Timing Correction and Optimization with Adaptive Delay Sequential Elements

Kambiz Rahimi, Seth Bridges, Chris Diorio
University of Washington, Seattle, WA
kamr@ee.washington.edu, seth@cs.washington.edu, diorio@cs.washington.edu

Abstract

This paper introduces Adaptive Delay Sequential Elements (ADSEs). ADSEs are registers that use nonvolatile, floating-gate transistors to tune their internal clock delays. We propose ADSEs for correcting timing violations and optimizing circuit performance. We present an ADSE circuit example, system architecture, and tuning methodology. We present experimental results that demonstrate the correct operation of our example circuit and discuss the die-area impact of using ADSEs. Our experiments also show that voltage and temperature sensitivity of ADSEs are comparable to non-adaptive flip-flops.

1. Introduction

Delay prediction and modeling are increasingly difficult tasks. Even with properly modeled delays, some percentage of fabricated circuits will fail due to delay variations caused by manufacturing tolerances. We propose Adaptive Delay Sequential Elements (ADSEs) for fine-tuning circuit timing after fabrication. ADSEs cannot prevent delay variations due to voltage and temperature changes, noise or crosstalk. However, ADSEs allow post-fabrication adjustment of setup/hold margins to accommodate these variations. ADSEs electrically tune their internal clock delays which enables clock skew optimization. Unlike previous methods [1]-[2], clock skew optimization with ADSEs does not require the addition of delay cells or exact manipulation of clock routing.

ADSEs’ tuned clock delays are nonvolatile and the tuning can be performed using minimal external control, without dedicated access to each floating gate. We build ADSEs by embedding floating-gate synapse transistors [3] within the sequential cell structure and using their adaptive properties to adjust internal clock delays. Researchers have previously used synapse transistors in many applications [4], [5]. ADSEs utilize synapse transistors to correct timing problems in mainstream digital designs.

A synapse transistor is a four-terminal MOS device comprising two pFETs with a common floating gate; the floating-gate charge represents a nonvolatile analog weight. One pFET adds electrons to the floating gate using impact-ionized hot-electron injection (IHEI) [6], and allows reading the stored value by measuring the pFET’s channel current. The second pFET has shorted drain and source, and removes electrons from the floating gate by means of electron tunneling [7]. Synapse transistors require a ~70Å oxide thickness for reliable charge storage. However, ADSEs can be fabricated in modern processes with thinner oxides by using the thicker oxide, required for I/O pads, within synapse transistors.

2. Adaptive Delay Sequential Elements

Adaptive versions of many static and dynamic sequential elements are possible. Here, we use the simple example of an Implicit-pulsed Master-slave Static (IMS) flip-flop to demonstrate our adaptive methods.

We augmented an IMS with our Adaptive Clock Generator (ACG). The resulting circuit, shown in Figure 1, is an Adaptive-delay Implicit-pulsed Master-slave Static (AIMS) flip-flop. AIMS is a scan testable, negative-edge triggered flip-flop. The ACG delay is controlled by the synapse transistor’s floating-gate voltage. We cannot measure the floating-gate voltage directly. However, when the injector pFET is saturated, we can infer the floating-gate voltage from the injector pFET’s drain current, which we refer to as the tuning current. During tuning operations, we control and read the tuning current with off-chip instruments and use it to adjust the floating-gate voltage. During normal operation, no tuning current flows and the floating-gate voltage sets the ACG delay.

We use IHEI to decrease the ACG delay in one of two distinct modes, GLOBAL and SELECTIVE. The GLOBAL mode is self-limiting and we use it to initialize ACGs to their minimum delays. We check for timing violations after initialization and then increase all ACG delays by tunneling electrons off the floating gates. After tunneling we use SELECTIVE injection to decrement individual ACG delays. SELECTIVE injection is possible because the logic value stored in each ADSE can control IHEI in its synapse transistor. We select individual cells for
injection by shifting in bit vectors through the scan chains.

The combination of concurrent delay increment and selective delay decrement enables us to set individual flip-flop clock skews without additional addressing circuitry. The ability to read tuning currents allows for closed-loop stepping of tuning currents and delays. Accurate delay control is possible with a small step size despite the non-linear dependence of delay on floating-gate voltage.

3. Experiments

We fabricated several ADSE test circuits in a 0.35 μm process. We used tunneling pulses to increase the Clk-Q delay of an AIMS flip-flop, followed by injection pulses to decrease the delay. Figure 2 shows how the Clk-Q delay of an AIMS flip-flop changes with cumulative injection time. We verified that applying short (1ms) injection pulses can adjust the tuning current with 1nA resolution corresponding to a delay change of approximately 1ps. We also show measured tuning currents after each injection pulse.

We experimentally verified that supply voltage and temperature sensitivity of ADSE delays are comparable to their non-adaptive counterparts and that tuned delay values are non-volatile.

ADSE layouts are larger than their non-tunable counterparts since they contain more transistors and tunneling devices that require wider spacing. AIMS is 39% larger than its non-adaptive counterpart (IMS). The impact of ADSEs on chip area we computed the area of ISCAS89 benchmark circuits [8] using layout areas from a typical 0.35-micron cell library. If we replace every IMS flip-flop with an AIMS the overall area increases from 9% to 17%. It is also possible to selectively replace non-adaptive flip-flops with ADSE to reduce the area penalty. ADSE have virtually no impact on power dissipation in normal operating mode.

References