

How can system level design solve the interconnect technology scaling problem?

Francky Catthoor^{1,2}, Andrea Cuomo³, Grant Martin⁴, Patrick Groeneveld⁵, Lauwereins Rudy^{1,2},
Karen Maex^{1,2}, Patrick van de Steeg⁶, Ron Wilson⁷

1 IMEC, Leuven, Belgium

2 Katholieke Universiteit, Leuven, Belgium

3 St-Microelectronics, Geneva, Switzerland

4 Cadence Berkeley Laboratories, Berkeley, California

5 Magma Design Automation, Cupertino, California

6 Philips Semiconductor, Eindhoven, The Netherlands

7 CMP Media, San Mateo, California

Abstract

The scaling of interconnect technology hits a red brick wall: interconnect delay and power do not follow Moore's law anymore. The use of new materials like Cu and low-k alleviated the problem temporarily, but physical limits are being hit. What does this mean for system level design? The session starts with an embedded tutorial, given by an interconnect semiconductor technology expert, explaining the physics behind the interconnect problem and the degrees of freedom semiconductor technology offers system designers. Panelists will then express their thoughts and discuss with you how the interconnect problem can be solved by taking these degrees of freedom into account at the system design level. Views from industrial designers, CAD vendors, IC manufacturers and researchers will be presented.

Tutorial: The future of a living apart together relationship between design and technology

Karen Maex

1. Introduction

The drive towards increasing density, complexity and frequency rapidly has required an enormous effort for technology to deliver transistors and wires with the right performance. Transistors are scaled and optimized for higher drive currents. Meanwhile, leakage currents in off state transistors are increasing. For the wires, the introduction of Cu and low-k dielectrics can provide a partial solution to the interconnect delay problem. Low k dielectrics are as well important, since they have an impact on power consumption. The impact of interconnects on electrical performance cannot be ignored any longer. Unfortunately, dielectric materials will reach a limit in k value within a few years, and the resistivity of Cu wire is expected to increase for the smallest dimensions. Improved circuit design and partitioning of the chip in modules may provide solutions, for future technology nodes. On-chip wafer-level packaging, provides a way to add more global interconnectivity on the chip.

In this tutorial an overview will be given of the recent technological evolutions for the fabrication of transistors and wires. The technological evolutions will be related to implications for performance and energy, where possible.

2. Technology challenges

2.1. Transistor performance

Scaling dimensions of transistors is accompanied by a drop in supply voltage: indeed the electrical field which can be sustained by materials is limited. The game is to maintain a right balance of electrical strength of materials within the device. Meanwhile as transistors are shrinking, higher switching speeds are obtained as well as a lower power consumption. This has so far been possible down to the 130nm node. However, for future nodes effects which could be ignored gain importance or even become dominant. In the following an overview of the trends is presented.

To maximize I_{on}

It is clear that the optimization of transistor performance is driven towards a maximal drive current. As the power voltage drops, the overdrive of the transistor becomes smaller, and measures have to be taken to lower the V_t . The thickness of the gate oxide is made as thin as possible and the mobility of the carriers in the substrate is optimized. The thickness of the gate oxide is not limited by the technological possibilities but by the characteristics of nm scale thin films. The mobility of the carriers is jeopardized among others by scattering with channel dopant atoms and the remote Coulomb scattering of the space charge region in the poly-Si gate. The latter can be avoided by dropping poly Si as a gate material and replacing it by a metal.

Increased mobility of the carriers in the channel can be achieved by using so called high mobility substrates such as strained Si and even Ge is possibly reconsidered instead of Si.

To minimize I_{off}

Although decreasing sizes of transistors and lowering voltages were accompanied in previous technology nodes by lower power consumption and higher speeds, in future technologies this is definitely not the case anymore. The current of the transistor while it is switched off is maybe small, but increases with each scaling node. In conjunction with the high density of the

transistors, the static power becomes dramatically high. This static power consumption poses problems both for the high performance application domain, in which the power dissipation and corresponding heat removal, gets out of control and in the so called low power applications, for which the lifetime of the battery is a crucial parameter.

How is technology addressing the static power consumption? Several components contribute to the leakage of the transistor and for each of them solutions are investigated:

Leakage between Gate and Bulk:

The nanometer thick gate oxides, which were thinned so dramatically to optimize the drive current, are not thick enough to withstand quantum mechanical tunneling. This leads to extremely high leakage current levels. The thin SiO₂ is therefore to be replaced by high k dielectrics. These are typically metal-oxides with a higher permittivity. Therefore one can achieve a similar drive current for thicker dielectrics, thereby getting rid of the tunneling problem: thicker dielectrics do not suffer from tunneling. There exists an exponential relationship between thickness and leakage current a few tenths of nm change in thickness results in a factor of 10 increase in leakage current. The introduction of high k dielectrics has been delayed, most probably to the 45 nm node, due to severe technological problems related to their implementation.

Leakage between Source/Drain and bulk:

The very shallow junctions used in the small transistors do exhibit much more leakage current in reverse bias. This is partly due to the higher background doping in the device, present to minimize short channel effects. Moreover new phenomena such as band to band tunneling become important. The new substrates which are used, as described above, do have a smaller bandgap and therefore intrinsically higher leakage currents for diodes.

Leakage between Source and Drain

One of the most severe problems is the increased leakage between source and drain for gate voltage below the threshold voltage. Supply voltages have become very low, V_t has followed that trend to maintain drive current.

One should realize that there is a significant temperature dependency of some leakage currents, namely whenever they are based on the Boltzmann statistics of the carriers in the conduction and valence bands. Indeed those leakage currents tend to increase very fast with temperature, and increase the risk of thermal runaway. New technologies based on SOI and FINFET gain importance since they minimize short channel effects and related leakage currents. The short channel effects in FINFETS are still being characterized.

3. Interconnects

The on-chip interconnect bottleneck is a very logical consequence of the progressive scaling of dimensions to achieve high transistor densities. The scaling does not result in smaller chips but in chips with higher complexity and, to the first order, a same chip size. It is clear that this approach leads to conflicting requirements.

The hierarchy of interconnects in local, intermediate and global wires has been a first attempt to cope with the technological limitations. Based on a different set of layout rules and dimensions, local, intermediate and global wires do have different RC delays (decreasing from local to intermediate to global), setting a hierarchy for length of wires. Signals can be transferred over a limited distance in one clock cycle. For future nodes and frequencies in the GHz range this distance is of the order of 1mm for synchronous subsystems.

A chip designed in technology node n, will be routed and wired based on the local, semi-global and global wiring hierarchy. Two nodes further, however, both x and y dimensions are scaled by a factor of two and the same chip can be fabricated in a quarter of the Si area. Because of the fact that chip size will not scale accordingly, but the free Si space will be used for more functionality, a problem of interconnectivity, or rather more generally, of communication, is created, which cannot be solved by the current length hierarchy, but should be taken care of by an "extended" wiring.

In the following section, we therefore distinguish between intra-tile wiring, which in principle scales also in length for future nodes and inter tile wiring serving the whole chip.

3.1. Intra tile wires

Cu wire

The scaling of the wire cross-section will lead to additional scattering effects for electrons. The barrier/Cu interface will contribute to inelastic electron-phonon scattering and reflection from the grain boundaries will result in a higher effective resistivity of the narrow Cu wires. These phenomena pop up at wire sizes of 100 nm and below. Even though it is not completely clear how much of this increase in resistance can be solved by technology optimization, it definitely will have an impact starting at the 45nm node.

Capacitance between wires

In the total capacitance between wires, the parallel plate capacitance between wires dominates for the sub 130nm nodes. Introduction of dielectrics with a lower permittivity than SiO₂ is being investigated.

Similarly, narrow spacings will show a higher k value than wider spacings. Any change by process of the uttermost atomic layers in a trench feature, can lead to a locally higher k value. The expectations of the positive effect of low k dielectrics on electrical performance will therefore become lower as we adopt a more aggressive technology.

3.2. Inter tile wires

The interconnect lines in the wafer-level redistribution layer have a much lower resistance than the BEOL interconnect wires. When using semi-additive copper plating techniques, the conductor width and spaces can be reduced to 10m, or even 5m, with a conductor thickness of 3-5m. When using a 5m thick low-k interconnect dielectric, the 5m thick, wide and spaced lines will have a similar capacitance per unit length. The resistance, however, will be more than 20 times smaller. This type of interconnect is therefore of interest for very long critical on-chip interconnect lines.

4. Variability

For the scaled technologies process variability will be more important. Thickness of films need to be controlled below the nm scale. The process induced variability becomes a matter of a few atoms more or less. This is the case for thin film growth and deposition. Most critical are gate dielectric thickness, dopant control for reducing short channel effects, gate electrode roughness due to patterning, wire cross sections due to patterning control and thickness of the wire etc. It should be noted that one has technological differences between densely patterned and sparsely patterned areas on the chip

In addition a variability is introduced by circuit layout: parasitic capacitances in the wiring are more prominent in dense wires than in semi-isolated and isolated wires.

5. Reliability

Let's add at least one note on the reliability. The current densities in the transistor and the wires are enormous. Current densities higher than 2MA/cm² will run through ultra-thin Cu wires even before we are at the 45nm node. More accurate test structures, probing the origin of electromigration are mandatory and will be helpful in optimizing the wire technology within progressively more stringent boundary conditions.

6. A perspective

Scaling imposes numerous boundary conditions on technology, resulting in devices with higher leakage currents, and interconnects with relatively high resistivities and capacitances that are not as low as one had hoped for. Major changes in materials and technology are needed for the scaled transistor to maintain its on-current and to limit as much as possible the increase in leakage current. For the wires, one can no longer blindly count on the positive effect coming from the materials to lower resistance and to lower capacitance as scaling proceeds. Most of the technological changes are driven, not so much for improvement, but to compensate for the drawbacks of scaling. It is clear that materials only can no longer compensate for the effects of scaling in view of static power consumption, and in view of wiring the transistors. So technological progress has to be considered as an other "compromise" in electrical performance, with different technological choices leading to different

compromises. That is where circuits and systems come into the picture. Moreover, the increase in variability, introduced by process and lay out, will significantly slow down the progress one can make using a worst case design based on a 6-sigma scenario in a more advanced technology node.

As a consequence the electrical characteristics of transistors and wires will change in the future and will become a complex function of technological parameters. These parameters will include the local environment of the transistor and the wire. Therefore, a design methodology based on partitioning transistors from wires and both of them from their local environment will become questionable.

Whereas the technology can be adapted, among others, for multiple threshold voltages and back gate biasing, it will be a challenge for circuit designers to cope with this new technological situation.

The fact that process control and knowledge of the manufacturing process is essential for differentiation, ceased to be true for digital CMOS at roughly 0.35 um generation, and has partly been the root for an industrial landscape of Fabless companies and Foundries, linked by the EDA software.

It is clear that the traditional one-sided system design and one-sided technology development in the "living apart together" relation, run out of steam for the future technology nodes.

7. A second opinion

The foregoing suggests that the traditional sign-off between system level design and technology cannot continue as a 'living-apart-together' relationship in which each "minds its own business". This relationship will hit a brick wall unless both start sharing intimate technology details. But are quarrels about leakage, power, variability and reliability enough reason to fundamentally change an otherwise successful relationship? Wasn't the 'living apart together' ideology not the secret to its enormous success in the first place? After all, a clear (mask-level) hand-off has enabled the cooperation of specialists from radically different fields and across vastly different levels of abstraction.

The ITRS roadmap for design technology correctly identifies design productivity as the key issue in keeping up with the **technology advances**. In practice this means that a small team of engineers must continue to be able to design a leading-edge chip in a short time (e.g. 9 months). There are two categories of productivity issues:

- **System complexity** issues that rise from handling the sheer size of the SoC. System complexity scales exponentially with Moore's law. Over the past decade, Electronic Design Automation tools such as RTL synthesis have improved designer's productivity from 4K gates per manyear to over 100K gates per engineer year. Although spectacular, this rate is insufficient sufficient to keep up with Moore's law. Design reuse and new hierarchical technologies are also required to maintain the productivity increase.
- **Silicon complexity** issues are related to the manufacturing technology. This includes the such as device and interconnect parasitics, physical and electrical design rules, device reliability and process variability. In a way, silicon complexity is the result of the underlying physics that

enable system complexity. In practice, silicon complexity increases the number of steps in the design flow.

Both system and silicon complexity are increasing, leading to a superexponential increase in overall design complexity. This is what puts a strain on a LAT relationship that was based on a clear division of tasks. To keep the design process of a billion transistor feasible, however, proper design abstractions and hand-off will remain key. As always, this might come at a price of a somewhat less efficient design. But we're used to such trade-offs in a relationship. It is just impossible for an engineer to overlook all aspects of the design trajectory while delivering the chip in time.

No doubt we will appropriately deal with the silicon complexity issues. Multi-VDD and Multi-Vt control will reduce leakage. Modern physical synthesis tools are dealing with the effects of wire parasitics successfully. Reliability can be designed into the chip through appropriate wire sizing and gate sizing. Most of these steps can remain somewhat transparent to the system designer as long as the block is not too big.

The major hurdle in dealing with system complexity is to enable effective reuse of IP modules. The current hierarchical design methodology requires an interface specification at the physical and electrical level that is extremely complex. A network-on-a-chip architecture can raise the abstraction level and make IP integration much simpler. I expect individual IP to be synchronous blocks in the order of 1-10 million gates. A Network-on-a-Chip architecture using such blocks will fit a GALS (Globally Asynchronous, Locally Synchronous) paradigm and simplify system level design.

Overall, the LAT relation between design and technology can survive the increase of silicon complexity, just as it did with previous technological advances. Living Apart Together (with a clear division of tasks) is the only way to keep this relationship sane and manageable. Dealing with the exponential increase in system complexity still remains the single biggest issue.

8. On the other hand ...

It may be that Living Apart Together will end not because it has become impossible, but because greater intimacy has become more desirable. For example, there are situations in which a close cooperation between system-level and circuit-level design yields valuable results. One of these is power management.

We have seen that a number of challenges are acting together to change the nature of the problem from a physical design point of view. These same challenges can be seen, however, from a system-design point of view. From that perspective, they could be formulated as follows:

1. interconnect scaling does not follow transistor scaling and hence long lines will cause severe delay and energy problems.
2. temperature-dependent leakage will dominate energy consumption, and hence the total power will not necessary scale by scaling the technology.
3. process variability will force worst-case design with enormous tolerances, like 6σ and more which happens now already in some critical memory designs.

If nothing happens at the system-level, the Living Apart Together relation that was used in the introduction to symbolically represent the cooperation between processing and design will risk a complete break-down. Several countermeasures are taken at the processing and circuit level abstractions but it is now relatively widely accepted that these will on their own not suffice to counteract the negative impact to a sufficient extent.

This is especially a problem in the area of power-sensitive designs where the above 3 factors will give a very negative combined effect. In those designs, the risk exists that scaling will not help reducing total system energy any further: on the contrary. To understand this, it is important to realize that currently system-level power reduction schemes are largely based on reducing dynamic energy, on largely ignoring the interconnect contributions and on a limited process variability so that tolerances are kept small. A few attempts are made to counteract part of the above new effects on an individual basis, like "variability tolerant communication design" [3] or "reducing leakage by powering down processing or memory modules" (e.g. [1], [4]). But no overall, global approach to deal with these effects is available yet. We believe that it is crucial to work on this and that indeed good opportunities exist to bring the processing, circuit level and design issues closer together again. That should happen preferably without giving up the relative decoupling that has allowed to reduce overall design complexity to a sufficient extent in the past decade to enable very complex SoCs.

A promising example of that - focused on the interconnect aspect for the moment - is proposed in [2] for embedded memories. Here several working points at the interconnect technology level are "propagated" over the memory circuit level to the system level. Exploiting the slack in system-level timing requirements for the memory accesses that are not in the overall critical path allows then to realize a significant extra energy reduction of 40%, even by only playing with one "technology knob". Larger gains are being explored by adding more of such knobs. The approach propagates the interconnect effects to the system level but each of the abstraction levels can still be solved individually, within the propagated constraints of the lower levels. Current work is starting to apply such principles also for the other technology effects in a combined solution. This approach should enable significant gains compared to the "brute-force" worst-case modeling approach of today's decoupled situation.

9. Memory design as an example

It is no coincidence that the above example involves a memory array. As has been pointed out previously, in deep sub-micron technologies interconnect scaling lags behind transistor scaling, even when using copper metallization. As the performance of interconnect scales much slower than, or in the wrong direction from, transistor scaling, it is becoming an ever more dominant part of the total performance of a given technology, both in terms of speed and power. This is nowhere more obvious than in memory design.

In memory design close cooperation between design and technology has been standard practice already for a number of generations. With smaller feature sizes the relation between memory design and technology will become closer. Careful

balancing of static noise margin/Ion/Ioff and area is required to have the most optimal memory cell in a given technology variant. This cooperation can be of great help for IP providers, like analog blocks and memories, but the extra effort this takes should not be underestimated.

A bottom-up look at interconnect related problems encountered within memories results in a number of solutions which could also be applied outside memories:

- Careful dimensioning of layout and for critical wires accepting non-minimum pitch routing reduces the overall delay (and power) of metal interconnects. Support from technology by making different flavors of metal interconnect available can increase the effectiveness of this method. As the number of metal layers increases in every technology node it is feasible to have a number of ‘fat’ metal layers that are exclusively used for global signal interconnect. Careful layout of interconnect can also improve the signal integrity of long wires.
- Bus repeaters and pipeline registers can increase the maximum performance over a given wire length.
- Some of the circuit techniques used in memories for signal transfer over long distances, could be applied to speed-up point-to-point busses or make them more noise resistant. Differential busses could be applied to reduce noise sensitivity. Current mode sensing schemes that are used in some high-speed memories could have a large impact on speed, as the wire load dependency is much smaller than when using voltage mode techniques. Reduced voltage swing busses can have a dramatic effect on power and also reduce delays. These techniques could result in new building blocks that become available in libraries.
- Reduction of the amount of capacitance that switches is common practice in low power memories. Outside memories, clock gating has the same goal to lower the power and noise.
- To realize the best possible performance, large memories are constructed in a hierarchical fashion of sub-arrays, arrays, and blocks with a subsequently increasing size. At each hierarchy level a different trade off can be made between the various implementation options. In order to benefit from scaling, one should aim for concurrent operation of the parts at the lowest possible level in this hierarchy. E.g. in stand-alone DRAMs this is implemented by allowing operations on banks, in a system-on-a-chip this could be realized a distributed-shared memory architecture.

A number of deep-sub micron effects will also be clearly visible to system level designers. To optimize memory performance, both in terms of leakage as speed, different Vt-variants are possible. It is now the task of system level designers to choose the correct memory type, or maybe a mix, out of a number of possible variants. When low leakage is important, having some form of power-down can help to reduce leakage power significantly. Of course the system architect should deal with such a provision efficiently since a power-down cycle consumes a lot of energy by its self. But that is basically the game of building high complexity SOC's.

10. From computing to communications bandwidth

These examples based on memory design suggest that there are significant advantages to be gained when system designers and memory designers talk. But there are ways in which system-level design itself is changing that will have a profound impact on the relationship between systems designers and the rest of the members in this relationship. To understand how changes in system design are redefining the chip design problem, we need to look at the way system-level design is evolving.

- SoC's designers need to leverage on pre-validated components and IPs such as processor cores, controllers and memory arrays. Design methodology will further support IP re-use in a plug-and-play fashion, including buses and hierarchical interconnection infrastructures.
- An example can be STBus Interconnect, a versatile, high performances interconnect IP from STMicroelectronics, allowing to specify the communication infrastructure in terms of protocol, interface and parametric architectures.
- SoCs will have to provide a functionally correct, reliable operation under data uncertainty and noisy signaling. The on-chip physical interconnection will be a limiting factor for both performance and energy consumption, also because the demand for component interfaces will steadily scale-up in size and complexity.
- Further technology requirements could be ultra low-noise RF performance, extremely high-speed digital performance, refined analog precision, energy management, very low bit-error rate memories (either SRAM, SDRAM or FLASH).

Although the main concepts and the terminology of Network-on-Chip design have been introduced quite recently, both the industrial and research communities have been starting to realize the strategic importance of refocusing the design paradigm of high-end digital IC from a deterministic, wire-based interconnection of individual blocks and IPs, to a thorough communication-based design methodology, aiming to face with data packetization and non-deterministic communication protocols in next generation's SoCs.

In fact, with the advent of 90nm and 65nm CMOS technology, the challenges to fix the Network-on-Chip (NoC) issue “by design”, will explode:

- To provide a functionally-correct, reliable operation of the interconnected components by exploiting appropriate network infrastructure and protocols, i.e. interconnections to be intended as “on chip micro-network”, which is an adaptation of the OSI protocol stack.
- To achieve a fluid “flexibility vs. energy-efficiency” system exploration, allowing an effective network centric power management. Unlike computation energy in fact, the energy for global communication does not scale down with technology shrinking. This makes energy more and more dominant in communications.

Reaching those goals will be crucial to the whole semiconductor industry in the next future, in order to face with the escalating range of signal integrity and physical wiring issues, who are

making the target IC reliability harder and exponentially expensive to achieve.

11. A manifesto

Thinking of an IC as a network of predefined components is a major step away from conventional systems design, and it clearly has major implications for chip design. We would argue that this is not a discreet step, but rather part of a continuing process—not of making some changes in system architecture, but in revising the way we think about IC design.

It's clear that technology scaling from 100 nm onwards is offering fewer answers, and more problems, to the questions posed by design. So, when we don't like the answers, it's time to change the question. That is, we need to successively relax the tight constraints imposed by today's design ideologies, and, by proposing alternative system level design strategies and methods, see whether technology evolution can take new directions and come up with new, more appropriate answers as it scales.

What are some of these ideological constraints and assumptions imposed by today's design approaches? A very brief survey of some of these issues is discussed in [7]. Let's identify and relax some assumptions in an order of increasing radicalism:

1. The assumption of synchronicity. Large chip-scale synchronous design is still the prevailing ideology despite the implications for power consumption, and difficulties in controlling clock skew and achieving high performance design and layout. The GALS paradigm has been much talked about and considerable research carried out, but not much used. It's time to push much more aggressively in this direction.
2. The on-chip bus model. Bus-based communication schemes are inefficient in design, power consumption, layout, and also carry with them the assumption that all communications must be perfect, predictable, complete, and invariant in latency. New models for network-style on-chip packet switching and routing (Network-on-Chip) have been proposed here - research and some experiments are happening but buses look to be with us for a long time.
3. All communications must be alike. That is, the vital control messaging occurring in a design must be treated exactly the same as the often somewhat redundant data being transported from processing element to processing element. Is this actually true?
4. No data can be lost in transport. We have seen lossy data communication schemes work in wireless communications and other signal and image processing schemes. Is this applicable only off-chip?
5. Control must be centralized, and control messages have to be sent as quickly as possible from one end of a design to another. The telecom and datacomm networks operate on a much more decentralized basis, and it is not always true that long latency for control messaging destroys the usefulness of the network. Call setup via geostationary satellites works about as well as for phones across the street, albeit with greater latency.

6. Compression is for air. Would it be possible to use on-chip compression, possibly lossy compression, to reduce the burdens placed on our interconnect media?
7. Computation must always work. Does every functional processing unit on-chip have to work perfectly? If one can yield a sufficient number of functional units, possibly with highly variant performance, can one still use such a device to do useful work?

So what might result if we relax these assumptions? We would have an on-chip decentralised network of possibly unreliable processing elements communicating over a variable interconnect network with compression, variability and the possibility of data loss. There are two key questions:

1. Does such a 'fabric' help resolve any of the technology scaling problems presented in the tutorial? We must assume that relaxing the requirements of design may allow the technology to evolve in different ways.
2. How do you map applications to such a fabric and get useful products?

This second question is the System Level Design problem expressed very succinctly. And it's not too different than figuring out how to offer value-added service functions on, for example, a wireless network! The fact that one is possible gives hope to the other.

If this direction does not work, the only alternative may be to go back to the small, and stick with older technologies - don't force evolution at such a rapid pace. Instead of large complex chips, go for the 'smart dust' - ad-hoc distributed networks of chips.

[1] L.Benini, A.Bogliolo, G.De Micheli, "A survey of design techniques for system-level dynamic power management", *{em IEEE Trans.\ on VLSI Systems}*, Vol.8, No.3, pp.299-, June 2000.

[2] A.Papanikolaou, M.Miranda, F.Catthoor, H.Corporaal, H.De Man, D.De Roest, M.Stucchi, K.Maex, "Global interconnect trade-off for technology over memory modules to application level: case study", *{em 5th ACM/IEEE Intl.\ Wsh.\ on System Level Interconnect Prediction}*, Monterey CA, April 2003.

[3] F.Worm, P.Thiran, P.Ienne, G.De Micheli, "An adaptive low power transmission scheme for on-chip networks", in *{em Proc.\ 15th ACM/IEEE Intl.\ Symp.\ on System-Level Synthesis}* (ISSS), Kyoto, Japan, pp.92-100, Oct.\ 2002.

[4] W.Zhang, J.S.Hu, V.Degalal, M.Kandemir, N.Vijaykrishnan, and M.J.Irwin, "Compiler-directed instruction cache leakage optimisation", *{em Proc.\ IEEE/ACM Intl.\ Symp.\ on Microarchitecture (MICRO-35)}*, Istanbul, Turkey, pp.208-218, Nov.\ 2002.

[5] R. Ho, K. Mai, M. Horowitz, "Managing Wire Scaling: A circuit Perspective"

[6] M.T. Bohr, "Interconnect scaling - The real limiter to high performance ULSI"

[7] Giovanni de Micheli, "Robust system design with uncertain information", *MEMOCODE 2003*, 24-26 June 2003, p. 283.