key in its supply chain are the semiconductor chip industry and the software industry. The semiconductor industry accounted for around US$ 273 Billion revenues in the 2007 calendar year.

The integrated circuits (ICs) are high tech intermediate industrial goods, essential for all electronics. The system drivers for the IC industry can be separated in the following categories: a) portable/consumer devices; b) networking and communication; c) medical electronics; d) office and departmental computing devices; e) automotive, and f) defense electronics. The ICT hardware is the most important driver for the semiconductor industry. The new market drivers for ICT are the devices for the ubiquitous, ambient-aware, permanently-on computing devices. Mostly of light computing load, but extreme mobility and energy autonomy. The intelligent objects of the future InterNet will require cheap, and yet powerful and communicating devices. These everlasting requirements mean that silicon circuits on planar wafers will not become economically unviable in the foreseeable future. Quite the contrary, the semiconductor industry will be a key enabler of the next generation computing – and in all spectrum of computing power. Uninformed is the confusion to be made with the trend to draw and fabricate nano-wires onto silicon substrates – the silicon manufacturing for CMOS transistors below 10nm is technically feasible, although such manufacturing is currently ruled out as being too costly. And not viable with photolithography tools that today use deep UV (ultraviolet) 193 nm light sources to fabricate 50nm device structures. Hence, the experts in technology presently assert that it may not be viable to assemble 10 Billion dollars fabrication lines for very high volume of silicon, but they do not contradict the fact that 45 nm to 180 nm CMOS nano-

devices are economically the viable integration choice, now and into future generations. In this reasoning, the CMOS silicon technologies and their derivatives will be mainstream for an industry that reached 270 billion dollar yearly, and it is bound to increase.

On the systems design area, the impact of complex systems integration in silicon has already established a global economic paradigm for systems development: the design teams in the nano-electronics industry are increasingly global, and they require a tight integration of software embedding techniques and hardware design. The latter enabling the former. The main impacts for the industry in the future are:

a) Managing complex systems design of ever increasing complexity hardware in global, multi-national engineering teams of software and hardware developers. The industry goal is to manage the high Non-Recurring Engineering (NRE) expenses, as well as to develop products that reach high volume scale quickly. General computing/communication devices are by definition high volume in the global markets. Embedded software is essential to enable services and other country-specific features, while the chip devices will have to truly global enabler.

b) The arrangement of 16nm-22nm CMOS factories around inter-company alliances for manufacturing. The complexity and challenges of CMOS manufacturing are to be overcome only to the extent that the integration benefits override the competition and can financially offset the large investments for CMOS wafer fabs. Such alliances are growing since the inception of Sematech in the 1980´s, an organization that brings together development teams from tens of different leading semiconductor companies to develop the key technologies for the future of semiconductor manufacturing.

c) Design methodologies will have to incorporate "more-than-Moore" heterogeneous device design. Integrating optics, sensors, actuator, and the like will require more complex and diverse CAD tools, dealing with other domains, not just digital systems design.

d) The integration technologies of importance are both on-chip planar tooling, as well as multi-chip 3-D stacking with through-silicon vias. This system-in-package trend is to use silicon thinned-wafers with chips on top, similar to what a PC board does with copper. Hence, silicon will be replacing board layers, instead of being replaced by emerging devices.

e) The trend towards miniaturized components on a packaged 3-D system, of about 1 to 5 cm3, will be the mainstream in the 2020´s, or even earlier. This is a transitional technology that will not drive silicon out of the computing – it will instead thrive on it.

f) 3-D integration is a key technology that will further even more the divergence between manufacturing technologies for memory on-silicon, and manufacturing technologies for processors, sensors, actuators and RF front-end circuits.

## 6. Conclusions

This paper discussed the challenges for ICT hardware integration technologies. In our scenario we dismiss the "beyond silicon" jargon commonly used as a synonym for silicon technology full replacement. Future disrupting technologies will have to find compatibility with computing-on-silicon, 3-D stacking of dies or otherwise they will be

ruled out as unviable by the leading hardware manufacturers. Transitional technologies are likely to emerge, in which at the nano-scale semiconductor materials other than silicon (Si) will be used. Si substrate will still hold submicron-scale devices. This is a reality today, with germanium 2-D ultra-thin layers (sub-10-nm thicknesses) as part of 20nm CMOS transistors. Future transitional technologies will include carbon on different forms and shapes (tubes or planes called graphenes), and even light-emitting devices with nano-particles. All being held by a silicon matrix. Transitional technologies will enable new systems integration on silicon, even electro-optics integration.

In this paper the authors proposed that: i) nano-scaled CMOS on silicon will remain the basic technology well into the 2020s, and will still be the key for electronics systems innovation; ii) non-silicon technologies that are compatible with the silicon substrate and can co-exist in the silicon fabrication line will be the winner technologies for the next massive hardware platform of commercial significance for general computing: heterogeneous system-in-package, with the most valuable general computing and memory still relying on silicon chips with complex 3-D integration between dies; iii) the CMOS technology evolution will diverge into two CMOS tiers: one technology for very dense memories, tera-scale bits on-the-same-chip, for main memory, and the other CMOS tier for multi-processors and x-level caches on a single chip. This latter CMOS tier will be used for programmable logic devices, which will continue to play a key role in computing systems design well into the future.

The architectures with over 1K or 10K processors per 3-D package (less than 1 to 3 cm3) will be the main high-end computing engines for servers in the 2020s. The issues that are relevant to the local Brazilian industry are: how to move forward as the leading companies land elsewhere the commercial opportunities in computer systems design; and how to create value and new technology jobs by fabricating the devices that embed ICT intelligence and bring to fruition the capacity to envision future ICT products. Brazil currently lacks the industrial and engineering resource base to locally design and fabricate the key devices for the net-of-things generation. At least to produce them on the scale that the economics of this industry dictates, as envisioning is easy, making them is very challenging. In the 2020s and beyond, ICT will still be relying on silicon technology.

In conclusion, as a contribution to the SBC Grand Challenges document, we propose that the Grand Challenge (number 3) in the SBC document should be re-worded to: "***The Grand Challenge in computer systems research for the next 10 years is the impact in computing due to the evolution and heterogeneity of hardware implementation at the giga-scale integration of basic components***."

## 7. Acknowledgements

The authors acknowledge the support of CNPq and FINEP for their research in micro- and nano-electronics, as well as the PGMICRO and PPGC graduate students and faculty members at UFRGS Federal University. This work is being supported by CNPq through the Millennium Institute NAMITEC (Network of Excellence Centers) and by the current National Institute (INCT) on Nano-electronics, which has more than 15

Universities working in challenging aspects of micro and nano-circuits integration in Brazil.

## 8. References

SBC (2006) Brazilian Computer Society. "Grand Challenges in Computer Science Research in Brazil 2006-2016", 25pgs. In: http://www.sistemas.sbc.org.br. Last accessed Dec 22nd, 2008. Brazil (2006).

ITRS (2007), International Roadmap Committee. "The International Technology Roadmap for Semiconductors - 2007". In www.itrs.net. Last accessed March 16, 2009.

Kilby, John (1959) US Patent, 1959.

Gartner Group (2008) In: "http://www.eweek.com/c/a/Desktop...Chip-Revenue....Semiconductor-Industry-. In-2008. Last accessed Feb. 09th, 2009.

Iwai, Hiroshi (2007) "Future of CMOS Technology and Manufacturing". EEE DL talk at Federal University of Rio Grande do Sul, mimeo, 105 p.