RETROSPECTIVE: Complexity-Effective Superscalar Processors

Subbarao Palacharla[#1], Norman P. Jouppi[$2], J. E. Smith[*3]
[#]Qualcomm, San Diego
[$]Google, Palo Alto
[*]University of Wisconsin-Madison, Madison
[1]subbap@qti.qualcomm.com, [2]jouppi@google.com, [3]jes@ece.wisc.edu

The 90s were exciting times for research on superscalar microarchitectures. Commercial microprocessors were transitioning from in-order execution to out-of-order execution [2, 3, 5]. Earlier research had established that out-of-order execution could extract significantly higher levels of instruction-level parallelism compared to in-order execution. What was still unclear and actively debated [1] was the impact of the complex logic needed for out-of-order execution on clock frequency and how well the logic scaled for wider and deeper out-of-order superscalar processors. We felt there was an immediate need to quantify the complexity so that architects could use that knowledge to design microarchitectures that achieved most of the benefits of out-of-order execution while still enabling a fast clock. We referred to such microarchitectures as *complexity-effective* microarchitectures.

While the goal of quantifying the complexity of a microarchitecture was a noble one, it was challenging to pursue for several practical reasons. Complexity can be quantified in many ways; how can we define complexity? Complexity is a significant function of the actual design; how do we put together a representative design? Complexity studies are hard to accomplish in an academic setting; how can we ensure credibility?

We ploughed forward with reasonable assumptions and best-known methods. We chose to measure complexity as the delay of the critical path through a piece of logic. Additionally, the exploration space was limited to paths originating from instruction decode, issue, and data bypass logic. Collaborating with engineers at Digital Equipment Corporation (DEC) provided us the opportunity to review representative implementations from the literature (mainly ISSCC papers) and from within DEC. The entire study was conducted in DEC to benefit from access to advanced implementation techniques and technology files for various process nodes. The methodology utilized used was derived from an earlier study of on-chip cache access timing in DEC by Wilton and Jouppi [4].

The second half of our paper proposed the dependence-based microarchitecture: a complexity-effective microarchitecture that uses multiple clusters to partition the issue and bypass logic so the resulting design can be clocked faster than an equivalent superscalar with a single window. To minimize slower bypassing across clusters, dependent chains of instructions were steered to the same cluster. Our results showed that the proposed microarchitecture can deliver higher performance than a centralized superscalar, despite IPC degradation, as the proposed microarchitecture enables a faster clock.

Our paper had identified window logic and data bypass logic as the most critical to execute dependent instructions in consecutive cycles. The quantitative results indicated that wire delays, caused by broadcasting values on long wires, will increasingly dominate total delay for the logic structures. These conclusions remain valid and applicable to today's processor designs. At the same time, it also seems like over the last two decades processor designers were able to use innovative engineering to scale window size and issue width without significantly impacting clock frequency. All this while still implementing a single unified window and avoiding clustering like in the dependence-based microarchitecture. Today's processors can issue as many as 12 instructions every cycle while maintaining 400 or more instructions in flight! Unfortunately, the secret sauce used to overcome complexity in these designs are proprietary and inaccessible to the wider architecture community.

This work inspired David Albonesi to initiate a workshop on complexity-effective design (WCED) that ran successfully over seven years (2000-2006). The workshop series expanded the scope of complexity-effective design to include verification complexity and power efficiency.

Perhaps the greatest contribution of this work was educational in nature. The paper made the topic of complexity of superscalar microarchitectures widely accessible to young architects (especially graduate students). They could clearly understand the role complexity plays in the tradeoff between IPC and clock frequency. The paper provided students with the knowledge to overcome limited exposure to actual design/implementation within the academic setting.

REFERENCES

[1] Linley Gwennap, *Speed Kills? Not for RISC Processors*, Microprocessor Report, 7(3), March 1993.

[2] Linley Gwennap, *Intel's P6 Uses Decoupled Superscalar Design*, Microprocessor Report, 9(3), March 1995.

[3] K. C. Yeager, *MIPS R10000 Superscalar Microprocessor*, IEEE Micro, April 1996.

[4] Steven J. E. Wilton and Norman P. Jouppi, *An Enhanced Access and Cycle Time Model for On-Chip Caches*, Technical Report 93/5, DEC Western Research Lab, July 1994.

[5] Jim Keller, *The 21264: A Superscalar Alpha Processor with Out-of-Order Execution*, October 1996, 9th Annual Microprocessor Forum.