

Reducing the Frequency Gap Between ASIC and Custom Designs: A Custom Perspective

Stephen E. Rich:
Matthew J. Parker:
Jim Schwartz:

Desktop Platforms Group
Intel® Corporation

(503) 613-5282 Design Automation Manager
(503) 613-5028 Design Automation Engineer
(503) 613-5286 Design Automation Engineer

RA2-401 2501 NW 229th Ave.
Hillsboro, OR 97124

(stephen.e.rich@intel.com, matthew.j.parker@intel.com, jim.schwartz@intel.com)

1 ABSTRACT

This paper proposes that the ability to control the difference between the simulated and actual frequencies of a design is a key strategy to achieving high frequency in both ASIC and custom designs. We will examine this principle and the methodologies that can be deployed to manage this gap.

2 INTRODUCTION

D.G. Chinnery and K. Keutzer's DAC paper entitled "Closing the Gap Between ASIC and Custom: An ASIC Perspective"¹ quantified the speed gap between ASIC and Custom designs to be approximately 8x. This paper proposes that the path to high frequency design requires the ability to apply costly local optimizations techniques to the critical paths of the design. Large microprocessor projects can afford to apply these techniques more generously and thus have been able to obtain higher frequencies.

Resources alone, however, cannot account for the frequency gap between ASIC, custom, and microprocessor designs. The means to higher frequency for all of these design styles is the ability to accurately determine which paths in the design are critical and to apply local optimization to limited parts of the design.

In order to differentiate the critical from the non-critical paths, design teams must reduce the amount of uncertainty in the design process. This will enable them to apply global optimizations to the non-critical portions of the design and limit the costs of local optimizations, such as dynamic circuitry or complex clocking, to the critical paths.

This paper proposes the most significant difference between the ASIC and Custom design methodologies is their ability to minimize and deal with uncertainty in the design process. Further, this paper develops a framework for optimizing designs by minimizing the amount of uncertainty in the design and provides simple tools to help guide the methodology development process for higher frequencies.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2001, June 18-22, 2001, Las Vegas, Nevada, USA.
Copyright 2001 ACM 1-58113-297-2/01/0006...\$5.00.

¹ D.G. Chinnery, K. Keutzer. Closing the Gap Between ASIC and Custom: An ASIC Perspective. Proc. of DAC '00.

After reading this paper the reader will walk away with:

- A process for over-constraining designs (section 8).
- A table of uncertainty sources (section 12).
- A process for developing methodologies to address design uncertainty (section 9).

3 ASSUMPTIONS

At the risk of stating the obvious, reducing the frequency gap between ASICs and Customs implies that ASIC designers: (1) Set significantly higher frequency targets and find value in developing the fastest design possible, and (2) Are willing to deploy new methodologies. This paper assumes you are either a designer looking for ways to significantly increase your frequency or a CAD developer looking for areas to focus on.

4 TERMS

Market Frequency: Customer committed silicon frequency.
Target Frequency: Frequency used by design tools/team.
Simulated Frequency: Simulator path predicted frequency.
Actual Frequency: Frequency measured in silicon.

5 HYPOTHESIS

The closer the correlation between the simulated frequency and the actual frequency the higher the actual frequency will be.

Any gap between the simulated and actual frequency implies a source of uncertainty in the design process. We propose that any uncertainty in the design process will limit the maximum potential actual frequency of the design and thus should be the focus of high speed methodology development.

6 UNCERTAINTY DEFINED

There are two major categories of uncertainty, or variations, in the design development process: Process uncertainty and design uncertainty.

6.1 Process Uncertainty

Process uncertainty is caused by variations in the manufacturing process. One example of process uncertainty is the in-die variation of transistor geometries. The transistor speed difference, caused by this variation, is generally addressed by guard banding the min delay analysis. Detailed knowledge of how process uncertainty is accounted for in the technology files and libraries is costly but vitally important. The ability to obtain this information may be one of the significant contributors to the ASIC vs. custom frequency gap. Although process uncertainty is a major cause of the gap between simulated and actual frequency, it is a well

understood concept that has been studied in detail. We will instead focus this paper on design uncertainty.

6.2 Design Uncertainty

Design uncertainty is caused by inaccuracy in the design analysis tools or unpredictable variations in the design between iterations. For example, assume an automated tool routes the clock without any special algorithms. The random nature of the routing would cause a large amount of variation in the RC delay to the sequential devices. If this variation is not analyzed, it becomes a source of uncertainty which is typically dealt with through guard banding. In addition, the random nature of the routing causes significant variations in the clock tree between iterations which may lead to reordering of the paths. This reduces the confidence that the critical path in any iteration will continue to be the critical path in future iterations. By using a better design practice, such as pre-routing the clock tree spines, the uncertainty of the clock routing and thus the arrival time of the clock edge can be significantly reduced between iterations. This helps to stabilize the path ordering attributed to clock variation. (See table in section 12: Sources of Uncertainty)

As the frequency of designs increase, the tolerance for design uncertainty drops dramatically as the uncertainty becomes a larger percentage of the total cycle time. The sources of design uncertainty grow as the frequency increases and second order electrical effects become first order effects. Thus, learning to efficiently deal with these types of design variations in a uniform manor becomes exceedingly important as the frequency of the designs increase.

The amount of uncertainty in the design practices can be determined by measuring the delay of as many paths as possible in actual silicon. The maximum difference between the simulated and actual frequencies is a good estimate of the total amount of uncertainty in the design.

Summary: *Uncertainty in the manufacturing and design process causes a gap between the simulated and actual frequencies.*

7 WHY UNCERTAINTY REDUCES THE MAXIMUM POSSIBLE FREQUENCY

Uncertainty in the design process causes the design team to waste resources and energy working on non-critical parts of the design. This ultimately results in either a slower part or an unnecessary delay in the time-to-market.

To illustrate this, let us first look at the typical slack graph shown in figure 6.1. In this graph, we will assume that the graph represents the simulated frequency as simulated with an ideal clock and no other forms of uncertainty.

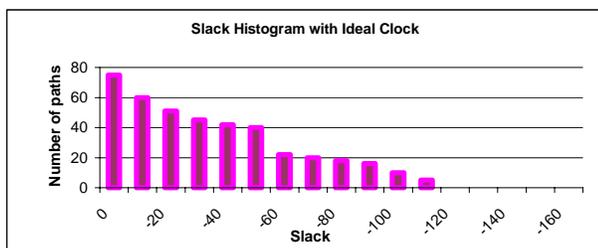


Figure 6.1

Let us inject some uncertainty in the design. In this example, we will use a worst case clock uncertainty value of 40ps (pick a meaningful number for your frequencies). This means that our design could run 40ps faster than expected or 40ps slower than expected. If the clock edge at the sampling sequential is 40ps late for the critical path, the design's actual frequency may be faster than the target frequency. In other words, the part would run faster than simulated.

However, if the clock at the sampling sequential is 40ps earlier due to the uncertainty in the design process, the actual frequency **would** be slower than the target frequency, which translates into a zero percent raise at the end of the year...(if you're lucky).

The simplest way to deal with this uncertainty, both in the design and in your raise, is to add 40ps of guard banding into the design flow to cover the worst case clock skew. The new graph in figure 6.2 shown below simply results in a shift of the slack histogram to the right. This will force the tools to work harder on at least the critical path thus limiting the risk of the actual frequency being slower than the target frequency.

Side Note: In real life it should be noted that there are several sources of uncertainty in the design process. Simply applying all of the uncertainty factors to every path usually leads to an unrealistic design window. Statistics are generally applied to determine how much guard banding should be added. We will keep our example simple, but it is important to note that, due to this practice, the probability that any path has more than the charged uncertainty is low but finite. The net result is that, even after guard banding the design, any given path could be hit by compounding uncertainty factors greater than the guard banding.

Now we are in danger of falling into a common design trap. Let us take a look at two groups of paths. Path (A) falls into the -160ps bucket and path (B) that falls into the -120ps bucket. It is very tempting to spend effort fixing the worst negative slack

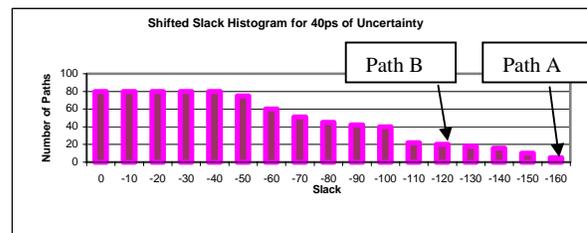


Figure 6.2

(WNS) of path A regardless of its complexity. However, since both paths A and B fall within the uncertainty range, either one could be the worst path in the design.

Assuming that fixing path A is significantly harder than fixing path B, the designer could focus their effort on path B ignoring A and still improve the frequency of the design. Thus, the slack graph can be very misleading when used to direct effort.

We propose that in the early stages of a project a modified slack graph should be used when directing resources. All paths that have a slack delta of less than the uncertainty in the design should be grouped into the same bucket. The modified slack graph is shown in figure 6.3.

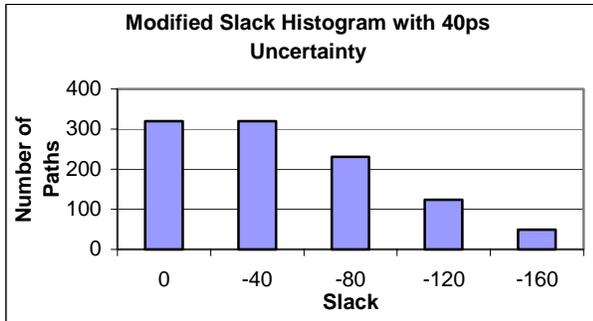


Figure 6.3

When viewed this way, the tail of the slack wall looks more like a wall; our WNS bucket went from having 5 paths to over 30 paths. This more accurately depicts the number of paths that need to be addressed to be certain of frequency improvements. Until all of the paths in the uncertainty window or WNS bucket are addressed, we can not be certain of actual frequency improvements. Conversely, any path that does not fall into the WNS bucket can not possibly effect the maximum frequency of the design and can be safely ignored (at least until paths in the WNS bucket are fixed).

Assuming we don't have the resources to fix all 31 paths within the WNS bucket, our frequency will be limited to -160ps of our target. Since we have no idea which of the path(s) in the first bucket is limiting the frequency of the design, any paths we fix may or may not improve the actual frequency of the part regardless of what our simulations are telling us. Therefore, if we are not fixing all of the paths in the WNS bucket, any effort we spend may or may not improve the design. Effort fixing any given path may be completely wasted fixing a problem that never existed. The larger the uncertainty, or guard banding, the higher the probability the design team is working on the wrong paths and effort is being wasted.

As the amount of uncertainty increases the number of paths in the WNS bucket increase. Thus, the effort to improve the frequency of the design also increases. The more accurately we can predict the critical path by reducing the uncertainty, the higher the return on investment(ROI) is of using more complex circuit topologies or methodologies to improve the frequency of the design. Designs with a large amount of uncertainty will have a much lower ROI for using these same techniques. In these designs, we believe that the more complex circuit topologies or methodologies don't get applied, resulting in a frequency gap between the ASIC and custom designs.

Let us examine this principle in more detail. It is typical for cell based static timing tools to propagate the worst case slope when timing through a path. This injects a large amount of uncertainty into the design process. For example, if a two input AND gate has an early arriving signal with a poor slope and a late arriving signal with a fast slope, many popular timing tools will combine the worst slope with the worst case timing. This is, of course, pessimistic.

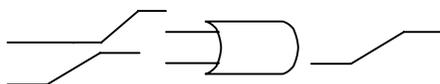


Figure 6.4

At first glance this does not appear to be a problem because the actual frequency will be higher than the targeted frequency. However, this violates our hypothesis that any practice that results in a gap between that actual and targeted frequency is bad. To illustrate this further let us look at the simplified design shown in Figure 6.5.

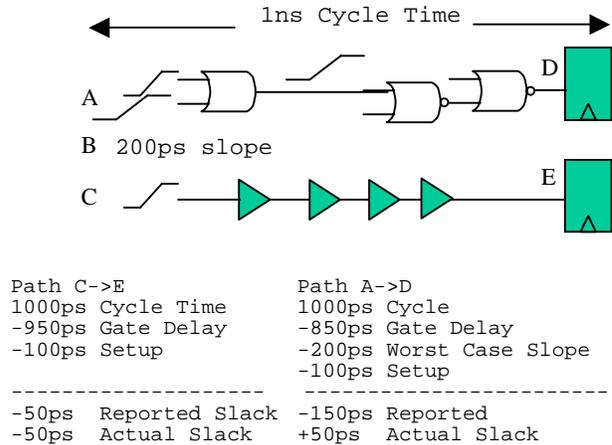


Figure 6.5

When timing the path from A->D, the typical timing tool will combine the worst timing from path A->D with the worst slope from B->D. As a result, the timing tools will show that the design is missing the timing targets by -150ps. The natural course is to spend resources fixing this path. In actuality, the path will run with +50ps of slack and any resources applied to it are simply wasted.

To make matters worse, the real critical path C->E largely goes unnoticed since it is not reported as the WNS path. If the designer focuses on path C->E any improvement made, up to 50ps, results in a higher frequency design and better bin splits. Another alternative is, of course, not to spend any additional resources and go straight to manufacturing. This is the kind of work that great raises are made out of.

Although we have chosen to use the example of worst case slope propagation, the principle holds true for all sources of uncertainty. The inevitable result of uncertainty is a gap in the simulated vs. actual frequency which results in the misdirection of resources that could otherwise be working on improving the time-to-market or ultimate frequency of the design.

Summary: Any uncertainty in the design process will cause a gap between the simulated and actual frequencies. This will cause the design team to work on the wrong paths increasing the effort required to obtain higher frequencies.

7.1 Revised Hypothesis:

The closer the correlation between the simulated frequency and the actual frequency for the paths in the WNS bucket the higher the actual frequency will be.

7.2 The Key is to Reduce Uncertainty

In order to correctly apply our resources, the uncertainty in the modified slack graph must be reduced. Figure 6.6 shows that, as the uncertainty window is decreased, the modified slack graph starts to look more and more like the actual slack graph proposed

in figure 6.1. The number of paths in the WNS bucket has now been decreased to 5 paths again. Any work applied to these paths will be more likely to increase the frequency of the design. The smaller the uncertainty, the more likely the effort we apply will affect the actual frequency of the part.

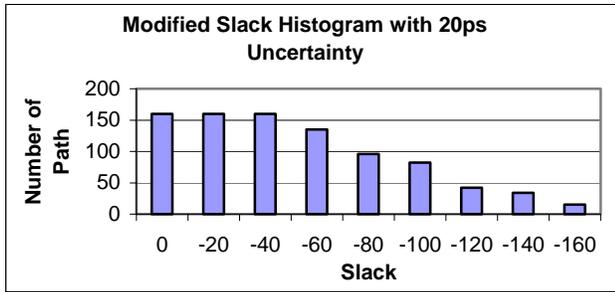


Figure 6.6

If the uncertainty is aggressively managed, the ROI of using local more costly methodologies such as dynamic circuitry or highly accurate clock analysis on just the paths in the WNS bucket is significantly increased. Reducing the uncertainty increases the probability of frequency returns for effort spent. Thus, early in the project, the design team should aggressively look for ways to reduce the level of uncertainty in the design in order to reduce the effort required to fix the critical paths during the later phases of the project.

Summary: Reducing the uncertainty will minimize the number of paths in the WNS bucket and thus the effort required to address them.

8 METHODS FOR REMOVING PATHS FROM THE UNCERTAINTY WINDOW

We have established that decreasing the uncertainty, i.e. moving paths out of the uncertainty window, is key to increasing frequency. There are three ways to remove paths from the uncertainty window:

1. Synthesis should use a total negative slack (TNS) cost function instead of WNS cost function. TNS is defined as the sum of negative slacks over the entire design. This will force the synthesis tool to optimize each path to as close to the target frequency as possible and will avoid a grouping of paths in and around the WNS path and thus in the uncertainty window. This is critical and can usually be done by simply toggling a few switches in most compilers.
2. The next approach is to over-constrain the design so that the synthesis tool or designer will push as many paths out of the WNS uncertainty window. Special care should be taken when over-constraining. If the over-constraining is not done consistently throughout the entire die and design process, it will actually introduce uncertainty into the design. It is much safer to simply adjust the target frequency rather than over-constraining. In order to properly constrain the design, the following should be followed:

- (1) Set the width of the WNS bucket to the uncertainty value.
- (2) All paths that fall within the WNS bucket should have negative slacks in order to over-constrain these paths.

(3) All paths that are not in the WNS bucket should have positive slacks in order to not over-design these paths.

Failure to follow these guidelines will result in over-design of the non-critical paths. This, in turn, will increase the design's area and power and may even adversely affect the timing of the critical path. Conversely, constraining any less than this will leave extra frequency on the table.

It is very important to note that the uncertainty window (i.e. WNS bucket) should *never* overlap the zero slack mark. If it does, the synthesis tools will group a large number of paths right at the zero slack mark which is in the WNS bucket. This increases the probability that a non-critical path in the design will become the actual frequency limiter. This happens when a path that could be easily sped up is hit with enough design/process variance that it becomes critical. The greater the number of paths in the uncertainty window, the greater the odds that one of them will be hit with compounding uncertainties. Thus, if the uncertainty window overlaps zero, the target frequency should be increased until all of the paths in the window have a negative slack.

Note: We are not suggesting to change the market frequency but rather the target frequency that the designers and synthesis tools use when making tradeoffs.

3. The third method for removing paths out of the uncertainty window is to simply reduce the uncertainty window. This is the subject of our final section.

9 THE UNCERTAINTY LIFECYCLE

When developing a plan to address the sources of uncertainty within a project, it is important to understand the basic lifecycle. Figure 9.1 shows the uncertainty life cycle. Let us examine it in the context of clock skew.

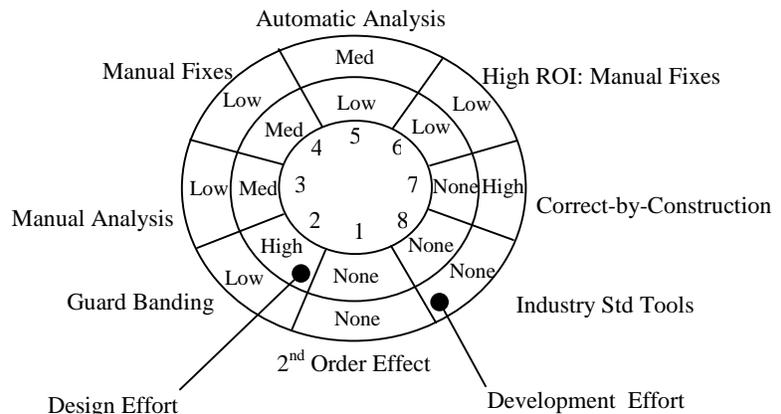


Figure 9.1

9.1 Stages of the Uncertainty Life Cycle

Stage-1, Second Order Effect: During Stage-1 of the life cycle, let us assume that the clock skew is less than 1% of the total cycle time. At this point in time, the clock skew can be safely ignored. Effort during this stage is zero.

Stage-2, Guard Banding: As the frequencies of the designs increase over time, we shift into Stage-2. Here the effects of clock skew can no longer be ignored so guard banding is used to

account for the clock skew uncertainty. The development effort is low, as most tools support guard banding in some form, but the design effort is very high. Every path not requiring guard banding results in effort wasted since guard banding is simply safeguarding for uncertainty and does nothing to reduce it.

Stage-3, Manual Analysis: As the frequencies continue to increase, it becomes critical to more accurately analyze the clock tree - at least locally for the critical paths. During this stage, it would be typical to override the clock skew guard banding on the critical paths after careful manual analysis (accurate device level analog simulation). This has the effect of reducing the uncertainty on the critical paths. The development effort is low; however, the design effort of performing the manual analysis on the critical paths is med-high. A common mistake is to assume that manual analysis is simply additional design effort. As previously noted, accurately analyzing the path can remove it from the WNS bucket just as effectively as fixing the path. Breaking through the global guard banding methodologies of stage-2 is difficult and can become a major barrier to many design teams. It is, however, critical to reducing the uncertainty in the design and, therefore, the frequency of the design. It is our belief that this is where many design teams get stuck due to limited resources.

Stage-4, Manual Fixes: Now that we can accurately analyze the clock skew on local portions of the design, the number of paths in the WNS bucket has been reduced. We can now apply manual fixes to the clock tree for the critical paths. For example, for the WNS paths, we may hand route the clock tree. It should be noted that, since we are applying the analysis and fixes locally, we have not reduced the uncertainty across the entire design. Thus, we may not be working on the actual critical paths. To avoid this, it is important to make sure that the local/manual analysis and fixes cover all of the paths in the uncertainty window. The development effort is, again, low and the design effort is med-high.

Stage-5, Automatic Analysis: As the frequencies of our designs increase, the number of paths in the WNS slack bucket will increase to the point where manual analysis is too costly. During this stage, we are forced to develop CAD tools and methodologies that reduce the uncertainty of the analysis of the entire clock tree. Reducing the uncertainty will reduce the number of paths in the WNS slack bucket, allowing us to continue using the local manual optimizations.

For example, the clock skew calculations may now take into account the point of divergence of the sampling and generating clocks instead of one simple guard banding value. This analysis would be done globally to reduce the uncertainty and thus reduce the number of paths in the WNS slack bucket. The development can be high depending on the skill set within the team. The design effort starts to decrease as the automatic global analysis reduces the uncertainty and accordingly the number of paths in the WNS bucket.

Stage-6, High ROI Manual Fixes: Now that the number of paths in the WNS bucket has been reduced, we can resume local fixes. The reduced uncertainty has improved the ROI as we are now more confident that we are working on the critical paths. The development and design effort at this stage are med-low.

Stage-7, Correct-by-construction: At this point the design team is using clock tree CAD tools that can automatically hit their skew targets. These tools are often developed within the company or

co-developed with an external EDA firm. The tool development effort is now extremely high but the design effort has fallen dramatically. It is important to note that correct-by-construction can be achieved by tool enhancements like better clock tree algorithms or via design methodology, such as using a latch-based design. Methodology solutions generally have lower development costs.

Stage-8, Industry Standard Tool: The ultimate goal is to have the clock tree algorithms which are capable of hitting the skew targets become part of the industry standard tools. This frees up resources to work on reducing the clock uncertainty even further or work on the next source of uncertainty. As the frequencies increase, the uncertainty introduced by tools, even those which are standard in the industry, will be too much and the uncertainty lifecycle will continue.

9.2 Developing an Uncertainty Plan

Now that we have a firm understanding of the uncertainty lifecycle and the effort profiles for each stage, we can develop a plan to address each source of uncertainty in our design. To do so, we must first list all of the sources of uncertainty in our design. Then, based on the availability and skill set of the design team, address each source of uncertainty by assigning it to a stage in the lifecycle. The key is to minimize the amount of uncertainty in the design with a plan that optimizes the combination of development vs design effort. Accurate analysis is key, even if the methodology does not apply fixes. Moving paths out of the guard banding stage will reduce the number of paths in the WNS bucket and the total design effort. Using these guidelines, it is possible for the design team to develop a plan for minimizing uncertainty in the design process based on the available resources. The following is a simple example of an uncertainty plan:

Source of Uncertainty	Value	Course of Action
Clock Skew	40ps	90% Guard Band Manual analysis and fix for WNS bucket
Worst Case Slope Propagation	400ps	No Guard banding Move to Depth First Search Timing Tool i.e. correct-by-construction
Fringe Effect	10ps	100% Guard Band
Inductance Extraction	1ps	Leave as 2 nd order effect

Uncertainty Plan Table 9.1

10 UNCERTAINTY AND THE PROJECT LIFE CYCLE

Deciding where to spend effort is an art form that is project as well as team specific. In this section, we will propose some simple guiding principles to help determine where effort may be best spent throughout the different phases of the project.

Beginning of the Project: In the early project development stage, emphasis should be placed on reducing the sources of uncertainty. As we have stated, reducing the uncertainty window is equivalent to removing the path from the uncertainty window through design effort. It should also be noted that manual analysis and/or fixing should not be attempted until the design has stabilized between

iterations. If the ordering of the paths is not consistent, any effort spent on manual optimizations would be wasted.

Middle of the Project: During this phase of the project, the methodology should be stabilized. Focus should shift to adjusting the target frequency as described in section 8. This will force the synthesis tool/design team to remove as many paths out of the uncertainty window as possible while avoiding over-design. Manual optimizations can start on the critical paths within the uncertainty window as soon as the design starts to stabilize between iterations.

End of the Project: During the final stage of the project, the methodology and target frequencies need to be fully stabilized. Any change in the methodology or target frequency at this point can cause a significant amount of re-work. The design team should instead focus their effort on all of the paths within one sigma of the simulated WNS path. The design team should treat all of the paths within this range as equal. Effort should be spent fixing not just the worst path within this range but *all* paths within this range. Each path fixed increases the probability of a higher frequency design.

11 CONCLUSION

We believe that the gap between ASIC and custom design is partly caused by the ability, or inability, of the methodologies to control the level of design uncertainty. We have proposed that the closer the correlation between the simulated frequency and the actual frequency for the paths in the WNS bucket, the higher the actual frequency will be. Any gap between the two frequencies implies uncertainty in the manufacturing and design process. This will cause the design team to work on the wrong paths, increasing the effort required to obtain higher frequencies.

12 SOURCES OF UNCERTAINTY (subset)

Categories of Uncertainty	Sources of Uncertainty	Details	Type of Uncertainty
In-Die Variance	Transistor geometry variations	Transistor geometry variation across the die causes uncertainty in transistor speed. Particularly important for min-delay analysis.	Process
	Temperature Variations	Hot spots cause parts of the die to run slower than expected effecting max-delay. Cold spots cause min-delay issues.	Design
	Power Delivery	IR drop across power distribution leads to slower device switching.	Design
Timing Analysis	Slope propagation	Worst Case Slope propagated by timing tools causes pessimistic timing. See example in figure 6.5.	Analysis
	Cell Based Characterization	Timing tool forced to interpolate tables causing error. Not as accurate as transistor level simulation.	Analysis
	Timing Arc Characterization	Multiple switching effects can not be accurately modeled with single input timing arcs.	Analysis
	Distributed vs. Lumped RC networks	Timing analysis inaccuracy introduced by using reduced distributed RC networks (or worse yet lumped RC).	Analysis
	Miller Effect	X-Cap should be increased for neighboring signal switching.	Analysis
Clock Analysis	Edge Skew	Differing RC delays between sampling devices.	Analysis
	Edge Jitter	PLL locking shifts between cycles.	Analysis
	Point of Divergence	Better than worst case guard banding, but still not as accurate as full analysis.	Process and Design
Extraction	Resistance	Reduction of network causing a large number of small resistors to be dropped.	Analysis
	Capacitance Fringe effect	Large signal spacing leads to increased orthogonal routing x-cap.	Analysis
	Inductance	Not accounted for in most tools/methodologies.	Analysis
Noise	Timing Push Out	Noise on sensitive node can cause setup failures.	Analysis

Reducing the uncertainty in the analysis of the design also opens the way to more costly local optimization techniques such as dynamic circuitry, datapath placement, and so on. The smaller the uncertainty in the design, the higher the return for these local techniques.

To control the uncertainty in the design process take the following steps:

- (1) List all sources of uncertainty in the design.
- (2) Develop a plan to reduce the uncertainty as much as possible using the modified slack graph and the process described in section 9.
- (3) Reduce as much guard banding as possible {remember: Having an actual frequency higher than the simulated frequency is still bad}.
- (4) Start using a TNS-based cost model.
- (5) Tune the target frequency as described in section 8.
- (6) Toward the end of the design, treat all paths within one sigma of the design as equal.
- (7) Push your CAD tool vendors to develop algorithms that reduce the design uncertainty in order to free up your resources.
- (8) Finally, view all of your methodologies decisions in terms of their effects on the gap between the simulated and actual frequencies: Any gap either positive or negative is costing you frequency.

Following this simple approach will help to improve the design speed of both custom and ASIC designs. We believe that, more often than not, the custom designs are already informally following this approach. Thus, any effort ASIC design teams can afford to spend on reducing the level of uncertainty in their designs will help to reduce the frequency gap now seen between ASIC and custom designs.