Design and Fabrication of a Microprocessor Using Adiabatic CMOS and Bennett Clocking

Ismo K. Hänninen¹, César O. Campos-Aguillón², Rene Celis-Cordova², and Gregory L. Snider¹(✉)

¹ Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
² Department of Electrical and Computer Engineering, Tecnológico de Monterrey, Monterrey, Mexico

Abstract. This paper will describe the design and implementation of a MIPS-based microprocessor using Bennett clocking to implement reversible logic. In Bennett clocking the clock signals form a “cascade” that moves information forward through logic gates in the compute phase, and then recovers energy during a decompute phase, forming a reversible logic circuit. New logic design and verification tools were developed, using structural Verilog and extensions to ModelSim to address the issues of adiabatic clocking, tools that are currently unavailable in commercial packages. The microprocessor is based on a simplified version of the MIPS architecture. After verification by our design tools it was then implemented using CMOS standard cells based on split-level charge recovery logic. The final design contains approximately 5700 transistors, and is currently being fabricated at MOSIS.

Keywords: Reversible microprocessor · Adiabatic CMOS · Bennett clocking

1 Introduction

In the 1960s computers dissipated a large amount of power and required significant cooling, so they were located in rooms with dedicated air conditioning units. This large power dissipation helped drive the transition in the 1970s from BJT-based computers to ones based on MOSFETs. BJTs were faster, but FETs produced less heat. Likewise power dissipation drove the transition in the early 1980s from NMOS to CMOS. NMOS was faster but CMOS produced less heat. In the 1980s and early 1990s the power requirements of computers were relatively low and significant computing resources could be placed in a closet with minimal cooling. During this time the need for processing speed was the paramount issue, and the exponential increases in power dissipation and heat each year were minor concerns. But by the late 1990s progress in devices had led to systems where power dissipation was again becoming a significant problem, and charts extrapolating the exponential increase in dissipation showed the heat production per unit area of a processor surpassing that of the sun, ~6 kW/cm², by 2017.
Clearly something had to change. Because of power dissipation the progress in device size that underlies Moore’s Law could not immediately be translated into a performance increase, undermining the business models of semiconductor manufactures. Unlike previously there was no new device or circuit primitive that could immediately offer lower dissipation. The solution needed to be at the system level, and the approach decided upon was multi-core architectures. Here, the clock frequency was held nearly constant to control power dissipation, but splitting up the computing task and spreading it among parallel computing cores produced overall performance gains. Through the 2000s the most important processor design constraint was to keep the power dissipation below 200 W/cm², the practical limit of air-cooling. To meet this constraint, it became necessary to turn off parts of the chip at certain times to reduce power, a practice known as Dark Silicon. It has been projected that by the 8 nm node 50-80% of a chip will be “dark” at any given time [1], which begs the question: why have that many transistors if they can’t be used?

Today, high-performance computers are back to requiring huge amounts of power and cooling. A number of data centers are being built where the waste heat can be used to heat buildings. Facebook recently opened a data center in Lulea Sweden, near the Arctic Circle, located in part to take advantage of the low ambient temperature for cooling. Data Centers consumed approximately 91 billion kilowatt-hours, 2-3% of electrical power in the US in 2013, a number that is expected to grow to 140 billion kW-hr by 2020 [2].

There are a number of factors contributing to the power dissipation in computation, but processing chips themselves contribute a significant fraction, and the power density is limiting chip development and utilization. Processor clock frequencies have not increased in a decade although the transistors used are nearly an order of magnitude faster. Without a significant change in the underlying approach to computation the rate of increase in computing performance will slow significantly. Moving forward requires an approach that combines devices and system architectures.

Today’s computers encode information with charge stored on capacitors, the CMOS gate and interconnect capacitors. Power dissipation for standard CMOS logic is given by the equation

\[ P_{Total} = N\left(\alpha CV_{DD}^2f + P_{Passive}\right) \]  

(1)

where \( V_{DD} \) is the supply voltage, \( C \) is the load capacitance at the output of each logic gate, \( N \) is the number of gates, \( \alpha \) is the activity factor, and \( f \) is the operating frequency. The first term represents the active power dissipation, i.e., the power dissipated in processing information. The second term, the passive power dissipation, is power that is simply wasted because a voltage is applied to the circuit.

Even modest projections of future device density and switching speed highlight the severity of the power issue. By lowering the supply voltage and using ideal switches with no leakage it may be possible to lower the switching energy to 100 kBT, considered a practical limit for error-free computing. However, a device density of \( 10^{11} \) cm² and a clock frequency of 100 GHz with \( \alpha=1 \) will give an active power of 4 kW/cm², close to that of the surface of the sun. Approaches to increase device density such as 3D integration will only exacerbate the power density problem.