Chapter 7

HIGH-PERFORMANCE ENERGY-EFFICIENT DUAL-SUPPLY ALU DESIGN

Sanu K. Mathew, Mark A. Anders, and Ram K. Krishnamurthy

Circuits Research Laboratories, Intel Corporation, Hillsboro, OR, USA

Abstract: This chapter describes the design of a single-cycle 64-bit integer execution ALU fabricated in 90 nm dual-Vt CMOS technology, operating at 4 GHz in the 64-bit mode with a 32-bit mode latency of 7 GHz (measured at 1.3 V, 25° C). The lower- and upper-order 32-bit domains operate on separate off-chip supply voltages, enabling conditional turn-on/off of the 64-bit ALU mode operation and efficient power-performance optimization. High-speed single-rail dynamic circuit techniques and a sparse-tree semi-dynamic adder core enable a dense layout occupying 280 × 260 µm² while simultaneously achieving (i) low carry-merge fan-outs and inter-stage wiring complexity, (ii) low active leakage and dynamic power consumption, (iii) high DC noise robustness with maximum low-Vt usage, (iv) single-rail dynamic-compatible ALU write-back bus, (v) simple 2Φ 50% duty-cycle timing plan with seamless time-borrowing across phases, (vi) scalable 64-bit ALU performance up to 7 GHz measured at 2.1 V, 25° C, and (vii) scalable 32-bit ALU performance up to 9 GHz measured at 1.68 V, 25° C.

Key words: arithmetic and logic unit (ALU); sparse-tree architecture; semi-dynamic design; dual-supply voltage design.

1. Introduction

Fast 32-bit and 64-bit arithmetic and logic units (ALU) with single-cycle latency and throughput are essential ingredients of high-performance superscalar integer and floating-point execution cores. Furthermore, in a typical
ALU operation, the lower-order 32 bits of the ALU output are required early for address generation and rapid back-to-back operations [1]. These constraints require a high-performance ALU with a compact layout footprint that minimizes interconnect delays in the core. ALU also contribute to one of the highest power-density locations on the processor, resulting in thermal hotspots and sharp temperature gradients within the execution core. The presence of multiple execution engines in current-day processors [1] further aggravates the problem, severely impacting circuit reliability and increasing cooling costs. Therefore, this strongly motivates energy-efficient ALU designs that satisfy the high-performance requirements, while reducing peak and average power dissipation. Traditional dense-tree adder architectures such as Kogge–Stone [2] use full binary carry-merge trees that result in large transistor sizes because of their high fan-outs and require wide routing channels for inter-stage wiring. Adder architectures such as Ladner–Fischer [3] address the wiring problem by reducing the number of inter-stage interconnects at the expense of exponentially increasing carry-merge fan-outs.

In this chapter a single-cycle 64-bit integer execution ALU [4] fabricated in 90 nm dual-Vt CMOS technology [5] is described. A sparse-tree adder architecture is employed to address the high fan-out issue mentioned above, as well as to reduce the inter-stage wiring complexity by up to 80% [6]. High-speed single-rail dynamic circuit techniques and migration of non-critical paths to fully static CMOS enable low carry-merge fan-outs, low active leakage and dynamic power consumption, high DC noise robustness, and a dense layout. The complete 64-bit ALU operates at 4 GHz in the 64-bit mode measured at 1.3 V, 25°C and consumes 300 mW total power. The corresponding 32-bit mode latency is 7 GHz, also measured at 1.3 V, 25°C. The lower- and upper-order 32-bit domains operate on separate off-chip supply voltages, enabling conditional turn-on/off of the 64-bit ALU mode operation and efficient power-performance optimization, resulting in up to 22% power savings. The 64-bit ALU performance is scalable up to 7 GHz measured at 2.1 V, 25°C, and the 32-bit ALU performance is scalable up to 9 GHz measured at 1.68 V, 25°C. Burn-in tolerant conditional keepers are inserted on all dynamic gates to enable full functionality under worst-case noise conditions and elevated supply/temperature stress tests [7].

The remainder of this chapter is organized as follows: Section 2 describes the organization of the ALU; Sections 3 and 4 present the key circuits that enable an energy-efficient single-rail ALU design; the ALU clocking scheme and timing plan are described in Section 5; Section 6 discusses the benefits of this design over a conventional dual-rail dynamic implementation; Section 7 presents the 90 nm dual-Vt CMOS implementation and silicon measurement results; the dual-supply voltage operation of the ALU is described in Section 8. Finally the chapter is summarized in Section 9.