Chapter 5

METHODOLOGY TO OPTIMIZE ENERGY OF COMPUTATION FOR SOCS

Jagesh Sanghavi, Eliot Gerstner
Tensilica
3255-6 Scott Boulevard
Santa Clara, CA 95054
sanghavi@tensilica.com, gerstner@tensilica.com

We present a novel energy optimization methodology based on processor customization. Unlike previous approaches focused either on behavioral-level optimization with approximate consideration for underlying hardware, or register transfer level (RTL), or gate-level power optimization with limited microarchitectural trade-offs, the new approach compiles cycle count reducing instruction extension description to synthesizable hardware and accurately estimates dynamic power at the register transfer level. For a sample set of digital signal processing (DSP) applications, we see energy reductions exceeding a factor of 10× compared to fixed instruction set processors.

5.1 INTRODUCTION

Power is an important design consideration for a range of battery-operated consumer electronic devices such as PDAs (personal digital assistants), cell phones, and digital cameras. To increase battery life during active use, the real metric to minimize is the energy of computation, i.e., the area under the power curve as a function of time. Secondly, these devices have bursty computation requirements during which a specific signal processing task is performed by a functional unit. Hence, it is important to effectively reduce the power dissipated by a functional unit when it is in the idle state. Finally, these devices must be programmable to cope with evolving standards requirements and provide feature evolution on the same hardware platform.

The dissipated power consists of three components: switching power, short-circuit power, and leakage power. The switching component is power dissipated by charging and discharging circuit nodes. The short-circuit component is due to short-circuit currents when both P-channel and N-channel
transistors are partially on during output signal transition. The leakage power is primarily due to gate leakage and subthreshold leakage. Although leakage power dissipation has received a lot of attention, the issue may be mitigated by process technology advances (multiple threshold voltages and high dielectric constant gate oxide), ASIC design methodology changes [14], and non-uniform scaling. Voltage scaling has been an effective technique to reduce the dynamic power (sum of switching and short-circuit power) with every new process technology due to quadratic dependence of power on the supply voltage. However, as the device geometries shrink further to 65nm and 45nm transistor gate lengths, the energy minimization will need to decrease its reliance on voltage scaling and will need to rely more on architectural and microarchitectural explorations, effective clock gating, and design methodology employing power rail shut-off techniques.

The impact of architectural and microarchitectural changes on power requires accurate estimation of the dynamic power. At the minimum, the dynamic power that depends on switching activity in the circuit requires an RTL description to estimate the power with reasonable accuracy [15]. However, it is extremely difficult to explore major architectural and microarchitectural changes while designing at the register transfer level. With time-to-market schedule constraints for a reasonably complex design, it is possible only to perform very limited design explorations. The previous work has focused on behavioral-level power optimization [1] with approximate consideration for the underlying hardware implementation and concomitantly inaccurate power estimates [7].

Extensible processors [5][19] have been proposed as a solution to dramatically improve the application performance. Application-specific processors can be extended by adding custom instructions to efficiently implement algorithmic kernels. By customizing the processor for a specific application or class of applications, extensible processors are able to drastically reduce the cycle count for a range of application benchmarks [4].

In this chapter, we propose a new methodology to optimize the energy of computation based on customizing an extensible processor. The new approach compiles the instruction extension description into synthesizable hardware and uses RTL power estimation to accurately focus on the dynamic power dissipation. The new approach reduces the area under the power curve over time by dramatically reducing the cycle count and shuts off the power to instruction extension units when they are idle.

The rest of the chapter is organized as follows. We define the problem and motivate the solution approach in Section 5.2. We present the energy minimization methodology in Section 5.3. We present the experiment results on a set of case studies in Section 5.4. Finally, we conclude the paper and provide directions for future research in Section 5.5.