TA 6.2: A 200MHz 64b Dual-Issue CMOS Microprocessor

Daniel Dobberpuhl, Richard Witek, Randy Allmon, Robert Anglin, Sharon Britton, Linda Chao, Robert Conrad, Daniel Dever, Bruce Gieseke, Gregory Hoeppner, John Kowaleski, Kathryn Kuchler, Maureen Ladd, Michael Leary, Liam Madden, Edward McLellan, Derrick Meyer, James Montanaro, Donald Priore, Vidya Rajagopalan, Sridhar Samudrala, Sribalan Santhanam

Digital Equipment Corporation, Hudson, MA

A RISC-style microprocessor chip, operating at up to 200MHz, implements a 64b architecture that provides huge linear address space without bottlenecks that would impede highly-concurrent implementations. Fully-pipelined and capable of issuing two instructions per clock cycle, this implementation can execute up to 400M operations per second. The chip includes an 8kB I-cache, 8kB D-cache and two associated translation buffers, a four-entry 32B/entry write buffer, a pipelined 64b integer execution unit with 32-entry register file, and a pipelined floating-point unit with an additional 32 registers. The pin interface includes integral support for an external secondary cache. The package is a 431-pin PGA with 140 pins dedicated to VDD/VSS. The chip is fabricated in 0.75µm n-well CMOS with 3 layers of metallization (Table 1). The die measures 16.8x13.9mm^2 and contains 1.68M transistors. Power dissipation is 30W from a 3.3V supply at 200MHz. A micrograph is shown in Figure 1.

The architecture includes comprehensive support for both 32 and 64b operations on an instruction-specific basis. It provides continuity with an earlier CISC architecture without sacrificing RISC performance characteristics. All operate instructions are register-to-register while memory operations are strictly load/store.

Instruction issue supports pair-wise execution among combinations of four basic units: load/store, integer operate, floating-point operate, and branch. The pipeline depth is seven cycles for everything, except ten for floating-point operate. Only integer multiply and floating divide are not fully pipelined. Integer latency is generally one cycle, load latency is three cycles, and floating-point latency is six cycles. Instructions are issued in order and under the control of register scoreboarding logic. The scoreboarding logic also controls result bypassing among the various units. The issue point is at pipeline stage four, beyond which the pipeline does not stall.

The floating-point unit is a fully-pipelined 64b floating-point processor which supports both VAX standard and IEEE standard data types and roundings. It is capable of generating a 64b result every cycle for all operations except divide. The pipeline stages include a 64b adder, a leading-one detector using input operands, two parallel exponent adders in stage 1, a radix-8 pipelined Booth multiplier organized as 8 odd and 8 even rows, a 64b shifter capable of shifting both left and right by up to 63bs, and a 64b double adder to enable parallel addition and rounding in the final stage.

Unique circuit structures and detailed analysis techniques are necessary to support the clock rate in conjunction with the complexity demanded by the concurrence and wide data paths. The clocking method is level-sensitive single-phase. Since there is no "dead time" with this scheme, it is imperative that clock integrity be assured to avoid race-through in latches. It is also important to avoid clock skew contribution to delay paths. A single-driver approach is used for clock distribution, with the third metal layer for the majority of clock routing (as well as VDD/VSS). Total capacitive load on the clock driver is 3250pF, requiring a final driver width of 25cm and 10cm for pMOS and nMOS respectively. The clock driver resides in the horizontal center of the chip and extends vertically from top to bottom of the core. A method to extract and display clock skew was developed to analyze the grid. Figure 2 shows a topographic representation of on-chip clock skew.

In contrast to a previous high-speed design, single-cycle latency is achieved in the 64b adder portion of the integer and floating-point ALUs by using a combination of logic and circuit techniques [1]. The logical scheme is a hybrid of two techniques: Manchester carries for the initial 8b groups, followed by a logarithmic carry-select tree [2].(Figure 3) The Manchester scheme is unique in that the nMOS chain is precharged low and is conditionally pulled high. This avoids threshold delays in the pass transistors and improves performance of the carry chain by 10% over the pull-down approach.

To provide maximum flexibility in applications, the external interface allows several different modes of operation. This includes choice of logic family (CMOS/TTL or ECL) as well as bus width (64/128b), external cache size and access time, and BIU clock rate. These parameters are set into mode registers during chip power-up. The logic family choice provides an interesting circuit challenge. The input receivers are differential amplifiers that utilize an external reference level. To maintain signal integrity of this reference voltage, it is resistively isolated and RC filtered at each receiver as shown in Figure 4. The output driver presents a more difficult problem due to the 3.3V VDD chip power supply. To provide a good interface to ECL, it is important that the output driver pull to the VDD rail (for ECL operation VDD=0V, VSS=-3.3V). This precludes using nMOS pull-ups. pMOS pull-ups have the problem of well-junction forward bias and pMOS turn-on when bi-directional outputs are connected to 5V logic in CMOS/TTL mode. The solution, a floating-well driver circuit, shown in Figure 5, avoids the cost of series pMOS pull-ups in the final stage [3].

Acknowledgments

The authors acknowledge the contributions of: H. Akhiani, B. Benschneider, L. Binder, D. Bertucci, V. Calderon, S. Carroll, G. Cheney, D. Conroy, B. Cooper, M. D'Addeo, T. Daum, M. Decker, T. Equi, S. Hassoun, B. Hicks, C. Holub, T. Jacobs, A. Jain, D. Jessel, M. Kantrowitz, S. Kreider, S. Kumar, A. Ladd, D. Lau, S. Lloyd, S. Lowell, R. Matthew, S. Meier, S. Morris, J. Pan, V. Rai, N. Rethman, S. Shah, D. Sites, D. Sleeper, C. Somanathan, E. StJohn, M. Tareila, M. Taylor, Lang Tran, Luan Tran, H. Tumblin, L. Wolf, Y. Yen.

References

[1] Conrad, R., et. al., "A 50 MIPS (Peak) 32/64b Microprocessor", ISSCC DIGEST OF TECHNICAL PAPERS, pp. 76-7, Feb., 1989.

[2] Sklansky, J., "Conditional-Sum Addition Logic", IRE Trans. Electron. Comput. EC-9, pp. 226-231, 1960.

[3] Lee, H., et. al., "An Experimental 1Mb CMOS SRAM with Configurable Organization and Operation", ISSCC DIGEST OF TECHNICAL PAPERS, pp. 180-181, Feb., 1988.

Figure 1: See page 256.

Feature size	0.75µm
Channel length	0.5µm
Gate oxide	10.5nm
VTXn/VTXp	0.5V/-0.5V
Power supply	3.3V
Substrate	p epi with n well
Salicide	Cobalt-disilicide
Burried contact	Titanium nitride
Metal 1	7.5kÅ AlCu, 2.25µm pitch (contacted)
Metal 2	7.5kÅ AlCu, 2.625µm pitch (contacted)
Metal 3	20kÅ AlCu, 7.5µm pitch (contacted)

Table 1: CMOS-4 technology.

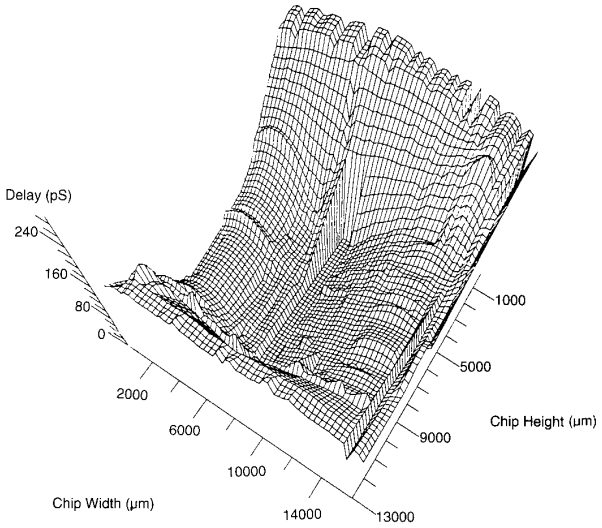

Figure 2: Clock delay vs position.

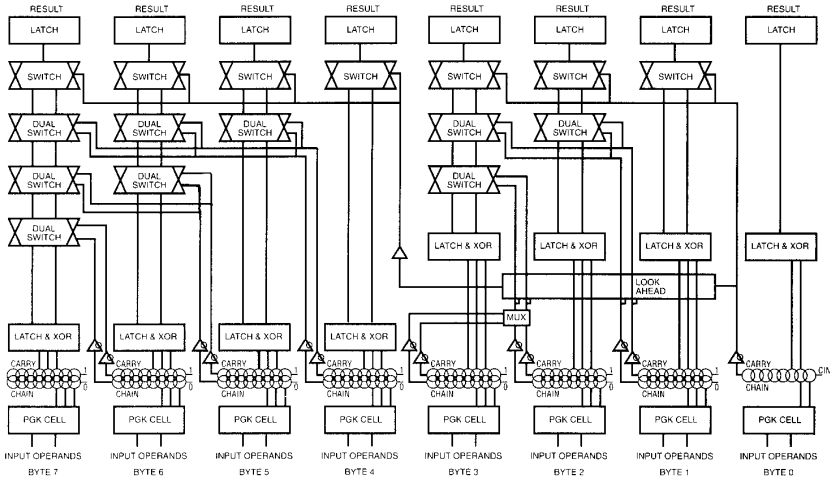

Figure 3: Diagram of 64b adder.

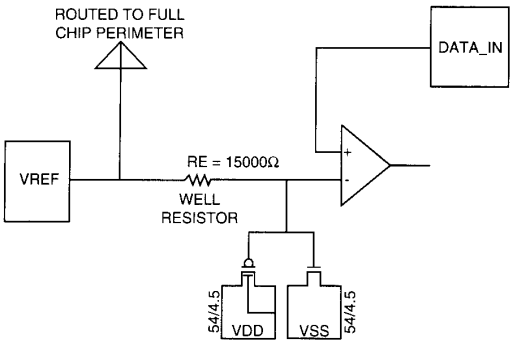

Figure 4: Reference voltage filter network at each input pin.

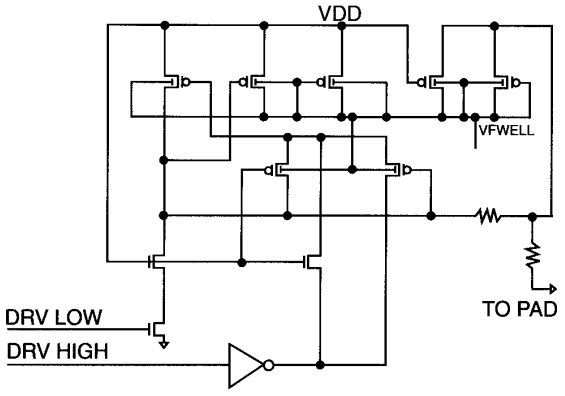

Figure 5: Floating-well output driver.