Injection-Locked Clocking: A Low-Power Clock Distribution Scheme for High-End Microprocessors

Hui Wu, Lin Zhang, Aaron Carpenter, Alok Garg, and Michael Huang
Department of Electrical and Computer Engineering
University of Rochester, Rochester, NY 14627
{hwu, linzhang, carpente, garg, huang}@ece.rochester.edu

Abstract

Distributing high quality clock signals is one of the most challenging tasks in high-performance microprocessors. Clocking circuitry accounts for an overwhelming amount of total power consumption in multi-GHz processors. Unfortunately, deteriorating clock skew and jitter make it very difficult to reduce power dissipation. We propose a new global clocking scheme, injection-locked clocking (ILC), to combat clock skew and jitter. This new scheme uses injection-locked oscillators as the clock receivers. It can achieve better power efficiency and jitter performance than conventional buffered trees with the additional benefit of built-in deskewing. Unlike other proposed clocking schemes, ILC is fully compatible with conventional clock distribution. In this paper, a quantitative study based on circuit- and microarchitectural-level simulations is conducted to verify the performance and power consumption improvements using ILC compared to conventional clocking. Alpha21264 is used as the baseline processor, and is scaled to 0.13µm technology and 3GHz. Two ILC configurations are constructed: one of them simply replaces the top-level global clock X/H-tree in Alpha21264; another one is optimized for ILC operation. Both circuit and architectural simulations show 20ps and 23ps reduction in jitter, 10.1% and 17% power savings, respectively, using the two ILC configurations in comparison to the conventional clocking.

1 Introduction

Clock distribution is a crucial aspect of modern multi-GHz microprocessor design. Conventional distribution schemes are more or less monolithic in that a single clock source is fed through hierarchies of clock buffers to eventually drive almost the entire chip. This raises a number of challenges. First, due to irregular logic, the load of the clock network is non-uniform, and the increasing process and device variations in deep sub-micron semiconductor technologies further adds to the spatial timing uncertainties known as clock skews. Second, the load of the entire chip is substantial, and sending high quality clock signal to every corner of the chip necessarily requires driving the clock distribution network “hard”, usually in full swing of power supply voltage. Not only does this mean high power expenditure, but it also requires a chain of clock buffers to deliver the ultimate driving capability. These active elements are subject to power supply noise, and adds delay uncertainty – jitter – which also eats into usable clock cycle. Jitter and skew combined represent about 18% of cycle time currently [25], and that results in indirect energy waste as well. For a fixed cycle time budget, any increase in jitter and skew reduces the time left for the logic. To compensate and make the circuitry faster, the supply voltage is raised, therefore increasing energy consumption. Conversely, any improvement in jitter and skew generates timing slack that can be used to allow the logic circuit to operate more energy-efficiently.

As commercial microprocessors are rapidly becoming multi-core systems, monolithic clock distribution will be even less applicable. In the era of billion-transistor microprocessors, a single chip is really a complex system with communicating components and should be treated as such. In communication systems, synchronizing clocks is also a rudimentary and crucial task. In this paper, we apply the concept of injection locking and the latest innovation in circuit implementation to clock distribution in microprocessor.

Injection locking is the physical phenomenon where an oscillator “locks on” to an external stimulus (a periodic signal) and fundamentally synchronizes with the input when the frequency of the input signal is close enough to the oscillator’s native frequency or its (sub)harmonics. Recent circuit implementation of injection-locked oscillators (ILO) not only demonstrated superb gain and noise rejection, but also showed flexible frequency multiplication and division capabilities and phase adjustment capabilities. Using ILOs, the (global) clock distribution of a microprocessor can be improved substantially. For example, all logic macro blocks can be clocked by independent ILOs connected to a low-swing global clock input.
signal. Compared to the traditional approach where a local clock buffer is being “powered” by a full-swing clock signal directly driven from a central source, ILOs enable much lower power expenditure on the global level and eliminate multiple levels of clock buffers, which in turn, reduces clock jitter. Additionally, we can further reduce clock skew time leveraging phase shift capabilities of the state-of-the-art design of ILOs.

Apart from the technical advantages, injection-locked clocking is also a non-intrusive technology. There is no need to change the processor architecture or the design methodology the way (partially) asynchronous designs do. All in all, ILOs promise to bring significant advantages to current and future high-speed microprocessors and open up opportunities to design novel clocking schemes. Given the significant investment in IPs and design tool chains in the synchronous regime and the fact that high-end microprocessors routinely spend 40% or more on clock distribution, injection-locked clocking (ILC) is a promising solution to meet the increasing challenge of clock distribution.

In this paper, we discuss a few possible scenarios of using injection locking for clock distribution. We also perform a detailed quantitative analysis comparing some options of ILC designs with conventional approaches in terms of power consumption. Due to the scarcity of detailed reports on processor clock distribution, especially its power consumption, in the public domain, our study is constrained to a few ILC options that are suboptimal. Even using these limited options, simulation results suggest that power consumption of a high-end processor reduces from 40.7W to 33.9W, a 17% reduction. This clearly shows the potential of ILC.

The rest of the paper is organized as follows: Section 2 analyzes the skew and jitter problem in conventional clocking; Section 3 discusses proposed clocking scheme based on injection-locking; Section 4 presents a case study and summarizes the experimental setup; Section 5 shows the quantitative analysis; Section 6 discusses related work; and finally, Section 7 concludes.

2 Challenges in Conventional Clocking

Figure 1 shows a typical conventional clock distribution scheme. The global clock is generated by an on-chip phase-locked loop (PLL) from an off-chip reference clock, usually a crystal oscillator at tens of MHz. The global clock is distributed using an H-tree, which consists of interconnect transmission lines and clock buffers, and then further distributed by local clock distribution networks. In order to minimize the global clock skew, the global clock-distribution network has to be balanced by meticulous design of the transmission lines and buffers.

This practice puts a very demanding constraint on the physical design of the chip. Even so, the ever-increasing process variations with each technology generation still results in greater challenges in maintaining a small skew budget. Another current practice is to use a grid instead of a tree for clock distribution, as shown in the upper-left local clock region in Figure 1. A grid has a lower resistance than a tree between two end nodes, and hence can reduce the skew. At the same time, a grid usually has much larger parasitic capacitance (larger metal layers) than an equivalent tree, and therefore takes more power to drive. Passive and active deskew methods [11, 21, 29, 34] have also been employed to compensate skew after chip fabrication. Apparently this approach increases the chip complexity, manufacturing cost, and in the case of active deskew, power consumption and jitter.

Figure 1. Conventional global clock distribution, showing an H-tree topology with interconnects and clock buffers [10].

Jitter poses an even larger threat to microprocessor performance and power consumption. The global-clock PLL and clock-distribution network generate noise, and hence contribute to global clock jitter. But the main culprit is usually the noise coupled from other circuits, such as power supply noise, substrate noise, and cross-talks. Short-term jitter (cycle-to-cycle jitter) can only be accounted for by adding timing margin to the clock cycle, and hence degrades performance. Unlike skew, jitter is very difficult to compensate due to its random nature. In order to reduce jitter, the interconnect wires in the global clock distribution network needs to be well shielded from other noise sources, usually by sandwiching them between Vdd/ground wires and layers. Shielding inevitably increases the parasitic capacitance of the clocking network, which means more and larger clock buffers, and hence larger power dissipation to drive them. In turn, having more buffer stages introduces another source of jitter, and the situation deteriorates quickly with faster clock speed. It is evident that current skew and jitter reduction techniques almost always result in higher power consumption. A better clocking scheme with less jit-
ter and skew directly translates into power savings for a given performance target.

3 Injection-Locked Clocking

3.1 Injection-Locked Oscillators

Injection locking [1, 22] is a special type of forced oscillation in nonlinear dynamic systems (also known as synchronization). Suppose a signal of frequency \( \omega \) is injected into an oscillator (Figure 2-a), which has a self-oscillation (free-running) frequency \( \omega_0 \). When \( \omega \) is quite different from \( \omega_0 \), “beats” of the two frequencies are observed. As \( \omega \) approaches \( \omega_0 \), the beat frequency \( |\omega - \omega_0| \) decreases. When \( \omega \) enters some neighborhood very close to \( \omega_0 \), the beats suddenly disappear, and the oscillator starts to oscillate at \( \omega \) instead of \( \omega_0 \). The frequency range in which injection locking happens is called the locking range (Figure 2-b). Injection locking also happens when \( \omega \) is close to the harmonic or subharmonic of \( \omega_0 \), i.e., \( n\omega_0 \) or \( \frac{1}{n}\omega_0 \). The former case can be used for frequency division, and the latter for frequency multiplication.

![Figure 2](image)

**Figure 2.** (a) Beat and injection locking phenomenon when an oscillator is driven by an input single-frequency signal. (b) locking range.

An injection-locked oscillator (ILO) can be considered as a simple first-order PLL (Figure 3-a), in which nonlinearity of the oscillator core functions as a phase detector. For example, in a typical divide-by-2 ILO (Figure 3-b) [28], the oscillator core (consisting of \( M_1 \), \( M_2 \) and \( M_3 \)) also serves as a single-balanced mixer for phase detection. Because of the simple structure, ILOs consume much less power than a full-blown PLL and can operate at extremely high clock speeds [37]. The fact that the built-in “phase detectors” are mixer-based also explains why ILOs can operate at the harmonic and subharmonic frequencies of the input signal.

Once locked to the input signal, the output of ILOs will maintain a determined phase relative to the input signal (Figure 4). The phase difference from the input signal to the output is determined by the injection signal strength, the frequency shift from its free-running oscillation frequency, and the frequency characteristics of the oscillator resonator. As shown in Figure 4, the phase shift \( \varphi \) is a monotonic function of the frequency shift \( \Delta \omega \), and the function is quite linear within the locking range except when close to the edges. By tuning the free-running frequency of the oscillator, we can tune the phase of the output signal [41]. Converting to time, this means tunable delay for the output signal. This phase transfer characteristics can be utilized to achieve deskew between different clock domains with no need for other deskew circuits.

![Figure 3](image)

**Figure 3.** (a) A generic model of an injection-locked oscillator (ILO). (b) a divide-by-2 ILO based on a common differential LC oscillator.

![Figure 4](image)

**Figure 4.** Phase transfer functions for divide-by-2 ILOs in Figure 3-b. \( \eta \equiv \frac{I_{\text{inj}}}{I_{\text{bias}}} \) is the injection ratio, \( \omega_1 \) is the free-running oscillation frequency, \( \Delta \omega \equiv \omega - \omega_0 \) is the frequency shift, and \( Q \) is the LC tank quality factor.

3.2 Clocking using ILOs

In this paper, we proposed a new clocking scheme as shown in Figure 5. Similar to conventional clocking, the global clock is generated by an on-chip PLL and distributed by a global tree. The difference is that we use injection-locked oscillators (ILOs) to regenerate local clocks, which are synchronized to the global clock through injection locking (see Section 3.1). Another difference is that most global clock buffers in conventional clocking are removed because the sensitivity of ILOs are much greater than digital buffers (see detailed discussion below). Essentially, we use ILOs as local clock receivers, similar to the idea of clock recovery in communication systems. Note that this is different from resonant clocking [8], where all the oscillators are coupled together.

2006 IBM P=ac2 Conference
(see Section 6). Further, ILOs can be constructed as frequency multipliers [20] or dividers [28, 39], and hence this scheme enables local clock domains to have higher \((n \times f_0)\) or lower clock speed \((f_0/m)\) than the global clock \((f_0)\). Such a global-local clocking scheme with multiple-speed local clocks offers significant improvements over conventional single-speed clocking scheme in terms of power consumption, skew, and jitter.

![Image](see Section 6). Further, ILOs can be constructed as frequency multipliers [20] or dividers [28, 39], and hence this scheme enables local clock domains to have higher \((n \times f_0)\) or lower clock speed \((f_0/m)\) than the global clock \((f_0)\). Such a global-local clocking scheme with multiple-speed local clocks offers significant improvements over conventional single-speed clocking scheme in terms of power consumption, skew, and jitter.

3.3 Power Savings

Injection-locked clocking can lead to significant power savings in high-speed microprocessors. The benefits come from several aspects. First, the possible combination of a low-speed global clock and high-speed local clocks can reduce the power consumption in the global clock distribution network. In the conventional approach, this would require multiple power-hungry PLLs for frequency multiplication. An ILO consumes much less power than a PLL because of their circuit simplicity [37]. This will become more evident in multi-core processors.

Second, ILOs have higher sensitivity than buffering inverters. As a synchronized oscillator, an ILO effectively has very large voltage gain when the injection signal amplitude is small, while the gain of an inverter is much smaller (Figure 6). This can be easily understood if we realize that synchronization in an ILO is usually achieved in tens to hundreds clock cycles, and hence in each clock cycle only a small amount of injection locking force is needed. While a digital inverter needs to change its state twice in every clock cycle. Therefore, the global clock signal amplitude can be much smaller in the new clocking scheme, which means less power loss on the parasitic capacitance and resistance of the global-clock distribution network. This will be increasingly attractive as the interconnect loss becomes dominant factors as the process technology scales further.

Further, the number of clock buffers in the global clock distribution can be reduced. In conventional clocking, in order to minimize jitter generated by digital buffers, the global clock signal needs to be driven from rail to rail throughout the whole network, and in turn many clock buffers are inserted. In injection-locked clocking, ILOs can achieve good jitter performance with small input signal amplitude (see Section 3.5). Therefore, the global clock signal amplitude no longer needs to be full Vdd swing, and few (or none at all) clock buffers are needed on the global tree. Reduced number of clock buffers directly translates into lower power consumption.

More importantly, because injection-locked clocking significantly lowers skew and jitter in the global clock, the timing margin originally allocated can be recovered, and used for circuit operation. This can enable faster clock speed. Or, we can trade it for lower power supply voltage (Vdd), and save power dissipation from not only clock distribution network, but all the logic gates on the chip. In Section 4 below, we demonstrate the power savings from all aspects using a quantitative case study.

3.4 Skew Reduction and Deskew Capability

Because the number of buffers is reduced in the new clocking scheme, skew due to mismatch of clock buffers is reduced compared to conventional clocking. More importantly, ILC provides a built-in mechanism for deskew. From Section 3.1, the phase difference between the input and output signals of an ILO can be tuned by adjusting its center frequency. This phase tuning capability enables ILOs to serve as built-in “deskew buffers”, and conventional deskew architectures can be applied directly. For example, similar to active deskewing in conventional clocking, phase detectors can be placed between local clock domains to check skew, and then tune corresponding ILOs. Removing dedicated deskew buffers not only saves power, but also reduces their vulnerability to power supply noise. Note that ILC deskewing is different from the distributed PLL approach [15, 27], where phase detectors have to be added between all adjacent clock domains for frequency synchronization, and then possibly for deskew. In injection-locked clocking, frequency synchronization is achieved by injection locking, and the phase detection is used for deskew only. In other words, injection-locked clocking with deskew tuning is a dual-
loop feedback system, and therefore provides both good
tuning speed and small phase error (residue skew). Be-
cause of the excellent built-in deskew capability of ILOs,
it can be expected that an injection-locked clock tree has
much more freedom in its physical design (layout).

3.5 Jitter Reduction and Suppression

Injection-locked clocking can significantly reduce jitter in
global clock distribution networks. First, reduced
number of global clock buffers also means less pick-up
of power supply and substrate noise, and hence less jitter
generation and accumulation. Second, because the de-
sign freedom in layout, clock interconnect can be placed
where there is minimal noise coupling from adjacent cir-
cuits and interconnects. In addition, similar to a PLL, an
ILO can suppress both its internal noise (low-pass filter-
ing) and input signal noise (high-pass filtering). So it can
possibly lower the input signal jitter at its output [16,37].
Overall, injection-locked clocking is likely to achieve
better jitter performance than conventional clocking.

3.6 Potential Applications

With the numerous technical advantages, ILC can be
used to improve high-end microprocessors and the de-
sign process in many ways:

First, ILC reduces jitter and skew compared to a con-
tventional clocking network. This reduces cycle time
and therefore allows a faster clock speed. As technol-
ogy scaling improves transistor performance but does
not reduce jitter and skew (which actually increase), the
improvement in clock speed will be more pronounced
over time. Although further increasing whole-chip clock
speed finds limited practical appeal in today’s setting, it
may still be effective in certain specialized engine inside
a general-purpose architecture.

Second, using ILC, clock distribution for a multi-core
system is a natural extension from a single-core system.
A conventional clocking scheme would require adding
chip-level PLLs. PLLs are bulky and particularly vulner-
able to noise and hence usually placed at the very edge
of a chip. In future multi-core systems, it represents a
significant challenge to place PLLs and route high-speed
clock signal to the destination cores. In contrast, in ILC,
a single medium-speed global clock signal is distributed
throughout the chip and locally, each core can multiply
the frequency according to its need.

Third, even in a single-core architecture, different mac-
roblocs can run at different frequencies. This is referred
to as the multiple clock domain (MCD) approach [18,
31]. Using ILC, we can locally multiply (or divide) the
frequency of the single global clock. One significant ad-

4 Case Study and Experimental Setup

4.1 Case Study

In this paper, we quantitatively demonstrate some ben-
efits of ILO in a most straightforward setting, a single-
core processor running at a single clock frequency. As
high energy consumption and the resulting heat dissipa-
tion issue become a dominant engineering challenge in
high-end microprocessors, we focus on the energy ben-
fit of using ILC in this case study. Our experiments
compare processors that only differ in the global clock
distribution, some using conventional clocking and some
using ILC. Due to the limited availability of detailed
characterization of clocking network in the literature, our
choice of the clocking network in ILC is limited and very
closely resembles that of the baseline processor. Note
that this is far from the optimal ILC design for the given
processor, but is sufficient to demonstrate the significant
benefit of ILC nonetheless.

Our baseline processor is Alpha21264, which has the
most details in public domain on its clock distribution
network [2, 3]. In this processor, an on-chip PLL drives
an X-tree, which in turn drives a two-level clocking grid
containing a global clock grid and several major clock
grids. The major clock grids cover about 50% of the
chip area and drive local clock chains in those portions.
The remaining part of the chip is directly clocked by the
global clock grid. The densities of the two levels of grids
are different. This configuration is illustrated in Figure 7-
a.

In the first configuration using ILC, we only replace the
very top level of the clock network. We remove most
of the buffers in the X-tree and replace the final level of
buffers (a total of 4) with ILOs. The rest of the hierarchy
remains unchanged (Figure 7-b). Note that in contrast to
the Alpha implementation, we send low-swing signals on
the global X-tree. Clearly, we can reduce the energy con-
sumption of the top level clock network. Furthermore, as
discussed before, clock jitter and skew will also reduce.
We convert this timing advantage into energy reduction
by slightly reducing the supply voltage to capitalize on the
timing slack.
While such a simple approach of using ILC as a drop-in replacement already reduces energy consumption, it is hardly exercising the power of ILC. As discussed before, numerous ILOs can be distributed around the chip to clock logic macro-blocks. Thanks to the built-in deskew capability, we can avoid using power-hungry clock grids altogether. However, to faithfully model and compare different approaches, we need parameters (e.g., capacitance load of individual logic macroblocks) for circuit-level simulation which we could not find in the literature. As a compromise, in the second configuration, we remove the global clock grid and use a set of ILOs to directly feed one single level of grids. The reason this is not done in the Alpha is that to meet the skew and jitter target, a single-level grid would need higher density, as well as stronger and more numerous drivers that together consume far more power than the two-level approach [2]. In an ILC, however, the skew and jitter performance is much improved. With this configuration, the clock network load can be derived based on results reported in [2, 3] and technology files. Since the chip areas not covered by major clock grids is directly clocked from the global clock grid, this single level of grids consist of all the major clock grids and the portion of global grid that directly feeds logic circuit (Figure 7-c).

To evaluate the benefits of injection-locked clocking, we perform both circuit- and architecture-level simulations on the baseline processors with each clock distribution configuration in Figure 7. In order to reflect the state of the art, we scale the global clock speed from 600MHz to 3GHz, and correspondingly the process technology from 0.35µm to 0.13µm. The validity of scaling is verified using Pentium 4 Northwood 3.0GHz processor as the reference.

4.2 Circuit Simulation Setup

At the circuit level, we use a commercial circuit simulator, Advanced Design Systems (ADS), to evaluate power consumption and jitter performance of the clock distribution network with different configurations. The simulations are based on extracted models of the clock distribution networks, including the buffer size, interconnect capacitance, and local clock load capacitance. Then the distribution network model is applied in the circuit simulation with ILOs and clock buffers constructed using SPICE models of transistors. The circuit model for the baseline chip (Figure 7-a) is shown in Figure 8-a. Capacitance values for global and major grids are calculated based on the reported chip dimension, grid structure and grid density. They are then scaled to 0.13µm technology. Clock load is calculated based on its reported power consumption, and also scaled to 0.13µm technology because it represents the logic transistors. All buffer sizes are derived from the reported power consumption. The model for the IGM configuration (Figure 7-b) is different from Figure 8-a only in the first stage, where the buffered X-tree is replaced by a passive X-tree driving four ILOs. Similarly the circuit model for IM’ configuration (Figure 7-c) is modified from IGM, by removing the global grid stage, and adjusting the capacitance of the remaining stages accordingly.

Since jitter is largely introduced by power supply and substrate noise through clock buffers, a noise voltage.

Figure 7. Illustration of the three different configurations of global clock distribution. Each configuration is designated according to its clocking network: XGM, IGM, and IM’. (a) XGM (b) IGM (c) IM’

Figure 8. Circuit-level simulation setup. (a) Clock distribution network modeling. (b) Jitter simulation.
source with a Gaussian distribution is inserted to the power supply node, as shown in Figure 8-b. Transient simulation is used to calculate the voltage and current waveforms along the clock distribution. Output clock waveform is analyzed statistically to get the distribution of the clock period. Jitter at the output is then calculated based on this distribution. Jitter is first measured in the baseline conventional clocking configuration, and the noise source amplitude is determined by matching measured jitter with reported value [21], 35ps. The same noise voltage source is then used in the subsequent jitter simulation for the ILC configurations, and the results are compared to the baseline configuration. We believe this approach is actually pessimistic considering the target jitter number (35ps) is among the lowest in conventional clocking reported [25]. The source jitter from on-chip PLL is represented using a built-in ADS model of clock with jitter, and the clock jitter is chosen to be 5ps, which is consistent with jitter of on-chip PLLs published.

### 4.3 Architectural Simulation Setup

For architectural simulations, we use a modified version of SimpleScalar [6] toolset simulating the Alpha ISA, modeling a chip with one core. The simulator models register files, ROB, issue queues, and load-store queues separately. Both dynamic and leakage power were investigated in detail. We use Wattch [4] for the dynamic energy component, and model the conventional clock tree in detail following the configuration of [2]. There is one important change to the simulator. When we use Wattch’s built-in scaling to scale to the target technology point, the global clock power (24W) is much higher than reported from our circuit-based power analysis and scaling (9.17W). At the time of this writing, we are unable to pin-point the reason for the apparently different scaling assumptions. To stay on the conservative side, we replaced Wattch’s clocking model with our circuit simulation-based results. This makes global clock distribution account for only 23% of overall power, which is significantly lower than reported results [2,17]. Therefore, the benefits of ILC reported in this paper are likely to be very conservative.

Leakage power is temperature-dependent and computed based on predictive SPICE circuit simulations for 0.13μm technology using BSIM3 [5]. We base device parameters, such as Vth, on the 2001 International Technology Roadmap for Semiconductors and IBM 0.13μm CMOS technology file. Temperature (for leakage calculations) is modeled with HotSpot [32] using the floorplan of our modeled single-core processor, adapted from the floorplan of Alpha 21364.

Table 1 lists all the parameters for processor (based on Alpha 21264) along with process specifications. The quantitative analysis use highly-optimized Alpha binaries of all 26 applications from the SPEC CPU2000 benchmark suite. 100 million instructions are simulated after fast-forwarding one billion instructions.

### 5 Experimental Analysis

#### 5.1 Jitter and Skew

In the circuit simulation, the PLL source jitter is set to 5ps, and the value of the added power supply noise source is chosen so that the output clock jitter for the baseline processor (Figure 7-a) is 35ps, as described in section 4. Apparently, there is 30ps jitter added along the clock distribution, which comes from the power supply noise coupled through the buffers. For the clock speed of 3GHz, the overall jitter in the baseline processor therefore corresponds to 10.5% of the clock cycle. In the case of ILC with IGM configuration (Figure 7-b), under the same power supply noise and source jitter, the output clock jitter is lowered to 15ps, i.e., 57% reduction. This translates into recovering 6% of a clock cycle at 3GHz, a significant performance improvement. As described in Section 3.5, the jitter reduction can be attributed to the reduced number of clock buffers and good noise rejection of ILOs. When ILOs are used to directly drive the local clock grids without the global grid as in IM’ configuration (Figure 7-c), thanks to the further reduction in the buffer stages, jitter is lowered to 12ps, or 66% lower than the baseline. Therefore, it clearly demonstrates that ILC can achieve better jitter performance than conventional clocking.

In the current study, it is assumed that built-in deskew capability of ILOs can reduce the skew to below 15ps, or 10ps savings in timing margin compared to the baseline processor (without any deskew). This estimate is con-
consistent with the results using existing deskew schemes [25], and hence quite reasonable. In fact, we believe ILC should lead to even lower skew as discussed in Section 3.4, which can be supported by a test chip measurement shown below.

**Test chip** A test chip was designed and fabricated to verify the jitter reduction and deskew capability of ILC [40]. As shown in the schematic of the test chip (Figure 9-a), a 3-section H-tree mimics the global clock distribution network in real microprocessors, and the leaves of the H-tree are four divide-by-2 ILOs, which divide the input 10GHz clock signal into 5GHz local clocks. The differential outputs of ILOs then drive four open-drain differential amplifiers, which are directly connected to output RF pads for measurements. The differential divide-by-2 ILO used in the test chip is shown in (Figure 9-b). NMOS transistors biased in the inversion region are used as varactors to tune the ILO center frequency, which in turn changes the phase of the local clocks for deskewing purposes.

The deskew capability is measured in the test chip by tuning the control voltage $V_c$ of two ILOs. The result is shown in Figure 10. The whole deskew curve shows that a wide skew range of up to 80ps can be compensated by the built-in deskew capability. Because of the continuous deskew characteristics of ILC, the deskew resolution of ILC depends on the skew measurement and control circuit. Under current skew measurement and control circuit the skew can be controlled at 7ps or less [33]. Thus, the assumption of 15ps skew in an ILC system is valid.

### 5.2 Chip-Wide Power Impact of ILC

**Baseline processor** The power consumption of the baseline processor ranges from 30.4W to 50.4W with an average of 40.7W. The power can be divided into three categories: global clock distribution power, leakage, and the dynamic power of the rest of the circuit. The breakdown of the power is visually shown in Figure 11. The global clock is unconditional and consumes 9.17W or about 23%.

![Figure 9. Schematic of (a) the test chip and (b) a divide-by-2 ILO used.](image)

![Figure 10. Deskew capability of ILC in the test chip, where $V_{dif} = V_{t1} - V_{t2}$.](image)

**ILC configurations** Now we analyze the power savings of ILC. For IGM (Figure 7-b), power savings come from two factors. First, the power consumed in the top-level X-tree is reduced from 1.72 to 1.56 Watt because the reduction of the total levels of buffers used and the lowered voltage swing on the X-tree. Second, as explained above, jitter and skew all improved when using ILC: a 20ps reduction in jitter and 10ps in skew are achieved in this configuration. These total savings of 30ps increases the available cycle time for logic from 273ps to 303ps. This, in turn, allows a reduction in Vdd without affecting the clock speed. We use the following voltage-delay equation from [30] to calculate the new Vdd, which is 1.415V.

$$
t = \frac{C}{k'(W/L)(V_{dd} - V_d)} \left[ \frac{2V_t}{V_{dd} - V_i} + \ln \left( \frac{3V_{dd} - 4V_t}{V_{dd}} \right) \right]
$$

The power reduction for the tested applications ranges from 3W to 5.2W with an average of 4.1W or 10.1%. The reduction is mainly due to the lowering of supply voltage. Considering the minimal change and the conservativeness in the entire evaluation process, the result is very encouraging.

Though using ILC as a drop-in replacement of the top-level clock distribution tree already shows notable power savings, as explained before, it is not fully exploiting the potential of the new clocking paradigm. The second ILC configuration, IM' (Figure 7-c), further reduces clock distribution power by reducing the size of the grid. For IM', the global clock power is reduced to 5.9W (from 9.17W in the baseline XGM) and the combined jitter and skew reduction is 33ps, which allows us to scale Vdd to 1.41v. The overall effect is an average of 6.8W (17%) total power reduction. Compared to IGM, IM' further reduces power by 2.7W or 7%.

The results of using different clocking structures are summarized in Figure 11. In this comparison, all configurations achieve the same cycle time. The density of the grids and the driving capabilities are determined using...
circuit simulation. We choose the design point where energy is minimized. For reference, we also show the result of replacing the two levels of grids by a single grid in the conventional configuration. Note that this grid is different from the M' grid as it needs higher density and larger buffers to achieve the same overall cycle time target. We designate this grid G', and the configuration XG'. We use the same methodology to compute its jitter performance, clocking load, and power consumption.

From the results, it is clear that ILC significantly improves power consumption. It is also clear that using a single-level grid per se is not the source of energy savings for IM': using a single grid in the conventional design leads to a significant 7.9W of extra power consumption.

Overall, we see that ILC can be introduced to a processor in various levels of ease. With minimum design intrusion, when only the very top level of the clock tree is modified to use injection locking, energy reduction is already significant (10%), thanks to the lowered jitter and skew. When we further optimize the clocking grid, the power savings become more pronounced (17%). All these are achieved without affecting performance or the design methodology of the processor.

6 Related Work

There have been intensive research efforts in recent years to address the challenges in high-speed clocking from different disciplines, including clockless design (asynchronous circuits), optical interconnect, and resonant clocking, to name a few. Each of these alternative solutions has its own technological issues to be addressed.

Optical interconnect potentially offers smaller delays and lower power consumption than electrical ones, and is promising for the global clock distribution network [13, 19, 25]. However, there are still great challenges in its silicon implementation, particularly for on-chip electrical-optical modulators [7]. Wireless clock distribution proposed in [23] [14] suffers substantial overhead in chip area and power consumption due to on-chip clock transceivers.

Among the proposed electrical solutions, a family of synchronized clocking techniques, such as distributed PLLs [15, 27], synchronous distributed oscillators [24, 35], rotary clocking [36], coupled standing-wave oscillators [26], and resonant clocking [9] have recently been proposed to improve the performance of global clock distribution. In [8, 9], on-chip inductors are added to all the local nodes of the global clock distribution tree, and hence turn it into a single large resonator. As we discussed in Section 3.1, resonance improves power efficiency. Therefore, this technique reduces dc power dissipation and lowers jitter in the global clock distribution network. It is a good step in the right direction. However, it does not provide skew capabilities like injection-locked clocking. The more stringent layout constraints due to on-chip inductors could even aggravate the problem of skew.

In [15, 27], an array of PLLs is constructed using a voltage-controlled oscillator (VCO) and loop filter at each node, and a phase detector between adjacent nodes. Each PLL generates the local clock in the particular clock domain, which is synchronized with others through the aforementioned phase detectors at the clock domain boundaries. Global clock as in conventional clocking is removed in this scheme, and hence it promises lower jitter. The drawbacks are that a) the global skew is still a problem since deskewing only happens locally, and b) the sensitive analog circuits in a PLL (phase detectors, loop filters, ring oscillators) are vulnerable to noise in the hostile environment of digital circuits.

In [24, 26, 35, 36], an array of oscillators are connected to the global clock distribution network, and thus are synchronized by coupling. The resulting oscillator array becomes a distributed oscillator. The difference is that in [36] the oscillator array is a one-dimensional loop, and the phase of oscillators change linearly along the array, similarly to a distributed VCO [38], which was based on traveling-wave amplification [12]. In [26], the oscillator array generate a standing-wave pattern on the network, i.e., each oscillator has the same phase. Essentially all these techniques use a distributed oscillator with interconnects as its resonator. A distributed oscillator suffers the problem of phase uncertainty due to mode locking [15, 27, 35]. This is evident in that similar topologies can be used for either traveling-wave [36] or standing-wave oscillation [26]. Another problem is that jitter tends to be worse than conventional clocking since the global clock is now generated on chip using lossy passive components, without the clean reference clock from the off-chip crystal oscillator. It is noteworthy that [8] unintentionally adds injection locking to distributed oscillator clocking and demonstrated good jitter performance.

Overall, all these promising technologies face signifi-
cant technical difficulties and require dramatic changes in processes technologies, design methodologies, or testing methods, and hence will face significant resistance in adoption. In comparison, injection-locked clocking is highly desirable because it is fully compatible with existing IC infrastructures as well as current design and testing methodologies.

7 Conclusions

Thanks to the high sensitivity, good noise rejection, and built-in deskewing capability of injection-locked oscillators, the proposed injection-locked clocking can significantly improve skew and jitter performance of a multi-GHz clock distribution network. Reduced number of clock buffers, and recovered timing margin from skew and jitter lead to substantial power savings for the whole processor. Initial results from circuit and architectural simulations confirmed our analysis. More detailed modeling and characterization is under way, particularly in skew simulation. A chip prototype has also been recently demonstrated [40]. We expect the benefits of this new clocking scheme will be even greater when it is applied to high-performance multi-core microprocessors and other high performance system-on-a-chip (SoC) systems.

References

[37] H. Wu and A. Hajimir. A 19 GHz, 0.5 mW, 0.35 μm CMOS frequency divider with shunt-peaking locking-range enhancement. In IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pages 412–413, 2002.