

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

Lin Zhang, Aaron Carpenter, Berkehan Ciftcioglu, Alok Garg, Michael Huang, and Hui Wu

Department of Electrical & Computer Engineering

University of Rochester

Rochester, NY 14627, USA

{linzhang, carpente, ciftciog, garg, huang, hwu}@ece.rochester.edu

Abstract—Recently we proposed a new clocking scheme, injection-locked clocking (ILC), to combat deteriorating clock skew and jitter, and hence reduce power consumption in high-performance microprocessors. In the new clocking scheme, injection-locked oscillators are used as local 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 alternatives, 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 improvements using ILC. Alpha21264 is used as the baseline processor, and is scaled to $0.13\mu\text{m}$ and 3GHz. Two ILC configurations are constructed: one of them replaces the top-level global clock X/H-tree in Alpha21264; another is optimized for ILC operation. Simulations show 20ps and 23ps jitter reduction, 10.1% and 17% power savings, respectively. A test chip is implemented in a standard $0.18\mu\text{m}$ CMOS technology. It has four divide-by-2 ILOs at the ends of a 3-section H-tree, generating 5GHz local clocks. Measured jitter of generated clocks is lower than that of the input signal. Two local clocks can be differentially deskewed up to 80ps relative to each other.

I. 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 a high quality clock signal to every corner of the chip necessarily requires driving the clock distribution network “hard”, usually in full swing of the 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 [24], 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

This paper is the full-length technical report version of the journal paper submitted to IEEE Tran. VLSI with the same title.

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 II analyzes the skew and jitter problem in conventional clocking; Section III discusses proposed clocking scheme based on injection-locking; Section IV presents a simulation-based case study, including the experimental setup and the quantitative analysis; Section V discusses the test chip; Section VI discusses related work; and finally, Section VII concludes.

II. CHALLENGES IN CONVENTIONAL CLOCKING

Fig. 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 Fig. 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], [20], [28], [33] 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.

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

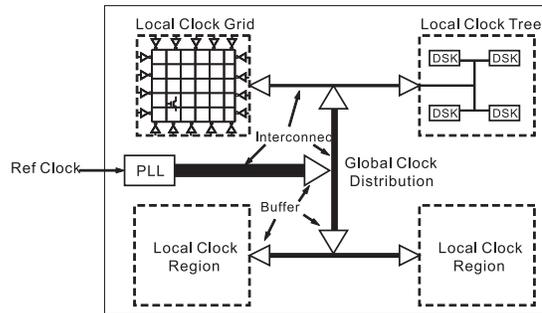


Fig. 1: Conventional global clock distribution, showing an H-tree topology with interconnects and clock buffers [10].

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 need 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 jitter and skew directly translates into power savings for a given performance target.

III. INJECTION-LOCKED CLOCKING

A. Injection-Locked Oscillators

Injection locking [1], [21] is a special type of forced oscillation in nonlinear dynamic systems (also known as synchronization). Suppose a signal of frequency ω_i is injected into an oscillator (Fig. 2-a), which has a self-oscillation (free-running) frequency ω_0 . When ω_i is quite different from ω_0 , “beats” of the two frequencies are observed. As ω_i approaches ω_0 , the beat frequency ($|\omega_i - \omega_0|$) decreases. When ω_i enters some neighborhood very close to ω_0 , the beats suddenly disappear, and the oscillator starts to oscillate at ω_i . The frequency range in which injection locking happens is called the *locking range* (Fig. 2-b). Injection locking also happens when ω_i is close to the harmonic or subharmonic of ω_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.

An injection-locked oscillator (ILO) can be considered as a simple first-order PLL (Fig. 3-a), in which nonlinearity of the oscillator core functions as a phase detector. For example, in a typical divide-by-2 ILO (Fig. 3-b) [27], the oscillator core (consisting of M_1 , M_2 and M_{tail}) also serves as a single-balanced mixer for phase detection. Because of the simple structure, ILOs consume much less power than a full-fledged PLL, and can operate at frequencies as high as tens of gigahertz [36]. The fact that the built-in “phase detectors” are

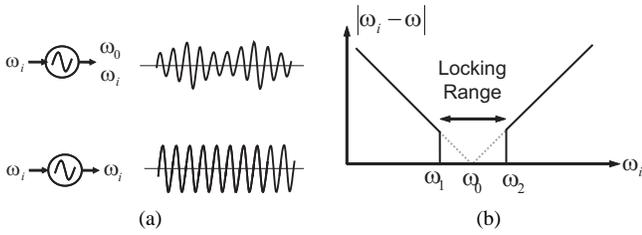


Fig. 2: (a) Beat and injection locking phenomenon when an oscillator is driven by a single-frequency input signal. (b) locking range.

mixer-based explains why ILOs can operate at the harmonic and subharmonic frequencies of the input signal.

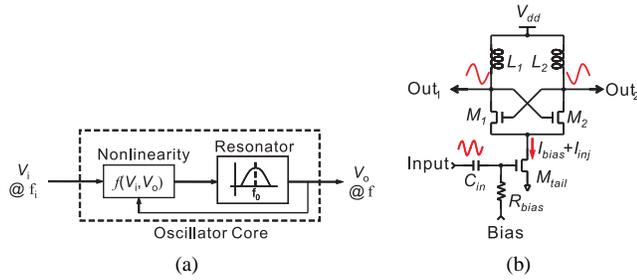


Fig. 3: (a) A generic model of an injection-locked oscillator (ILO). (b) a divide-by-2 ILO based on a common differential LC oscillator. The input signal is injected into the oscillator core through the tail transistor M_{tail} . This topology exhibits good injection locking efficiency because of the built-in single-balanced mixer structure.

Once locked to the input signal, the output of ILOs will maintain a determined phase relative to the input signal (Fig. 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 Fig. 4, the phase shift φ 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 [40]. This phase tuning characteristics can be utilized to achieve deskew between different clock domains with no need for other deskew circuits.

B. Clocking Using ILOs

Recently, we proposed a new clocking scheme as shown in Fig. 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. 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,

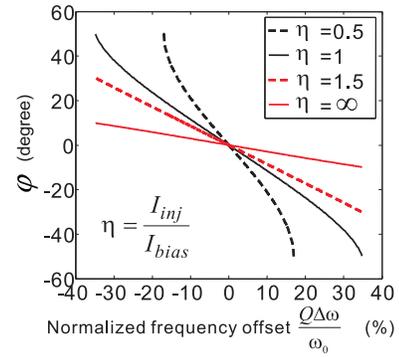


Fig. 4: Phase tuning characteristics for a divide-by-2 ILO in Fig. 3-b. $\eta \equiv I_{inj}/I_{bias}$ is the injection ratio, ω_0 is the free-running oscillation frequency, $\Delta\omega \equiv \omega - \omega_0$ is the frequency shift, and Q is the LC tank quality factor.

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 (see Section VI).

In addition to acting as clock receivers, ILOs can be constructed as frequency multipliers [19] or dividers [27], [38], and hence this scheme enables local clock domains to have higher ($n.f_0$) 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.

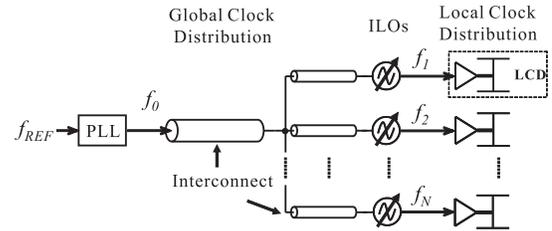


Fig. 5: Injection-locked clocking (ILC).

C. Power Savings

Injection-locked clocking (ILC) can lead to significant power savings in high-performance microprocessors. The benefits come from several sources.

First, using injection locking, it is straightforward to use a low-speed global clock combined with high-speed local clocks. This 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 [36]. The benefit will become more evident as future processors incorporate more and more cores.

Second, ILOs have higher sensitivity than clock buffers (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 (Fig. 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 an inverter needs to change its state twice in every clock cycle. As a result of this difference, the signal amplitude of the global clock can be much smaller in the injection-locked clocking scheme, which results in 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 a dominant factor as the process technology scales further.

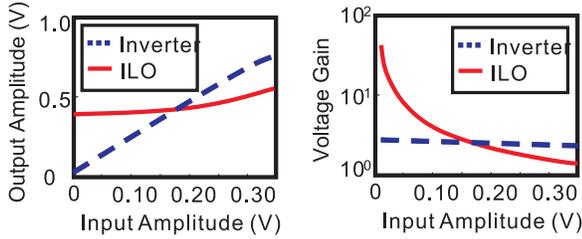


Fig. 6: Voltage gain of an inverter and an injection-locked oscillator at different input signal levels.

Further, the number of clock buffers in the global clock distribution can be reduced. In conventional clocking, in order to minimize jitter generated by clock 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 III-E). Therefore, the global clock signal amplitude no longer needs to be full 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 (V_{dd}), and save power dissipation from not only clock distribution network, but all the logic gates on the chip. In Section IV below, we demonstrate the power savings from all aspects using a quantitative case study.

D. 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 III-A, 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], [26], 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). Because 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).

E. 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 of the design freedom in layout, clock interconnect can be placed where there is minimal noise coupling from adjacent circuits and interconnects. In addition, similar to a PLL, an ILO can suppress both its internal noise through high-pass filtering and input noise through low-pass filtering, and hence can possibly lower jitter at its output. Using a differential structure, an ILO can be less insensitive to the common-mode power supply and substrate noise than an inverter by design. Therefore, injection-locked clocking is likely to achieve better jitter performance than conventional clocking.

F. Potential Applications

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

First, ILC reduces jitter and skew compared to a conventional clocking network. This reduces cycle time and therefore allows a faster clock speed. As technology 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 vulnerable 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 can be distributed throughout the chip and locally each core can multiply the frequency according to its need.

Third, even in a single-core architecture, different macroblocks can run at different frequencies. This is referred to as the multiple clock domain (MCD) approach [17], [30]. Using ILC, we can locally multiply (or divide) the frequency of the single global clock. One significant advantage of using ILC to enable multiple clock domains is that the local clocks have a well-defined relationship as they are all synchronized to the global clock. As a result, cross-domain communication can still be handled by synchronous logic without relying on asynchronous circuits. Note that although ILOs are not as flexible as PLLs in frequency multiplication, they are sufficient for MCD processors as only a few frequency gears are needed for practical use [41].

IV. CASE STUDY

In this paper, we quantitatively demonstrate some benefits 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 dissipation issue become a dominant engineering challenge in high-end microprocessors, we focus on the energy benefit 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 Fig. 7-a. The three planes X, G, and M represent the three layers of clock distribution networks the X-tree, the global clock grid, and the local clock grid respectively.

In the first configuration using ILC (Fig. 7-b), we only replace the very top level of the clock network. We remove all 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. Note that in contrast to the Alpha implementation, we send low-swing signals on the global X-tree. This reduces the energy consumption 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 not fully exploiting the benefits 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 ILC configuration, we remove the global clock grid and use a set of ILOs to directly feed one single level of grids. Note that this is not done in the Alpha, because 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]. Using ILC, however, the skew and jitter performance is much improved.

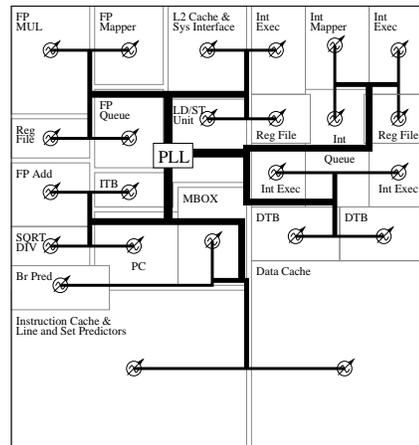


Fig. 8: A possible layout floorplan for the ILC based global clock distribution in Alpha 21264.

With this configuration, the load of the clock network 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 (Fig. 7-c). Finally, thanks to the deskew capability of ILOs, there is no need to use a balanced global clock tree. In Figure 8, we show an example clock tree design. In this example, each macroblock in the floorplan is driven by an ILO which is at the leaf of the global clock tree.

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 Fig. 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 0.35 μ m to 0.13 μ m. The validity of scaling is verified using Pentium 4 Northwood 3.0GHz processor as the reference.

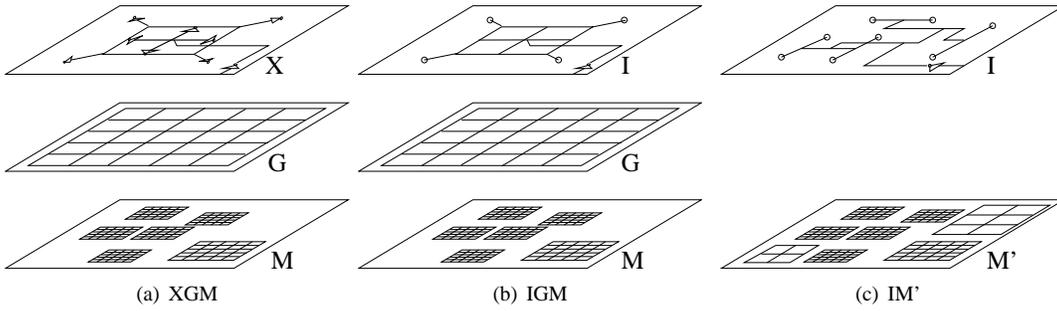


Fig. 7: Illustration of the three different configurations of global clock distribution. Each configuration is designated according to its clocking network: XGM, IGM, and IM'.

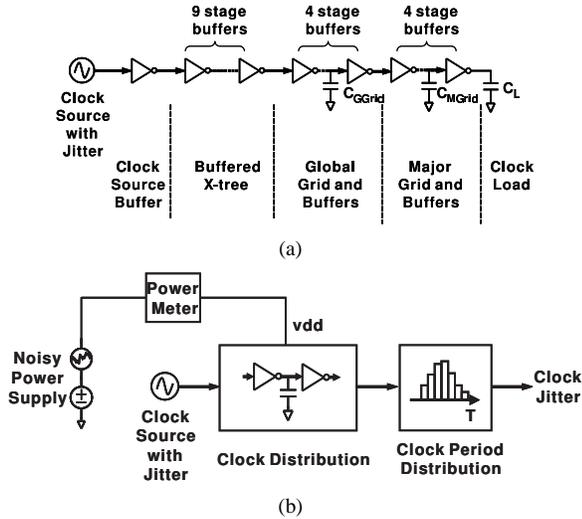


Fig. 9: Circuit-level simulation setup. (a) Clock distribution network modeling. (b) Jitter simulation.

A. 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 (Fig. 7-a) is shown in Fig. 9-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 $0.13\mu\text{m}$ technology. Clock load is calculated based on its reported power consumption, and also scaled to $0.13\mu\text{m}$ technology because it represents the logic transistors. All buffer sizes are derived from their reported power consumption. The model for the IGM configuration (Fig. 7-b) is different from Fig. 9-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 (Fig. 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 source with a Gaussian distribution is inserted to the power supply node, as shown in Fig. 9-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 in [20], 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 [24]. 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.

B. 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]. Leakage power is temperature-dependent and computed based on predictive SPICE circuit simulations for $0.13\mu\text{m}$ technology using BSIM3 [5]. We base device parameters, such as V_{th} , on the 2001 International Technology Roadmap for Semiconductors and IBM $0.13\mu\text{m}$ CMOS technology file. Temperature (for leakage calculations) is modeled with HotSpot [31] using the floorplan of our modeled single-core processor, adapted from the floorplan of Alpha 21364.

Table I 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

Processor core	
Issue/Decode/Commit width	4 / 4 / 4
Issue queue size	16 INT, 16 FP
Functional units	INT 4+1 mul/div, FP 1+2 mul/div
Branch predictor	Bimodal and Gshare combined
- Gshare	1K entries, 10 bit history
- Bimodal/Meta table/BTB	4K/4K/1K (2 way) entries
Branch misprediction penalty	7+ cycles
ROB/Register(INT,FP)	64/(80,72)
LSQ	32, 2 search ports, 1-cycle port occupancy, 2-cycle latency
Memory hierarchy	
L1 instruction cache	64KB, 2-way, 32B line, 2 cycles
L1 data cache	64KB, 2-way, 32B line, 2 cycles
L2 unified cache	2MB, 4-way, 32B line 15 cycles
Memory access latency	250 cycles
Process specifications	
Feature Size / Frequency	0.13 μ m / 3GHz
V_{dd} / V_t	1.5V / 0.43V

TABLE I: System configuration.

applications from the SPEC CPU2000 benchmark suite. 100 million instructions are simulated after fast-forwarding one billion instructions.

C. Analysis of 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 (Fig. 7-a) is 35ps [20]. 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 (Fig. 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 III-E, 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 (Fig. 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 consistent with the results using existing deskew schemes [24], and hence quite reasonable. In fact, we believe ILC should lead to even lower skew as discussed in Section III-D, which can be supported by a test chip measurement shown below.

D. Chip-Wide Power Impact of ILC

a) 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 Fig. 10. The global clock is unconditional and consumes 9.17W or about 23%.

b) ILC configurations: Now we analyze the power savings of ILC. For IGM (Fig. 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 [29] to calculate the new Vdd, which is 1.415V.

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

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' (Fig. 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%.

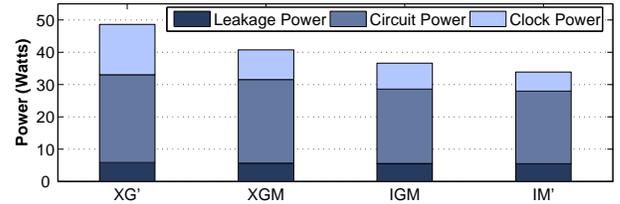


Fig. 10: Breakdown of processor power consumption with different clock distribution methods.

The results of using different clocking structures are summarized in Fig. 10. 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.

V. TEST CHIP

A. Test Chip Implementation

A test chip was designed and fabricated to verify the jitter reduction and deskew capability of ILC [39]. As shown in the schematic of the test chip (Fig. 11-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 root of the H-tree is directly connected to a ground-signal-ground (GSG) pad to facilitate testing (Fig. 12). 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 [27] used in the test chip is shown in (Fig. 11-b). This is essentially a differential LC oscillator, with the input signal injected into the gate of the tail transistor. We chose this ILO circuit for the test chip because of its well-understood operation and good performance. The inductors are implemented as on-chip spiral inductors, which are made on metal 5 with a quality factor about 4 at 5GHz. Such low Q is not a problem for ILO operation and actually helps increase the locking range. If better metal is available, the power efficiency can be further improved. 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.

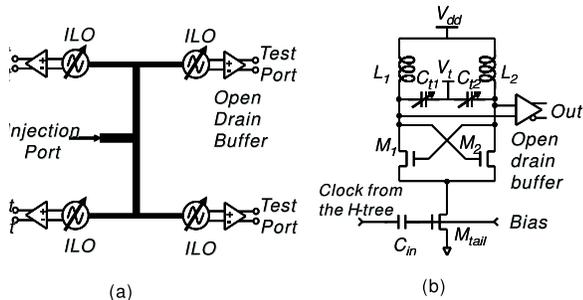


Fig. 11: Schematic of (a) the test chip and (b) a divide-by-2 ILO used.

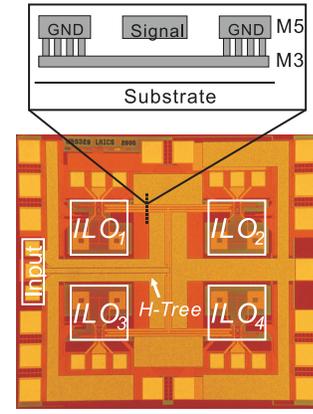


Fig. 12: Chip micrograph of the test chip. The whole chip size is $1.5\text{mm} \times 1.3\text{mm}$, and each ILO occupies $0.25\text{mm} \times 0.22\text{mm}$. The H-tree sections measure $500\mu\text{m}$, $280\mu\text{m}$, and $290\mu\text{m}$, respectively, from root to leaves.

The H-tree is constructed using coplanar-waveguide (CPW) transmission lines. Bottom shield is used to reduce substrate coupling in a real microprocessor environment. This limits the maximum characteristic impedance of the transmission line to be just over 40Ω in this technology. So the transmission lines from the H-tree leaves to the root are designed to be 40Ω , 20Ω and 10Ω , respectively, in order to achieve impedance matching at all junctions. Width of signal and ground lines, spacing between them, and choice of metal layers are also optimized for minimizing the clock propagation loss. The tree dimension is about $800\mu\text{m} \times 600\mu\text{m}$.

B. Measurement Results

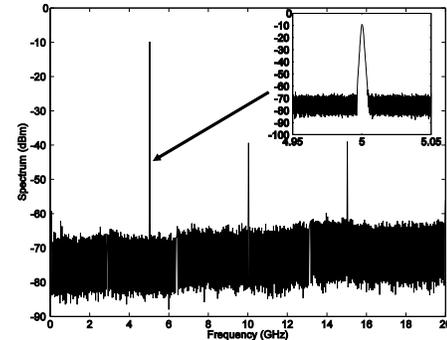


Fig. 13: Spectrum of the generated local clock signal from ILO1, identical to that from other ILOs on-chip.

The test chip is measured using an RF probe station. The input is a sinusoidal signal from a continuous-wave (CW) signal generator. The power supply voltage is 1.4V. The spectra of the local clock signals generated by the four ILOs are almost identical, and one of them is shown in Fig. 13.

The locking range of ILOs on the test chip is found to be identical, and that of ILO1 is shown in Fig. 14. The injection signal amplitude is calculated from the measured incident power and reflection coefficient (S11) at the root of the H-tree.

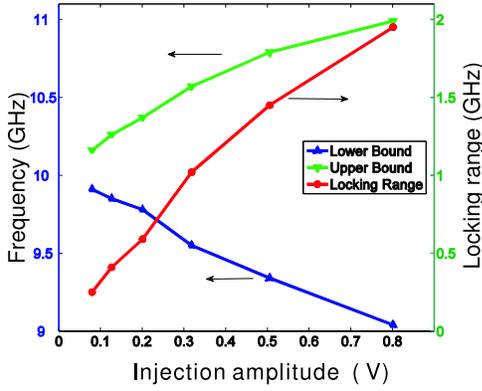


Fig. 14: Locking range of ILO1, identical to other ILOs.

It can be seen that when the input signal has rail-to-rail swing (1.4V), the locking range is about 17%, which is sufficient for both accommodating process/temperature variation and deskew tuning (see below).

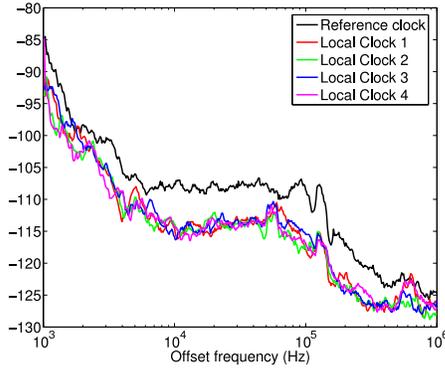


Fig. 15: Phase noise of reference clock and 4 output clocks at different positions on chip.

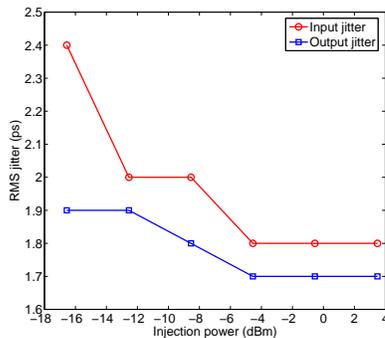


Fig. 16: Jitter characteristics.

Phase noise of both the input and output clock signals are shown in Fig. 15. The 6dB reduction (up to about 500kHz offset) because of the divide-by-2 operation is evident, which shows that the internal ILO noise is suppressed by injection locking. Fig. 16 shows the long-term RMS jitter of both the input and output signals measured using a self-referenced jitter measurement method with a sampling oscilloscope [16]. The output timing jitter is even less than that of the input

signal. Considering the frequency division, this result clearly demonstrates that ILOs can serve as a PLL and clean up the clock signal.

The deskew capability is measured in the test chip by tuning the control voltage V_t of two ILOs. The result is shown in Fig. 17. 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 [32]. Thus, the assumption of 15ps skew in an ILC system is valid.

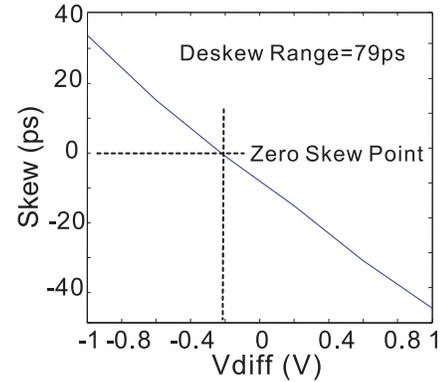


Fig. 17: Deskew capability of ILC in the test chip when tuning ILO1 and ILO2 differentially ($V_{diff}=V_{t1}-V_{t2}$). The skew is measured between the two output clock signals of ILO1 and ILO2. Note that there is some imbalance between ILO1 and ILO2 caused by mismatch in the clock distribution tree and measurement system.

The test chip consumes a total power of 52.8mW, where 45.3mW comes from the 1.8V-supplied open-drain buffers. The ILOs core circuitry working under 1.4V Vdd only consumes 7.3mW when biased low and injection signal is 6dBm. The bias circuitry consumes 0.2mW.

VI. 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], [18], [24]. However, there are still great challenges in its silicon implementation, particularly for on-chip electrical-optical modulators [7]. Wireless clock distribution proposed in [22] [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],

[26], synchronous distributed oscillators [23], [34], rotary clocking [35], coupled standing-wave oscillators [25], 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 III-A, 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 deskew capabilities like injection-locked clocking. The more stringent layout constraints due to on-chip inductors could even aggravate the problem of skew.

In [15], [26], 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 [23], [25], [34], [35], 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 [35] the oscillator array is a one-dimensional loop, and the phase of oscillators change linearly along the array, similarly to a distributed VCO [37], which was based on traveling-wave amplification [12]. In [25], 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], [26], [34]. This is evident in that similar topologies can be used for either traveling-wave [35] or standing-wave oscillation [25]. 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 significant 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.

VII. 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 [39]. 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.

ACKNOWLEDGEMENT

The authors wish to thank P. Holloway, B. Chatterjee, J. Yu, A. Shah, and V. Abellera of National Semiconductor for their support in chip fabrication. The work is also supported in part by Department of Energy, and National Science Foundation, through grants 0509270 and 0719790.

REFERENCES

- [1] R. Adler. A Study of Locking Phenomena in Oscillators. *Proc. IRE*, 34:351–357, June 1946.
- [2] D. W. Bailey and B. J. Benschneider. Clocking Design and Analysis for a 600-MHz Alpha Microprocessor. *IEEE Journal of Solid-State Circuits*, 33(11):1627–1633, Nov. 1998.
- [3] W. J. Bowhill et al. Circuit Implementation of a 300-MHz 64-bit Second-generation CMOS Alpha CPU. *Digital Technology Journal*, 7(1):100–118, 1995.
- [4] D. Brooks, V. Tiwari, and M. Martonosi. Wattch: A Framework for Architectural-Level Power Analysis and Optimizations. In *International Symposium on Computer Architecture*, pages 83–94, June 2000.
- [5] BSIM Design Group. *BSIM3v3.2.2 MOSFET Model - User's Manual*, Apr. 1999.
- [6] D. Burger and T. Austin. The SimpleScalar Tool Set, Version 2.0. Technical report 1342, Computer Sciences Department, University of Wisconsin-Madison, June 1997.
- [7] K. Cadien et al. Challenges for On-Chip Optical Interconnects. *Proc. SPIE*, 5730:133–143, Nov. 2005.
- [8] S. Chan, K. Shepard, and P. Restle. 1.1 to 1.6GHz Distributed Differential Oscillator Global Clock Network. In *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pages 518–519, 2005.
- [9] S. Chan, K. Shepard, and P. Restle. Uniform-Phase Uniform Amplitude Resonant-Load Global Clock Distributions. *IEEE J. Solid-State Circuits*, 40(1):102–109, March 2005.
- [10] E. Friedman. Clock Distribution Networks in Synchronous Digital Integrated Circuits. *Proc. IEEE*, 89(5):665–692, May 2001.
- [11] G. Geannopoulos and X. Dai. An adaptive Digital Deskewing Circuit for Clock Distribution Networks. In *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pages 400–401, 1998.
- [12] E. Ginzton, W. Hewlett, J. Jasberg, and J. Noe. Distributed amplification. *Proc. IRE*, 36:956–969, Aug. 1948.
- [13] J. Goodman, F. Leonberger, et al. Optical Interconnections for VLSI Systems. *Proc. IEEE*, 72:850–866, July 1984.
- [14] X. Guo, D. Yang, R. Li, and K. K.O. A Receiver with Start-up Initialization and Programmable Delays for Wireless Clock Distribution. In *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pages 386–387, 2006.
- [15] V. Gutnik and A. Chandrakasan. Active GHz Clock Network Using Distributed PLLs. *IEEE J. Solid-State Circuits*, 35(11):1553–1560, Nov. 2000.

- [16] A. Hajimiri, S. Limotyrakis, and T. Lee. Jitter and Phase Noise of Ring Oscillators. *IEEE J. Solid-State Circuits*, 34(6):896–909, June 1999.
- [17] A. Iyer and D. Marculescu. Power-Performance Evaluation of Globally Asynchronous, Locally Synchronous Processors. In *International Symposium on Computer Architecture*, pages 158–168, May 2002.
- [18] E. Kaimiley, P. Marchand, et al. Performance Comparison between Optoelectronic and VLSI Multistage Interconnect Networks. *J. Lightwave Technol.*, 9:1674–1692, 1991.
- [19] K. Kamogawa, T. Tokumitsu, and M. Aikawa. Injection-Locked Oscillator Chain: A Possible Solution to Millimeter-Wave MMIC Synthesizers. *IEEE Trans. Microwave Theory Tech.*, 45(9):1578–1584, Sept. 1997.
- [20] N. Kurd, J. Barkatullah, R. Dizon, T. Fletcher, and P. Madland. A Multigigahertz Clocking Scheme for the Pentium 4 Microprocessor. *IEEE J. Solid-State Circuits*, 36(11):1647–1653, Nov. 2001.
- [21] K. Kurokawa. Injection Locking of Microwave Solid-State Oscillators. *Proc. IEEE*, 61(10):1386–1410, Oct. 1973.
- [22] R. Li, X. Guo, D. Yang, and K. K.O. Initialization of a Wireless Clock Distribution System Using an External Antenna. In *IEEE Custom Integrated Circuits Conf. Dig. Tech. Papers*, pages 105–108, 2005.
- [23] H. Mizuno and K. Ishibashi. A Noise-Immune GHz-Clock Distribution Scheme using Synchronous Distributed Oscillators. In *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pages 404–405, 1998.
- [24] A. Mule, E. Glytsis, T. Gaylord, and J. Meindl. Electrical and Optical Clock Distribution Networks For Gigascale Microprocessors. *IEEE Transactions on Very Large Scale Integration Systems*, 10(5):582–594, Oct. 2002.
- [25] F. O’Mahony, C. Yue, M. Horowitz, and S. Wong. A 10-GHz Global Clock Distribution Using Coupled Standing-Wave Oscillators. *IEEE J. Solid-State Circuits*, 38(11):1813–1820, Nov. 2003.
- [26] G. Pratt and J. Nguyen. Distributed Synchronous Clocking. *IEEE Trans. Parallel Distributed Systems*, 6(3):314–328, March 1995.
- [27] H. Rategh and T. Lee. Superharmonic injection-locked frequency dividers. *IEEE J. Solid-State Circuits*, 34(6):813–821, June 1999.
- [28] P. Restle et al. A Clock Distribution Network for Microprocessors. *IEEE J. Solid-State Circuits*, 36(5):792–799, May 2001.
- [29] A. S. Sedra and K. C. Smith. *Microelectronic Circuits*. Oxford University Press, 2004.
- [30] G. Semeraro et al. Dynamic Frequency and Voltage Control for a Multiple Clock Domain Microarchitecture. In *International Symposium on Microarchitecture*, pages 356–367, Nov. 2002.
- [31] K. Skadron, M. Stan, M. Barcella, A. Dwarka, W. Huang, Y. Li, Y. Ma, A. Naidu, D. Parikh, P. Re, G. Rose, K. Sankaranarayanan, R. Suryanarayan, S. Velusamy, H. Zhang, and Y. Zhang. HotSpot: Techniques for Modeling Thermal Effects at the Processor-Architecture Level. In *International Workshop on THERMal Investigations of ICs and Systems*, Oct. 2002.
- [32] S. Tam, R. Limaye, and U. Desai. Clock Generation and Distribution for the 130-nm Itanium 2 Processor With 6-MB On-Die L3 Cache. *IEEE J. Solid-State Circuits*, 39(4):636–642, April 2004.
- [33] S. Tam, S. Rusu, U. Desai, R. Kim, J. Zhang, and I. Young. Clock Generation and Distribution for the First IA-64 Microprocessor. *IEEE J. Solid-State Circuits*, 35(11):1545–1552, Nov. 2000.
- [34] H.-A. Tanaka, A. Hasegawa, H. Mizuno, and T. Endo. Synchronizability of Distributed Clock Oscillators. *IEEE Trans. Circuits Syst. I*, 49(9):1271–1278, Sep. 2002.
- [35] J. Wood, C. Edwards, and S. Lipa. Rotary Traveling-Wave Oscillator Arrays: a New Clock Technology. *IEEE J. Solid-State Circuits*, 36(11):1654–1665, Nov. 2001.
- [36] H. Wu and A. Hajimiri. 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 412–3, 2001.
- [37] H. Wu and A. Hajimiri. Silicon-Based Distributed Voltage Controlled Oscillators. *IEEE J. Solid-State Circuits*, 36(3):493–502, Mar. 2001.
- [38] H. Wu and L. Zhang. A 16-to-18GHz 0.18 μm Epi-CMOS Divide-by-3 Injection-Locked Frequency Divider. In *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pages 602–3, 2006.
- [39] L. Zhang, B. Ciftcioglu, M. Huang, and H. Wu. Injection-Locked Clocking: A New GHz Clock Distribution Scheme. *IEEE Custom Integrated Circuits Conf. Dig. Tech. Papers*, pp.785-788, 2006.
- [40] L. Zhang and H. Wu. A Double-Balanced Injection-Locked Frequency Divider for Tunable Dual-Phase Signal Generation. *IEEE Radio-Frequency Integrated Circuits (RFIC) Symposium Digest of Papers*, pp.137-140, 2006.
- [41] Y. Zhu, D. Albonesi, and A. Buyuktosunoglu. A High Performance, Energy Efficient, GALS Processor Microarchitecture with Reduced Implementation Complexity. In *International Symposium on Performance Analysis of Systems and Software*, pages 42–53, Mar. 2005.