Gigascale Integration (GSI) Interconnect Limits and N-Tier Multilevel Interconnect Architectural Solutions

Jeff A. Davis, Raguraman Venkatesan, Keith Bowman, and James D. Meindl
Georgia Institute of Technology
School of ECE
Atlanta, GA 30332-0250
404-894-4770
jeff.davis@ece.gatech.edu

The International Technology Roadmap for Semiconductors (ITRS) projects that by 2011 over one billion transistors will be integrated into a single monolithic die. The wiring system of this billion-transistor die will deliver power to each transistor, provide a low-skew synchronizing clock to latches and dynamic circuits, and distribute data and control signals throughout the chip. The resulting design and modeling complexity of this GSI multilevel interconnect network is enormous such that over 10^{17} coupling inductances and capacitances throughout a nine-to-ten level metal stack must be managed. This discussion session will focus on system-level limits that on-chip signal interconnects place upon a GSI product design in the 21st century.

Interconnect limits potentially threaten to decelerate or halt the historical progression of the ITRS because miniaturization of interconnects, unlike transistors, does not enhance their performance. A current solution to this interconnect dilemma is to reverse scale longer semi-global and global interconnects such that they have “fat” cross-sectional dimensions. This strategy can enhance interconnect performance, but at the expense of wire density. Therefore, the central hypothesis of this discussion is that on-chip wiring will limit die size, clock frequency, and power dissipation in future designs thereby forcing the use of n-tier multilevel interconnect architectures that utilize repeaters to increase wire speed and/or density.

To investigate this hypothesis, a complete stochastic wire-length distribution model, technical projections from the ITRS, and interconnect performance and noise models are used to construct the architecture of a GSI multilevel wiring network. In this multilevel network it is assumed that interconnects on adjacent metal levels are routed orthogonally, and the wire dimensions on each orthogonal wiring pair are calculated to ensure that the latency of the longest interconnect does not exceed ninety percent of the clock period. Each multilevel wiring tier, which is defined as a collection of pairs with the same cross-sectional wire dimensions, is occupied with interconnects such that the required interconnect area is exactly equal to the available area.

Using these models, the limitations of historical approaches to microprocessor and ASIC design are investigated. Starting with the assumption that one million highly-connected logic gates are contained in a logic megacell for 1999, the number of metal levels is projected over the next fifteen years by doubling the number of highly-connected logic gates in a megacell every two years. A highly-connected logic megacell is defined as a statistically homogeneous array of logic gates in which Rent’s Rule describes the I/O requirements of any arbitrary logic partition within the megacell. Logic megacell areas, clock frequencies, and technologies for projected designs are taken directly from the ITRS. As seen in Figure 1, without the use of repeaters (scenario 1), the number of required metal levels approaches unrealistic values beyond 2005. In fact, the number of projected levels at 2014 is almost an order of magnitude larger than the number of levels prescribed by the ITRS.

The required interconnect area beyond 2005 for the first scenario exceeds the available area given by the ITRS. To avert this situation, one possible solution is to limit the size of each highly-connected megacell to reduce the overall wiring demand. Therefore, as an alternative to the historical approach of doubling the number of logic gates, Figure 1 also shows that restricting the maximum number of highly-connected gates to a value of around 10M gates (scenario 2) keeps the number of metal levels per megacell to a controllable number through 2014. Larger digital systems would be constructed out of megacells that are loosely connected so as to deal with limited wiring area resources.

Another possible solution to this problem is to utilize interconnect repeaters to increase wire density. Figure 1 shows the results of repeater insertion in an n-tier multilevel interconnect architecture (scenario 3) for megacells that double in size every two years. The results in Figure 1 suggest that significant repeater insertion throughout a logic megacell could control the number of wire levels until 2011. Extensive repeater use, however, would significantly change current chip design characteristics. For example, as seen in Figure 2 beyond 2005 the number of transistors required for repeater insertion ranges from 4-13% of the total logic transistors; the amount of area required for repeaters is roughly 40% of the total megacell area; and the power dissipated by repeaters ranges from 20-30% of the total power dissipation.

On-chip interconnects could severely limit system characteristics of future digital systems. The primary question for this discussion session is: How extensively will interconnects limit future designs and what are the most feasible solutions to circumvent these limitations? Limiting the complexity (e.g. the number of highly-connected logic gates) of a megacell appears to be one alternative. In situations, however, where designs call for massive integration and complexity, significant chip resources need to be allocated for repeaters. Therefore, a future GSI product could have a significant number of active elements dedicated to logic, memory, *and* interconnects.

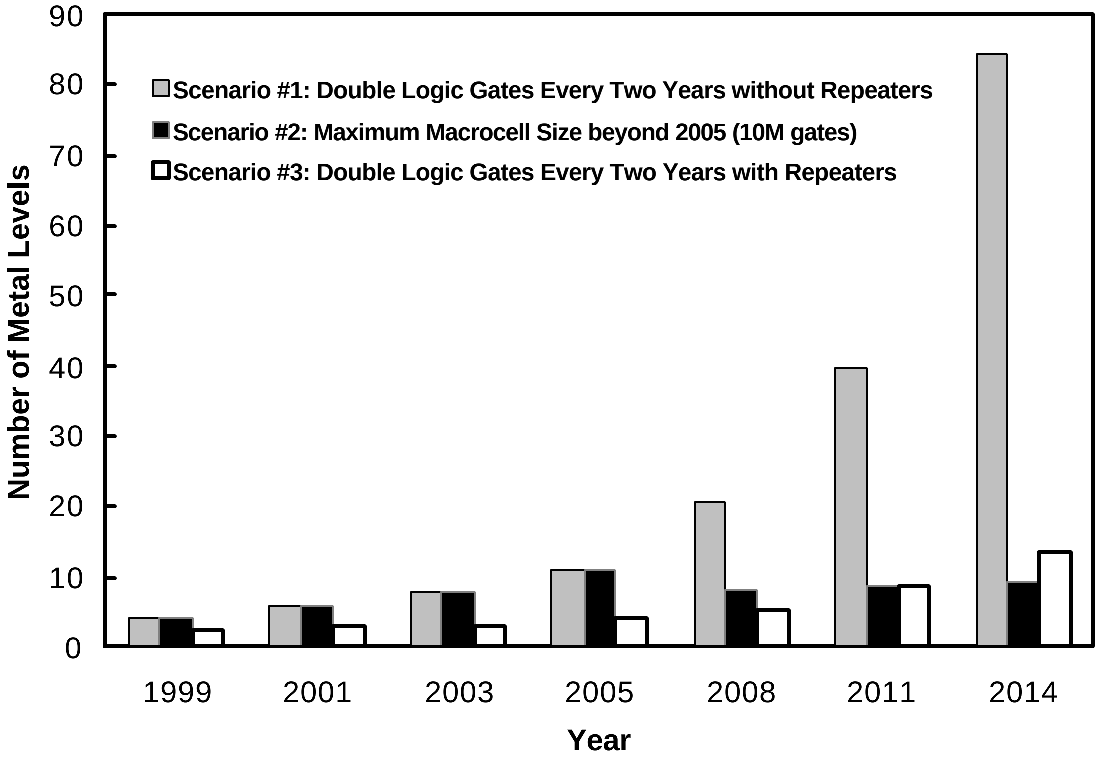

Figure 1: Projection of the number of metal levels for highly-connected logic megacells from 1999 to 2014