Timing Closure for Low-FO4 Microprocessor Design

David S. Kung
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598
914 945-3183
kung@watson.ibm.com

ABSTRACT
In this paper, we discuss timing closure for high performance microprocessor designs. Aggressive cycle time and deep sub-micron technology scaling introduce a myriad of problems that are not present in the ASIC domain. The impact of these problems on floorplanning, placement, clocking and logic synthesis is described. We present ideas and potential solutions for tackling these problems.

Categories and Subject Descriptors
B.6.3 [**Hardware**]: Logic Design – *design aids, logic and physical synthesis, design methodology.*

General Terms
Algorithms, Performance, Design.

Keywords
High Performance, FO4, Synthesis, Placement.

1. INTRODUCTION
It is a popular belief that timing closure is a mature field and no longer requires research focus. In the light of aggressive cycle time approaching 10GHz and beyond and relentless technology scaling, this complacent myth needs re-examination. The impact of cycle time and technology scaling on timing closure for microprocessor designs is particularly prominent. Traditionally, the divide and conquer approach using a relatively large design team is adequate for microprocessor designs. Each core is divided into units and each unit is partitioned into macros. Timing closure is applied to each macro independently and conventional synthesis and placement technology is sufficient to achieve cycle time target. This paper shows that going forward, new ideas and approaches must be implemented to meet the technology and cycle time challenges. In the following sections, the problems that are derived from these challenges and their potential solutions will be discussed.

DAC 2004, June 7–11, 2004, San Diego, California, USA

2. PROBLEMS AND SOLUTIONS

2.1 Macro Partitioning and Floorplanning
The cycle time puts a constraint on the maximum separation between latches. For a design at 10 GHz using a 65 nm technology, the maximum distance is around a few hundred tracks which is less than the linear dimension of a macro. This constraint has far reaching impact on partitioning and floorplanning of macros. One immediate consequence is that no latch to latch path can have more than one macro crossing. Moreover, interconnects between macros must be under a specific length. Violations must be fixed by inserting latches to break up the offending paths, which can potentially degrade chip performance. One solution is to perform simultaneous macro partitioning and latch insertion as in [1][2]. The cost function space is cycle time, power (number of latches) and chip performance. A holistic approach of combining macro partitioning, floorplanning, power and performance analysis would definitely be beneficial.

2.2 Routability
In 65nm technology and beyond, manufacturability and reliability are major concerns. The impact on cell layout and interconnect design has resulted in challenges in routing. Since some of the routing violations are caused by topological obstructions, the correlation between wire length and routability is increasingly tenuous. As a result, a wire length based placement engine might not deliver a routable solution.

The routabiltity problem can be approached from several different angles. The placement engine must understand the pin densities and pin access geometry. The goal is to generate a placement that is easier to route by incorporating the above cost functions. A complementary approach is to generate a faithful congestion picture during physical synthesis and incrementally fix potential routing violations. This approach requires integration of physical synthesis with a router which has fast incremental capability.

2.3 Shrinking Feasible Design Space
The aggressive cycle time and technology scaling has substantially reduced the size of the feasible region in the design space. We discuss several examples in the following subsections.

2.3.1 Logic Structure
The cycle time also dictates the maximum number of levels of logic in a latch to latch path. Trading area to reduce the number of levels of logic must be done judiciously. The area increase could exceed the power budget and potentially lead to placement and routing problems. We employ a sophisticated Boolean restructuring [3] that can target aggressive reduction of logic depths on critical paths and recover area on the non-critical regions. The net area impact has generally been benign.

A more subtle impact of shallow logic depth is in the domain of parametric yield. Assuming a more or less random distribution for parametric variations, the variance of the path delay as a percentage of the path delay increases as the number of levels of logic decreases[4].

2.3.2 Noise Margins

In order to maintain a healthy margin for signal integrity, the global slew limit is set to be a fraction of the cycle time. With shrinking cycle time, the gap between the typical output slew of a gate and the global slew limit has narrowed dramatically compared to previous generations. This poses a challenge to area recovery and results in an unfavorable area impact. The algorithms that perform delay-area tradeoff must be much more precise and effective.

2.3.3 Placement Structures

Another feasible space constriction derives from constraints in placement geometry. One such example is the clock driver-latch geometry constraint. To minimize clock skew and power consumption, the relative positions of a clock driver and the latches that it drives must adhere to some rigid rule. This distorts the placement from the "rule-free" configuration and degrades timing as well as routability. A placement engine that can handle various types of ad hoc constraints is needed to deliver quality results.

2.4 Diminishing Tolerance for Errors

Technology scaling has raised the bar for modeling accuracy of delays, particularly for interconnects. Inaccuracy in timing models leads to erroneous sorting of critical paths. The quality of gate type and size implementation might be lacking. Perhaps more importantly, underestimating slews results in re-emergence of slew violations that were presumably fixed during timing closure. These violations need to be fixed manually as a post process and add unnecessarily to turn around time.

3. CONCLUSIONS

Timing closure for high performance microprocessors is frothed with interesting challenges in the field of synthesis, placement, routing, clocking, design methodology and delay modeling. Instead of having passed its prime, I see timing closure as a driver of innovative tool research and development in the foreseeable future.

4. REFERENCES

[1] Cong, J., Sung, K. L. Physical Planning with Retiming. *ICCAD 2000*, 2-7.

[2] Cocchini, P. Concurrent Flip Flop and Repeater Insertion for High Performance Integrated Circuits. *ICCAD 2002*, 268-273.

[3] Kravets, V.N., Sakallah, K.A. Resynthesis of mult-level circuits under tight constraints using symbolic optimization. *ICCAD 2002,* 687-693.

[4] Kerry Bernstein, private communications.