Post-Silicon Tuning with the Aim of Clock Skew Reduction

Fariborz Dadnam
fariborz@ucla.edu

***Abstract*—In this paper, various techniques for reducing the clock skew due to inter- and intra-die PVT variations will be examined. First, the components that are used in these techniques will be evaluated and later the performance and objective of these methods will be explained. The advantages and disadvantages of each technique are also going to be explored later is this paper.**

***Index Terms*—Post-silicon tuning, clock skew, PVT variations**

I. Introduction

POST-SILICON tuning is a manner is used to adjust and optimize the circuits' performance after design stage and fabrication when any change to the circuit is not possible or unpredicted variations may alter the circuit operation. Clock skews and jitters are usually caused by PVT variations. Most techniques follow the similar procedure to reduce the clock skew. First, the phase or delay difference between clock reference and clock at a node is required to be detected. Then, proper adjustments will be applied to reduce this timing difference. However, different techniques may utilize different algorithms, following different goals, to overcome this issue. Further, employing novel devices to sense clock skews or using specific devices to tweak the clock skews may diminish the area, cost and power consumed by these extra components and increase accuracy and speed, instead.

II. PVT Variations

Process, voltage, and temperature (PVT) fluctuations cause inevitable variations in fabricated chips.

Process variations can be classified into two types: inter- and intra-die variations. Inter-die variations model the variations over different dies while intra-die variations model the individual and local variations within the same die (e.g. the random mismatches due to doping fluctuations). Process fluctuations cause considerable clock uncertainty in sub-micron technologies. Particularly, since the clock signal has to be distributed all over the chip synchronously, any dissimilarity over the die may cause significant clock skew [1].

Voltage variations due to power supply or coupling at different locations of the die can initiate timing uncertainly in clock buffers and consequently produce clock skews or jitters.

Temperature variations are usually initiated by heat generated in dense and most active spots of the die. These variations may affect the performance of the transistors and therefore their response to clock signal.

As a result, it is desired to suppress and overcome these fluctuations by tuning the chip either before fabrication – during design and layout stage – or after fabrication.

III. Silicon Tuning

As was described, PVT variations can have a significant impact on circuit timing of a die. For this reason numerous pre- and post-silicon methods have been proposed to reduce the effect of these variations.

Some variations due to PVT can be predicted and considered during the design and layout phase. For example, more active circuits may be placed far away from each other to distribute heat across the chip as uniform as possible. However, these pre-silicon techniques are usually effective for expected and known variations. As combining the aggressive circuits and shrinking processes, the influence of variations over the chip and intra-die increases [2] [3].

This high-range of variations is not always possible to predict during the design and layout stage but after fabrication. For this reason, by collecting data from automatic test equipment (ATE) from the test wafers, some information will be fed back to design process. Nevertheless, not every measurement is precisely possible during ATE process. Especially, it is almost impossible to monitor all clock signals over the entire chip. Furthermore, many physical parameters of the chip may change during its life. As a result, post-silicon techniques suggest that some adaptive circuits to be implanted in the chip in order to self-monitor the chip after fabrication and tune the chip accordingly to keep its performance stable.

Since adaptive post-silicon tuning (PST) methods allow each die to be adjusted independently, even dies with strongly skewed process conditions can be tweaked to meet power and delay specifications. Hence, PST techniques provide the opportunity for almost all manufactured chips to exactly meet their constraint [4].

IV. Various PST Approaches for Clock Deskewing

Numerous PST approaches have been proposed in order to achieve best accuracy and desired speed of clock deskewing with minimum number of extra components. These

This work is supported in part by the Electrical Engineering Department of the University of California, Los Angeles.

approaches are based on the idea of detecting the clock skew between a reference clock and the target node and then apply proper signals to a clock buffer to adjust the clock signal of some part of the chip including the target node.

A phase or delay detector can detect the arrival time (phase) difference between two clock signals. These signals have to be as accurate as possible to avoid any uncertainly.

A tunable clock (delay) buffer is a clock buffer that can delay the clock signal of some part of a chip according to its controlling inputs.

In addition to the goal of optimizing these major components used in PST methods, different methods try to find the best topology and distribution of these PST components over a clock network. These approaches may follow some direction according to a specific chip design (e.g. in a dual-core Intel® Itanium® processor [5]) or just follow a general topology to be used in many chips (e.g. in QRT method [3]).

In this paper, first, various novel elements used in PST methods will be studied and then different methodologies will be evaluated.

A. Phase Detectors

Phase or delay detectors (PD) are used to monitor the target node clock and compare it with the reference clock. If any clock skew detected, a proper signal is sent to the clock buffer to adjust the target clock, accordingly.

Up/Down Detector [3]

The up/down detector (UDD) measures the skew between the reference clock and the node clock being tuned. Any skew of greater magnitude than UDD's threshold will trigger either an UP or DOWN signal to indicate that the tunable buffer (TB) should be adjusted. The schematic of a UDD is shown in Fig. 1.

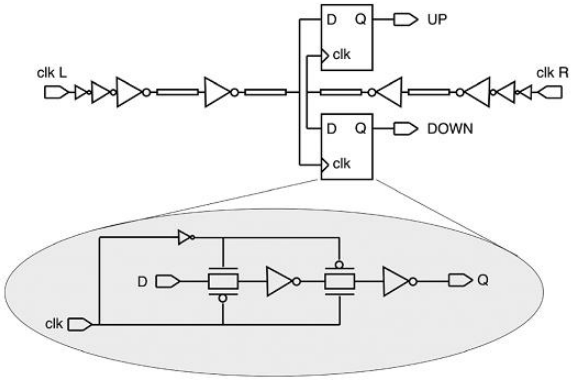


Fig. 1. UDD schematic with detailed DFF [3]

There are two important design considerations regarding UDDs: 1) delivering accurate clock signals to a UDD for comparison and 2) setting the up and down thresholds of a UDD precisely.

First, the two clocks being compared here are not available at the same location on the chip. Therefore, path lines from the two clocks (the reference and the node) must be carefully routed to UDD in order not to alternate the difference in timing between these two clocks. Furthermore, all of the inverters used in this top schematic are single-edge clocking (SEC) buffers (to be discussed later), since the SEC method results in reduced latency and skew, therefore higher accuracy.

Second, a UDD must detect the skews greater than a designed threshold accurately. There are two skew thresholds defined for this design at upper and lower bounds. This definition has to be symmetrical about the zero skew axes, such that the absolute value of negative skew threshold is equal to the positive skew threshold. Furthermore, the setup time of a simple dynamic D-flip-flop (DFF) is used as UDD's threshold for accuracy of UDD. The setup time of this DFF can be precisely controlled by using the different sizes of the input transmission-gates.

Fig. 2 represents the assignment of UP/DOWN signals with respect to the skew thresholds.

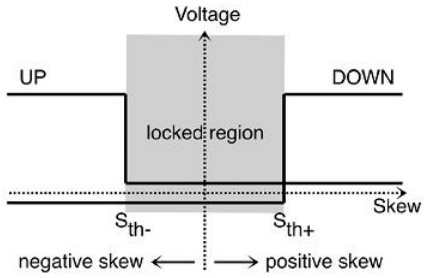


Fig. 2. The UDD's UP/DOWN signal assignments with respect to the skew thresholds [3]

There are three regions defined for UDD actions. When no skew is detected, the system is in the locked state and no adjustment signal will be sent. Otherwise, either an UP or DOWN signal is being sent to the next stage.

The skew threshold of a UDD is the design parameter that will ensure stable locking of whole system over PVT variations. It is proven that the best result is yield when $S_{th} = 1.5ss$. ss is defined as the distance between any two adjacent tuning steps. Mathematically, ss is obtained as follows,

$$ss = tuning\_range/(\#bits - 1)$$

Here, $tuning\_range$ is the tuning rage introduced by the delay buffer and has to span the whole range of PVT variations.

B. Tunable Clock Buffers

PST methods utilize tunable clock (delay) buffers to control the clock signal received by the target nodes. These clock buffers have to be very precise in order not to produce any additional clock skew and, in addition, get activated at a right time.

Single-Edge Clocking Delay Buffer [6]

Single-edge clocking (SEC) is a fairly new technique that can reduce clock skew and jitter by up to 20% and 30%, respectively, in contrast with traditional clocking methods.

The tunable clock buffers have traditionally been intended

to generate equal rise and fall times. They propagate almost constant clock pulse width throughout the clock network while only one edge is the critical edge to be used for timing purposes. Conversely, SEC systems distribute asymmetrical clock signals and only the clock's critical edge is being utilized while abandoning its other edge. However, this neglected edge may not go beyond the clock cycle bounds. In fact, this critical edge is the one that ultimately activates the target node flip-flops.

Fig. 3 shows a symbolic difference between traditional clocking (a) and single-edge clocking (b).

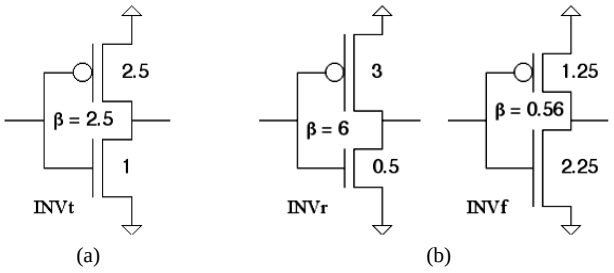


Fig. 3. (a) Traditional clocking and (b) single-edge clocking [6]

In this novel SEC system the critical edge clock signal propagates via alternating strong pull-up (INVr) and strong pull-down (INVf) buffers while the neglected edge is calm and lets the clock cycle grow. By simple equations, the allowable degradation bound of neglected edge for a given clock period and distribution network is being calculated.

The size of buffers is defined through relation, $\beta = W_{pullup}/W_{pulldown}$. In this design, it is tried to keep the combined NMOS+PMOS gate area the same size while shifting β. The reason is to keep the input and output capacitances relatively constant. This allows the critical edge faster and sharper by overweighting it and, conversely, underweighting neglected edge. Fig. 4 illustrates this difference.

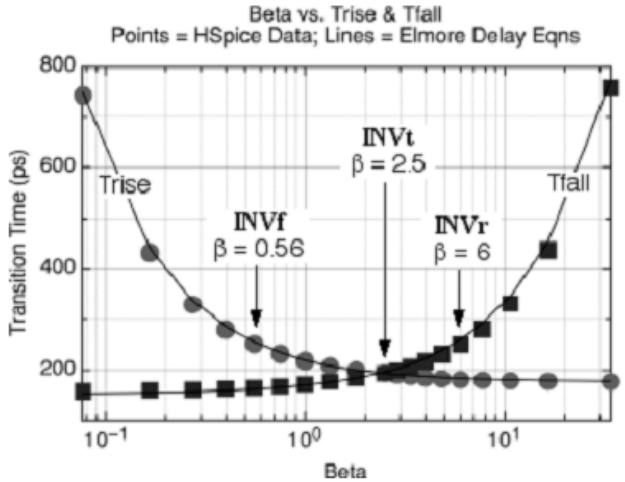


Fig. 4. Buffer rise and fall times versus β [6]

A delay buffer is typically created by delaying a buffer or inverter. There are three general methods to delay an inverter: 1) adding load capacitances to the inverter's output (tcp-INVf), 2) limiting the inverter's power and/or ground supplies (tsi-INVf), and 3) feeding back the inverter's output (e.g. clock vernier device (CVD), to be discussed later). The first two methods are being employed in this paper while the third one is being used in industry. Fig. 5 represents these two delay buffers.

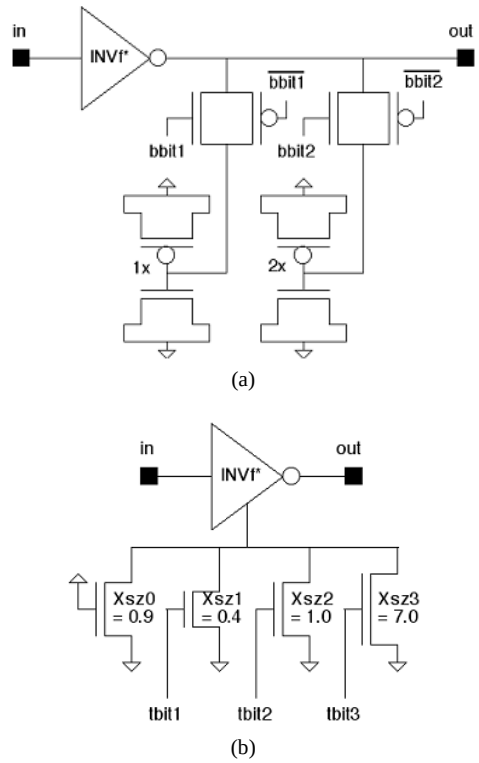


Fig. 5. (a) tcp-INVf, (b) tsi-INVf tunable delay buffers [6]

In case of the shunt capacitors, as shown in Fig. 5(a), tcp_INVf can provide four selectable delay steps by different $bbit1$ and $bbit2$ combinations, $\{00(min), 10, 01, 11(max)\}$. By adding a capacitor, controlled by these bits, the clock signal will slow down.

The goal of the limited (starved) inverter method is to control the delay of the critical edge while allowing the neglected edge to pass unchanged. As shown in Fig. 5(b), this is done by adding NMOS transistors to ground. In this case, there would be four delay steps by combining $tbit1$, $tbit2$, and $tbit3$ bits as follows, $\{111(min), 110, 100, 000(max)\}$.

Fig. 6 shows transient response of manipulating these two delay buffers.

As seen in these graphs, the tsi-INVf buffers delay only the critical edge, thus all four of the neglected edges are matching. On the other hand, the tcp-INVf buffers delay both the critical and neglected edges. Because of this property of the tsi-INVf buffers, they are able to function at shorter clock periods (higher clock frequencies) than the tcp-INVf buffers.

Because of wider pulses in tcp-INVf, consumption would be larger in comparison with tsi-INVf case, the power case. Also, fabricating MOS transistors rather than capacitors, in tsi-INVf buffers, take smaller area and this area is almost

(a)

(b)

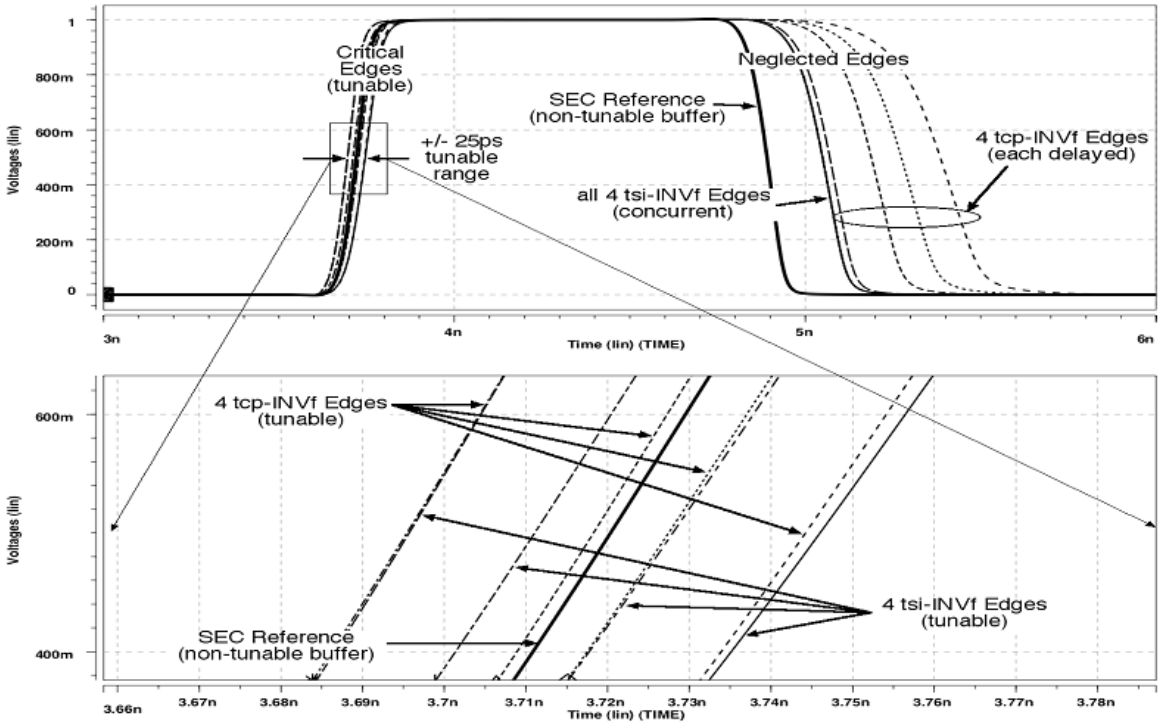


Fig. 6. SEC transient response; (a) node clocks (full period), (b) critical edge (zoom) [6]

invariant to delay values.

As a result, tsi-INVf buffers are more appropriate to be used in the SEC systems than tci-INVf ones.

Fig. 7 shows the PST improvements by this method in case of applying different PVT variations.

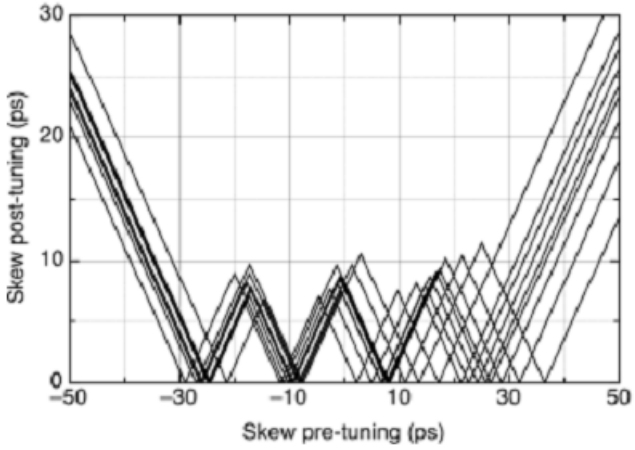


Fig. 7. Skew compensation over PVT [6]

As a limitation of the tsi-INVf delay buffers is that they can only be controlled by the thermometer code method (not binary), so having more than 4 or 5 steps per buffer could be difficult. Though, two or more of these buffers could be used within a particular clock tree branch to provide more tuning steps and a wider tuning range.

Clock Vernier Device Delay Buffer [7]

A clock vernier device (CVD) delays a received reference clock signal ($SLCBO$) in a controlled manner and then send it out ($CVDO$) to the final buffering stage before the target node. Fig. 8 shows a CVD delay buffer schematic.

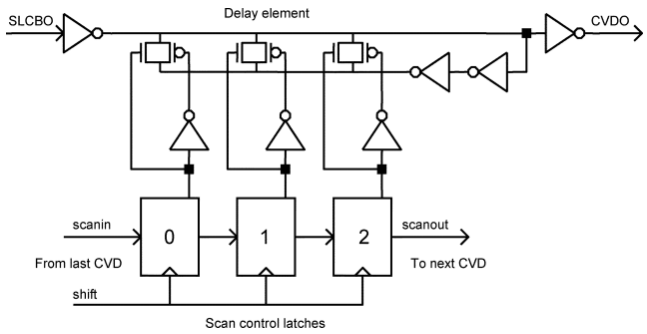


Fig. 8. CVD delay buffer schematic [7]

Using this configuration with three scanned control laches, totally eight delay steps are allowable. Also, the power consumption of these CVDs proportionally related to the delay time.

In practice, CVDs have proven to be very helpful. On initial Montecito® silicon, CVDs were utilized to make a workaround for an electrical silicon bug which would have made the OS boot time longer than expected. This was done by patching the firmware in which would modify the settings of scan chain of CVDs. This is an example that delay buffers were controlled manually rather than by monitoring circuits such as PDs. According to [7], a total of 6417 CVDs are present in each core of Itanium® family processors. Each CVD controls an average of 100 target nodes.

Another benefit of using CVDs, which is proven by experiment, is that the performance of them is improved by

increasing the frequency. This advantage has provided the opportunity of using CVDs in processor industry, widely.

Hot-Carrier Injection Delay Buffer [8]

Hot-carrier injection (HCI) provides a low-cost PST method that can adjust the V_{th} of the fabricated transistors in each tunable clock buffer individually. Fig. 9 illustrates the HCI mechanism.

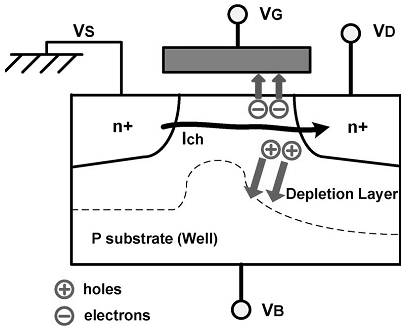


Fig. 9. HCI mechanism on an NMOS transistor [8]

In order to stress HCI transistors, there are two conditions to be satisfied: 1) a large current flows through the channel, 2) a high V_{ds} exists. The first condition cause some electrons get energized by the horizontal channel electric field and may become hot-carriers and therefore injected into the gate oxide. The second condition causes acceleration of the channel carriers of the drain. Consequently, ionization collision and avalanche multiplication can occur and generates electron–hole pairs. Some of these pairs are also injected into the gate oxide. The result of HCI stressing is the increase of V_{th}. Because a large amount of carriers can be trapped for the device's entire lifetime, this V_{th} increase can be memorized.

Fig. 10 shows the employment of HCI method in an HCI trimmed clock buffer (HTCB).

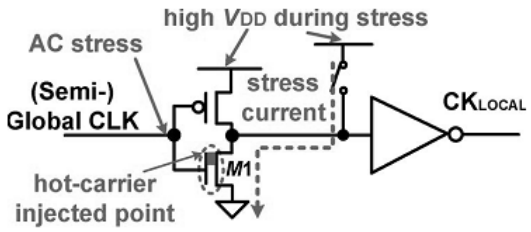


Fig. 10. HTCB schematic [8]

M1, in this figure, is stressed by HCI method and, in fact, is triggered by positive clock edges. In addition, the right inverter can reshape the delayed clock signal to improve the swing rate.

The switch is "on" during stressing to apply a high V_{dd}. Concurrently, an AC or a DC signal is applied to the gate for a brief time. Thereafter, when stressing is done, the switch turns back "off" and V_{dd} becomes normal. As mentioned, the V_{th} increase is memorized by M1 and hence delay will increase as well. As a result, by controlling the switch, the clock signal can be adjusted. Fig. 11 presents the impact of HCI trimming on the CK_{LOCAL} at the target node.

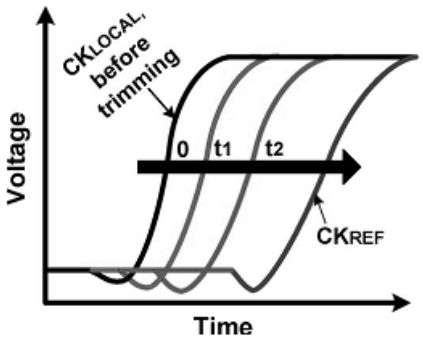


Fig. 11. Impact of HCI trimming on CK_{LOCAL} [8]

This proposed HTCB is fabricated in a 40-nm high-k CMOS. Fig. 12(a) and (b) show the measured clock rise edges and skews at $0.5V$ V_{dd} before and after trimming, respectively. Fig. 12(a) is superimposed on Fig. 12(b) in order to examine whether other logic gates – that share the same power lines – are influenced by imposing a high stressing V_{dd}. This result is sketched on Fig. 12(c). As seen, the rise time of CK_{REF} is almost unaffected. Fig. 12(d) shows the result when CK_{LOCAL} is delayed.

As discussed, this method is low-cost and consumes a small area of the die. The effectiveness of HCI method has been confirmed as employed by HTCB. Especially, in sub-micron chips this method can be easily used. The key here is to increase the chance of injection of hot-carriers into gate oxide.

A drawback of this mechanism is the use of large amount of current through the channel. In addition, as a future consideration, the time of stressing can be optimized to get a faster response off this method.

C. PST Methodologies and Structures

Several novel methodologies to employ PST components in various topologies have been proposed. In this paper, I will just go over a few of them that have provided the best overall result over various PVT fluctuations.

Quadrantal Ring Tuning

By increasing the size of clock networks and consequently growing levels of intra-die variations, it will require a finer resolution of adjustability consisting of many more tunable clock zones. Quadrantal Ring Tuning (QRT) methodology provides a symmetrical way to distribute clock signals throughout the chip. In addition, this proposed methodology is easily expandable to as many zones as needed to achieve higher accuracy.

This technique employs UDDs and SEC delay buffers for clock deskewing the chip. A unit block of these components is called a delay-locked loop (DLL). These components are arranged in a quadrantal clock H-three.

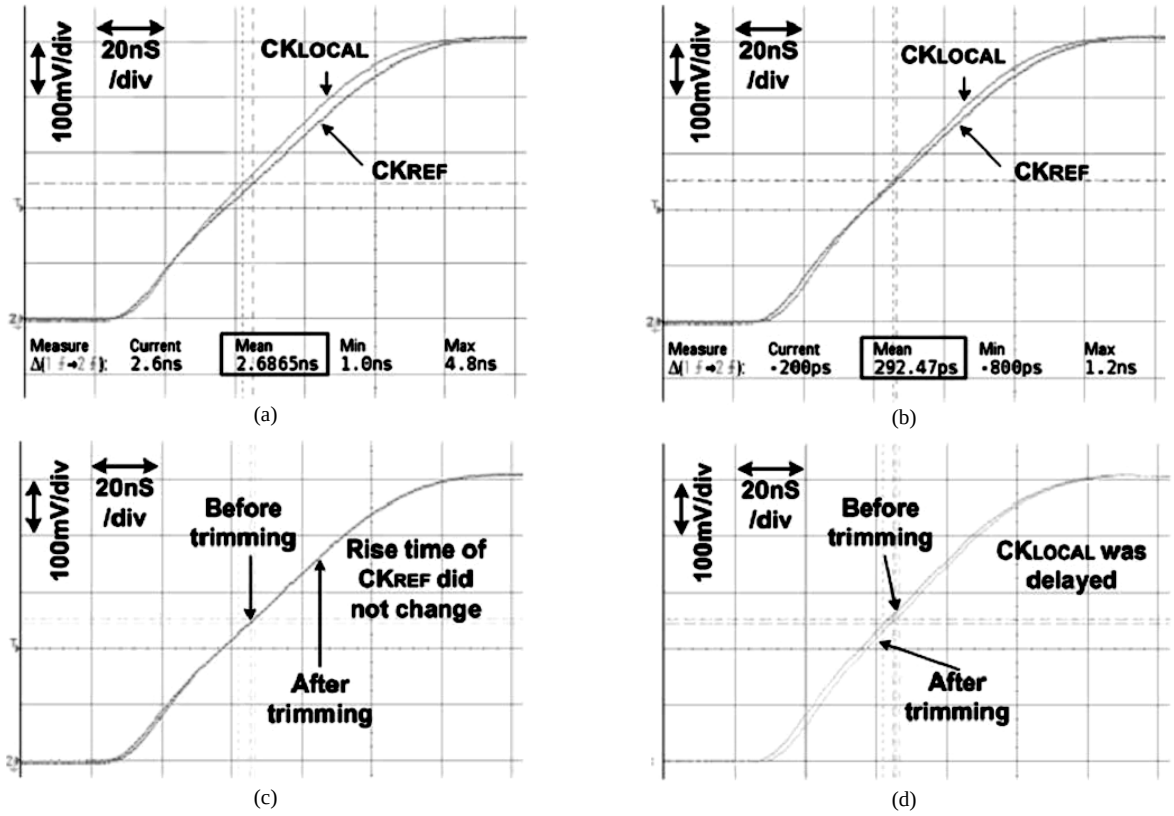


Fig. 12. Measured clock waveforms at $0.5V$ V_{dd}; (a) before trimming, (b) after trimming, (c) superimposing CK_{REF} of (a) on (b), (d) delayed CK_{LOCAL} [8]

Fig. 13 shows such a system.

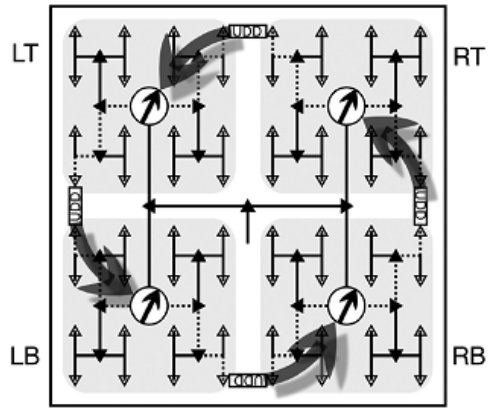


Fig. 13. Signle-level QRT diagram [3]

As seen on the figure, the four quadrants are linked in a manner of counterclockwise. As a matter of fact, each delay buffer (indicated by an arrow inside a circle) is controlled by a UDD at the border of the same quadrant and the quadrant before in clockwise (as the dark arrows present).

For accuracy of clock signals received by UDDs, they are placed close to the clock sources. Here, the reference clock – for example for the top UDD – comes from the right quadrant and the target node is on the left quadrant. These UDDs are placed at the edge of the die since the variations at the edge are larger.

As discussed in SEC delay buffers section, these buffers can only use thermometer codes. In this case, for every delay buffer a thermometer code generator (TCG) is proposed that converts DC signals from UDD to 3-bit codes. It is imperative to place these TCGs close to delay buffers to avoid any extra clock skew due to long distance. However, DC signals from UDD (UP/DOWN) can routed a longer distance without distortion. These TCGs are just simple clocked NAND pulse generators.

The selection of tuning zones must balance between the desire for *simpler control of fewer, large tuning zones* with the desire for *higher precision/accuracy of more, small tuning zones*. In Fig. 13, a single-level QRT is shown at which the tunable buffers are placed at the level-2 of the H-tree. A multi-level QRT with level-2 and 3 tunable buffers is shown in Fig. 14.

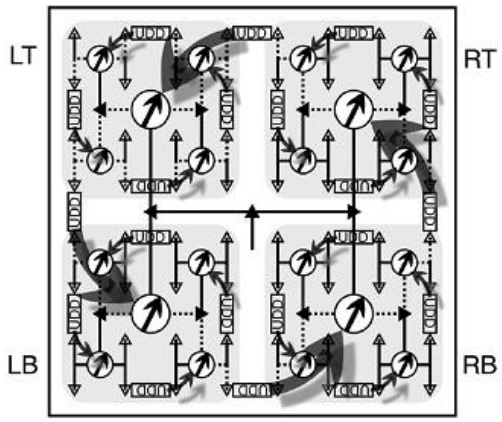
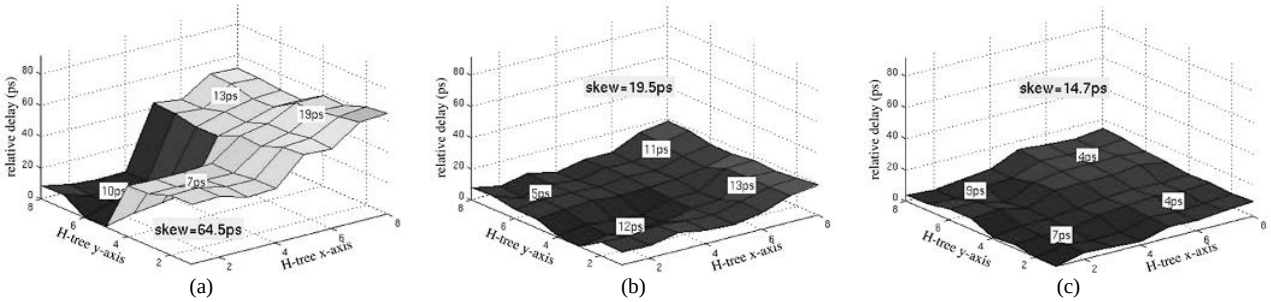


Fig. 14. Multilevel QRT with delay buffers at level-2 and 3 [3]



(a) (b) (c)

Fig. 15. Surface plots of the relative clock delays; (a) no tuning, (b) QRT level-2 only, (c) QRT level-2 and 3 [3]

In this configuration, sixteen additional DLLs are used for further precision and accuracy.

For a simulation test, a level-2 QRT and a level-2 and 3 QRT have been examined. Fig. 15 shows the result over PVT variations.

This result is very promising. These simulations have gone over twelve spatial variations along with random variations to simulate intra-die fluctuations due to PVT.

As mentioned, this method is symmetrical and could be useful for any chip. Moreover, this symmetric property provides this opportunity that makes this design easily expandable for more accuracy. Also, it provides a very promising clock deskewing over a wide variety of PVT variations.

As limitation of this technique is that, the tunable buffers can be adjusted only one step at a time. Also, as mentioned, the tuning range of buffers is proportional to this step size.

Statistical Timing Analysis for Unsymmetrical Structures [5]

Many design PST methodologies usually use a tunable clock buffer for each flip-flop or an entire level of a clock-tree (e.g. QRT method). This can cause significant design overhead and long tuning time. In this proposed approach, the clock buffers are inserted at both internal and target nodes of a clock-tree and used a bottom-up algorithm to reduce the number of candidate tunable clock buffer locations. In this paper, a statistical timing analysis has been used to find critical locations to put tunable clock buffers and assign appropriate tuning range to them.

For this method, any delay buffer discussed before can be used and there is no assumption of using any specific ones. This method has followed two different goals: minimizing area and minimizing the number of clock buffers. These two goals require two different optimization algorithms, but are driven with the same timing yield model.

This method uses previous data from statistical researches to model delay model. Hold-time and setup-time are the design constraints. An optimal timing yield model has been developed in this paper by studying the effect of clock buffers on the timing yield. Please note that the optimal timing yield is based on the assumption that all clock buffers are adjustable and have infinite tuning ranges.

To minimize the total area, a Simultaneous Perturbation (SP) algorithm is used to reduce the time for gradient approximation of the timing yield function. In this algorithm, the tuning ranges of the candidate clock buffers – that do not provide improvement to the timing yield – are forced to be squeezed down toward zero.

To minimize the number of tunable clock buffers, a greedy algorithm is utilized. However, because of the runtime issue of the Greedy algorithm, a Batch selection algorithm is also proposed to speed up the process.

In order to test these algorithms, an H-tree has been assumed. For the initial point of algorithm, it has been assumed a clock buffer at every branching point and terminal of the H-tree. Fig. 16 illustrates such an H-tree.

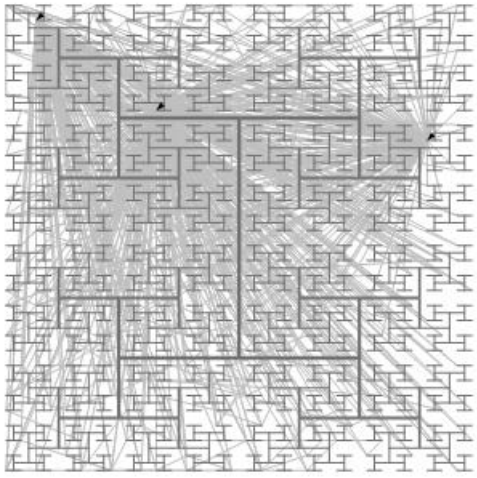


Fig. 16. The PST clock-tree of S35932.The arrows indicate the tunable clock buffers and timing critical paths [5]

The result from this method (IterSP) for total area improvement has been compared with regular method in which all terminals have identical tunable clock buffers. TABLE I shows data from five circuits.

As seen from this table, IterSP achieves greater than 65% improvement in area reduction for all cases. The reason for these significant improvements is that IterSP assigns the tuning range to what is required and not more than that. The other reason is that IterSP distributes the total tuning range among all candidate clock buffers in different levels.

Although this method shows a significant improvement in results, it takes an enormously long runtime for larger designs. In order to overcome this issue, for large problems, the number of candidates has to be limited.

TABLE I
COMPARISON ON TOTAL AREA BETWEEN A REGULAR METHOD AND THE ITERARIVE SP ALGORITHM [5]

	Regular		IterSP			
Circuit	Area	CPU (m)	Area	Reduction (%)	Steps	CPU (h)
S9234.1	24.94	1.9	8.51	65.9	52	1.1
S13207.1	68.60	2.8	11.33	83.5	63	3.4
S15850.1	104.85	7.5	10.99	90.4	107	6.5
S35932	101.23	42.2	1.90	97.4	73	44.3
S38584.1	238.26	36.7	17.19	92.8	189	30.8

TABLE II
COMPARISON ON NUMBER OF TUNABLE CLOCK BUFFERS AMONG THREE DESIGN METHODS [5]

	Number of candidate buffers	Number of tunable clock buffers					CPU Time		
Circuit		Levelized	Greedy	Reduction (%)	Batch	Reduction (%)	Levelized	Greedy	Batch
S9234.1	149 (75)	32	8	75	8	75	$1.7m$	$16.9m$	$4.6m$
S13207.1	417 (210)	256	16	94	18	93	$1.9m$	$35.8m$	$6.8m$
S15850.1	641 (321)	128	17	87	21	84	$13.2m$	$1.4h$	$21.6m$
S35932	541 (271)	16	3	81	3	81	$1.5h$	$14.7h$	$3.1h$
S38584.1	1591 (796)	512	–	–	162	68	$14.3m$	$> 2\ days$	$8.9h$

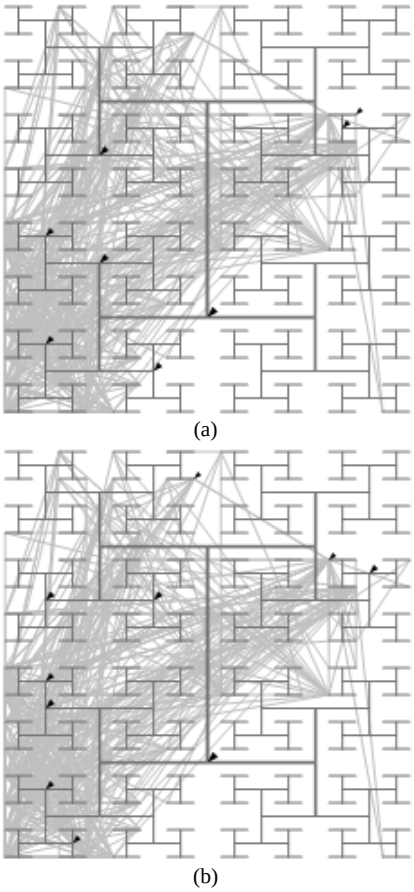
The results for the number of tunable clock buffer reduction are given in TABLE II. Here, the data have been compared for the methods of Levelized, Greedy, and Batch. The Levelized algorithm inserts tunable clock buffers at a single level as close as possible to the root clock as long as the target timing yield is favored.

As seen, in general, Levelized uses four times more tunable clock buffers than Greedy and Batch selection methods. The reason is that the tunable clock buffers are only placed on one level of clock network. The comparison between Greedy and Batch selection methods shows that the improvement is almost identical; however, the runtime is much less in case of Batch selection method.
Fig. 17 shows the difference between Greedy and Batch selection methods on S9234.1.

For cases of large circuits, the algorithms' runtime is significant. The majority of the runtime is spent on Monte Carlo simulation for estimating timing yields. The runtime can be greatly reduced, in practice, by parallel computing to minimize the impact on the design turn-around-time (TAT).

A brute-force method may assume tunable clock buffers at every terminal of the clock tree and try to design the tunable clock buffers to have a wide tuning range. Nonetheless, this can add a large overhead to the design. To optimize the usage of extra components, this statistical method has been proposed. In this approach, there were two objectives to be followed: either reduce the total area used by clock network or reduce the number of tunable buffers. These objectives were achieved by two different algorithms: iterative SP method and Batch selection method, respectively.

This method is very promising as it can be used simultaneously with other methods to reduce the number of used components or the area used by these components. However, the optimization time is a big issue in this method. Also it doesn't consider the PVT and random variations that might be introduced after fabrication.



(a)

(b)

Fig. 17. Clock trees of S9234.1 generated by (a) Greedy and (b) Batch methods [5]

V. CONCLUSION

In this paper, the purpose of using post-silicon tuning

devices has been deeply explained. The major elements of a typical PST unit were considered. We selected phase detectors and tunable clock buffers from various papers to explain these components thoroughly and at the end of each section the advantages/disadvantages of each method were reviewed.

Several approaches of PST clock network topologies were proposed in this paper. Each of these approaches was developed to obtain different goals. For instance, the QRT method tries to achieve the best improvement over a variety of PVT variations. On the other hand, the statistical method pays more attention on the area used by clock networks and the number of extra elements added by clock deskewing circuits.

REFERENCES

[1] X. Li, B. Taylor, Y. Chien and L. T. Pileggi, "Adaptive Post-Silicon Tuning for Analog Circuits: Concept, Analysis and Optimization," *IEEE*, pp. 450-457, 2007.

[2] J.-L. Tsai, L. Zhang and C. C.-P. Chen, "Statistical Timing Analysis Driven Post-Silicon-Tunable Clock-Tree Synthesis," *IEEE*, pp. 574-580, 2005.

[3] Y. Pu, X. Zhang, K. Ikeuchi, A. Muramatsu, A. Kawasumi, M. Takamiya, M. Nomura, H. Shinohara and T. Sakurai, "Post-Silicon Clock Deskew Employing Hot-Carrier Injection Trimming With On-Chip Skew Monitoring and Auto-Stressing Scheme for Sub/Near Threshold Digital Circuits," *IEEE*, pp. 294-298, 2011.

[4] S. Naffziger, B. Stackhouse, T. Grutkowski, D. Josephson, J. Desai, E. Alon and M. Horowitz, "The Implementation of a 2-Core, Multi-Threaded Itanium Family Processor," *IEEE*, vol. 41, no. 1, pp. 197-209, January 2006.

[5] J. G. Mueller and R. A. Saleh, "Autonomous, Multilevel Ring Tuning Scheme for Post-Silicon Active Clock Deskewing Over Intra-Die Variations," *IEEE*, vol. 9, no. 6, pp. 973-986, June 2011.

[6] J. G. Mueller and R. Saleh, "A Tunable Clock Buffer for Intra-die PVT Compensation in Single-Edge Clock (SEC) Distribution Networks," *IEEE*, pp. 572-577, 2088.

[7] Z. Lak and N. Nicolici, "A New Algorithm for Post-Silicon Clock Measurement and Tuning," *IEEE*, pp. 53-59, 2011.

[8] S. H. Kulkarni, D. Sylvester and D. Blaauw, "A Statistical Framework for Post-Silicon Tuning through Body Bias Clustering," *ICCAD*, pp. 39-46, November 2006.

[9] M. Y. Kao, K.-T. Tsai and S.-C. Chang, "A Robust Architecture for Post-Silicon Skew Tuning," *IEEE*, pp. 774-778, 2011.