DesignCon 2004

Practical Multi-Gigahertz Clocks for ASIC and COT Designs

Haris Basit, OEA International, Inc.
basit@oea.com

John Wood, MultiGig Inc.
John.wood@multigig.com

Ken Pedrotti, University of California Santa Cruz
pedrotti@ee.ucsc.edu

Abstract

This TecForum begins by providing background on existing methods employed by full-custom designers for multi-gigahertz clocks. In addition, it explores in some detail the inductive nature of GHz clock networks. A novel clocking methodology termed "rotary clocking" will be proposed that provides: near zero skew; greatly reduced noise generation; insensitivity to process, voltage and temperature variations; and very significant power savings. The presenters provide evidence that this method would allow ASIC and COT designers to utilize far more of the available process performance, narrowing the gap to full-custom performance. The TecForum will conclude with a look at additional tools and techniques that need to be developed to fully automate this approach.

Author Biographies

Haris Basit worked on the design of surface emitting ultra bright LEDs using advanced eptiaxial growth techniques while a graduate student at the University of Illinois, Urbana. Later, at IBM he worked on Hetrojunction Bipolar Transistor (HBT) technology in a joint development project with Rockwell. Following this he joined Rockwell to work on HBT and MESFET circuits and processes. During his tenure at Rockwell, Mr. Basit also managed a DARPA program for the design of advanced EDA software focused on High Speed Design. Subsequent work at Bell Labs involved introduction of Formal Model Checking software. Mr. Basit is currently VP of Business Development at OEA International where he manages development of EDA software targeted for RF and analog design.

John Wood is the Engineering Director of MultiGig, Ltd., a U.K. technology startup specializing in multi-gigahertz circuit design I.P. Previously, he has worked as a consultant design engineer on multi-domain design projects in mechanical, power electronics, infrared optics, and software development roles. He holds a number of patents which have been licensed for manufacture in the fields of infrared plastic welding and high-speed digital signaling. His technical interests include all areas of engineering design, but particularly electromagnetics, VLSI circuit design, and high-speed analog techniques.

Kenneth D. Pedrotti is Associate Professor of Electrical Engineering at University of California, Santa Cruz.   His technical interests are in the areas of high-speed electronics for lightwave systems, optical and optoelectronic components for all-optical networks and solid-state visible and infrared imaging. Prior to joining the faculty in Sept. 2000 he was with the Rockwell Science Center in Thousand Oaks, CA and with what is now Conexant Systems in Newbury Park, CA serving in a variety of positions in both research and management. There he worked on the development of high-bandwidth WDM components and switching systems, preformed early investigations into all-optical networks for which he was recognized as a co-recipient of a 1996 R&D 100 award, and developed monolithic optoelectronic integrated receivers with world record performance. Additionally he has worked on AlGaAs/GaAs Heterojunction Bipolar Transistor (HBT) circuits for optical transmitters, receivers, switches, clock recovery and data regeneration, as well as MOCVD crystal growth of quantum structures in III-V compound semiconductors for optoelectronic devices, including lasers, detectors and modulators. From 1995-1999 he served on the Board of Governors of the IEEE Solid State Circuit Society.

Introduction

Clock frequencies exceeding 1-GHz have been available in full-custom integrated circuits such as processors for several years. Intel introduced a 1-GHz Pentium® III processor in March 2000 using a 0.18 μ process and six layers of metal. A similar process allowed the Pentium® 4 architecture to attain 2-GHz one year later. Pentium® 4 chips built on a 0.13 μ process currently operate above 3-GHz. In stark contrast clock frequencies for standard cell based ICs designed on 0.13 μ processes have great difficulty exceeding 700 MHz. Even using more advanced nanometer processes current results for standard cell based ICs top out at a little above 1-GHz [1]. While much has been published about this performance gap [2, 3, 4] there has been little progress. This TecForum proposes a practical approach that will significantly narrow this gap.

The goal of higher clock frequencies is so alluring that many novel approaches are under investigation. These include: optical distribution of global clock signals [5]; routing global clocks on package level wiring [6 Chapter 9]; wireless clock distribution [7]; standing-wave clock distribution [8, 9]; asynchronous designs [10]; and digital deskewing circuits [6 Chapter 5]. Techniques such as digital deskewing circuits are currently in use on first generation IA-64 processors. However, most of these techniques will remain impractical for the vast majority of standard cell based designs where resource and time constraints are far more limiting.

Furthermore, even a brief analysis of existing approaches used for obtaining multi-gigahertz designs in full-custom processors shows that many of them are impractical for standard cell designs due to purely economic reasons. Nearly forty Watt power consumption by the clocked components can be justified on an Itanium®; however, a standard cell based graphics chip with similar power dissipation would not be commercially viable. Economic forces dictate design teams for standard cell based ICs that are a small fraction of the size used for processor design. Furthermore, the time allocated for standard cell design is generally far less than full-custom design. The luxury of additional time and manpower allows full-custom design teams to hand craft registers, size individual transistors, use dynamic logic, carefully design pipelines and to rely extensively on time borrowing (cycle stealing, useful skew, slack) techniques. Finally, performance binning of parts, a time honored technique for processors, is often not useful for standard cell designs that must operate at a specific frequency.

For purposes of this paper we assume, that standard cell designs above 1 GHz will be deemed practical if the following criteria are met: 1) Power dissipation is kept far below current processor levels. 2) The size of the design teams remains relatively small. 3) The design cycle must be reasonably short and predictable. 4) No changes would be necessary to existing fabrication technologies.

We show that to meet the first criteria of keeping power dissipation low a method must be devised to circumvent the switching power (CV^2F). This method must also be amenable to a high degree of automation in order to permit small design teams to implement multi-gigahertz designs in a reasonable time period. Furthermore, to truly lay claim to the adjective 'practical' the automated tools must ensure timing closure in the harsh parasitic environment of multi-gigahertz frequencies with few iterations and little or no manual intervention. The techniques put forth in this TecForum address all of these criteria.

Timing Uncertainties and Clock Distribution

The key to obtaining multi-gigahertz clocks is to reduce timing uncertainty. Timing uncertainty is the sum of spatial variation (skew) and temporal variation (jitter). Contributors to skew include: buffer

sizing, fanout and parasitic differences between clock paths; process variations; supply voltage variations; and temperature variations. Contributors to jitter include: noise on the power or ground supplies; capacitive or inductive coupling; and jitter from the PLL itself. Even paths with nominally zero skew can have significant timing uncertainty when on-die process, voltage and temperature (PVT) variations are taken into account. With existing methods these PVT effects can only be reduced by increasing power levels, by reducing latency or some combination of the two.

Clock distribution networks on existing standard cell designs are typically built using balanced RC trees. These trees can take the form of binary trees, H trees, X trees or arbitrary matched tree networks. At a given value of process, voltage and temperature (and in the absence of inductive effects) these trees can be automatically balanced to have nominally zero skew. However, they often include a dozen or more levels of buffering which introduces a large PVT sensitive latency. Furthermore, current clock synthesis tools frequently balance the various clock paths by trading buffer delay in one path against RC delay in another path leading to a precarious balancing of the paths that is very sensitive to PVT variations. A clock spine driven by a very large buffer is sometimes used to replace the first few levels of clock buffering, thus reducing the latency and its attendant uncertainty at the expense of additional power dissipation. Clock spines, being wider interconnects, often require an analysis that includes inductive effects.

While generally not used on standard cell designs, clock grids [11] have been used on a variety of processors since their original use on the DEC Alpha processor. At the expense of greatly increased power dissipation, clock grids reduce latency and skew and also result in less PVT dependence. PVT variations while inherently large in active circuits are inherently small for interconnects of sufficient width and spacing. As can be seen in Figure 1, accurate analysis of clock grids requires extensive parasitic analysis including inductive effects of the clocks and all return paths.
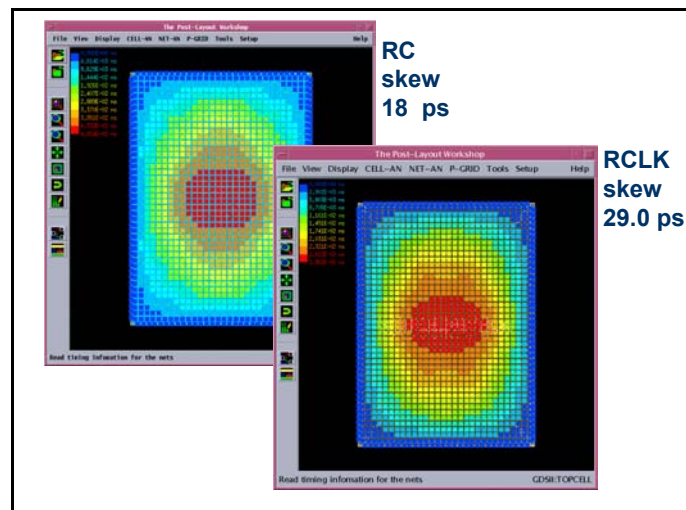

Figure 1: Clock grid with buffers placed around the periphery and driven at 2 GHz. Clock loads are uniformly distributed throughout the area. Blue represents the minimum delay and red represents maximum delay. An RC analysis shows 18 ps of skew while a properly modeled RCLK analysis shows 29 ps skew.

Due largely to PVT variations timing uncertainties have scaled poorly with successive generations of processes. Any proposed solution for practical multi-gigahertz designs must minimize PVT variations since this is a fundamental cause of timing uncertainties. Trees with numerous levels of buffering will have great difficulty avoiding PVT variations while spine or grid alternatives result in greatly increased power dissipation.

Common Misconceptions

Before exploring the patented rotary clock approach it is instructive to address common misconceptions regarding standard cell designs at gigahertz speeds. It is often assumed that something mysterious happens when the time for the signal to propagate across the chip equals or exceeds one clock period. In fact, in existing designs, even 300 MHz clock signals take many cycles to reach their destination flip-flops. The registers might be clocking off of a signal that had been launched from the PLL four or more cycles earlier.

Chip-size is not necessarily a constraint on data path frequency either. Many engineers work under the false constraint that data signals must be able to span the chip within one clock period and if not, the clock frequency must be reduced to enforce this constraint. However, pipelining of wires allows for full-speed throughput no matter how large the chip might be, at the expense of increased latency. The Pentium® 4 has many places where the wires are pipelined. Even large off-chip latencies can be tackled by using multi-threaded processor architectures that are able to absorb latency by task switching to a new or previously stalled task when requested data for the current task may have a long pipeline delay.

Another erroneous assumption is that dynamic (domino) logic is the secret to fast full-custom designs. Domino logic is certainly faster, probably twice as fast as static logic [12], but why should this matter if interconnect delays are said to dominate? This apparent contradiction is resolved by realizing that interconnect delay does not dominate in a full-custom hand crafted layout of optimized transistor geometries. Often the combined advantages of dynamic logic, transistor sizing and hand crafted layout are lumped under "dynamic logic" when only one-third to one-half of the benefit comes from the dynamic nature of the logic. Hand crafted full-custom designs are faster mostly because hand-placements minimize wire lengths and interconnect parasitics, transistor sizing is optimal, clock skew is controlled far more accurately allowing very fine-grained pipelining designed by hand to take advantage of time borrowing where latches or multiphase clocks are used [13].

The fact that timing uncertainties, including both skew and jitter, are almost always below 15% of the clock period, lead some to assume that they can have no more than a 15% impact on performance[2]. This reasoning is somewhat circular since the clock frequency is purposely chosen so that uncertainties are small. Performance has been shown to increase monotonically with pipeline depth provided that the latency of the pipeline is not systematically exposed [14]. The minimum clock period of a pipeline is set by the sum of the timing uncertainty, pipeline overhead and allowed time for useful work. As a minimum we may set the allowed time for useful work at about seventeen FO4 delays. (An FO4 delay is the delay of an inverter driving four identical inverters.) To this we add a pipeline overhead of three FO4 delays giving a total of twenty FO4 delays. On a standard 0.18 µ process with an FO4 delay of 25ps this yields a minimum clock period of 500 ps (2.0 GHz) if timing uncertainties could be eliminated. Reducing timing uncertainties will lead to greatly increased clock rates and deeper pipelines. In contrast, performance can never be high where clock uncertainty is large without resorting to multiphase custom circuit techniques such as skew-tolerant design [13].

Inductive Modeling of Interconnects

Most existing standard cell design flows model interconnects as distributed resistive and capacitive (RC) networks – completely ignoring inductance effects. Much has been written about including inductance effects in high frequency designs [15, 16, 17, 18, 19, 20]. Ismail and Friedman [18] propose utilizing

inductance to sharpen rise and fall times thus decreasing crowbar current. Repeater insertion methodologies in the presence of inductance have also received significant attention [21].

Often the phrases "transmission line effects" and "inductance effects" are used interchangeably. In the context of this TechForum it is important to note that transmission line effects refer to the case where inductive effects combine with capacitance to make an LCR line that propagates a signal. For this to happen, resistance must be low [16], rise time must be fast and the physical layout must permit a fairly well controlled impedance along the entire length of the line.

True transmission lines are not generally part of a digital designer's experience leading to a perception that 'inductance does not matter'. Generally, digital place and route tools force this perception to be true. Auto-routers understand only the RC model of interconnect and construct wires accordingly. Inductance effects are often mitigated by repeater-insertions that break long lines into shorter pieces where the inductive effects are not seen. The same cannot be said of long wires such as clock or power distribution wires which must be analyzed for inductance to get working gigahertz silicon.

The onset of inductive effects can be modeled in a number of ways. One very simple but useful method is to compare the resistance of a wire to its inductive impedance. If the inductive impedance is a significant percentage of the resistance one may assume that a full RCLK model is needed. To simplify the resistance calculation we will ignore frequency dependent resistance (skin and proximity effects) thus calculating only DC resistance. To simplify the inductance calculation we will assume only the partial inductance of the line thus ignoring the return path. The reader is advised that these simplifications are justified only for the purpose of a rough initial comparison. A proper calculation of inductance must include return paths. To calculate the inductive impedance we will also need to determine a 'significant' frequency. The significant frequency is not directly related to the clock frequency but to the fastest rise or fall time [22]. However, if we assume that the fastest rise and fall times are about 10% of the clock period we can calculate that the significant frequency is approximately three times the fundamental clock frequency. Since our goal is to have clock frequencies in excess of 1 GHz the minimum significant frequency is 3 GHz.

Inductive impedance as a percent of DC resistance is given in Figures 2 and 3. Clock frequencies from 1 to 3 GHz are shown (significant frequency range from 3 to 9 GHz). Figure 2 is for a copper interconnect on a typical interconnect routing layer assumed to be 0.35 μ thick. Figure 3 is for a copper interconnect on an upper metal layer and is assumed to be 0.90 μ thick. Inductance effects are least noticeable on thin narrow lines at low frequency. Conversely, thicker and wider lines at higher frequency can behave much more like inductors than resistors. However, even for the narrowest lines on thin lower level metal layers inductive impedance represents several percent of the DC resistance at gigahertz frequencies.
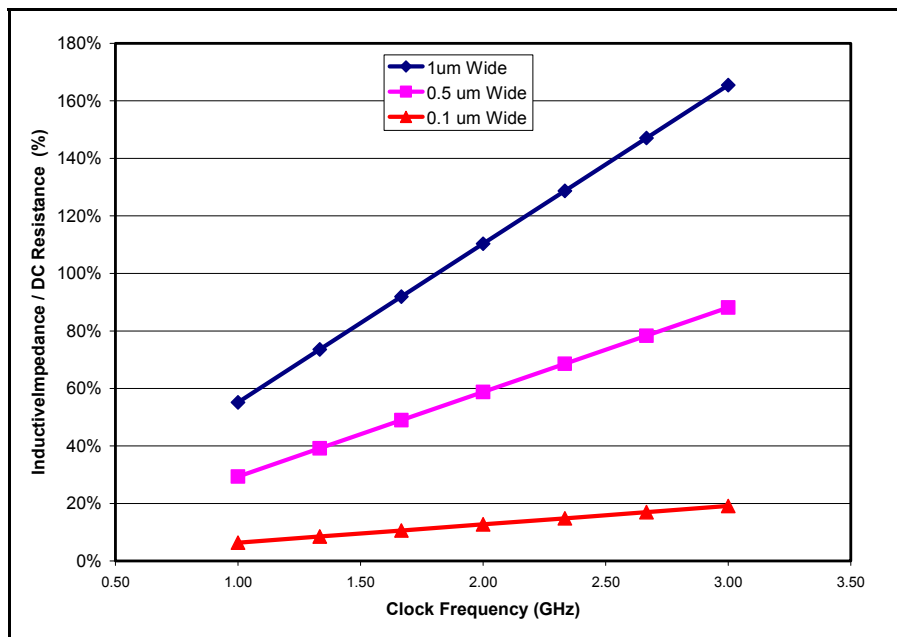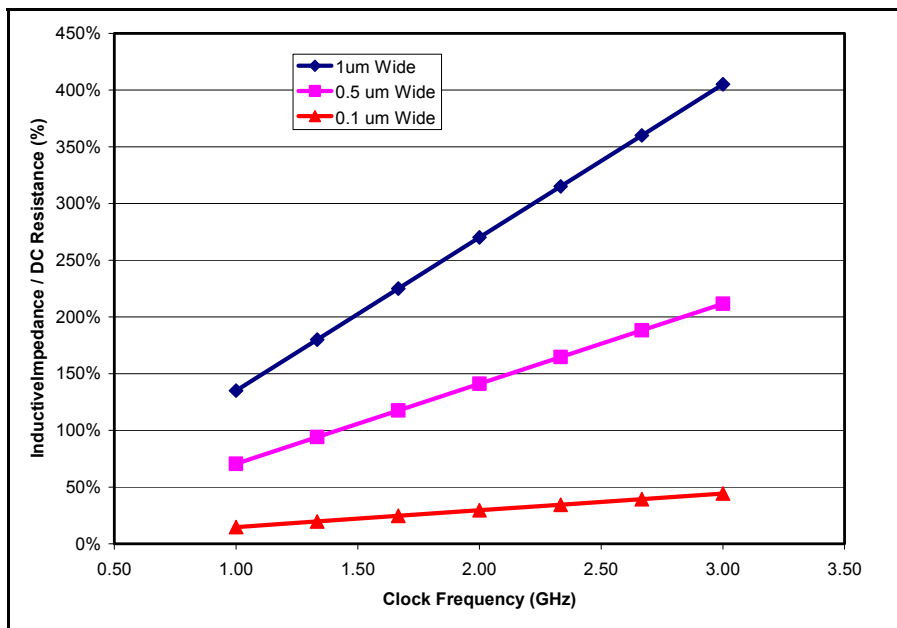
Figure 2: Inductive impedance as a percent of DC resistance for a 500 µ long 0.35 µ thick interconnect at three different line widths. Inductive impedance (= jωL) is calculated using a significant frequency that is three times the clock frequency.

Figure 3: Inductive impedance as a percent of DC resistance for a 500 µ long 0.90 µ thick interconnect at three different line widths. Inductive impedance (= jωL) is calculated using a significant frequency that is three times the clock frequency. Note that for a 0.5 or 1.0 µ wide line inductive impedance is larger than DC resistance at most frequencies of interest. Note that thick top level metals usually do not permit line widths as narrow as 0.1 µ.

As shown by Figure 2 and Figure 3, inductive effects play a significant role in modeling of interconnects for gigahertz clock distribution networks. This is especially true if non-minimum width wires or thick upper level metals are used. As shown in Figure 3, on 0.90 µ thick copper interconnects inductive impedance often exceeds DC resistance.   Modeling gigahertz clock lines without including inductive effects can lead to serious errors. Furthermore, coupling between the clock and signal nets can lead to significant jitter. While capacitive coupling can be effectively shielded by routing power and ground around the clock, inductive coupling cannot be shielded against so easily. Also, one is faced with

significant inductive coupling between nets that are physically distant from each other – a problem not encountered in capacitive coupling. For example, two parallel wires that are 500 µ long, 0.5 µ wide and spaced 100 µ apart have an inductive coupling coefficient that is still 25% of the value of those same wires placed 0.5 µ apart (K=0.2 versus K=0.8). Thus, lines that are hundreds of wiring tracks apart could have significant inductive coupling.

One final interconnect complication is that inductance and capacitance can work together to store a substantial amount of energy. This is especially troublesome at high speeds where undesirable inductive behavior such as ringing and overshoot can affect the timing by changing the transition points.

The radically increased complexity of analyzing clock nets with the inclusion of inductive effects could present a severe obstacle to our stated desire of automating the design flow. While a full RCLK network of the clock and nearby nets can be extracted using existing commercial tools [23], existing delay calculators do not support inductance thus requiring delay simulation be done in spice. Running spice on large RCLK networks is often prohibitively slow.

Some of the added complexity can be avoided by building the clock as a controlled impedance transmission line. If the transmission line is also differential then inductive coupling is reduced to a local problem. A differential clock carries equal and opposite currents on each wire, thus to distant signal lines there appears to be no current in the clock. This zero-effective-current effect appears in wires that are at a greater distance from the clock than the spacing between the differential clock wires – typically just a few microns. A clocking scheme composed entirely of differential transmission lines would allow one to ignore coupling to all wires more than a few wiring tracks away thus greatly simplifying the modeling of associated interconnects.

Exploiting On-Chip Inductance

We perhaps labored the point about the inductive nature of high-frequency interconnect. The reason is that, inevitably, clock networks become transmission-lines at high speeds and while Analog and RF designers have been comfortable designing on-chip transmission lines for years, it is still a new concept to the ASIC designer. Given that clock networks are inevitably transmission lines at high frequencies we look at ways to not only accurately model them but to possibly benefit from their inherently different nature.

When modeling the clock distribution network as transmission-lines, there are only three possible methods of managing energy on the transmission lines:
1. Absorb it with terminators at the end of the transmission-line. This is what is done in a clock tree.
2. Reflect the energy to form standing waves [5, 8, 9]
3. Rotate it – the new Rotary Clock concept presented in this TecForum.

Method 1 is rugged and reliable but essentially wastes all the clock energy as heat. Method 2 results in standing wave patterns of variable, low-amplitude sine waves. These waves are wholly unsuitable for direct use in digital circuits. Method 3, rotation, is a patented concept which supports almost ideal rail-to-rail square wave clocking of single or multiphase circuits.

Now we describe the new Rotary Clocking Methodology and the beta-stage CAD tools and flow that implement practical GHz clocking for standard cell based designs.

Description of Basic Rotary Clock Ring

To understand the rotary clock it is useful to begin with a thought experiment. Assume that a pulse is placed on one end of a differential transmission line. This pulse then travels down the transmission line at a velocity defined by 1/sqrt(LC). The inductance in this equation can be calculated purely by reference to the geometry of the differential transmission lines. The capacitance also would normally be assumed to be composed of the parasitic capacitances of the lines. However, lumped intentional capacitors could also be used provided that they are uniformly distributed along the differential line with spacing that is small in comparison to the wavelength. Thus, we may artificially control the velocity of propagation by increasing or decreasing the capacitive load – independent of the line geometry.

Once created, the energy of this traveling pulse must go somewhere, it can either be absorbed or reflected if the line is terminated, or if the line is very lossy dissipated by the line itself. An alternative is to form the differential transmission line into a loop that includes a Möbius like twist. In this last case the energy of the pulse continues to travel around the line incurring only losses due to the resistance of the wire and negligible dielectric losses. Each rotation of a differential pulse will result in a phase inversion of the signal caused by the physical translation of the two wires due to the Möbius effect.
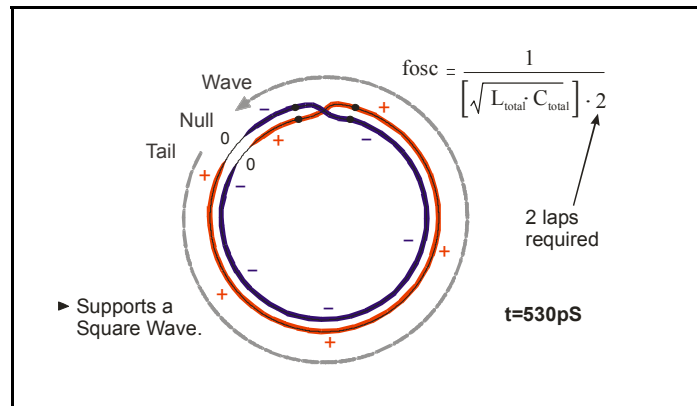

Figure 4: A differential transmission line connected in a loop with a Mobius like twist forms the basic element of the rotary clock.

Thus two rotations form a complete cycle giving a frequency as shown in Figure 4. This is the key to the oscillation process. A crucial result is that the capacitive loads being 'driven' by the transmission line are not dissipating power but part of the transmission line itself. If these capacitive loads are clocked elements we can successfully eliminate CV^2F power from the entire clock distribution up to and including the input capacitance of the latches or flip-flops.

The resistive losses in the transmission line are compensated by placing back to back invertors as shown on the left side of Figure 5. These inverters are all in shunt configuration. An extraction of the basic ring with all inverters would appear to be one very large inverter with its output and input shorted together as shown on the right side of Figure 5. Thus, at DC the entire clock distribution is represented as a single net and two transistors. But at RF frequencies it behaves like a transmission line obeying all the usual transmission-line theories.
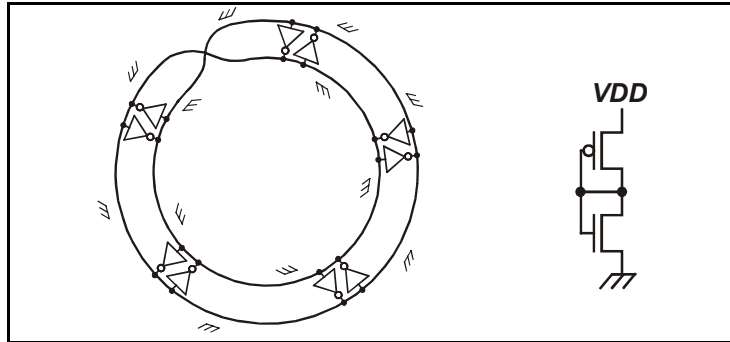
Figure 5: CMOS inverters add energy to overcome losses, initiate start-up and ensure rail-to-rail voltage swings. At DC or low frequencies the ring appears as a large inverter with the input and output shorted together. The inverters are in shunt configuration unlike a ring-oscillator.

Start-up of the rotating wave is spontaneous on power-up with the slightest amount of noise because the self-biased loop is highly unstable until rotation begins. Circuits can be added that enforce a particular preferred rotation direction. Resistive losses are small, and the inverters help to establish a constant clock voltage in the presence of resistive lines. To minimize resistive losses thick upper metal layers should be used for rotary clock lines.

Although we have an oscillator, the Rotary Oscillator is not a resonant device in the usual sense of an LC tank. The circuit produces rail-to-rail square-waves directly at all points of the loop, only the phase varies from place to place.

In brief, the building block of a Rotary Clock circuit is a differential transmission-line loop which is closed on itself and has a Möbius twist in the wire. When activated by transistor circuits this structure becomes a Rotary Oscillator. The closed loop supports rotation of transmission-line signals without significant absorption or reflection thus allowing most of the energy to continue circulating around the loop. These structures are physically small and can be fabricated on-chip as coplanar transmission-lines on the upper metal layers – resembling a clock grid.

Connecting Multiple Clock Rings into a Grid

Having established a basic rotating wave on a single ring, we can connect multiple rings together into a grid like arrangement with the 'corners' hard-wired together. This allows one to cover any arbitrary shaped area of an IC with a regular grid like pattern. Figure 6 shows one possible configuration of four interconnected rings.
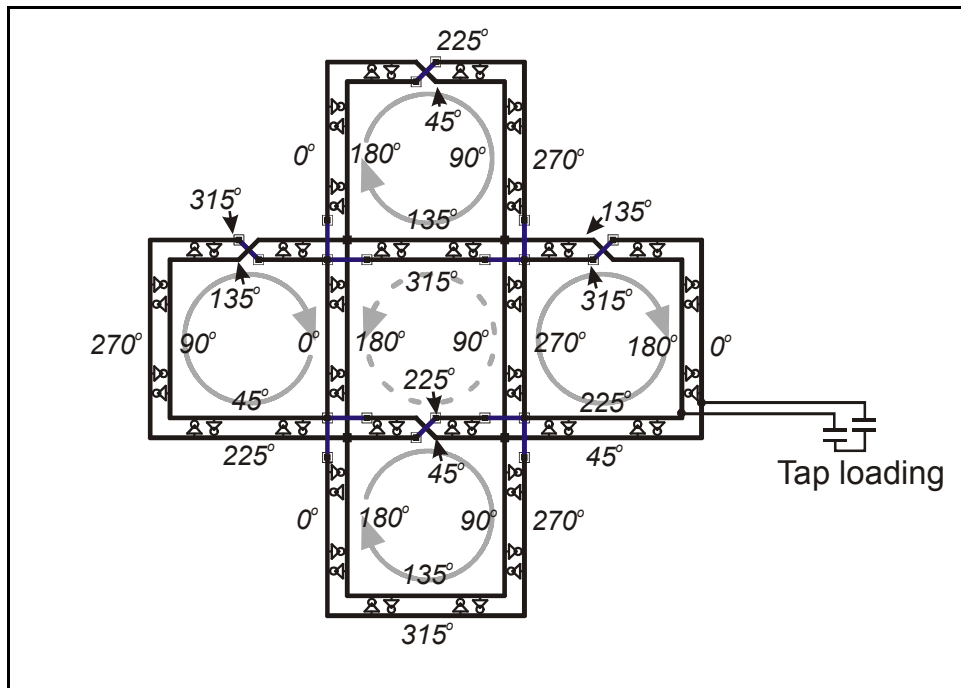
Figure 6: Multi-ring operation can be created by connecting the rings at the 'corners'. The resulting structure resembles a grid and can cover arbitrary shaped regions of the chip. While shown as squares above, rectangular shaped rings with the long dimension built using the thickest metal layer work best. The four explicit outer rings have clockwise rotation while the implicit center ring has a counterclockwise rotation.

A fundamental aspect of the rotary clock approach is that these interconnected rings must have mutual locking of the various rotary clock domains into a define phase relationship. It is this locking that allows accurate timing to be established across the integrated circuit. All oscillators can have their phase and frequency influenced by the injection of small amounts of signal near their natural frequencies of oscillation [24]. This phenomena is known as 'injection locking' and is widely observed in natural [25] and artificial systems [26]. Injection locking can be used to advantage, as in some high powered lasers [27] or phased array antenna systems [28, 29]. In these cases, stability and spectral purity is controlled by the injection of a small signal from a very high quality source. Injection locking can also be problematic and something to be avoided as with VCOs in clock and data recovery systems in which a phase locked loop is desired to lock onto an underlying clock signal but direct injection locking of the VCO itself is undesired. The above examples assume that the injection signal proceeds from one source to an oscillator that is perturbed. If however many independent oscillators are coupled, synchronous behavior of the ensemble can frequently occur as with the synchronization of applause [30], flashing by fireflies or, more importantly to us, the synthesis of coherent, low phase noise, reliable cardiac rhythms by the cross injection locking of many low quality cellular oscillators in cardiac muscle [31, 32].

The rotary clock distribution approach undertaken here includes the development of a thorough understanding of the effect of coupling strength, coupling topology and tunability on aspects relevant to timing control such as frequency tunability, effect of oscillator mismatches, and phase stability of the resulting system. In this TecForum we will discuss the fundamental phenomena of injection locking of oscillators, the synchronization of many oscillators by mutual or cross injection locking and the expected improvements in jitter realized in cross injection locked systems.

Power Dissipation in Clocked Storage Elements

Moving on from clock distribution, we look at the components which use clock signals, the flip-flops and latches. Latches are in general level-sensitive devices and maintain transparency during the entire time that clock is active. Flip-flops are edge sensitive devices and are often formed by clocking two latches back-to-back with non-overlapping phases of a clock. Full-custom designers make extensive use of latches while most standard cell designs are largely based on flip-flops. The Rotary Clock approach supports both methods.

As can be seen on the right side of Figure 7, internal dynamic power dissipation caused by CV^2F within high speed flip-flops accounts for far more power dissipation than the global clock signal distribution itself. Many of the older papers on clock distribution were based on analysis of systems where this was not the case.
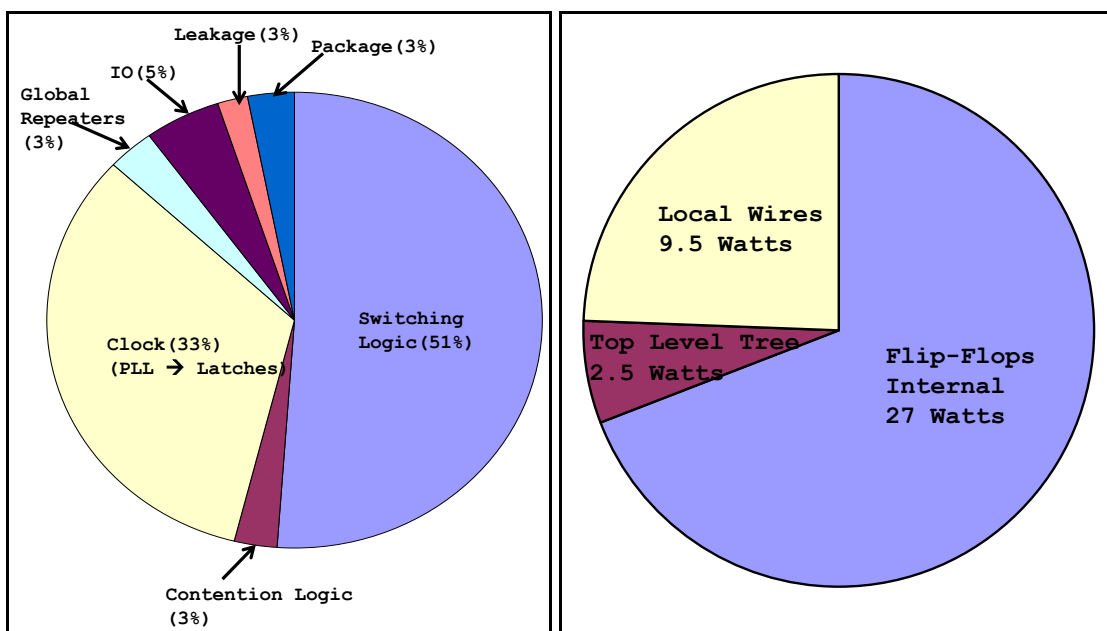

Figure 7: Pie chart on left shows the percentage distribution of power on the Itanium 2 processor. The pie chart on the right shows the further breakdown of 33% of power identified as clock power in terms of watts. [33]

Internal power dissipation of the synchronization elements is therefore the most fruitful line of attack on clock power. Even in ASICS upwards of 10nF of capacitance can be directly attributable to internal clock node capacitance. Doing the math, at 1-GHz and 1 volt this equals 10 Watts using CV^2F. The local wiring capacitance load includes the input pin capacitance of the flip-flops.

Capacitive Loading on Rotary Grids

In the Rotary Clock approach the internal capacitance of the latches or flip-flops becomes an inherent part of the distributed transmission line capacitance. Thus, only a very small amount of power is dissipated due to this internal capacitance. Instead the energy in the capacitors is exchanged with inductive energy of the transmission line as the pulse travels past the flip-flop. A full analysis is outside the scope of this TecForum but it can be shown that adding distributed capacitive loading to a rotary clock network does not negate the energy saving characteristic of Rotary Traveling waves. Although the transmission line energy increases with added capacitive loading, still, most of the power is recirculated

in the loops, and energy exchanges between magnetic and electric fields in the inductors and capacitances.

Latch Design for Rotary Clocks

Analysis and design of latches and flip-flops has become a field in its own right [34]. The most common design tradeoff for a storage element is the dynamic power consumption versus setup and delay time. Other tradeoffs such as clock loading, data path power and cell size give a wide range of performance metrics by which alternative architectures have been evaluated.

Clocked storage elements are essentially differential by nature – they have an "N" and a "P" sampling transistor driven from an internally derived differential (2 phase) clock. This is the case even when the input clock pin to the flip-flop is single-ended. Analog engineers will recognize these circuits as sample-and-hold devices but here they only need to support digital signal levels.

An edge triggered flip-flop can be built using just two latches placed in series – with each latch operating on a different phase. Non-overlapping differential drive clocks are essential to maintain reliable operation of the flip-flop since any overlap of the active times of the two phases will cause a flush-through of data. In order to tackle clock power, a more efficient flip-flop is required that exposes the N and P gates for direct connection to the inherently non-overlapping Rotary Clock.

Over 75% of the dynamic power of a latch or flip-flop can be saved by directly driving these internal latch or flip-flop nodes with the Rotary Clock and letting the inductance of the clock provide the charge/discharge energy to the storage element. Furthermore, the design of optimal storage elements is eased considerably when clock loading and dynamic power are no longer primary issues.

The affect on frequency and phase of the Rotary Clock due to non-uniform capacitive loading must be compensated by the CAD tooling during the process of placement.

Scheduled Skew and Pipelining With Rotary Clocks

A rotary clock makes a full 360 degrees of clock phase available on the chip. Many strategies are known to exploit useful skew [35]. Pipeline stages can be balanced by choosing the correct phase of clock for each part of the pipeline. A tool that understands the highly predictable phase delay of the Rotary Clock can place flip-flops to maximally exploit skew. Exactly such tools are currently in development and will be discussed at the TecForum.

To illustrate the opportunity for useful skew, Fig. 8 shows the post-placement slack available in a synchronous design. Clocking this circuit with a single-phase clock, even if it is zero-skew is sub-optimal. Many of the paths with a lot of slack could pass time to the tight paths in the pipeline. Slacks can also be used in other ways, for example, to reduce power or area on non-critical paths.
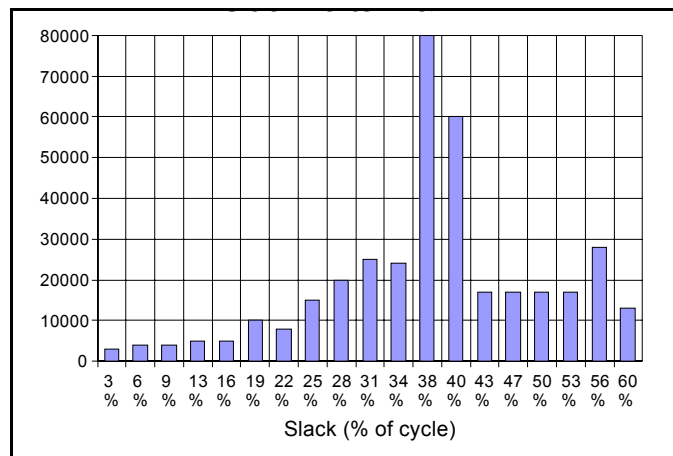
Figure 8: Number of paths at each slack point after placement [39].

As new process technologies come online and clock frequencies increase, the amount of useful work done in one cycle will be reduced, forcing standard cell designers to increase the level of pipelining. This may require modifications to the style of RTL and eventually to improved synthesis tools that automatically implement pipelining [36].

Interconnect dominated designs require the acceptance of latency at design time. Latency tolerant designs will scale using pipelining, latency intolerant designs will not. Clock control is crucial to achieving fine pipelining and the also necessary for exploiting time borrowing techniques.

With current design techniques there is a significant power penalty for deeper pipelines since each additional stage increases the total number of clocked cells. However, a primary benefit of rotary clocking is that it greatly reduces the power consumed by clocked cells thus removing this barrier to increased pipeline depth.

The Asynchronous Alternative

The problems associated with high speed clocks and synchronous system design have led many researchers and startups to investigate use of 'self-timed' logic circuits. Most of the justification for these efforts is based around comparisons to single-phase clock systems which cannot balance pipelines correctly. The Rotary Clock provides a natural and efficient scheduled-skew system that delivers all of the benefits of asynchronous techniques at a much higher speed. True asynchronous systems must operate with a two-way handshake protocol which effective doubles the interconnect delay versus the 'open-loop' synchronous paradigm. Typically an asynchronous gate will signal a "ready" signal to the next gate in the pipeline and wait for an "acknowledge" back before transmitting a new data bit. This is very wasteful of time, area, wire and dynamic switching power.

Testability Issues

The existing clocking methodologies have developed a large base of testing techniques that allow for functional testing. Many of these techniques rely upon single stepping the clock. Since it is not appropriate to stop the rotating pulses another method was devised to allow full built in scan and test capabilities. This method is compatible with existing techniques but does not require the rotary clock itself to be stopped or single stepped.

Tools and Methodology for GHz Rotary Clocked Design

With minor modifications existing commercial tools and flows can be used to generate a Rotary Clock based design for a fully synchronous chip of any size.   To fully benefit from rotary clocks, the design effort should be much more front-end loaded. This added effort pays off handsomely in reduced back-end design work and is much more scalable with new processes. Also, front-end architecture investigation and design is less costly and risky than back-end solutions such as custom layout flow. A set of beta-level EDA tools will be discussed that supplement existing commercial tools.

Given the freedom to rework the design flow the following should be considered: At the architectural level one should favor latency insensitive algorithms and communication protocols. RTL should be written for explicit support of fine-level pipelining. As the cost of adding flip-flops is very low, automatic pipelining tools should be configured with the correct tradeoff settings. Physical design should commence with the placement of the rotary clock grid and sequential elements. Highly predictable skew and very low jitter can be used to full advantage by synthesizing multiphase deeply pipelined designs.   The above techniques can result in a latency insensitive design that is scalable to higher frequencies. An interesting alternative that will also be discussed at the TecForum is 'pipelining for low power'.

Summary

At this point, we restate some of the assertions we made in the first part of this TecForum:
- Reduction of timing uncertainty can result in dramatic increases in clock frequency. An ASIC with low timing overhead could in theory run much faster, we have calculated that a 0.18u CMOS standard cell ASIC should be capable of running at up to 2 GHz.
- At higher clock frequencies pipelining is the key to raising throughput [14, 37, 38] and creating designs which will scale and be insensitive to latency.
- Standard cell designers cannot currently use most of the tricks full-custom designers can use because extra flip-flops for pipelining incur a power penalty, clock skew control is too difficult and expensive in terms of power, and custom layout productivity is too low.
- Clocking is a significant fraction of chip power consumption. Nearly all of this energy is expended driving CV^2F in the pass-elements of the latches. Power constraints can also constrain the operating speed.
- Standard cell designs are also further inhibited from exploiting some full-custom techniques due to the lack of automated CAD tools to support them. These include: aggressive clock scheduling, skew tolerance through latch based designs and deep pipelining.

This TecForum proposes a practical fully-synchronous multi-gigahertz Rotary Clock technique to help close the gap between standard cell and full custom design. This technique achieves greatly reduced timing uncertainty, dramatically reduced clocking power, simplified interconnect modeling and is also ideal for deep pipelining. The practicality of this technique makes it economical for standard cell designs as well as full-custom. The addition of front-end EDA tools designed to further exploit the capabilities of Rotary Clocks will bring multi-gigahertz designs within the reach of the typical ASIC design team.

References

[1] Stephan Held, Bernhard Korte, Matthias Ringe, Jens Vygen, Jens Maβberg. "Clock Scheduling and Clocktree Construction for High Performance ASICS," ICCAD 03, November 11-13, 2003

[2] D. G. Chinnery, B. Nikolic, K. Keutzer. "Achieving 550 MHz in an ASIC Methodology," DAC 2001, June 18-22, 2001

[3] D.G. Chinnery, K. Keutzer. "Closing the Gap between ASIC and Custom: AN ASIC Perspective," DAC Proceedings 2000

[4] Stephen E. Rich, Matthew J. Parker, Jim Schwartz."Reducing the Frequency Gap Between ASIC and Custom Designs: A Custom Perspective," DAC 2001, June 18-22, 2001

[5] "Optical and Electrical High-Speed Digital Clocking," Interconnect Focus Center Quarterly Workshop, Stanford University, December 7, 2002

[6] Quing K. Zhu. "High-Speed Clock Network Design," Kluwer Academic Publishers, 2003

[7] B. Floyd, X. Guo, J. Caserta, W. Bomstad, T. Dickson, J. Mehta, C.-M. Hung, and K. O. "Wireless Distribution," (Invited), ACM/IEEE International Workshop on Timing Issues in the Specification and Synthesis of Digital Systems, Dec. 2002, Monterey CA

[8] V.L. Chi. "Salphasic distribution of clock signals for synchronous systems," IEEE Trans. Comput., vol. 43, pp.597 602, May 1994.

[9] Mahony, et al. "10GHz Clock Distribution Using Coupled Standing-Wave Oscillators," ISSCC 2003, Paper 24.4

[10] G. M. Jacobs, R. W. Brodersen. "A Fully Asynchronous Digital Signal Processor Using Self-Timed Circuits," IEEE Journal of Solid-State Circuits, Vol. 25, No. 6, pp. 1526-1537, Dec. 1990.

[11] D. W. Bailey and B. J. Benschneider. "Clocking Design and Analysis for a 600-MHz Alpha Microprocessor," IEEE Journal of Solid-State Circuits, Vol. 33, Nov. 11, pp. 1627-1633, Nov. 1998

[12] Kerry Bernstein, Keith M. Carrig, Christopher M. Durham, Patrick R. Hansen, David Hogenmiller, Edward J. Nowak, Norman J. Rohrer. "High Speed CMOS Design Styles."

[13] David Harris, Mark A. Horowitz. "Skew-Tolerant Domino Circuits," ISSCC 1997

[14] Eric Spangle, Doug Carmean. "Increasing Processor Performance by Implementing Deeper Pipelines," Proceedings ISCA 2002, the 29th Annual International Symposium on Computer Architecture

[15] O. E. Akcasu. "Case Study of On-Chip Inductance Effects (Extraction and Analysis)."

[16] A. Deutsch, P.W. Coteus, G.V. Kopesay, H.H. Smith, C.W. Surovic, B. L. Krauter, D. D. Edelstein, and P. Restle. "On Chip Wiring Design Challenges for Gigahertz Operation," Proc. Of the IEEE, Vol. 89, No. 4, 2001

[17] Y. I. Ismail et al. "Figures of Merit to Characterize the Importance of On-Chip Inductance," T-VLSI, Vol. 7, No. 4, pp. 442 - 449, December 1999

[18] Yehea I. Ismail, Eby G. Friedman. "On-Chip Inductance in High Speed Integrated Circuits," Kluwer Academic Publishers, 2001

[19] A. Deutsch et al. "When are Transmission-Line Effects Important for On-Chip Interconnections?" T-MW Theory, Vol. 45, No. 10, PP. 1836-1845, October 1997

[20] O. E. Akcasu, Kerem Akcasu. "Impact of the On-Chip Inductive Effects on the Power Distribution Networks for Simultaneous Switching Noise and Ground Bounce Analysis for High-Speed Processor Design."

[21] Y. Ismail and E. Friedman. "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits."

[22] Howard Johnson, Martin Graham; "High-Speed Signal Propagation Advanced Black Magic," Pearson Education, Inc. 2003

[23] O. E. Akcasu et al., "NET-AN a Full Three Dimensional Parasitic Interconnect Distributed RLC Extractor for Large Full Chip Applications."

[24] Adler, R. "A study of Locking Phenomena in Oscillators," Proc. Of the IRE, vol. 34, pp. 351–358, June 1946

[25] Strogatz, S. H. "Norbert Wiener's Brain Waves" Lecture Notes in Biomathematics," Vol. 100 Springer, 1993

[26] Kurokawa K. "Injection Locking of Microwave Solid-State Oscillators," Proc. IEEE Vol. 61 no. 10 Oct. 1973 pp 1386-1410

[27] Buczek C. J.; Freiberg R. J., Skolnick M. L. "Laser Injection Locking," Proc. IEEE Vol. 61, No. 10 Oct. 1973 pp 1411-1431

[28] Stephan K. D., Morgan W. A. "Analysis of Interinjection -locked Oscillators for Integrated Phased Arrays," IEEE Trans Antennas and Propagation Vol. AP-35, No. 7, July 1987

[29] York, R. A. "Quasi-Optical Power Combining using Mutually Synchronized Oscillator Arrays," IEEE Trans. Microwave Theory and Techniques, Vol. 39, No. 6 June 1991

[30] Neda, Z.; Ravasz, E.; Vicsek, T.; Brechet, Y.; Barbasi, A. L. "Physics of rhythmic applause, Physical Review," E Vol. 61, No. 6 June 2000 pp. 6987-6992

[31] Clay J. R.; DeHaan R. L. "Fluctuations in Interbeat Interval in Rhythmic Heart-Cell Clusters, Role of Membrane Voltage," Biophys. J. Vol. 28, December 1979 pp. 377-390

[32] Enright, J. T."Temporal Precision in Circadian Systems: A reliable Neuronal Clock from Unreliable Components?" Science Vol. 209, Sept. 26, 1980 pp. 1542-1545

[33] Samuel D. Naffziger, Glen Colon-Bonet, Timothy Fischer, Reid Riedlinger, Thomas J. Sullivan and Tom Grutkowski. "The Implementation of the Itanium 2 Microprocessor, IEEE Journal of Solid-State Circuits," Vol. 37, No. 11, November 2002

[34] Vojin G. Oklobdzija, Vladimir M. Stojanovic, Dejan M. Markovic, and Mikola M. Nedovic. "Digital System Clocking High-Performance and Low-Power Aspects," John Wiley and Sons, Inc., 2003

[35] Ivan S. Kourtev, Eby G. Friedman. "Timing optimization through clock skew scheduling," Kluwer Academic Publishers

[36] Maria-Cristina Marinescu and Martin Rinard. "High-level Automatic Pipelining of Sequential Circuits," Proceedings of the 14th International Symposium on System Synthesis (ISSS 2001), Montreal, Canada, October 2001

[37] A. Hartstein and Thomas R. Puzak "The Optimum Pipeline Depth for a Microprocessor," Proceedings ISCA 2002 The 29th Annual International Symposium on Computer Architecture

[38] M.S. Hrishikesh , Norman P. Jouppi , Keith I. Farkas , Doug Burger , Stephen W. Keckler, Premkishore Shivakumar . "The Optimal Useful Logic Depth per Pipeline Stage is 6-8 FO4," Proceedings ISCA 2002, the 29th Annual International Symposium on Computer Architecture

[39] Brad Marshall, Juergen Koehl, and Tilman Wagner. "A New ASIC Timing Signoff Methodology," IBM Micronews, Second Quarter 2002, Vol. 8, No. 2