POWER, INTERCONNECT AND COMPLEXITY CRISIS IN FUTURE VLSI: FROM A DESIGNER’S POINT OF VIEW

Anuradha
(ECE Department, Ajay Kumar Garg Engineering College/ U.P.T.U., India)
Anuradha.pughat@gmail.com

ABSTRACT

The VLSI design productivity crisis, that is, the fact that the number of available transistors grows much faster than the ability to design them meaningfully, has become the greatest threat to the growth of semiconductor industry. The cost-performance is rapidly improved as devices are scaled down. This is the reason why the miniaturization has been pursued persistently for these thirty years and will be pursued in the future. As semiconductor geometries continue to shrink and the chips themselves become ever-more complex, the cost of research and development continues to skyrocket. Recently, however, unfavorable effects caused by the scaling have become eminent. First, the power density increases. Secondly, interconnect related quantities increase such as interconnect delay, current density, and noise. Lastly, since the number of devices on a chip increases, the design and test of VLSI’s become more difficult. There are certain key issues that serve as active areas of research and are constantly improving as the field continues to mature. In this paper, power crises, interconnection crises and complexity crises and their possible solutions are described together with the possible view of the future VLSI’s.

Keywords – VLSI crises, interconnection problems, complexity in VLSI design, low power, More-than-Moore, complexity crises

1. INTRODUCTION

There have been several technological revolutions over the last 2000 years, including the invention of printing press in the 15th century, currently we are at the information technology revolution that started in about 1960, made commercially successful by the invention of transistors and integrated circuits. One of the big issues now facing the industry is how to best spend limited funding for research and development, when so much needs to be done. Continued scaling to 20nm and below, a transition to 450mm, 3D integration, device new structures, new materials with unknown integration challenges - these are all needed. Since 1970, the number of components per chip has doubled every two years and new design challenges are introduced. This historical trend has become known as “Moore’s Law” [1]. As the number of components (i.e. transistors, bits) per chip increases, the total chip size has to
be contained within practical and affordable limits (typical chip sizes should be <145 mm$^2$ for DRAM devices and <310 mm$^2$ for microprocessor units (MPUs)). This can be achieved by a continuous downscaling of the critical dimensions in the integrated circuit, which can be expressed in terms of Moore’s Law as a scaling by a factor of 0.7 ($\frac{1}{\sqrt{2}}$) every 2 years, where “critical dimension” is understood as “half pitch”, as defined in the International Technology Roadmap for Semiconductors” (ITRS) [2]. It’s now 2012 and, two years later, Moore’s Law is holding firm. With very large scale integrated (VLSI) circuit fabrication entering the deep sub-micron (DSM) era, devices and interconnection resources are scaled down to smaller sizes and placed at an ever increasing proximity. Most companies already lack the resources for true innovation, and the requirements will only increase. Will the cost of device development finally derail the predictions of Moore’s Law?

The automotive electronics and semiconductor market sees enormous worldwide demand for its products, which brings tremendous opportunity for semiconductor companies. The growing demand for higher speed and more energy-efficient electronics has forced IC designers to put considerable effort in decreasing the delay and energy consumption of VLSI systems using different circuit and architectural techniques. This increasing pressure to keep pace with the competition results in a reduction development time and the constant need for new and innovative products. A designer must be able to do physical design, synthesis, validation and testing. After an introduction about sources of variation, the next section 2 of this paper introduces More-than-Moore revolution, section 3 briefs interconnection crises, section 4 talks about power crises and finally, Section 5 discusses the challenges of different techniques under the extreme process variation effects of complexity crises.

2. **More-than-Moore**

Design and realization of microelectronic systems using VLSI/ULSI technologies require close collaboration among scientists and engineers in the fields of systems architecture, logic and circuit design, chips and wafer fabrication, packaging, testing and systems applications. It is clearly stated in the introduction to the ITRS executive summary: “a basic premise of the Roadmap has been that continued scaling of electronics would further reduce the cost per function […] and promote market growth for integrated circuits” as shown in fig.1.

![Fig. 1 The virtuous cycle of the semiconductor industry](image)

From a VLSI designer’s point of view, three crises exist in realizing VLSI’s in coming years, namely power crisis, interconnection crisis, and complexity crisis. In this paper, these crises and possible solutions are described together with the possible view of the future VLSI’s. However, semiconductor companies deal with the triple challenges of increasing complexity, shorter lead times and growing competition. In fact, the need for enhancing quality, creating product differentiation while meeting safety and other
regulations, is becoming more challenging. The industry is now faced with the increasing importance of a new trend, “More than Moore” (MtM), where added value to devices is provided by incorporating functionalities that do not necessarily scale according to "Moore's Law". These technologies typically enable non-digital functionalities, such as wireless communication, power management, sensing and actuation, and represent a growing part of the total silicon-based semiconductor market. The second trend is characterized by functional diversification of semiconductor-based devices as shown in fig. 2. These non-digital functionalities do contribute to the miniaturization of electronic systems, although they do not necessarily scale at the same rate as the one that describes the development of digital functionality. Consequently, in view of added functionality, this trend may be designated “More-than-Moore” (MtM). "More-than-Moore" based technologies have already made a considerable contribution to the worldwide microelectronics market, and the opportunities are huge. As system component integrates, drivers change according to the impact of More-than-Moore’s Law as shown in fig. 3.

It is believed that More-than-Moore will add value on top of and beyond advanced CMOS with fast increasing marketing potentials with perspective materials for future THz devices. The materials system for embedded sensors and actuators, perspective materials for monolithic heterogeneous integration, material systems for embedded innovative memory technologies, development of new materials with customized characteristics, through silicon via (TSV) approaches and 3D integration schemes are important key challenges for the realization of the "More than Moore" strategy in future.

Fig. 2 The combined need for digital and non digital functionalities in a integrated system is translated as a dual trend in the International Technology Roadmap for semiconductors: miniaturization of the digital functions (“More-Moore”) and functional diversification (“More-than-Moore”) [10].
3. INTERCONNECTION CRISES

Currently, increasing the number of metal layers in a multilevel metallization as the device size decreases increases wiring connectivity. In the future, however, designers will have difficulty catching up with the rising demand for higher wiring connectivity by merely increasing the number of metal layers. A new three-dimensional integration technology to overcome future wiring connectivity crises is proposed [3-4]. The huge amount of interconnection lines in VLSI makes the interconnect delay and crosstalk noise more dominant factors in the overall circuit speed [5]–[7]. Not only transistors but interconnects will be determining cost, delay, power, reliability and turn-around time (TAT) of the future VLSI’s. Some of the design issues of the deep submicron interconnects are as follows. The higher current gives rise to static and dynamic IR voltage drop problems and reliability degradation due to electro-migration. The smaller geometry and denser pattern lead to RC delay increase and signal integrity problems such as high crosstalk noise and large delay fluctuation due to capacitive coupling among adjacent lines. The number of metal layers and the interconnects be it global and local also tend to get messy at such nano levels. The higher speed causes inductance related issues and electro-magnetic interference (EMI) problems. The interconnect crisis is depicted in Fig 4.

Fig. 4 Interconnect determines cost and performance

Today, Automotive & Semiconductor industries are facing a very difficult economic environment, together with heterogeneity increasing product quality, increased dependence on outsourcing, growing competition, and an endless list of business partners. The huge
operation current is required in the future for high-performance VLSI’s. This type of high current develops voltage drop due to the resistance of power supply lines. Even for intermediately power-consuming chip needs very thick metal like 10µm to keep the IR drop within an acceptable level. This type of thick metal will be implemented in a package. In the future, area pads and co-design of a VLSI and a package will become necessary. The several vertically stacked chip layers in 3D LSI chips or 3D multichip modules (MCMs) are fabricated using a new integration technology [8]. More than 105 interconnections per chip form in a vertical direction in these 3D LSI chips or 3D MCMs. Consequently, wiring connectivity can be increased dramatically while reducing the number of long interconnections.3D-integration technology can evolve into a new system-level integration technology that combines LSI and future package and board-level integration technologies.

The copper interconnects cannot keep pace with the IC interconnect requirements as the feature size continues to scale down to nano-scale. Theoretical works predicted that carbon nano-tube (CNT) is more superior than copper for future VLSI interconnects in terms of electrical conductivity, thermal management and reliability. The experimental efforts on the controlled growth of aligned CNTs; the integrations of CNT interconnects with IC technology, the electrical characterization of the CNT interconnect and electro-migration test result of CNT-based interconnects are described [9]. Metallic carbon nano-tubes (CNTs) have received much attention for their unique characteristics as a possible alternative to Cu interconnects in future ICs. At 1000-µm global or 500-µm intermediate level interconnects, the delay of MWCNT interconnects can reach as low as 15% of Cu interconnect delay. The MWCNTs are easier to fabricate with less concern about density control, they can be attractive for immediate use as horizontal wires in VLSI, including local, intermediate, and global level interconnects [11]. The problems of scaling interconnects to nano-metric dimensions in future VLSI applications should be addressed when copper interconnects are compared to innovative interconnects made by bundles of metallic carbon nanotubes. A possible future scaled CNT bundle microstrip is analyzed and compared to a conventional microstrip [12].

Single-walled carbon nano-tube (SWCNT) bundles have the potential to provide an attractive solution for the resistivity and electro-migration problems faced by copper interconnect as process technology scales [13]. The results indicate that SWCNT interconnect bundles can provide significant improvement in delay over traditional copper interconnect depending on bundle geometry and process technology. In modern very large scale integration (VLSI) technology, efforts have been devoted to reduce metal wiring pitch to increase chip density and to save silicon budget [3]–[5]. The huge amount of interconnection lines in VLSI makes the interconnect delay and crosstalk noise more dominant factors in the overall circuit speed [6]–[8]. This makes metal wiring line resistance and line-to-line capacitance, thus the resistance-capacitance delay (RC delay) and interline crosstalk noise, increase. An accurate closed-form models have been developed for wire capacitance, wire delay and crosstalk noise [20]. In this model of metal interconnection, the closed-form formulas are derived for the wiring capacitance, delay and crosstalk noise, all as explicit functions of the wire thickness, dielectric thickness, interwire spacing and wire width. The capacitance model gives line-to-line and line-to-ground capacitances separately as shown in fig 5,6 and lead to precise delay and crosstalk estimations. This model is useful for VLSI design and process optimization.
The efforts are made to lower the resistance and capacitance of interconnects as shown in Fig. 6. Still, the interconnect delay is a big headache in designing a scaled-down interconnect system. The delay can be reduced by the buffer insertion technique but the power increases by about 70% due to the inserted buffers. In the optimally buffered interconnect, the capacitance of the system increases due to the inserted buffers. The total capacitance is increased by 73% compared with the system without buffers. The increase in capacitance in turn increases power consumption. Another way to decrease the interconnect delay without increasing power is to use a thicker and wider metal layer using a superconnect technology as shown in Fig. 7. The super-connects in a package used in cooperation with on-chip interconnects will solve the IR drop problem, the clock distribution problem and other problems of the future VLSI’s. Ultra-sensitive and flexible, graphene could also make it possible to create superior touch screens and detectors capable of diagnosing diseases much sooner than existing testing devices. Leading consumer electronics makers anticipate that graphene-based devices will be easily available within few years.
4. POWER CRISSES

With shrinking technology reducing power consumption and overall power management on chip are the key challenges below 100nm due to increased complexity. Leakage current is also becoming an increasingly important fraction of the total power dissipation of integrated circuits. Unfortunately designing for low power adds another dimension to the already complex design problem and the design has to be optimized for power as well as Performance and Area. Low power VLSI designs can be achieved at various levels of the design abstraction from algorithmic and system levels down to layout and circuit levels. Operation at such low power/voltage poses very interesting challenges and offers new opportunities to develop emerging applications, as well as to stimulate and enable new technologies and markets. Process designers scale both the applied voltage and the oxide thickness to maintain the same electric field [27,28]. This approach reduces power by about 50% with every new technology node. The technology variant with the thicker gate oxide aims for low-leakage design and must support a higher voltage to achieve a reasonable performance [29]. When selecting a technology to optimize the power for a given design, you must take both aspects into consideration: the need to use a smaller geometry to reduce active power and the need to use a low-leakage variant to reduce leakage. The International Technology Roadmap for Semiconductors (ITRS) forecasts an exponential increase in power dissipation for System-on-Chip (SoC) systems over the next 15 years. Power dissipation has a considerable impact on the future development of technology. Fig. 1 illustrates the required power dissipation levels for SoC devices, in order to sustain current growth levels. A single chip processor in production is consuming more than 100W of power and the power will be further increasing up to about 150W in the next decade. Thus, power consumption is very important figure-of-merit in predicting the technology trend.
Today the feature sizes being downscaled from 0.35µm (in 1996) to 20nm. It is predicted that feature sizes will continue shrinking to less than 45nm in the few years to come. With the increase in die dimensions, more functions are integrated into one chip. There is also a trend of continuous increase in the number of devices within a chip. Modern day chips target over a billion transistors inside a die. Not only the heat generated by the consumed power but also the huge current needed to operate such VLSI is an issue. Reduction in transistor switching delays results in faster signal transition times and higher clock frequencies. The current increase is due to the supply voltage (VDD) decrease down to about 0.3V in 15 years. The most effective system level power management technique is recognized as dynamic voltage/frequency scaling (DVFS). In this technique, the power delivered to each functional module is varied, based on the instantaneous workload [21–23]. Conventionally, circuits are designed to accommodate the worst case delay and are therefore becoming very limited in their performance advantages. While some designers use techniques to reduce the timing margins and improve speed, others use the same techniques to save dynamic energy by reducing $V_{dd}$ and working at nominal or slower speeds. Research has shown that minimum energy is typically achieved when $V_{dd}$ enters the near/sub-threshold region, where about a 10x reduction in energy per computation is achievable [24]. As $V_{dd}$ is lowered, the power is reduced effectively because the power is quadratically dependent on $V_{dd}$ but the delay increases.

Speculative speed-up techniques reduce voltage and frequency margining in an application where the speculative circuit is usually correct, speed-up or energy reduction is achieved [25]. ROI beyond 32nm may be questionable as velocity saturates, mobility degrades with power improvement. The delay can be shortened by decreasing $V_{th}$ but the low $V_{th}$ induces large sub threshold leakage current, which prevents to decrease the total power. Consequently, a trade-off between speed and power should be considered to decrease power. Operation in the sub threshold region most often is synonymous to minimum-energy operation. Minimum energy operation for digital circuits typically requires scaling the power supply below the device threshold voltage. Minimum-energy point (MEP) is very expensive in terms of performance; just like minimum delay point (MDP) is in terms of energy. We can gain 10-times in performance by increasing energy by 20% above MEP [26] as shown in fig 9. The traditional operation region is shown around minimum-delay point (MDP) while Ultra low-energy region is around minimum-energy point (MEP).
Power and energy consumption of digital systems may increase significantly during testing. Moreover, it may be responsible for cost, performance verification as well as technology related problems and can dramatically shorten the battery life when on-line testing is considered. This extra power consumption due to test application may give rise to severe hazards to the circuit reliability. Designers must consider the effects of increased sub threshold leakage, and should use circuits that are robust in the presence of threshold variations. When the activity level is quite low, as for a RAM, it may be more appropriate to choose a higher operating voltage, and higher thresholds, to minimize leakage currents. Different strategies are available at different level in VLSI design process for optimizing the power consumption as shown in Table1. Designers need an intelligent approach for optimizing power consumptions in designs so that effective power management is possible by using the different strategies at various levels in VLSI Design process.

Table -1, Strategies for Low Power Designs

<table>
<thead>
<tr>
<th>Design Level</th>
<th>Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating System</td>
<td>Portioning, Power down</td>
</tr>
<tr>
<td>Level</td>
<td></td>
</tr>
<tr>
<td>Software level</td>
<td>Regularity, locality, concurrency</td>
</tr>
<tr>
<td>Architecture level</td>
<td>Pipelining, Redundancy, data encoding</td>
</tr>
<tr>
<td>Circuit/Logic level</td>
<td>Logic styles, transistor sizing and energy</td>
</tr>
<tr>
<td>Technology Level</td>
<td>recovery</td>
</tr>
</tbody>
</table>

In the "home of the future" smart solutions will be required with energy efficient intelligent subsystems for all kind of appliances. There is a need for high voltage capabilities in power management, power conversion and power distribution in general. This scenario will require innovation in power electronics and will drive technical requirements for the underlying high voltage and power technologies. It is expected that the non-digital / non-memory part of integrated systems will play an increasing role in the future putting more emphasis on the MtM domain [10]. The effective use of multiple $V_{dd}$, multiple $t_{ox}$, and multiple $V_{th}$ is the key to prevent the power consumption from exploding.
5. COMPLEXITY CRISIS

After 1990s, circuit innovation moved toward dealing with the new problems that arose due to the increases in complexity enabled by scaling. For example, many designers dealing with issues such as signal integrity, power supply quality, and the distribution of precise timing references and many of these issues continue to be explored today. Based on our beliefs that the only effective way to improve productivity is to maintain design complexity at a manageable level, and that formal methods are the right techniques for complexity control. As the number of transistors increase, the power dissipation is increasing and also the noise. If heat generated per unit area is to be considered, the chips have already neared that of the nozzle of a jet engine. At the same time, the Voltage scaling of threshold voltages beyond a certain point poses serious limitations in providing low dynamic power dissipation with increased complexity. Silicon complexity, which refers to the impact of previously ignorable physical phenomena, is at the center of this crisis. As the complexity of chips grew during the digital circuit designer’s focus moved to address system-level issues that had now become critical: supply distribution, clock distribution and latch design, and noise robustness. Power supply distribution grew to be a huge issue as the rising complexity and performance of digital systems coupled with decreasing supply voltages led to large increases in supply current. Lower voltage and higher current required that the supply impedance decrease even more rapidly. In 2002, Rahal- Arabi presented the design and validation of the supply network in a couple of Intel processors [30]. More recent work has focused on trying to mitigate the effect of supply noise on performance [31]. STT-RAM is a next-generation, MRAM technology. Researchers sponsored by Semiconductor Research Corporation (SRC) announced that they have successfully created contact hole patterns for a wide variety of practical logic and memory devices using a next-generation directed self-assembly (DSA) process. Applying a relatively simple combination of chemical and thermal processes to create their DSA method for making circuits at 22 nanometers (nm), the research team at Stanford University projects that the nanofabrication technique will enable pattern etching for next-generation chips down to 14nm.

![Fig. 10: contact hole patterns by directed self-assembly (DSA) process](image)

All these bring a great promise of “system-on-a-chip”, but also introduce many new issues in the design process. Crosstalk and variability are two important issues among them. Due to the local and global process variation effects in deep sub-micron technologies, the speed of the transistors can vary dramatically from die-to-die or device-to-device. High speed clocks used now make it hard to reduce clock skew and hence putting timing constraints. This has opened up a new frontier on parallel processing. Even on the fabrication front, we are soon approaching towards the optical limit of photolithographic processes beyond which the feature size cannot be reduced due to decreased accuracy. This opened up Extreme Ultraviolet Lithography techniques. And above all, we seem to be fast approaching the Atom-Thin Gate Oxide layer thickness where there might be only a single layer of atoms serving as the oxide layer in the CMOS transistors. New alternatives like Gallium Arsenide technology are becoming an active area of research owing to this.
In the MtM WP we may want to emphasize that “a single integrated circuit” is in fact monolithic (single die) and that, consequently, all components (functions) have to be manufactured in a single (CMOS-compatible) process technology. It is plausible that new nanotechnologies can also be used to complement or replace CMOS. In the current and coming decades VLSI design - which currently enables us to build million-transistor chips will become Gigascale (GSI) design and Terascale Scale Integration (TSI) design, respectively. Thus, manufacturing defects will increase, devices will get less reliable, interconnect will be orders of magnitude slower than transistors, new nanotechnologies will emerge, and signal and power management issues will be aggravated. Some of the unique system design challenges posed by anticipated nanoscale CMOS and molecular electronics technologies are presented at system level approaches to deal with these challenges [33]. Nitridation can convert a thin surface region of SiO$_2$ into a nitroxide film which is a diffusion barrier that allows the use of thin dielectrics in MOS structures and a variety of gate metals without contaminating the interfacial region. The use of nitrided SiO$_2$ for very large scale integration (VLSI) applications is becoming increasingly attractive nowadays. A two-activation-energy model of nitridation and a structure for MOS gate insulator applications is proposed [32]. This structure is achieved using rapid thermal nitridation at 1300°C for 20 s in 1 atm. of ammonia.

Some of the issues in System-on-a-chip (SoC) are huge initial investment for masks and development, undistributed IP’s (i.e. CPU, DSP of a certain company), process-dependent memory IP’s, IP testability, upfront IP test cost, difficulty in embedding high precision analog IP’s due to noise, and process incompatibility with non-Si materials and/or MEMS. Another way to cope with the complexity crisis is to design an electronic system by assembling the real components, that is, we can design in higher abstraction and thus escape from the complexity crisis. A new type of assembly technology called System-in-a-Package (SiP) has been proposed as shown in Fig.11. SiP simulation requires many disciplines including RF circuit design, packaging design, thermal design. the SiP can be much more compact in the system. Because we can integrate all the RF, including the antenna switch and power amplifier, and because we can integrate high-Q passive components, we can have a single package with an antenna signal going in and digital data coming out.

![Fig. 11: System on chip (SoC) vs System in Package (SiP) Technologies](image)

**CONCLUSION**

The job of the digital circuit designer has grown with the chips, moving from optimizing and validating gates, to working on functional units, to now designing complete
systems. Digital designs have gone through dramatic changes over the past two decades—moving from chips that contained tens of thousands of devices to today’s chips that may contain over a billion transistors. While the progress in digital design has clearly been tremendous, tackling current and future system issues and power challenges will lead to significant further innovation. We look forward to seeing continued reports of these digital circuit design advances over the next two decades of the conference. VLSI’s, packages and the higher assembly structure are co-designed to make the electronic system higher in performance. For those applications requiring the lowest power, highest clock rates and lowest unit costs, the SoC seems to be a good match. For those applications that require a fully functional, highly specialized module that can be easily integrated into a system, the SiP seems most appropriate. Making SiP more widely usable requires an ecosystem of components, which means standards like, interfaces, testability, power management and validation.

REFERENCES


