UNIVERSITY OF CALIFORNIA
Santa Barbara

Realization and Formal Analysis of Asynchronous Pulse Communication Circuits

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

by

Merritt Philip Miller

Committee in Charge:

Professor Forrest Brewer, Chair

Professor Luke Thogarajan

Professor Li-C. Wang

Professor Tevfik Bultan

March 2015

The dissertation of
Merritt Philip Miller is approved:

Professor Luke Thogarajan

Professor Li-C. Wang

Professor Tevfik Bultan

Professor Forrest Brewer, Committee Chairperson

December 2014

Realization and Formal Analysis of Asynchronous Pulse Communication Circuits

by

Merritt Philip Miller

Dedicated to the person crazy enough to follow in my footsteps

Acknowledgements

There are many people I wish to thank, without whom my research and this dissertation would not be possible. While an exhaustive list would be prohibitive I would like to thank a number of people who were instrumental in this work and me completing my degree. Forrest Brewer, as an advisor and mentor, who's guidance and input made possible so many developments. Luke Thogarajan, Li-C. Wang, Tevfik Bultan, for serving as my committee and all of the useful advice. Joseph Incandella and Guido Magazzu for the research opportunities and introduction to the world of electronics for high-energy physics. I would like to thank everyone I have worked with in the lab, Greg, Nittin, Amitabh, Kunal, Wei, Joseph, Ethan, Alec, Dan, Di, and Carrie. Most importantly I owe an amazing debt of gratitude to Mom, Dad, Sab and Whitney; With out your love and support it would not be possible to finish this life-consuming endeavor.

I would like to thank the Department of Energy for their generous funding of the grant DOE–CERN CMS SLHC "Protocols and IP-Cores for Control and Readout in Future Higher Energy Physics Experiments" Which provided needed support for much of the presented work.

Curriculum Vitæ

Merritt Philip Miller

Education

2015	Ph.D. in Electrical and Computer Engineering (Expected), University of California, Santa Barbara.
2009	M.S. in Electrical and Computer Engineering, University of California, Santa Barbara.
2007	B.S. in Electrical Engineering, University of California, Santa Barbara.

Publications

Miller, M.; Hoover, G.; Brewer, F., "Pulse-mode link for robust, high speed communications," *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on* , vol., no., pp. 3073-3077, 18-21 May 2008
doi: 10.1109/ISCAS.2008.4542107

Miller, Merritt; Brewer, Forrest, "Formal verification of analog circuit parameters across variation utilizing SAT," *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013* , vol., no., pp. 1442-1447, 18-22 March 2013
doi: 10.7873/DATE.2013.294

Miller, M.; Brewer, F.; Magazzu, G.; Wang, D., "Multi-gigabit low-power radiation-tolerant data links and improved data motion in trackers" *Journal of Instrumentation* vol. 9, no. 12, pp. C12011 December 2014
doi: 10.1088/1748-0221/9/12/C12011

Abstract

Realization and Formal Analysis of Asynchronous Pulse Communication Circuits

Merritt Philip Miller

This work presents an approach to constructing asynchronous pulsed communication circuits. These circuits use small delay elements to introduce a gate level sense of time, removing the need for either a clock or handshaking signal to be part of a high-speed communication link. This construction method allows the creation of links with better than normal jitter tolerance, allowing for simple circuit architectures that can easily be made robust to radiation induced soft error.

A $5Gbps$ radiation-hardened link, targeted at use in detector modules at the LHC, will be presented. This application presents a special challenge due to both very high radiation levels ($1 + MGy$ life time dose) and the demand for minimum resource (area, power, cable cost) use. The presented link, realized in 130nm technology, is unique in that it has low power ($\tilde{}50mW$ end to end) and very low area $0.12mm^2$ including electrostatic discharge protection, and I/O amplifiers. Due to its asynchronous construction and the gate design style, the link has essentially zero power dissipation when idle, and enters and exits its idle state with no delay.

In addition to the construction of the link, this presentation covers the design and analysis methodology that can be used to create other asynchronous communication circuits. The methodology achieves higher performance than conventional static technology but needs only a reasonable design effort using tools and strategies that are only mildly extended versions of those familiar to digital static designers. It is used to construct the serializer, deserializer, and

self-test circuitry for the presented link. In this case, a 5Gbps SER/DES and a 2GHz parallel pseudo-random number generator are implemented in 130nm CMOS technology using a gate design style that does not dissipate static power.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation presents a method of describing, constructing and analyzing fast, practical, circuits for communication systems. The motivating application of the developed techniques and circuits is serial communication streams used in the instrumentation of high-energy physics detectors – one of the highest radiation environments where micro-electronics are operated. Radiation environments are a unique application due to radiation induced soft-error. Large physics experiments are additionally a unique application due to a high demand for data. Data links in high energy physics must also consume as little power and chip area as is practical to ease integration. Asynchronous pulse mode circuits and the formal design style demonstrated herein allow efficient construction of a high performance, low power, communication scheme tailored for such extreme environments.

Pulse-mode asynchronous circuits operate without a clock, instead using brief pulses to mark time within a construction. A brief pulse is a useful marker of an event, since there is little uncertainty in the time it marks; this uncertainty is bounded by its duration. This limited ambiguity in pulse arrival can be used to create circuits where uncertainty in time is one of the

major limiting factors for system performance. The nominal pulse width of gates within a pulse circuit becomes a quanta of time upon which the circuit operates. This version of asynchronous pulse circuits does not provide delay insensitive operation, instead, sections of circuitry operate without stage-to-stage feedback in feed-forward mode. The lack of feedback enables potentially faster operations for communications circuits, but comes with two additional costs: the need for timing verification to confirm correct operation, as well as a higher-level description of a protocol that will prevent faster circuits from outstripping the capacity of slower ones.

Described here is a class of asynchronous pulse circuits, deliberately limited to ease analysis and practical construction complexity. A formal method for describing this class of circuits is presented, including a means of verifying correct latch function via a series of timing checks. A method of examining gate behavior by translating the problem into one of Boolean satisfiability (SAT), and thereby avoiding exhaustive simulation is presented. This is followed by an analysis of an example cell set for asynchronous pulse logic. Finally an example of a 5Gbps serial link that utilizes the asynchronous pulse technology and is specifically hardened to radiation will be described.

Figure 1.1 is an asynchronous pulse stream. The pulses in this figure mark the occurrence of an event, This dissertation will present a means of generating systems that can use such pulse streams as a core means of operating.

1.1 Asynchronous Pulse Logic

There are many styles of circuit design that utilize pulses. Some of the most famous of these are SRCMOS[1, 2], a clocked logic family, and the asP*[3], GasP[4], and single track

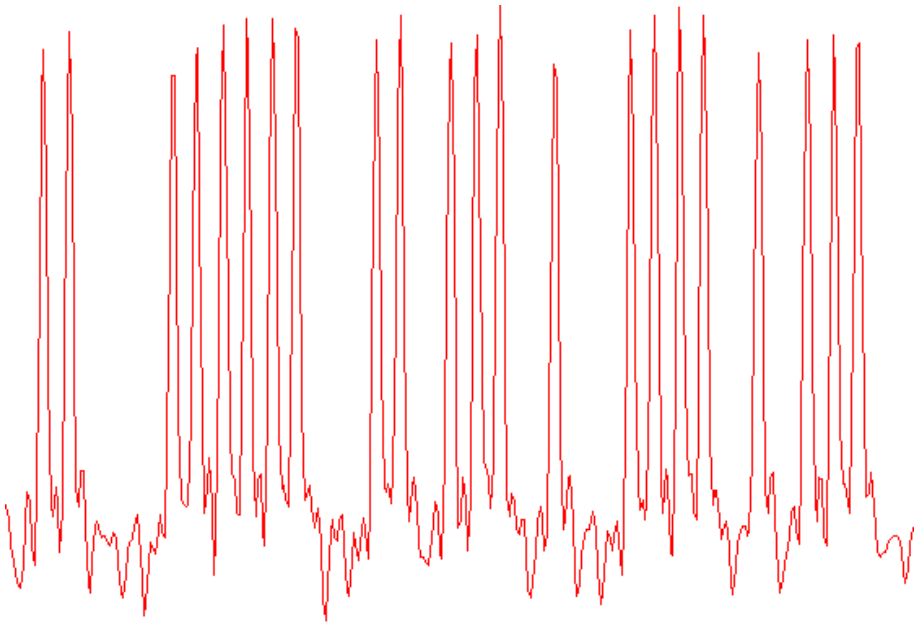


Figure 1.1: An asynchronous pulse stream. Pulses such as these drive asynchronous pulse gates and are the markers of time in pulse logic. The pulses are of uniform shape, yet the spacing (relative timing) of pulses can vary as long as there is a minimum space.

asynchronous[5] styles of asynchronous circuitry. The prime unique feature of the design style discussed here is that it is both asynchronous and feed forward, that is it neither uses a clock to create timing within the construction, nor does it use a handshake between data source and data sink to regulate the motion of data. This work is also distinct from the feed-forward method of surfing[6, 7], which uses a timing signal much like a phased clock. This paradigm is unlike hand-shake based construction in that timing analysis and margining is used to set the rate at which the system operates. This means that the time for hand-shake times can be removed from the system specification, a small gain for short-range communications, but more important for the long-range case. The removal of a clock from the construction, on the other hand, allows savings in power and area, as there is limited need for feed-back controlled timing recovery.

This work covers a technology for producing high event-rate digital circuits. This technology uses pulses to indicate events which advance the state of the realized machine. By constraining state advance to the existence of a pulse in the correct circumstances, the technology allows for the construction of circuits without a clock or hand-shake required for operation. This allows the designer the freedom to express systems that are delay insensitive and use a hand-shake, and systems that are synchronized by a single timing source using similar circuits and terminology.

The power of this technology is demonstrated by implementing a self-testing high-rate data link. This data link is specifically designed for use in a very high radiation environment, where low-jitter, low-drift timing information comes at a high cost in terms of power and area. This alternate paradigm of pulsed based signaling places timing information explicitly, rather than implicitly, in the data stream – allowing for simple circuits to perform the duties of clock and data recovery from the serial data stream.

The link is constructed to reduce the presence of high-speed clocks within the system, another source of power dissipation that is difficult to harden against radiation-related upset. Data generators and consumers in a system rarely operate one bit at a time, and the circuit interfaces of LHC[1] front-end chips, one of the target applications, are no exception. The link takes advantage of the fact that it only needs to maintain timing accuracy with a slower, parallel bus. This takes the form of small bursts of operation where timing is derived from an uncalibrated delay-line, and then intermittent small pockets of delay as the link re-synchronizes to the slower system clock. In this way the link can operate its internal interface with a 625MHz clock while providing a 5Gbps serial stream, and since there is no continuous tuning operation running (for a calibrated delay line) or high frequency clock the overall system has numerous savings in power and complexity.

This work will present the technology, a useful means of representing events as pulses and constructing a working system. The description covers the definition of a pulse, pulse detection, pulse circuit construction, formalism in expressing a correct pulse system, considerations in physical realization of a pulse system, and an example demonstration of a system that uses both pulses and a clock to construct a 5Gbps serial link and a random number source that implements a 17bit LFSR fast enough to feed the data link. Additionally a discussion of the considerations of radiation damage in such a circuit will be addressed, and the implementation of radiation hard by design in the technology will be demonstrated

[1]The (LHC) Large Hadron Collider is currently (2014) the highest energy, and one of the highest luminosity particle accelerator on Earth - and is the design environment of the described link.

1.2 Philosophy

The core goal of this work is to ease the process of correct-by-construction design of practical high-speed circuits. While somewhat nebulous, the term practical high-speed circuit is taken to be a circuit where data rate is of key concern, but trade-offs must be made to meet other system requirements and goals. In the case of the radiation data link, the other goals sought are reductions in power, layout area, and designer effort. Power and area concerns are met by banishing high-speed clocks from the design style, using asynchronous circuits for all high rate interactions. A high-rate high-accuracy clock is deemed too costly for low power and low area efforts.

The effort exerted to specify, construct, and verify a circuit is not frequently touted along with circuits with superlative performance, but is a crucial aspect of design. Simply put, faster design processes allow more options to be considered along the way, and have the chance to provide early feed-back as to the feasibility of implementing new concepts. To this end, formal specification and ease of checking are directly sought as an enabling aspect to the technology.

High radiation environments are studied as an application because they present a unique set of design trade-offs. Device degradation in the face of radiation induced damage is logical concern, with a test-based solution: do not use devices that cannot survive the environment, characterize the parametric drift of devices that are used. Characterization can handle parametric drift, but due to high-energy particle strikes, microelectronic circuits also experience an elevated level of soft-errors; errors in function that can be cleared or corrected. The philosophical approach here is to distribute error detection and correction functionality throughout the circuitry used. Recovery time for a circuit becomes the metric for hardness in this approach,

as during the time period of error, the circuit has lost some measure of correction and is more sensitive to the next error.

1.3 Radiation Hard By Design

Microelectronics are sensitive to ionizing radiation as well as high-energy ions. Radiation effects impact integrated circuits in two basic ways: it can change the properties of the materials that the circuit is made of, as well as depositing charge within the circuit, directly impacting circuit function. This damage manifests it self as both chronic degradation of devices used in circuitry, as well as two acute effects: soft-errors, and hard-errors or device destruction.

Two approaches are used to approach the manufacture of electronics for radiation environments: selection of components for ones that have demonstrated tolerance of radiation environments, and to design systems that have appropriate behavior in the face of radiation even if device failure is a possibility. The whole-system approach involves a designer taking into account radiation effects and adds radiation hardness to the specifications that steer engineering decisions. To achieve a working system considering radiation damage, the devices used may have property drift and intermittent failure in a radiation environment, but if the underlying devices are likely to fail completely, there is little a designer can do to make a working system. Any approach to radiation tolerance must include some work at the device level, but a good system design can extend devices that would other-wise be marginal to be successful components of a fully-functioning system.

The technology described herein was driven by a need for creating data links for sensors at The Large Hadron Collider. The sensors at the LHC produce an exceedingly large amount of

data, and with each upgrade cycle, more and more data becomes part of the design. Sensors in high energy physics are exposed to high radiation levels by design – a sensor that saw no radiation saw nothing in the experiment. Portions of the LHC reach radiation levels that exceed, by many fold, the expected doses seen in space service equipment – a common radiation environment target for micro-electronics. For comparison, many low earth orbit certified components are rated for total dose of 3kGy, while lifetime dose for the highest exposure levels of the LHC are expected to exceed 1MGy - a factor of 300 more lifetime absorbed energy, over a much shorter lifetime.

1.3.1 Options not explored

The application assumes no shielding and no special device-level processing.

Shielding is not a real option for electronics involved in the LHC for two reasons: most important the goal of detecting particles, which would be negatively impacted by shielding, and the second is that at the energy levels of particles of the LHC shielding would be counter-productive. There are two components to this lack of efficacy. First, the very high particle energy means that stopping one particle causes a shower of lower energy particles making the total radiation levels higher. The radiation due to stopping a particle is known as bremsstrahlung. Shielding that would protect against both primary and secondary particles would be impractically thick requiring meters of lead in some cases. The second reason that high energy particles preclude shielding is the notion that particles traveling quickly are unlikely to interact with the material that they are passing through – a concept similar and related to tunneling. There is a peak momentum for absorbed energy for any given particle type – the Bragg peak. Effective

shielding has to reduce the momentum of the particle below this peak, otherwise it is merely increasing the amount of energy deposited in the protected circuit.

There exist microelectronics processes are specially designed for high-radiation environments, yet this work does not require or assume the existence of any special process or device. The rationale behind this decision is multi-faceted. The reasons behind selecting a process tend to center on price and availability. More specialized processes tend to cost more and have the information less-widely distributed. Additionally processes that are less popular have lower guarantee of sustained availability. Making the fewest assumptions about the underlying technology gives a design the best chance to last the test of time and to be broadly used in a collaborative effort, such as in the large experiments at the cutting edge of high energy physics research. The utility of common CMOS processes at the current radiation doses allows their use and hence the benefits in savings and general applicability.

1.3.2 Radiation environment issues in sub-micron CMOS

This work was developed with common planar silicon CMOS processes in the 65nm-250nm process node range. Many of the radiation effects in these processes will be present in other integrated circuit process nodes, and many of the techniques developed to mitigate them will carry forward. For processes in these nodes there is a trend to increased tolerance of total dose in processes with smaller feature size. Starting around the 45nm process node the physical structure of common silicon processes is altered.At the $45nm$ process node, Intel announced the use of high-k gate material[8], using the heavier-than-silicon metal hafnium in the oxide. Hafnium has a neutron cross section of roughly 100 barns[9], while silicon is only $\sim .7$ barns[10]. While the same issues should appear, yet there is the possibility that the dominant damage

mechanisms will change, as is evidenced by the hardness of sub-micron FETs to charge trapping in their gate oxide, a disqualifying issue for higher-voltage FETs with thicker oxide layers.

Current implementations of sub-mircon CMOS processes have a number of similar properties, and, as a whole, are reasonable candidates for radiation hardened electronics. CMOS processes give very high performance field effect transistors, and these processes, highly studied materials are commonly used to construct these field effect transistors. Every device will suffer damage in the face of enough radiation; in the case of field effect transistors the insulating layers are considered the most sensitive to damage. The high dopants densities and thin insulator thicknesses in these processes lead to an attainable level of radiation tolerance. In silicon CMOS processes in the 65nm-250nm scale range SiO_2 is the most common insulator. It is known that silicon dioxide at a thickness of 5 or less it is less susceptible to damage on a per-volume basis[11]. SiO_2 Thicknesses in these process nodes is available at thicknesses of 5 or less.

1.3.2.1 Approach To hardening

Radiation environments on the scale being discussed here are harsh on microelectronic devices. The individual components of electronic circuits being constructed are subject to both life-time damage and temporary malfunction. This breaks the mitigation strategy into two fronts: aging effects (changes in behavior on a time-scale much longer than that of the circuit) and soft errors (errant behavior that is on the time-scale of the circuitry involved).

Long term damage sensitivity removes some devices from the realm of practicality, and leads others to require circuits tolerant of their property changes. An easy way to remove the physics from the design equation is to build test structures of all desired devices, and characterize these test structures in the face of radiation. These radiation tests are similar to accelerated

aging tests and are subject to an art and science beyond the scope of this work. The resulting information gathered will be information on lifetime degradation and damage models. Devices that see experience too wide a variation are not used; while devices with some induced damage are used in circuits that have broader tolerance for variation. More adaptive or tolerant circuit designs will always fare better in a radiation environment with regards to total dose since lifetime damage is a form of parametric shift – the more shift tolerated the more damage tolerated.

1.3.2.2 High-power acute issues

There are two classes of acute radiation damage issues: upset and latch-up/burn-out. Latch-up and burn-out are both severe, destructive issues in acute radiation damage. In both cases a particle that strikes the circuit causes an excess of charge in a sensitive region of the struck device. The device is set into a positive feed-back mode where more, rather than less, charge flows through the device; causing a device failure. Latch-up is a potentially destructive condition where an electrical feedback mechanism is triggered. The feed-back path is consists of a parasitic SCR within an integrated circuit[12] that is triggered in to its on condition. In theory latch-up, if caught early, can be stopped prior to device destruction (by powering the device off). Burn-out refers to a similar positive feed-back mechanism, where the critical difference is that the positive feed-back loop includes heating of the device. A radiation strike creates a highly conductive path, which quickly heats, and in the right circumstances, this heating frees carriers making the path more conductive. This means that the mechanism can involve more structures than the SCR, any material that can be coaxed into conduction, has a positive thermal feedback, and is normally operated with a field across it, is potentially vulnerable. When the burn-out

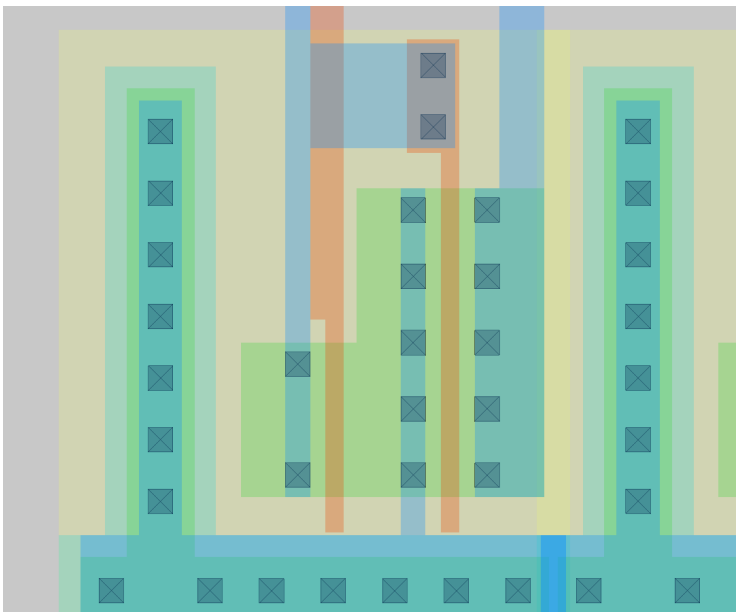


Figure 1.2: Layout of a cell with additional contact regions. Full contact rails surround logic on three sides to help reduce the risk of latch-up.

event includes puncturing the gate oxide of a FET this becomes known as single event gate rupture, a problem most associated with high-power devices[13].

Single event burnout issues are much more common with high-voltage devices[14], and are not expected to be a dominant factor in the systems studied. This leaves latch-up as a key concern for the system. Latch-up issues, being primarily electrical, can be combated by increasing charge collection centers in the layout[15]. To achive this goal dramatically more aggressive a more aggressive than usual substrate and well contact scheme was developed. Figure 1.2 shows a cell with this extra contact layout.

1.3.2.3 Low-power acute issues

Radiation induced charge deposition, so called single-event upset (SEU), is the remaining acute effect of importance. SEU is caused when charge due to a radiation strike merely corrupts information within the integrated circuit. SEU is the largest challenge of deep submicron radiation hard design[16], and the key motivator of the technology developed in this research work. The corruption of information can have a number of impacts from errors in computation to errors in timing and errors in continuous voltages created by bias circuitry. Much of this study will be dedicated to dealing with single event upset.

1.3.2.4 Chronic Issues

There are three key chronic issues due to radiation, both relating to the slow degradation of the materials that make up an integrated circuit. These are total irriadated dose(TID), mostly moderated by photons, lattice damage, mostly moderated by protons and neutrons, and doping inversion, also moderated by heavy particles.

Both TID and doping inversion are limited concerns in deep sub-micron technologies. The primary impact of TID is charge trapping in oxide layers. As already mentioned, the thin oxides common in deep sub-micron technologies have a very low limit to the amount of charge that can be trapped[11]. Because of this hardness, the behavior of thin oxide FETs have been a device of interest and their behavior is well studied [17, 18, 19]. In Effect, wide-channel FETs are radiation hard and suitable for use in the 130nm process[20]. Doping inversion is a damage mechanism where impinging particles change the chemical make-up of the materials in a given silicon wafer. This is a real concern for electronics in the LHC since many such particles are expected to strike electronics within the detector. The high doping levels common to small-scale

planar CMOS technologies offer a good amount of protection against complete type inversion; the doping levels are typically very high in the $10^{17}/cm^3$ range and greater, while the number of 1MeV n^0 equivalents is limited to $10^{15}/cm^2$ over the service lifetime. The final issue at hand is lattice damage, cause by collisions between heavy particles and the semi-conductor lattice. Lattice damage reduces the conductivity of field effect transistors, and alters the behavior of bulk devices like bipolar junction transistors and diodes – usually for the worse.

Given the expected chronic issues, Field effect transistors realized in a deep sub-micron process are expected to be much more radiation tolerant than the available alternatives. Thus the decision is made to rely chiefly on FETs as the semiconductor device of choice.

1.4 Permissions and Attributions

The contents of chapter 3 was co-authored with Forrest Brewer and was presented at and published in the proceedings of the 2013 Design Automation & Test in Europe in Grenoble, France conference under the title "Formal verification of analog circuit parameters across variation utilizing SAT"[21]. Additionally the content of chapter 5 was originally presented at the 19th Real Time Conference 2014 in Nara, Japan under the title "5GB/s Radiation Hard Low Power Point to Point Serial Link"; a summery version is in the conference record, and the material in full is currently pending in the Transactions in Nuclear Science edition associated with the conference. This material is used under the terms of the IEEE author re-use policy allowing primary author re-use in subsequent work.

Chapter 2

Systematic Asynchronous Design

This Chapter presents a design methodology for creating asynchronous circuits for practical high-performance design; circuits where rate, power, area, and designer effort are all considered for trade-off. The design style works in concert with a logical circuit implementation to enable the description and realization of high-speed asynchronous circuits. The methodology is an adaptation of existing techniques for creating asynchronous, but not delay insensitive circuitry. It is specifically targeted to the production of asynchronous circuits for data link and high-speed communication systems and is an enabling technology for the links developed later in this work.

2.1 Introduction

High-speed link design is a common problem in integrated circuits where interface and communication speeds frequently exceed core clock rates. For example, in the current Intel Xeon E7 lineup no processor has a bus speed less than 6.4GT/s, and no processor has a core

frequency greater than 3.7 GHz[22] and at the opposite end of the power spectrum Atmel offers an 84MHz processor family with USB at 480Mb/s[23]. Complex communication link circuits are found on medium to long wires in integrated circuits, longer PCB traces, and transmission near the dispersion limit for copper cables. These systems involve operating a metal wire near the rate where substantial amplitude loss and symbol timing jitter are serious issues. *Correction of signal amplitude is not particularly difficult, however, symbol-dependent timing errors are far more difficult to mitigate.*

Synchronous circuits have limited capacity to handle timing issues. Specifically, time domain synchronization is difficult, and is commonly relegated to specialized blocks, such as DLLs, PLLs, and skew compensators, whose behaviors are outside of the synchronous domain. Conventional design methodologies create a complex circuit topology where the link correctness is not guaranteed by the timing closure of the synchronous design. For example, valid sampling of a signal by a deserializer is dependent on the behavior of the associated PLL or timing circuit, not a local property of the deserializer. This issue makes the design of such circuits expensive and complex and slows the adoption of link alternatives because of the perception of risk.

Instead, we implement asynchronous logic blocks to handle the slow parallel to fast serial domain interface. Asynchronous systems are inherently tolerant of timing variance. However, an unconstrained asynchronous design would have potential for high design complexity. Instead, constrained composition rules and a variety of pulse-logic gates are chosen that allow a limited set of classical timing constraints to close both the low-speed and high-speed design behaviors. The high-rate system can then be designed in a coherent manner escaping the complexity of forcing a conventional synchronous timing paradigm to accommodate high variance, high speed signals. Unlike many asynchronous systems, the proposed scheme uses feed-forward

construction. Feed-forward logic design does come at a cost: the timing of the system needs to be verified as part of the construction procedure – system timing is not safe by construction. An important part of the methodology are means to limit the complexity of the timing verification.

2.1.1 Related Work

There are a few asynchronous design paradigms meeting the system rate requirements. At such rates, classical feed-back based delay independent techniques are problematic. In the case of physically long transmission media (starting at the mm scale for multi-GHz signals) time of flight for the electromagnetic wave carrying the signals adds substantially to the system delay, reducing performance. On-die scale structures (100μ scale) have substantial propagation delay and at lengths of 1mm has potential signal integrity issues. These considerations defeat design styles based on feed-back such as GasP[4]. Instead, the circuits presented here are a limited sub-class of self-resetting CMOS circuits – a design style that has a reset circuit assigned to small clusters of domino-like logic. SRCMOS circuits work in an inherently pulsed manner. The down side of this is that SRCMOS circuits need pulses to arrive nearly simultaneously for proper function[24]. Work on timing analysis of SRCMOS includes [25, 1]. SRCMOS circuits have seen application in asynchronous circuits in [5, 26] but asynchronous use is commonly restricted to systems that have feed-back to confirm correct behavior.

Pulse reset timing is also noted as reset interference in [5], where the locally reset signal creates a timing constraint in an otherwise timing independent gate.

The notion of using a pulse to synchronize other pulses has been explored in [6] one means of solving the overlap problem of SRCMOS. The implementation comes at the cost of

noise margin due to the conditional lowering of switching thresholds, though conceivably a similar design could be realized that created the necessary timing regime.

2.1.2 This Work

Pulses are used for timing critical communication in a similar fashion to the logic of [24] and [6] . For moving data in situations where the interconnect medium is limiting, especially in terms of jitter, the differences between pulses and edges are minimal, as demonstrated in section 2.2. The logic gates in this work are SRCMOS style gates, with restrictions to simplify analysis. System construction rules partition signals into two classes, to identify signals containing events (e.g. clocks) from others. Static timing verification is then applied relative to system events to confirm correct function. Pulsed signals, with a single characteristic pulse width are used for communicating events in the system. The single pulse-width allows the use of pulse gates, such as in [27, 28, 5, 6] and similar to[4]. These gates are known to maintain a stable pulse width, where logic without feedback would accumulate uncertainty, leading to the eventual decay of short pulses. In the long-range signaling application, high-speed serial streams are assumed to be a pseudo-complimentary pair of wires carrying pulses to mark 1 and 0 bits as events in the bit stream. This encoding creates events for each data bit, simplifying detection and processing of the stream.

2.2 On-Chip Signaling: Pulse VS. Edge

A pulse consists of a both a rising and a falling edge and at first glance it would seem that any information carried by a pulse could be done by an edge, with another edge left to spare. Sending information as a pulse has the advantage that a pulse unambiguously marks the

Table 2.1: Propagation times for a slow edge (100ps rise), and a 165ps pulse.

	Edge		165ps Pulse	
Coupling noise	Average	σ	Average	σ
No coupling	251.3 ± 1.0	8.6 ± 0.7	239.5 ± 0.9	7.7 ± 0.7
Fastest	215.6 ± 0.9	7.5 ± 0.7	208.5 ± 0.8	6.9 ± 0.5
Slowest	285.9 ± 1.1	9.5 ± 0.7	276.7 ± 1.0	8.7 ± 0.6

All values shown with 95% confidence interval marked. Both the pulse and edge have similar propagation delay and arrival uncertainty. The pulse width is small to not limit the signaling rate.

arrival of an event while simultaneously allowing consecutive events to have the same transition characteristics. This is in contrast to edge-based signaling, where a rising edge is followed by a falling edge, which usually requires separate detection of rising and falling edges. It has been shown though, that edges can succeed as a marker of events[29].

Because edge-based communication has a theoretical advantage in power and bandwidth, it is important to demonstrate that pulsed signaling does not come at a high cost relative to edges in practical integrated design. A case study using a smaller 130nm process node interconnect wire is used to test this hypothesis. This wire is chosen for two reasons: first, the 130nm node is well characterized, and process variation figures can be produced with confidence; second, it is around this process node where wire dimensions became a limiting factor[30] in signaling rate, making it likely that a similar length wire in finer scale will be implemented in a similar thickness of metal.

2.2.1 Case Study 5mm wire 130nm process node

A 5mm wire in the thin metal layer of a 130nm process node is used as an example to compare pulsed and edge encoding for an event signal. For this metal a conductor thickness of $.3\mu m$ and a inter-layer dielectric thickness of $.3\mu m$ is typical. The wire width and spacing is chosen to optimize the cost function of $delay \times wire\ pitch$ giving a wire width of $.55\mu$ and a wire spacing of $.38\mu$. This configuration gives a fringing capacitance of $100fF/mm$, a side coupling capacitance of $31fF/mm$, and, assuming a copper conductor, a resistance of $133\Omega/mm$. In this process node an inverter with 1μ wide NMOS, and an appropriately matched PMOS will have roughly $2k\Omega$ equivalent drive and $5fF$ of gate capacitance. Additionally such an inverter should have about $5ps$ of intrinsic delay due to self-loading. Delay optimal sizing [31] of the inverters used for wire repeating gives an inverter size of 26μ NMOS. With these parameters, minimal worst-case delay (with an even number of stages) occurs with 4 internal repeaters (5 long wire segments) with a 1mm spacing. The single-stage worst-case delay time constant in this configuration is 68.4ps.

2.2.2 Edge-communicated signaling

Arrival jitter is approximated by using Monte-Carlo simulation consisting of 1000 runs, sufficient to gain 95% confidence values for most measures. Process variation is taken from a vendor model for the 130nm process. Power variation is estimated to have a global, correlated variation of 30mV power to ground, modeling power regulator noise. In addition to global power noise, a local, uncorrelated variation of 30mV is added to each power and ground node, modeling IR noise internal to the IC.

When there is a single fast edge ($<100ps$) of a slow signal ($f < 200MHz$) the average (across process and voltage variation) propagation time is projected to be $251.3 \pm 1.0ps$, close to the value that $\#stages \times stage\ delay \times ln(\frac{1}{2})$ predicts. The sample standard deviation (σ) of the arrival time in this experiment is $8.6 \pm .7ps$. Assuming a Gaussian distribution, a 5σ interval gives a delay between $203.5ps$ to $298.5ps$ for a 5mm wire, considering only process and voltage variation.

Due to the high coupling capacitance (38% of the total) the impact of neighbor wires must be considered. The worst-case jitter occurs when both neighbors are correlated in the same or opposing direction as the main signal. Same direction switching has delay $215.6 \pm .9ps$ with a sample σ of $7.5 \pm .7ps$. The 5σ fast arrival time under these circumstances is $173ps$. Opposite direction switching delay is $285.9 \pm 1.1ps$ with $\sigma \approx 9.5 \pm .7ps$, giving a 5σ slow case of $338ps$. This is a range of $165ps$ of environmental jitter that an ideal latching strategy cannot compensate for. A more common single clock phasing strategy would have a cycle time grater than $338ps + \tau_{su} + \tau_{clk-Q}$, approximately $700ps$ in the 130nm process node yielding a maximum rate of $1.43Gbps$.

2.2.3 Pulse-communicated signal

Two features of the above analysis suggest that a narrow pulse might work well. First, the single stage delay for this setup was projected to be roughly $68ps$, worst case (5 stages, 338 ps longest projected delay), allowing a pulse on a $70ps$ time scale. Second, the arrival jitter is $165ps$, thus a worst-case environment is unlikely to erase a wider pulse. Table 2.1 compares the propagation times for edges and pulses. Using 165ps as the full-width, half-maximum measure of the pulse, there are no extinguished pulses observed, with propagation times very similar to

edges. The minimum pulse period is twice the pulse width – 330ps, marginally faster than the non-skew-compensated rate of edges; pulses have the potential to operate faster than a clocked edge system.

Using self-resetting gates to create a regenerative buffer, the performance of the pulsed line can be improved. Self-resetting gates protect pulse widths, and thus jitter cannot destroy a pulse, allowing a falling edge closer to the corresponding rising edge. For systems using self-resetting buffers, two conditions must be met: First, the pulse width and its reset time are obeyed. Second, the pulses must arrive in order. The pulse width is set locally within the buffer since it is self-timed. Native pulse width2.5 in this technology was determined to be $64ps$ with $\sigma \approx 4ps$, giving $84ps$ for a maximum pulse width, and $168ps$ for a safe pulse interval. Jitter for the self-timed buffer case is slightly higher than for the inverter buffer case at an estimated total jitter (5σ+pattern dependence) of $188ps$. This extra jitter, as compared to the inverter buffer case, is due to the extra logic required in a self resetting buffer. This extra logic also extends the maximum propagation time, in this case it is $405ps$. The arrival order uncertainty limit of $188ps$ is the dominant of the two restrictions.

2.2.4 Pulse vs. edge for marking an event

The study shows the relative propagation timing of pulses and edges are very similar, even when the width of the pulse is small. A key conclusion from this case study is that jitter can dominate gain/bandwidth in limiting performance. A pulse, because it is atomic and unambiguous in its arrival, need not be correlated to any other signal. Thus pulse-based event detection can operate as fast or faster than edge event detection correlated to another signal or state.

Table 2.2: Maximum event rates for alternative signaling methods on a 5mm wire

Method	Period	Notes
Handshake	597p	Interleaving forward & backward wires (minimizes worst-case propagation)
Clocked Edge	338p	No Phase compensation
Pulse	*330p*	*Not using self-resetting buffers*
Edge with DLL or PLL	223p	Requires DLL/PLL of ~20mW (assumes 15ps RMS jitter ≈ -112dBc/Hz phase noise)
Pulse	*188p*	*With self-resetting buffers*

Table 2.2 shows a number of cases, and the associated minimum period of operation for a BER of 6×10^{-7} (corresponding to $\pm 5\sigma$ variance). For the purposes of comparison, the minimum latch timing is left out of the presented period. In a clocked system, the latch sample interval (setup and hold) would add 50-200ps depending on design style. In asynchronous systems, both pulsed and handshake, latch sampling time does not necessarily add to the minimum period, as both techniques have sampling times built into their respective operating mechanisms. Pulsed communication systems offer a viable alternative to clocked edge systems as shown in Table 2.2. Pulsed systems not only can be faster, but they have the construction simplicity of asynchronous systems, not requiring high-rate, low-jitter clocks.

2.2.5 Long Range Perspective

Data transmission in integrated circuits is a known problem for designers[30] and one of the chief limiters in integrated circuit performance. It is currently common for transmission circuits to be capable of much faster operation that the media that they are connected to, that is to say an inverter with less than 50ps of intrinsic delay may drive a bus with 300ps. When local actions can happen much faster than longer range actions, more complex designs at the ends of a transmission media become practical.

Some initial examples of using more complex transmission and reception logic to combat media limitations came with inter-chip interconnect, leading to the use of pre-emphasis filters such as in[32], but the limitations of slow wires cause the same pattern-dependent problems between circuits whether or not those circuits are part of the same IC chip. Pulsed interconnect offers the chance to solve some these problems.

Pulses can have an intrinsically beneficial frequency spectrum. Return-to-Zero, or RZ modulation is a known method of helping cope with dispersion on long lines, such as fiber optic cables at their upper limit [33]. The signaling method of returning to zero limits the amount of information stored at lower frequencies making dispersion less disruptive to data.

Asynchronous detection, especially the kind that pulses enable, helps with transmission in a lossy media. High data rate signals suffer both attenuation and jitter at the hands of lossy transmission, as demonstrated by the example in section2.2.4. Signaling techniques that attempt to compensate for some of this effect have been developed but come at a cost in power and complexity[34]. An asynchronous system has the advantage of being able to tolerate timing uncertainty without extra complexity, and thus offers a signaling gain over a synchronous system.

2.3 Composition Rules

This class of systems are not delay insensitive, because it is impractical to have a feedback path in long-distance communication circuits – that is, it is less costly to design margins to tolerate variance than it is to communicate timing information. The lack of delay insensitivity means timing verification is required for system closure. Static timing analysis is a common part of verifying clocked systems and numerous extensions, such as yield estimation, have been formulated[35]. It is advantageous to use a constraint system that inherits the same analytic methodology, as was done for clocked SRCMOS[1]. The obvious draw-back of an asynchronous, event-driven system is the potential complexity of the set of timing constraints. Described below are a set of constraints targeted directly at providing adequate design flexibility while minimizing the checking and validation complexity.

2.3.1 Terminology

Data Signal class that cannot, on its own, cause the system state to change. Communicated by electrical high and low levels.

Event Signal class than can, on its own or in conjunction with Data, cause the system state to change. Communicated by short pulses.

Gate A basic unit of system description, either an Event producing P-Gate or a Data-output D-Gate. A gate is triggered by an event dependent on a data condition.

D-Gate The gate type with memory. Output signal type is Data.

P-Gate The gate type without memory. Output signal type is Event. Implemented in self-resetting logic.

2.3.2 Timing Check Complexity

Static timing for a clocked system involves summing inertial delays along signal paths to constrain each latch-latch path. For acyclic logic, the complexity of the check is simply bound by the number of delays along the path and thus scales as $O(n)$ given n delay bearing nodes including latches. For a circuit with m unconstrained events $m!$ constraints are needed in worst case although this is practically limited by gate fan-in. Unfortunately, behavioral issues such as meta-stability and non-determinism prevent inertial models from applying in any case. In the interest of simplified design effort, we choose to place construction constraints on gates so ensure predictable (inertial) delays apply and have complexity $O(m^2 \cdot n)$. This has the consequence that some kinds of circuits cannot be built within these constraints, e.g. fair arbiters, however, all circuits necessary for communication links can easily be constructed.

2.3.3 Logical Restrictions on Construction

Signals are partitioned into two classes or types: Events and Data. The Event class serves to mark time, and is analogous to a clock. The state of a gate can only change with an event. Events are communicated by pulses of finite duration fixed for a system technology. The other class, Data informs state updates in the presence of an event. Data is communicated as traditional digital levels. Correctness for Data signals works as in a clocked system with members of the event class serving as clocks. As usual, the value of a Data signal must be stable between the setup and hold time for each gate relative to the arrival of an updating event.

The second constraint is that a gate may have only *one active event at a time*. One-at-a-time events prevent complex timing issues from arising within the relative timing check. In order for two events to be processed simultaneously they must act on separable parts of the system. For two events to inform interacting parts of the system they must have a fixed timing relation ensuring one event at a time. This includes re-timing schemes where such event ordering is enforced. This does place a burden on the designer to describe any two-event behavior as a set of one-at-a-time actions. In this view, arbitration is not a feature of the methodology, instead the designer must describe the arbitration method within the language of the system.

These two rules create a paradigm where timing verification can can happen in $O(m^2n)$ time complexity, involving two different timing checks limited by this complexity class. Data timing is checked on $Event \rightarrow Data \rightarrow Event$ sequences, similar to a clocked system where the sequence is $Clock \rightarrow Data \rightarrow Clock$. Since events are one-at-a-time, a single event sponsors the data change, and a single event causes the result to be sampled. All data transitions are checked between any pair of events giving a check that is square in the number of events and linear in the number of gates in the data path.

2.3.4 Event timing

A known problem for SRCMOS pulse logic is the relative timing of rising and falling edges within the system so that pulses can appropriately overlap[1]. The one-at-a-time pulse model makes the event timing check the dual of the SRCMOS timing check. Here *pulses must not overlap* for correct behavior. Having a single characteristic time of the event class simplifies the computation, and using as small of a pulse as practical maximizes the systems event handling, either in rate, or in tolerance to variable timing. The event timing model is created with the

electrical realization of D-Gates and P-Gates in mind. Section 2.5 describes the transistor level design. Conditional Events are implemented with a self-resetting gate structure (P-Gates), while memory is stored in set-reset latches (D-Gate). In both cases, the pulse arriving at the gate serves, effectively, as the sampling aperture for the pull-down network; this sets the data hold window to the actual pulse width of an event. The actual minimum value will be set by the need to reliably sample, and thus by the maximum complexity in the pull-down network. Figure 2.6 shows an Event timing diagram for a D-Gate, marked with some of the critical timing considerations.

A special case requires the pulse width to be part of the timing check. For circuits where the output of a latch is logically feed-back into its input, the input pulse width and the latch delay must be timed such that the pulse-to-Q time of the latch exceeds exceeds the input pulse width. This ensures that a transiting output does not corrupt its own input signals.

2.4 System Description

To limit ambiguity, we use a simple formalism to describe an asynchronous pulse circuit. The formalism and use is similar to a guarded action language[36, 37], as the actions of gates are undertaken only when the required event occurs and the pre-conditions are satisfied. The linguistic form allows non-deterministic and parallel behavior, enabling checking that does not explicitly consider execution order. The designer is responsible for creating a system that functions correctly given this order independence.

2.4.1 Language

For simplicity, discussion of behavior will take the form of:

$$Event(guard) \rightarrow action$$

D-Gates have both set and reset behaviors denoted:

$$Data = \begin{cases} Event(guard) \rightarrow Set \\ Event(guard) \rightarrow Reset \end{cases}$$

P-Gates can be triggered by more than one $Event(guard)$ in the case of combination this will be denoted:

$$\left.\begin{array}{l} Event(guard) \\ Event(guard) \end{array}\right\} \rightarrow Event$$

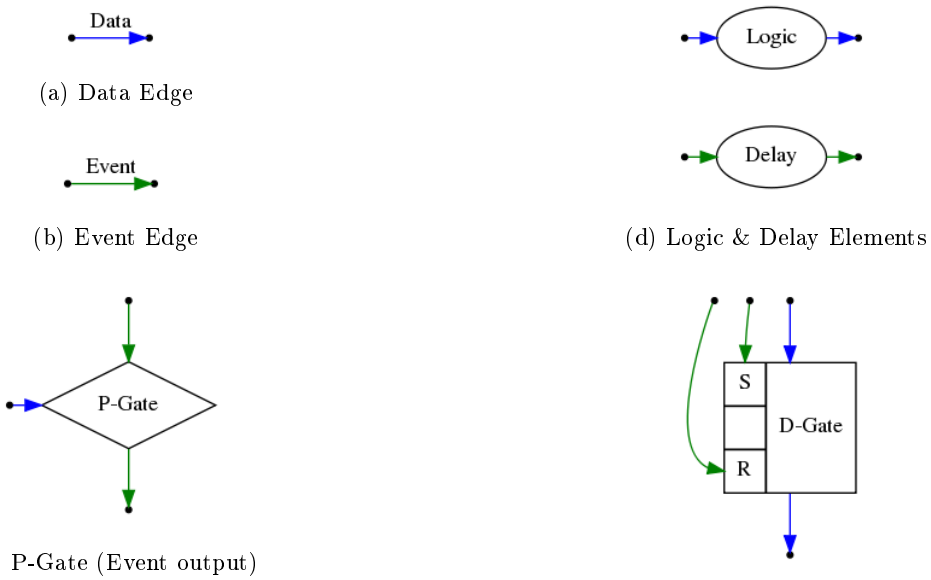
2.4.2 Control Data Flow Graph

A control data flow graph is a useful structure for visualizing the Event/Data motion within the system. Figure2.1 shows the basic elements of such a graph: Data and Event edges, D-Gate, P-Gates, Logic and Delays.

2.4.3 3-bit Counter Example

To demonstrate the system we formulate a 3-bit binary counter with asynchronous reset. The counter can be fully described in its latch behavior since there are no conditional events:

$$Bit0 = \begin{cases} Count(\overline{Bit0}) \rightarrow set \\ \left.\begin{array}{l} Count(Bit0) \\ Reset \end{array}\right\} \rightarrow reset \end{cases}$$



(a) Data Edge

(b) Event Edge

(c) P-Gate (Event output)

(d) Logic & Delay Elements

(e) D-Gate (Data output)

Figure 2.1: Control Data Flow Graph elements for pulse logic analysis

$$Bit1 = \left\{ \begin{array}{c} Count(\overline{Bit1} \wedge Bit0) \rightarrow set \\ \left. \begin{array}{c} Count(Bit1 \wedge Bit0) \\ Reset \end{array} \right\} \rightarrow reset \end{array} \right.$$

$$Bit2 = \left\{ \begin{array}{c} Count(\overline{Bit2} \wedge Bit1 \wedge Bit0) \rightarrow set \\ \left. \begin{array}{c} Count(Bit2 \wedge Bit1 \wedge Bit0) \\ Reset \end{array} \right\} \rightarrow reset \end{array} \right.$$

The CDFG for this system is shown in Figure 2.2.

If we choose to construct the system with gated events, there is the opportunity to simplify the design, at the cost of a more complex timing verification. A description of such a system is:

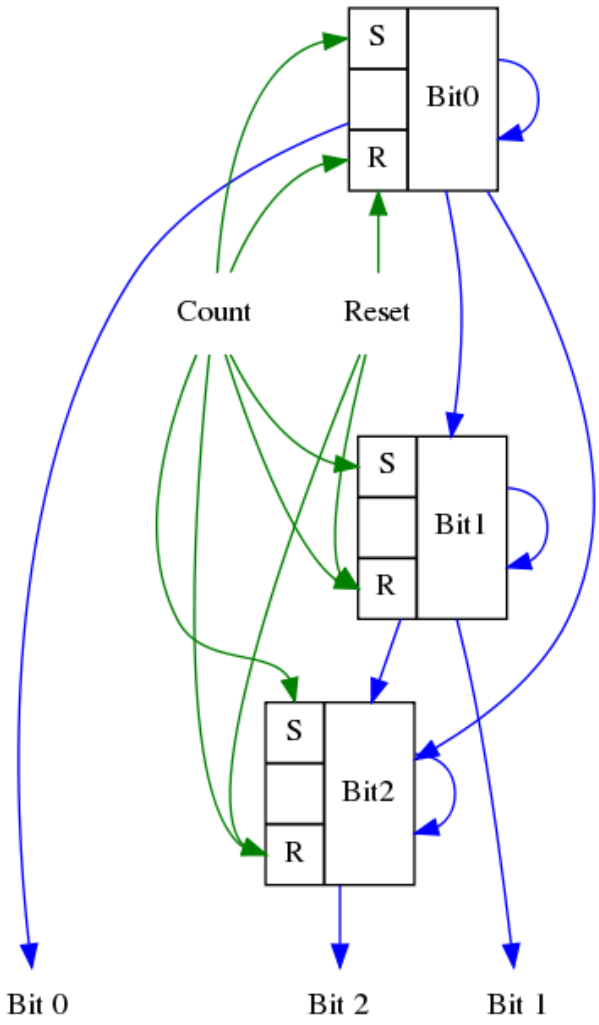


Figure 2.2: 3-bit counter without gated events. Simple structure compared to alternative

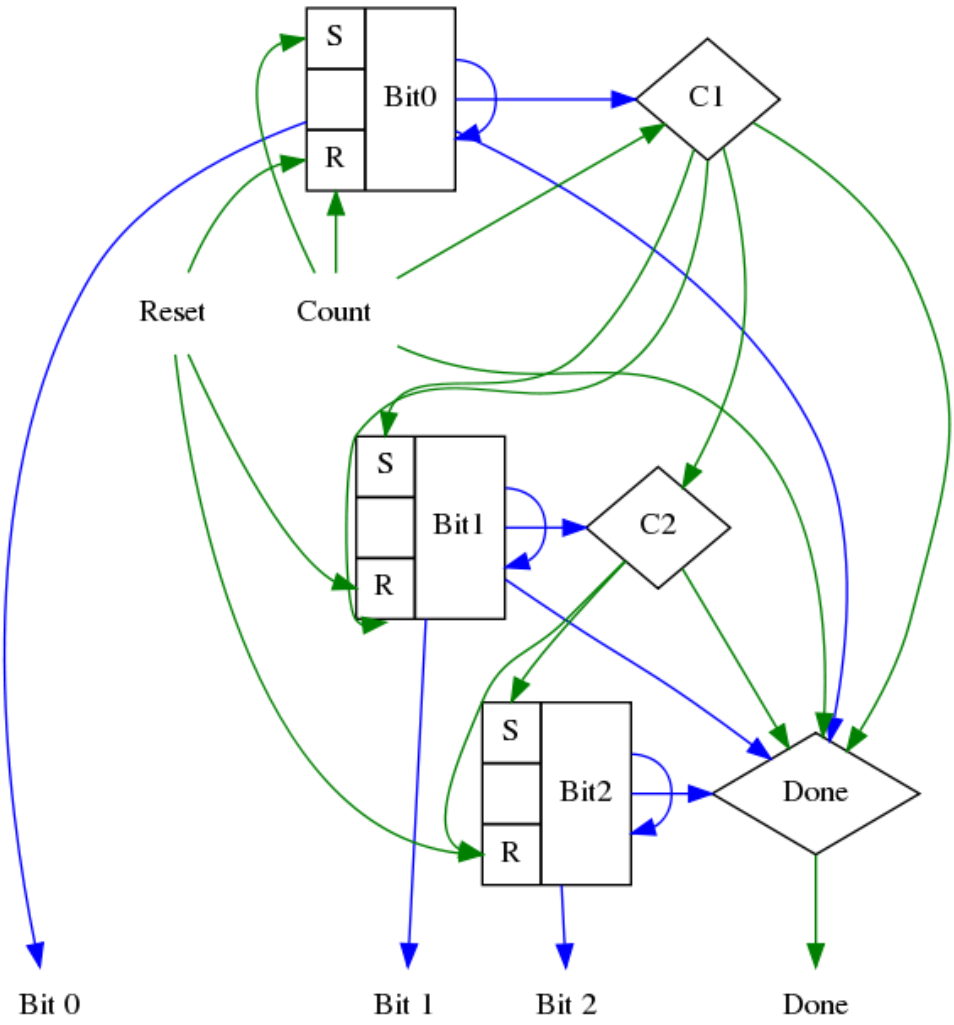


Figure 2.3: 3-bit counter with gated events. Higher levels of design re-use, and uses faster gates with fewer and terms in their pull-down networks

$$Bit0 = \left\{ \begin{array}{c} Count(\overline{Bit0}) \rightarrow set \\ Count(Bit0) \\ Reset \end{array} \right\} \rightarrow reset$$

$$Count(Bit0) \rightarrow C1$$

$$Count(\overline{Bit0}) \rightarrow Done$$

$$Bit1 = \left\{ \begin{array}{c} C1(\overline{Bit1}) \rightarrow set \\ C1(Bit1) \\ Reset \end{array} \right\} \rightarrow reset$$

$$C1(Bit1) \rightarrow C2$$

$$C1(\overline{Bit1}) \rightarrow Done$$

$$Bit2 = \left\{ \begin{array}{c} C2(\overline{Bit2}) \rightarrow set \\ C2(Bit2) \\ Reset \end{array} \right\} \rightarrow reset$$

$$C2(\overline{Bit2}) \rightarrow Done$$

This system is equivalent to chaining toggle latches to handle the task of counting, and like a series of chained toggle flip-flops, the timing signal is a derived waveform, rather than a phase of the clock, requiring extra verification steps to confirm functionality. In the first case, Reset must not be concurrent with count. In the second case, Reset must not be concurrent with three different signals (Count, C1, C2), and the timing of data is now dependent on multiple signals, creating a timing uncertainty from a static timing analysis point of view, i.e. without knowing the state of Bits 1, 2, and 3 there are a number of different possibilities of when it is safe to reference Bits 1, 2, or 3 as compared to Count. In true fashion for a functioning asynchronous system, though, a done signal can be produced that reliably indicates an appropriate time to sample the output. This is the CDFG for this alternate realization is in Figure 2.3.

2.5 Gate construction

The strong typing of signals not only simplifies timing checks, it allows the construction of timed gates using pull-down networks. This structure allows a gate to act on a given event given a set of guards, allowing delay tolerant operation. Because the pull-down network can be complex, logic functions are incorporated into the front-ends of D-Gates and P-Gates, giving fast, small designs.

There are two classes of gate in this construction paradigm, pulse *P-Gates* and *D-Gates*. P-Gates, Fig 2.4, have a pulsed output, and thus are used for conditional Events, while D-Gates, Fig. 2.5, have a level output, and are used for Data signals. P-Gates use a local self-resetting logic, and hence are related to the self-resetting logic families.

2.5.1 Pull-down network timing

The structure of the pull-down network, combined with the typing rules for Events and Data, create a timing constraint set for each pull-down network. The behavior of these networks are similar for pulse gates and latches, and the analysis holds for both. Correct functioning of the pull-down network determines the values for the timing constraints in the composition rules (section 2.3). The prohibition on event overlapping (the phasing constraint in Fig. 2.6) ensures that the action due to any event is not preempted by another event, a critical requirement for timing verification to not be forced to check event combinations.

Consider the timing of a D-Gate, like the one shown in figure2.5. Event A triggers setting this latch, while event B triggers reset. In both cases, the action is contingent on the logic of Data A, B, and C encoded into the respective pull-down networks. A timing diagram

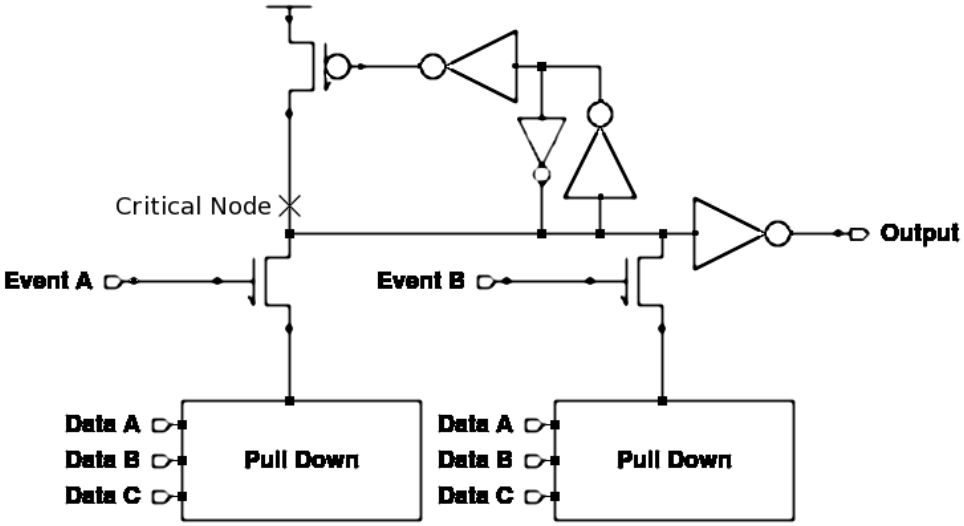


Figure 2.4: Drawing of a P-Gate. This class of gate has a pulse as an output, and thus is used for gating events in the system. This gate class would perform the miltiplex and de-multiplex operations in a SER/DES in a high-rate signaling system

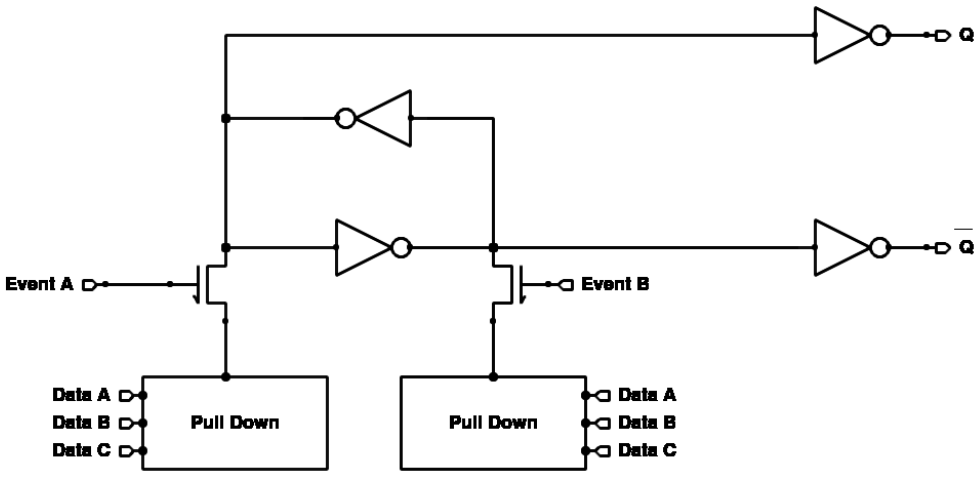


Figure 2.5: Drawing of a D-Gate for this technology. This gate has data as the output, and two different input behaviors (a set condition and a reset condition). If a D-Gate has the same event for both set and reset, and complimentary trigger conditions it can behave exactly like a pulse-triggered D flip-flop.

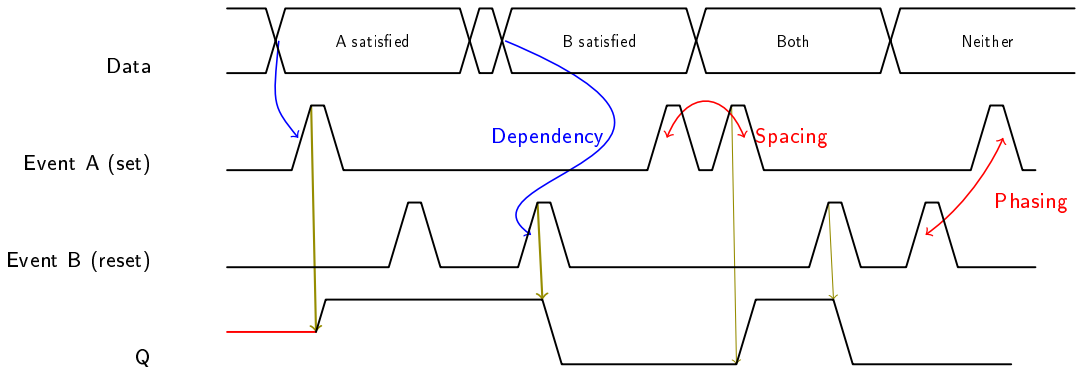


Figure 2.6: Timing Diagram for the SR latch of Fig. 2.5. Critical timing relationships are shown, including: (Olive)Event->Data causal relationships, (Red) relative event timing both same event timing and event-to-event phasing, and (Blue) Data, Event relationships for action enabling.

for this latch is shown in figure 2.6. Electrically, the pull-down network is assumed conducting or non-conducting when the associated event pulse arrives. In the case of the D-Gate, the set condition must be stable for the set pulse and the reset condition must be stable for the reset pulse.

The pull-down network impacts the timing equation in two ways: intrinsic delay, and its impact on pulse timing. First is the intrinsic delay, the amount of time for the switch between conduction and non-conduction. Static timing analysis avoids the state of the system, thus all of the turn-on and turn-off combinations will be converted to minimum and maximum propagation times. Typically the turn-on time associated with the pull-down network very limited, for example the pull-down net for the set pull down network for the first 3 bit counter from Section 2.4.3 has an estimated intrinsic delay range of up to $20ps$ in the 130μ process, less than the transition time of signals in the system. The second impact the pull-down network impacts timing is by limiting current through the pulse signal's transistor. In the case of the 3-transistor pull-down from the 3^{rd} bit of the counter the expected source degeneration of the pulse transistor is $\frac{1.5k\Omega \cdot \mu}{width}$ to $\frac{3.k\Omega \cdot \mu}{width}$ in the same process.

2.5.2 Pulse timing

Pulse width is a critical property for correct gate operation; too narrow a width and the gate will not reliably function, too long a pulse and performance is compromised. Both D-Gate and P-Gates have similar pull-down behavior, and a study of minimum width can be done with either. P-Gates are the only type that produces pulses and thus set the functional pulse width in the system.

Correct P-Gate behavior requires that input pulses are capable of triggering the P-Gate. A P-Gate is considered triggered when it's critical node, marked in Figure 2.4, is pulled down during an active event. Phrased in another way, the total charge that flows through the pull-down network during a pulse must be sufficient to overcome the keeper and flip the critical node. The keeper circuit will establish a minimum voltage required to start removing charge from the critical node, creating a threshold below which any signal in the event line is ignored. This will be the minimum input voltage V_{Imin}. The strength of the Event input is characterized by its transconductance, g_m. This transconductance takes into account the size of the associated transistor, the effects of the resistance in the pull-down network it is attached to, and the effects of the keeper circuit. The value of g_m should be selected to be valid over the range of inputs between V_{Imin} and V_{DD}. Given g_m the triggering condition is:

$$\int_{pulse} gm \cdot min(0, V_{in}(t) - V_{Imin}) \geq V_{DD} \cdot C_{crit}$$

Where $V_{in}(t)$ is the input waveform and $C_{crit}$is the critical node capacitance. This equation gives us an important property of pulse detection: a lack of pulse amplitude can be corrected with a longer duration pulse – the area between the pulse and the minimum input threshold is the key property that indicates detectability; this property is illustrated in Figure 2.7. For long range communication, issues of attenuation can be off-set by increasing pulse duration. In the case of lossy wires, increasing pulse width also reduces attenuation, thus increasing pulse width is doubly effective.

In the case of the 3^{rd} bit of the 3 bit counter, the g_m is approximated to be: $\frac{170\mu A}{\mu m \cdot V}$ on average giving a input threshold voltage calculated to match at $300mV$. This means that, with

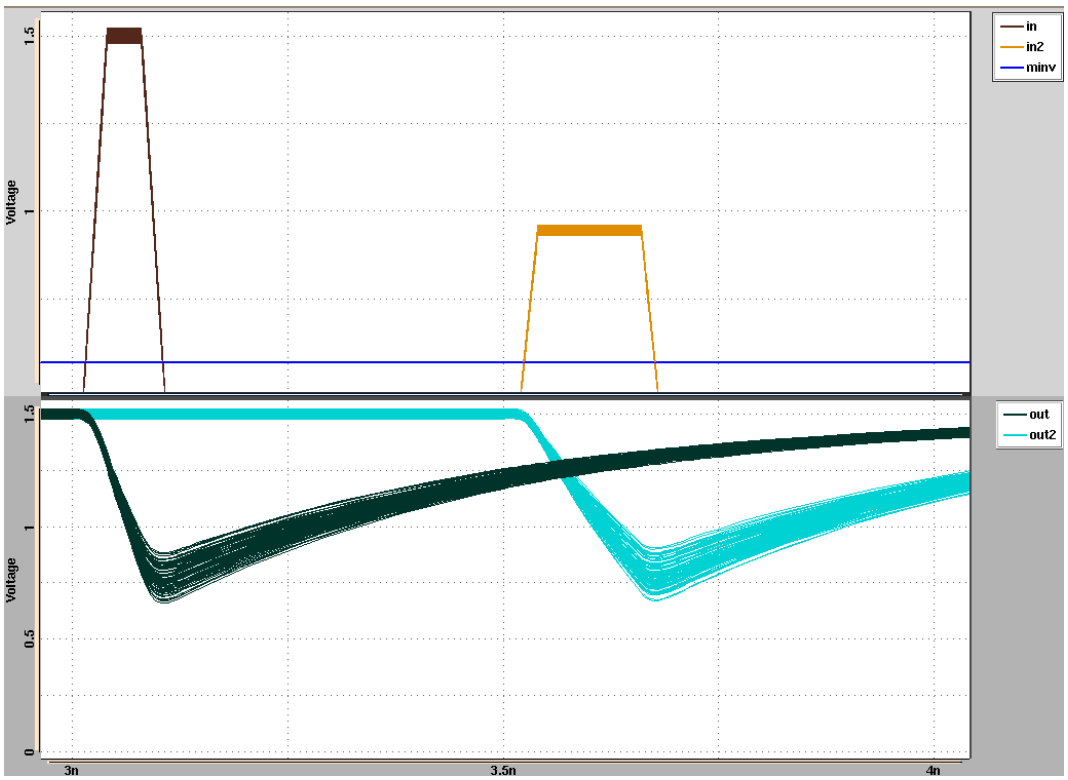
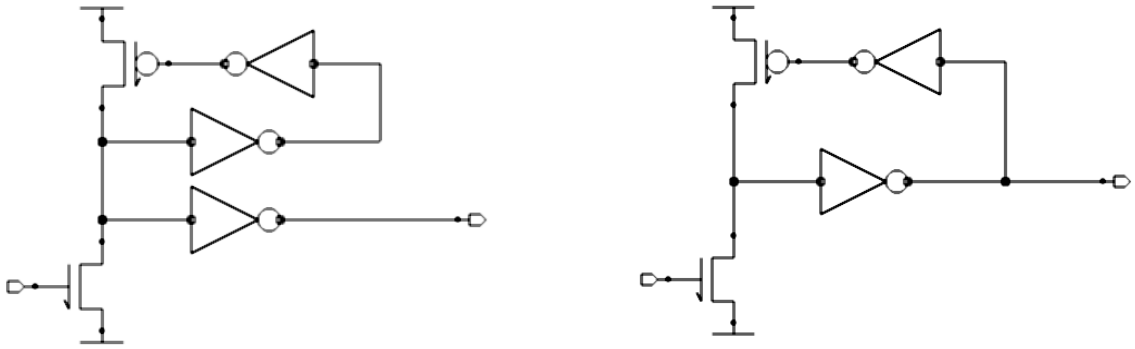


Figure 2.7: Two pulses of different volatage and duration demonstrating similar detectability. Pulse voltage directly impacts detectability, which determines the required minimum pulse width for a functioning system



(a) Load isolated feed-back network. Preferred design for maximum rate, low-load situations

(b) Load sensitive feed-back system. More robust to high-load situations; helps protect against unexpected pulse attenuation

Figure 2.8: Two self-reset feed-back network options.

this setup, for every fF of C_{crit}roughly $9ps \cdot V$ of pulse area is needed to trigger the system using 1μ wide transistors in the pull-down network.

2.5.3 Drive and Feed-back network

The characteristic pulse width of a self-resetting gate is set by the propagation delay through the feedback path. The output amplitude is dependent on both the output driver and its load. Pulse detection requires that the delay of the feedback path must match or exceed the amount prescribed by driver and load environment.

The feed-back network for a pulse gate can be connected or, alternately, isolated from its load environment, as shown in Figure 2.8. The load isolated model (Figure 2.8a) can be faster than the load sensitive feedback and has more reliable timing, as it does not have a longer interconnect impinging on feed-back. This feed-back network is well suited for situations where the loading is either constant and known, so that a correct pulse width can be selected, or

for small fan-out situations where pulse attenuation is unlikely and speed and repeatability are more important. Load sensitive feedback (Figure 2.8b) on the other hand can be useful in a production environment where cell design and cell use are not concurrent. The load sensitive gate is slowed by placing it in a high-load environment. This is undesirable from a performance perspective, but by slowing the feed-back loop, the circuit can improves detectability and thus reduces the odds of a unexpectedly high load causing pulse attenuation. Figure 2.9 shows a comparison of the isolated feedback system to the sensitive feedback system. In the lightly loaded case, with $7fF/\mu_{driver}$ of load capacitance the two feedback networks perform similarly. In the heavily loaded case, with $50fF/\mu_{driver}$ of load, the sensitive feedback produces a slower, more detectable result than the isolated feedback. If speed is more critical than detection, the isolated network is nearly 200ps faster in this example. If detection is more critical, the sensitive feedback has roughly 4 times the detectability in the heavily loaded case.

2.6 Data Link Performance Estimation

Here we will investigate a hypothetical serializer and deserializer structure to demonstrate the verification procedure an the potential power of the construction methodology. The serializer-deserializer pair can be used to create a structure that acts like a wide wire bus using fewer wires, simplifying processor construction. Using the copper wires already studied, and a two pulse communication system, the maximum safe operation rate with self-resetting buffers is estimated to be 188ps (from Sec. 2.2.3), which translates to a bit rate of 5.3GHz, 2-10 times faster than the clock rate of logic in this node. If we realize the SER/DES interface as a 4-bit wide DDR interface that would give a clock rate of no more than 664MHz, an thus produce a

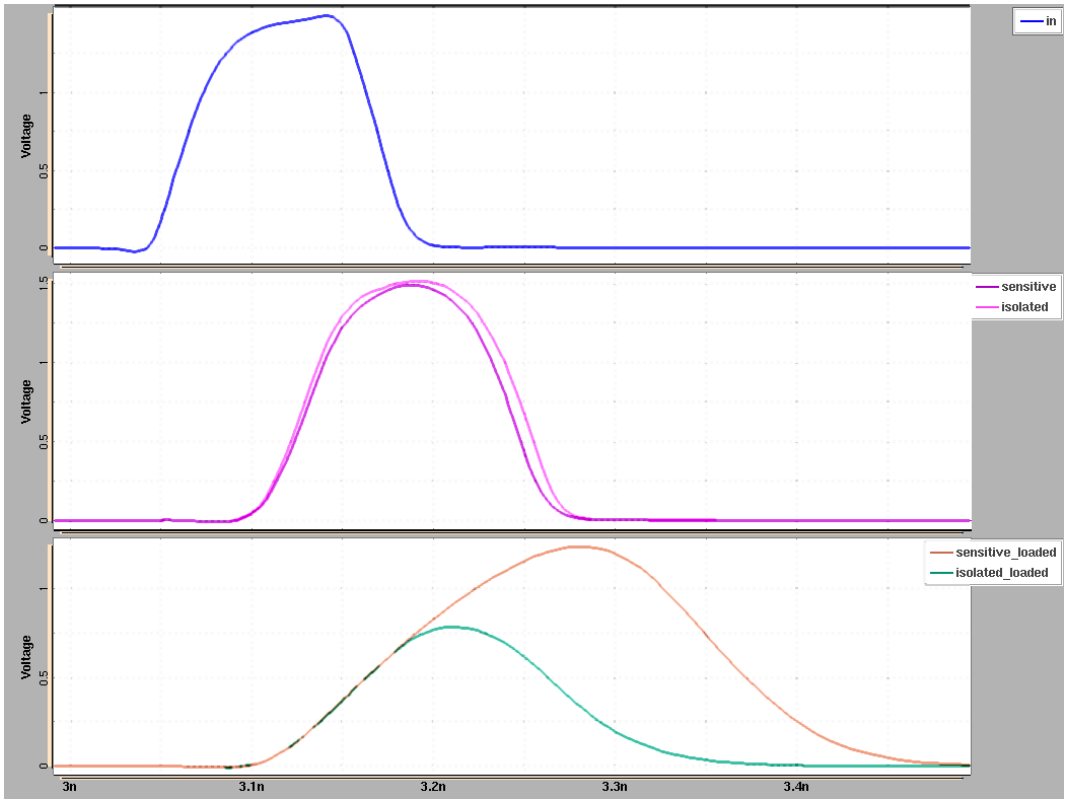


Figure 2.9: Voltage traces for isolated and sensitive feedback networks, both lightly loaded ($7fF/\mu_{driver}$) and heavily loaded($50fF/\mu_{driver}$). In the lightly loaded case the isolated feedback and sensitive feedback systems operate similarly, while in the heavy load case, the sensitive feedback produces a slower, more detectable result than the isolated feedback.

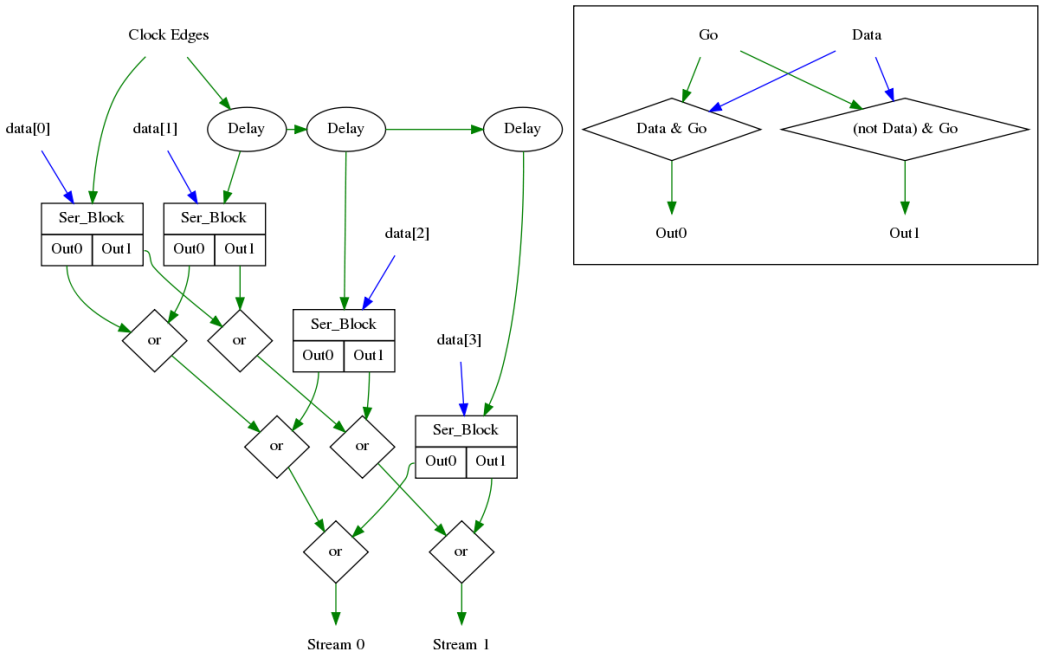


Figure 2.10: Simple Serializer Design. Delay line based timing possible due to high jitter tolerance of the system as delays have a lower bound for safety but not an upper bound.

component that could be timed off of the clock of a processor core. It is assumed that the rising and falling edges of the clock are translated into events (pulses) for both edges. The deserializer will produce a data clock on the output; a FIFO can be easily designed in this technology, but is not part of this study.

The serializer structure is shown in Figure 2.10. This serializer uses a delay line to time the individual serializer cells; a 4-input or operation merges all of the 1 and 0 events. The delay line must be timed for safety – there will be a minimum delay, but all delays above this value will be considered safe until the total delay time exceeds the clock period. This constraint shows itself in the formal verification path as a consequence of the one-at-a-time input requirement into the 4-way or gates. The matching deserializer is shown in Figure 2.11. The arc from the

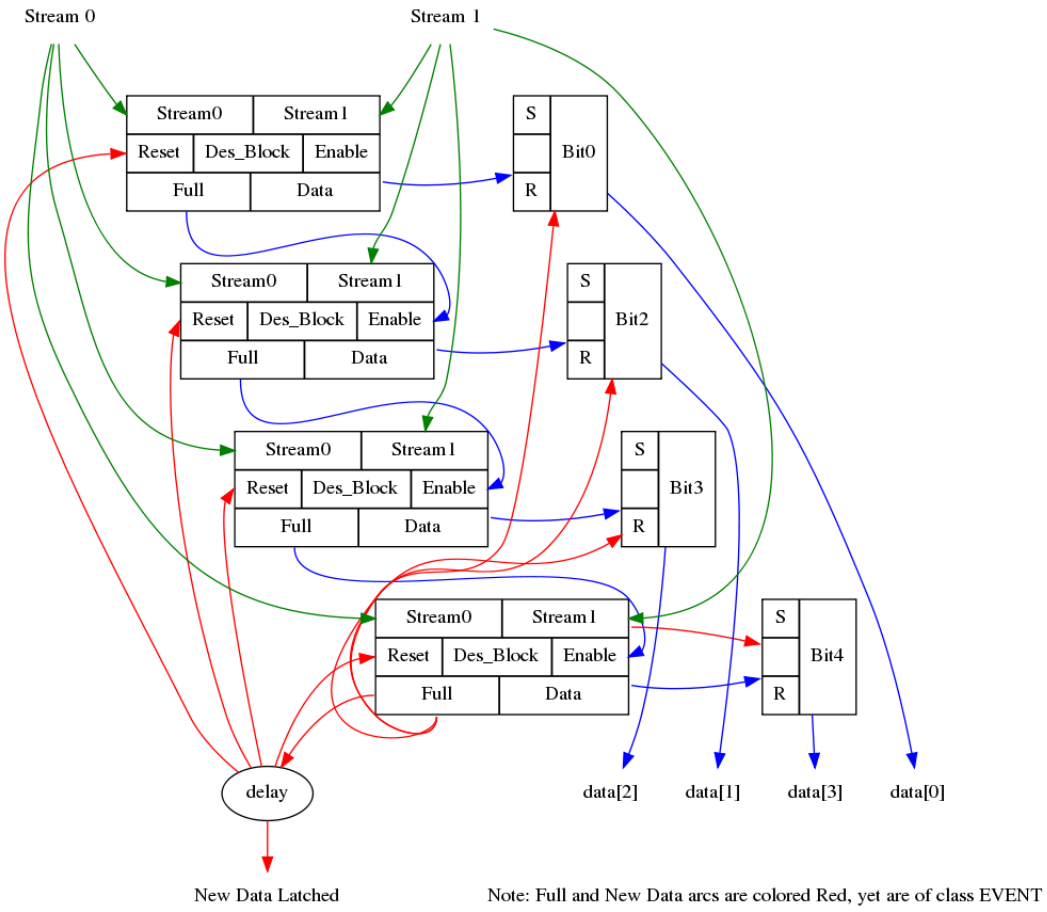


Figure 2.11: Deserializer matching the serializer in Figure 2.10.

full bit of the 4th deserializer element through the delay and associated reset elements is the longest (delay-wise) arc in the deserializer in Figure 2.11. Formally the need to reset before the new stream bit will set the rate limit on the system; while the latch sampling time will set the delay – roughly one pulse width of time.

Critical Constraints For the SERDES (assuming that event-event times must be separated by a pulse-width and take a pulse-width to transpire):

1. Bit - Bit timing $>$ max(Line Jitter, $2 \times pulse\ width$

   - Sets SER delay minimum time

2. Trigger Time for worst-case gate

   - Sets system pulse width

3. Bit4 - Bit1 timing $>$ Deserializer Set-Full time + latch aperture (1 Pulse Width)+1 Pulse width

   - Sets Reset delay minimum time (equivalent to 3x pulse width)

The Maximum clock rate tolerated by the SER/DES system is thus:

$$4 \times Max(SER\ Period) + 4 \times Pulse\ Width$$

Table 2.3 shows the computed figures for 130nm and projected figures for 65nm and 45nm. Transistor specifications come from foundry data. The interconnect wire is the same dimensions for each case, the assumption made is that similar thickness interconnect would be chosen for similar wiring distance. In 130nm the projected bit to bit time is 157% of the nominal

Table 2.3: Performance Estimates for 4-bit SER/DES system

	130nm	65nm	45nm
Trigger Time (System Pulse Width)	93.8ps	52.1ps	23.0ps
Max Bit-Bit Time	295ps	240ps	211ps
Computed Minimum DDR Clock Period	320MHz	428MHz	534MHz
Max Data Rate	2.56Gbps	3.42Gbps	4.27Gbps

minimum bit-bit time to accommodate $\pm 5\sigma$ variation. This gives a cycle time of 295ps. The slowest gate in the system (the 4-way or gate) has a pulse time of $70.3ps$ nominal with a σ of $4.7ps$ giving a system pulse width of less than $93.8ps$. The total period is thus $1.56ns$ giving a core clock rate (assuming DDR of $320.5MHz$ and a bit rate of 2.56Gbps. The system design scales well, and by the 45nm node, 4.27Gbps can be squeezed through reliably.

2.7 Design Methodology Conclusion

A method to efficiently construct self-timed feed-forward gate circuits was demonstrated. Circuits are designed using a few simple construction rules, and a small number of static timing checks to verify functionality. The methodology supports data dependent operation and concurrent signaling, though correct sequencing is left up to the designer to specify.

Chapter 3

SAT based steady state analysis

In the parlance of electrical engineers, Analog circuits are circuits that are described in a domain with continuous time and continuous voltage behavioral descriptions. Additionally, engineers imply the use of high-order analysis models when describing a circuit as analog. Due to the potential complexity, typical manual analysis strategies focus on limiting a analog circuit to a handful of modes of operation, allowing for linear approximations around the characteristic operating points of these modes. Thus, in practice, the design and verification of analog circuits are heavily human-guided where designers identify appropriate model reductions, intended (and error) modes of operation, and critical measures of functionality.

The characterization of high-speed logic cells frequently requires the detailed analysis that is traditionally associated with analog circuits. The equivalence between high-voltage, low-duration pulses and low-voltage, high-duration pulses presented in Chapter 2 requires an analysis that is detailed in time voltage and initial conditions. Automated analysis (behavioral verification) of pulse-base or high performance digital cells would thus ease the design process of cell sets for a design flow of asynchronous pulse logic.

The analytic methods of chapter 2 use a continuous time, discrete value model for analysis – an event's timing is not bound to a discrete set; yet the event is understood to be atomic, of a sufficient voltage to function. Removing exhaustive cell characterization in the voltage domain would greatly simplify the analytic process. The methodology described in this chapter enables finding steady-state solutions for a number of circuits that traditionally fall within the class of circuits treated as analog.

3.1 Formal Verification of Analog Circuit Parameters and Variation Utilizing SAT

Analog circuit verification has been a topic of rapidly increasing interest in recent years with verification strategies based on improving satisfiability modulo theories (SMT), interval analysis engines, as well as simulation mixed with other solution strategies [38, 39, 40, 41]. The vast majority of these techniques seek to create provable models that describe properties of the time evolution of analog circuits. This has a much more humble goal: quickly determining circuit steady-state operational bounds over transistor and environment variation. Classically, analog designers have relied on the tried and tested Monte-Carlo characterization of the operating space and its accelerated variants [42, 43] that depend on assumption on the statistics of the variations and the simplicity of the boundary topology. For complex circuits, however, or wide ranges of device variability, both assumptions are problematic, and the results are only stochastically conclusive.

The early work analog model checking was done by Kurshan [44], who formalized arguments for discrete proof spaces over continuous functions. Hartong [45] took a representation

tack including modeling Ids(Vds, Vgs) from data as we have also done. Similar transistor modeling approaches were used by Little, Meyers [46] and Yan [47, 48]. All of these works used specialized techniques to create discrete atomata for the purpose of forward projection of the analog circuit state. Hedrich did tolerancing verification based on polynomial bounding and projecting symbolically, however, his transistor models were linear approximations [40]. Work on variance estimation by Nassif [49] abstracted the sources of variance for timing and other parameters and performed stochastic modeling. Alternative simulation acceleration approaches by Signhee [42] improved the ability to simulate close to the bounds of circuit margins.

Validation by Simulation has advantages and difficulties. Simulation readily accepts circuit scales far larger than any verification scheme known to the author. On the other hand, it does not readily link to formal analysis or property proof. To manage device and environmental variations with simulation, Monte-Carlo methods (or similar means of extracting statistics) are typically used, requiring a large number of configurations to be checked. This process tends to be very time consuming if indeed feasible at all. Additionally, due to its stochastic nature and the complexity of analog circuit behavior, such methods are usually incomplete. Abstractly, simulation verifies behavioral properties for single parameter selections and initial conditions on each simulation run. In contrast, the methods proposed in this work describe properties known to hold over sets of intervals in the parameter space, providing formally supported bounds on circuit parameters.

Practical use of analog verification techniques requires methods that efficiently scale to typical problem sizes with accuracy sufficient to at least match that of the device models. They must provide concrete bounds on behavior regardless of whether the nominal behavior or statistical distribution can be compactly expressed. We believe that although the safety

property is formally simple, it is nonetheless quite important in practical variance modeling and design for variance compensation. Its simplicity allows for practical scale robust verification in reasonable time.

This work describes an approach to proving properties on steady-state circuit behavior that scales to practical circuit size and complexity while using practical (non-symbolic) device models. The approach, by design, only utilizes methods that can easily be mapped to a boolean satisfiability (SAT) problem, a problem for which high quality solvers exist. Despite the limitation to steady-state properties, several valuable properties such as operation points and margins, headroom can be described and verified in very reasonable run times.

It is important to note that a similar effort was made by Tiwary et al. [50]. The general concept of modeling method and behavior discovery were the same, on the other hand a SMT solver was used instead of a ILP or SAT solver, and hence different assumptions about the bounds of the problem were made as well as having a slightly different formulation. This work also has device variation considered within the presented model.

3.2 Verification

This work targets the verification of steady-state properties of analog circuits. The verification process considers both a circuit description and a set of imposed constraints. A circuit is defined as a bi-partite graph of nodes and devices. All edges common to a node share a common potential. In general, voltages are defined as differences in potential between two nodes. All voltages are expressed as relative to a single ground (GND) node. The modeled behavior is time invariant (steady state), hence the circuit state is precisely the set of voltages on all

circuit nodes. This representation implicitly enforces Kirchoff's voltage law (KVL), requiring the total voltage drop around a loop to be zero. There are additional constraints on nodes (enforcing current balance) and on devices which map voltages to currents that must also hold for a valid circuit model. In general, a circuit may have multiple or even infinite sets of states that meet all constraints, these states are called 'operating points'. Our interest in verification is to determine exterior bounds on node voltages for which no operating points exist given the circuit and imposed constraints.

An example analog circuit is shown in Fig. 3.1. In addition to containing the state of the circuit, nodes also serve connect currents for the devices. In the example circuit VDD, GND, A, and B are all nodes of interest and interact with at least two devices each. Kirchoff's current law (KCL) requires that nodes have currents balance into and out of the node. Fig. 3.1's node A has its currents marked by small arrows. The currents flowing in the direction of these arrows must sum to 0 for the circuit to represent a physical system. This gives a constraint on valid circuits – Eq. 3.1.

$$\forall n,\, n \in Nodes \sum_{i \in n's\, currents} i = 0 \tag{3.1}$$

Devices are connected to nodes and direct currents between nodes. Fig. 3.1 has four devices N1, N2, P1,and P2 representing two types of transistors (NMOS & PMOS) as well as a source device. Fig. 3.2 shows the critical voltages and currents for a field effect transistor (NMOS in this case). In simplest form, FET transistors have a single dominant current flowing between source (S) and drain (D); The current is a function of two voltages: the difference between the gate (G) and the source, and between the drain and the source. For FETs, ignoring

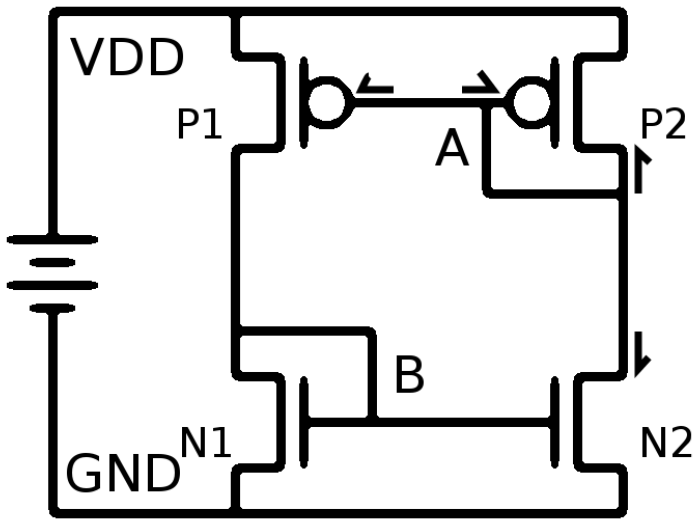
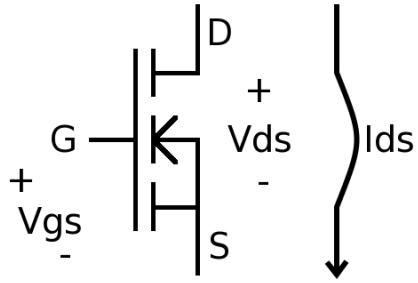
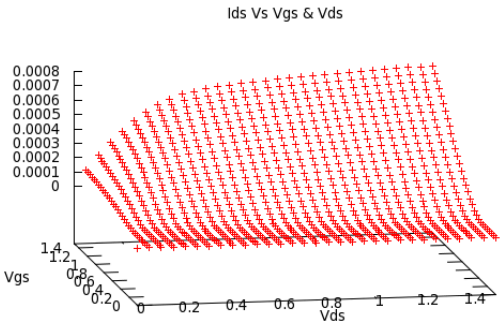


Figure 3.1: Simple circuit showing nodes (A, B, VDD, GND), devices (transistors p1, p2, n1, n2 and a voltage supply) and branch currents for node A.



(a) Transistor with model critical voltages and current marked. Vgs is the gate-source voltage



(b) I_{DS} vs V_{GS}, V_{DS}, the characteristic behavior curve for this transistor.

Figure 3.2: NMOS transistor schematic and characteristic behavior curve

the influence of the back , two differential voltages (Vds, Vgs) determine the current flow (Ids). While the methodology can fully support models of body bias, for simplicity this was left out of the models treated here.

Device modeling is a critical part of the process. A fully specified device in a design can have a range of different behaviors in implementation due to manufacturing variation as well as aging processes. This work uses a bounded behavior model; such a model is very general – agnostic to the nature or cause of device variation. The model can be built with test data and no need for other known physical parameters Fig. 3.4 shows a slice of a transistor model that can be used for bounded behavior proofs. The model, by definition, captures the outer bounds of possible behavior for the device. Bounds that are conservative can still be used without invalidating the proof, mathematically expressed as Eq 3.2.

$$f_{lower}(\{nodes\}) \leq i_{device} \leq f_{upper}(\{nodes\}) \tag{3.2}$$

The computed bounds on circuit behavior, as compared to monte-carlo simulation is depicted in Fig. 3.3 with lines representing computed bounds and points representing simulation. This is a result generated from the circuit shown in Fig. 3.1 over all potential variations in the FET device models. This simple bound will be used later in the paper to determine verifiable properties of interest to the user. The discrete solver implementation to efficiently and conservatively determine these bounds is discussed below.

3.3 Mapping to SAT

The verification approach uses a simple discrete SAT solver to handle the proof elements of the procedure. To simplify the translation of constraints, the problem is first cast as a 0-1 ILP, also known as Psuedo-Boolean, problem (a known NP complete problem [51]). There are a number of translation techniques that allow such a problem to be easily re-cast as a SAT problem [52, 53]. The high quality of existing SAT solvers is what makes them attractive for practical application. The solver used here is a version of MiniSat+, described in [52], modified to take SAT clauses in addition to Pseudo-Boolean constraints.

Node voltages are represented as fixed-point binary numbers stored as an unsigned magnitude and a sign bit.[1] The solver has a single scale factor for all voltages to be represented. The proof will be invalidated if values that are unrepresentable are needed creating a lower bound on the scale factor. A scale factor that is larger than needed will require more bits in representation to achieve the same precision as one that is sufficient yet smaller. The single discrete value of a voltage will be used to represent an entire range of voltages in the continuous version of the problem. Mathematically, the range of possible values for a representation V is expressed in Eq.3.3, where n is the number of representation bits. The smallest difference between adjacent binary representations is called the Least Significant Bit Value or LSB for short.

$$voltage \in scale\,factor \times (V + [-0.5, 0.5]);\ V \in \mathbb{B}^n \tag{3.3}$$

[1] Several other representations were explored before choosing this simple one - in particular, the representation of 2-d dense parameter arrays seemed to call out for alternative representation. Although some progress was achieved, performance was reduced by the additional complexity of comparison and numerical operation functions.

Device currents are handled in a distinct manner from the node voltages. Devices are specified by an upper and lower bounding function on currents given voltages. This means that a binary representation of a current can have exactly one value when mapped back into continuous space; the model bears the requirement of providing a range of possible values by providing a lower and upper bound value; this constraint is expressed in Eq. 3.4.

$$repr(i_{lower}) < i_{lower} \leq i_{upper} < repr(i_{upper}) \tag{3.4}$$

The implementation of KCL in the system requires the current into and out of a given node to balance. This constraint is implemented by taking a sum of all of the currents into and out of a node and then constraining it to have a small absolute value. Two constraints, Eqs. 3.5 and 3.6, are created to implement this.

$$i_{node} = \sum_{devices} i_{device} \tag{3.5}$$

$$-\varepsilon < i_{node} < \varepsilon; \varepsilon = scale\,factor \times \#currents_{node} \tag{3.6}$$

Eq. 3.5 creates a requirement that each node have a list of currents going into and out of it. During circuit translation the software keeps such a list while implementing the constraints due to device models. As a final step the currents associated with the nodes are then summed to create this constraint.

Eq. 3.6 uses a small absolute value to prevent two specific cases from failing. One case arises for nodes where devices have no tolerance in their current representations (for example when the model is built expecting higher resolution than what the solver chooses to use). Another case occurs when a scaling factor within the system (eg $I_A = 2I_B$) is used and can cause

a similar problem. Due to the discretization of circuit state and device models, a sum may have as much as half a bit of error (relative to the underlying model) per device current so an amount of slop equal to this potential error is allowed. Increasing the solver resolution reduces this tolerance relative to any absolute value presented, hence the resulting error can be made arbitrarily small.

The discretization of the state creates subsets of the continuous state space that contain potential operating points of the circuit. The conjunction of the above Pseudo-Boolean constraints cannot be satisfied for any discrete state whose patch does not contain an operating point. Thus the solver finds all states in the desired exterior bounds as unsatisfiable. Bounds on operation are found by iteratively looking for the largest satisfying value for any voltage of interest. By construction, these bounds can be determined to any level of accuracy by simply decreasing the granularity of the discrete space. In practice, FET models derived from measurements provide sensible limits to increasing accuracy of the analysis.

The circuit introduced in Fig. 3.1 has two nodes that are non-trivial to specify: A and B. In actual operation the voltage of these nodes is highly dependent on device models. Validation by simulation gives a set of operating points that is graphed in Fig. 3.3. This figure shows a region where the behavior of circuit A is likely to be and a large region where no simulation found operating points. For any interval or value of one parameter, the verification above can quickly find upper and lower voltage bounds on the other parameter providing an absolute limit to potential operating points since one or more circuit constraints must be violated. The value of this is that circuit properties are complex in general and there is no simple way to determine the practical bounds on operating regions (i.e. extrema of Monte-Carlo searches). This technique can find such bounds directly.

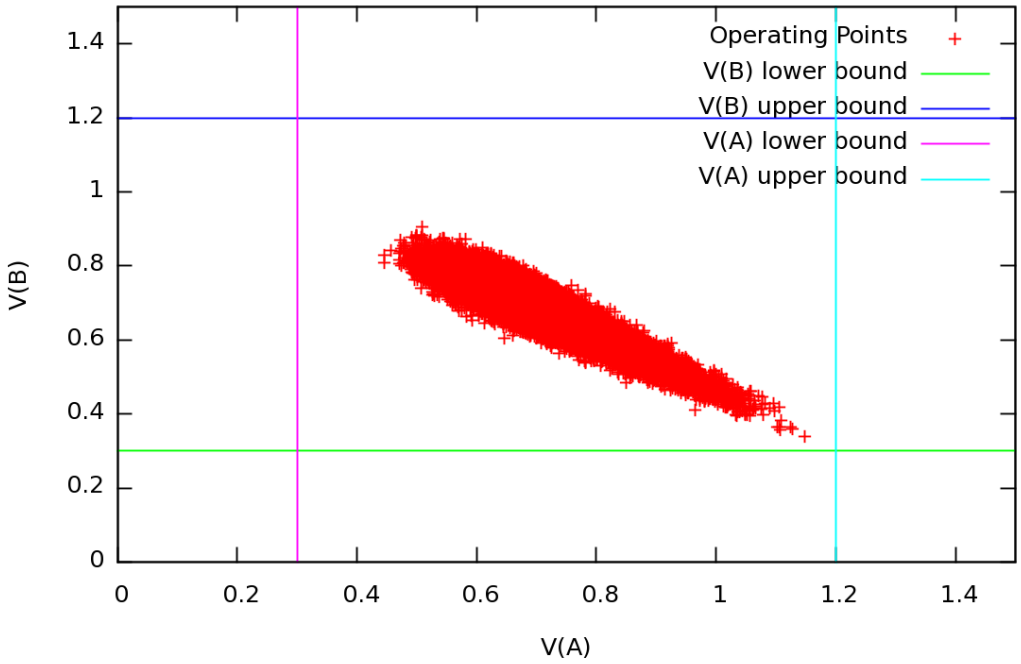


Figure 3.3: Map of verification results for circuit in Fig.3.1

3.4 Device Models

Device models are the functions used to map nodal voltages into device currents. The models specifically map into a range of currents representing the possible upper and lower limits on device current for a given circuit state. The voltages at the nodes the device is attached to are the dependent variables in this mapping and the resultant device current is the independent variable. The device model must obey Eq.3.4 so the solver can correctly operate on the exterior bounds of circuit operation. The device model creates a solver variable, I_{device}, and places bounds on that variable. The bounds created may or may not be dependent on the voltages at the devices terminals.

3.4.1 Linear devices

Linear terms are particularly easy to express in the pseudo-Boolean language making the modeling of linear devices easy. There are three kinds of linear device modeled in this work: voltage sources, current sources, and resistors. Each device adds simple constraints to the system. All three kinds of device have only two terminals. For two terminal devices there is one voltage of interest; the difference potential between the two nodes. Since each voltage number actually represents a range of voltages, this difference has an associated range of values. The magnitude of this range is the combined ranges of the two underlying voltages. To limit error a differential voltage is not computed directly, instead the software tracks the difference and incorporates it algebraically in to subsequent calculations.

3.4.1.1 Current sources

Current sources have a single systemic constraint; they must have a specified current flowing through them. This is modeled as two constants, a lower bound on current and an upper bound on current. This creates a simple device constraint defined in Eq. 3.7. The lower and upper bound of operation are constants given as part of the definition of the source.

$$I_{lower-bound} \leq I_{device} \leq I_{upper-bound} \tag{3.7}$$

3.4.1.2 Voltage sources

Voltages sources also have add a constraint; they limit the difference in voltages between their terminals. The voltage source constraint is in Eq. 3.8. The differential voltage value must also be calculated for this constraint to be valid. there are two options for this

calculation. The first is to create a new constraint to express the difference (Eq. 3.9, where V_a and $V_b$are the voltages at the two ends of the voltage source). The second option is to re-write the voltage source constraint as two, more complex single ended constraints (Eq. 3.10). Since the solver takes only single-ended constraints the trade-off is between three constraints and two. The benefit for the 3-constraint system is if multiple devices are in parallel, the value $V_a - V_b$ can be reused between devices and possibly speed up solution. The benefit for the 2-constraint system is more rapid processing of the constraint library, which is a small bonus for any given constraint, but can add up to a significant part of the computation.

Mathematically, an ideal voltage source can provide infinite current to keep a fixed voltage between its terminals. The solver must be able to recognize that the current flow through the device is effectively unbound. The KCL constraint system would assign no current flowing through a device, unless a current variable and bounds were created for that device. One option is to add a constraint like that of Eq. 3.11 to the list, where I_{max} is some value that is so large that exceeding it represents a computational failure somewhere in the system. A second option is to identify nodes with ideal voltage sources attached, and remove the KCL constraints for that node; that is allow each device to sink or source any amount of current to that node, knowing that the voltage supply will accommodate any difference.

$$V_{lower-bound} \leq V_{diff} \leq V_{upper-bound} \tag{3.8}$$

$$V_{diff} = V_a - V_b \tag{3.9}$$

$$V_a - V_b - V_{lower-bound} \geq 0$$
$$V_{upper-bound} - V_a + V_b \geq 0 \tag{3.10}$$

$$-I_{max} \leq I_{device} \leq I_{max} \tag{3.11}$$

3.4.1.3 Resistors

Resistors have a current that is proportional to the voltage difference between their two terminals. Like the current source they add a relation to current but now the upper and lower bounds are dependent on the terminal voltages. The resistor is specified as having a minimal and maximal resistance R_{min} and R_{max} respectively. The difference in terminal voltage is computed as in Eq. 3.9, one of the options for solving a for a voltage source. The resistor model is implemented with upper and lower bounds on the resistance value setting upper and lower bounds on the device current, as described in Eq. 3.12. The utility of having a specific calculated differential voltage for resistances is that there is one variable that can be scaled twice for the two bounds.

$$R_{min} \times V_{diff} \leq I_{device} \leq R_{max} \times V_{diff} \tag{3.12}$$

3.4.2 Non-linear devices

There is one type of non-linear device of interest for this work: the transistor. Mapping non-linear devices in 0-1 ILP requires multiple linear constraints to create a non-linear one. Due to the nature of 0-1 ILP (and ILP in general) polytopes are easy to express; the models used are expressed as a union of polytopes. While the solver can understand models of this generality, efficiently creating them is difficult. On the other hand, generating meshes with bounded error

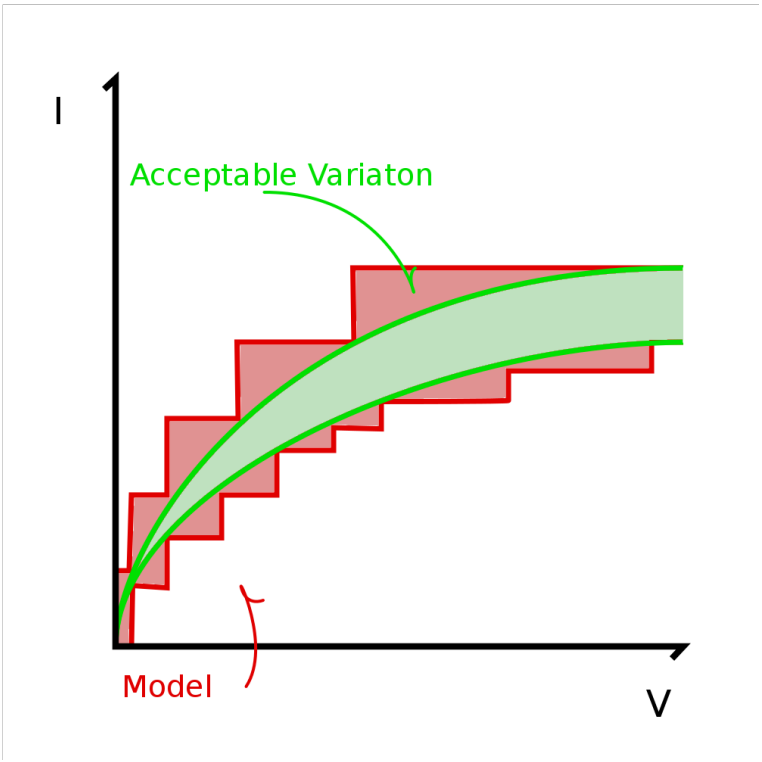


Figure 3.4: Transistor model, showing natural gutter of acceptable behavior and conservative discrete union of polytopes model

has a number of heuristics [54, 55]. For simplicity, this work chooses all edges to be parallel to one of the variables defining the space. Because of the underlying SAT solver used, the union is written as a series of constraint cubes, one of which must be obeyed. This can be visualized in the model shown in Fig. 3.4. This figure shows the continuous space boundaries of behavior and the step like polytope approximation.

Models of this type have proven useful in other applications. For example, sub-division of continuous space is a useful method for solving for time domain transitions[56, 45]. Piece-wise transistor models have also been used to make computation easier in simulators[57]. Union of polygons was also used in the interval based solver in [40]

3.4.2.1 Spice/Monte-Carlo Extraction based models

The first set of transistor models were based on an underlying data set created by Monte-Carlo simulation preformed in spice. These models are based on a .13μ Technology. The bounding region was determined by the central 99.9% of 30,000 Monte-Carlo runs performed in spice ($\pm 3\sigma$ variance, 97% confidence). The monte-carlo runs took approximately 15 minutes of computer time to execute and a further minute or so to reduce to an acceptable model. A number of models were built from this data set. The main PMOS and NMOS models were created with 50mV resolution in Vgs and Vds and a 100nA resolution in Ids. This yielded models that consists of 230 and 188 cubes in Vgs,Vds,I. A version limited to 40 cubes in each using a 150mV step was also created for faster approximations. Another much higher resolution model set with more than 10,000 cubes per transistor was created for scalability testing. This more accurate model required much more data; it was built with an expanded 100,000 point run, and still had insufficient data in some scattered regions. The high cube count model could be improved with a larger data set as many cubes were needed to cover model irregularities.

3.4.2.2 ASU PTM Corner-case transistor curve based model set

The models generated for the examples in this section were created form the ASU PTM models from their nano-spice tool. 3σcorner models were downloaded for the 130nm process node and considered the outside bounds of operation for the transistors in question. The devices were crafted with 5% variation in L_{eff} and 15mV in V_{th} variation. These models were subdivided using regular steps then these regions were merged as to maintain an error less than a given threshold a model using 20mV steps in V_{GS} and V_{DS} and 2% accuracy in I_{DS}. This gives a model with approximately 2,000 cubes for each of NMOS and PMOS. Because of

Table 3.1: Table of models. The amp.sp test has 14 transistors and finds 8 bounds and is discussed more fully in Sec. 3.5.2

Source	Monte #1	Monte#2	ASU PTM
Vgs steps	125mV	50mV	20mV
Vds steps	375mV	Acc. Dependent	
I_{dev} accuracy	Fixed Vds	10%	2%
NMOS cubes	40	188	1936
PMOS cubes	40	230	1936
Runtime (amp.sp)	1.7s	16.2s	397s
#Vars	4402	21062	190210
#Constraints	5090	26510	243986

the smooth and simple nature of the PTM model re-casting these models is easy by comparison to the Monte-Carlo method.

3.4.3 Model impacts on problem complexity

The design of optimal models is beyond the scope of this work but remains important. A device model is needed for each device in the circuit. Clauses that represent the model are added to the constraint set for each device instance that uses that model. Thus the SAT problem complexity will roughly grow linearly with the number of polytopes used in the device models.

3.5 Applied cases

3.5.1 Resistor divider – output voltage

The resistor divider is a very basic case to test some of the concepts of the system. As a basic test a divider is constructed using two resistors each $1k\Omega \pm 5\%$ with a voltage source of 1.5V. The solver is asked to find the minimum and maximum exterior bounds for the possible values of the output voltage. For this example case the solver quickly gives the answer: the output value is bound on the upper side by 0.7875V and the lower side by 0.7125V. Due to the finite nature of the solver these values will be rounded to the nearest least significant bit of the representation. The maximum value is rounded up and the minimum is rounded down because the exterior bound is being discovered, if the actual bound cannot be represented a safer approximation is used.

3.5.2 Differential amplifier – minimum bais to support drive current

The differential amplifier test is a more complicated case. The circuit shown in figure3.5 is translated into a spice file along with some designer constraints. first a output drive requirement:$|V(on) - V(op)| < .05$; $I_{load}(on, op) = 50\mu A$. Next a limit on the required input to create that drive:$|V(ia) - V(ib)| < .1$. Finally a restriction that the answer must be close to centered relative to the supply:$.45 \times V(vdd) < V(ia) < .55 \times V(vdd)$

The more complex constraints added allow for setting up the exterior conditions on the amplifier. In this case the delivered current is $50\mu A$ and the inputs are within 5% of the mid-way point between the power rails and are offset less than .1v. The bias voltages deemed viable are the ranges of bias voltages that allow the circuit to meet the specification given the

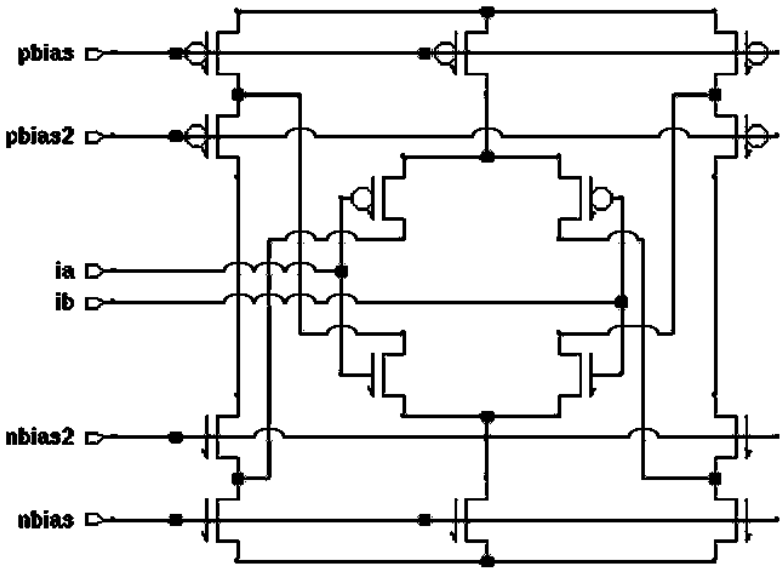


Figure 3.5: Differential amplifier with multiple bias points

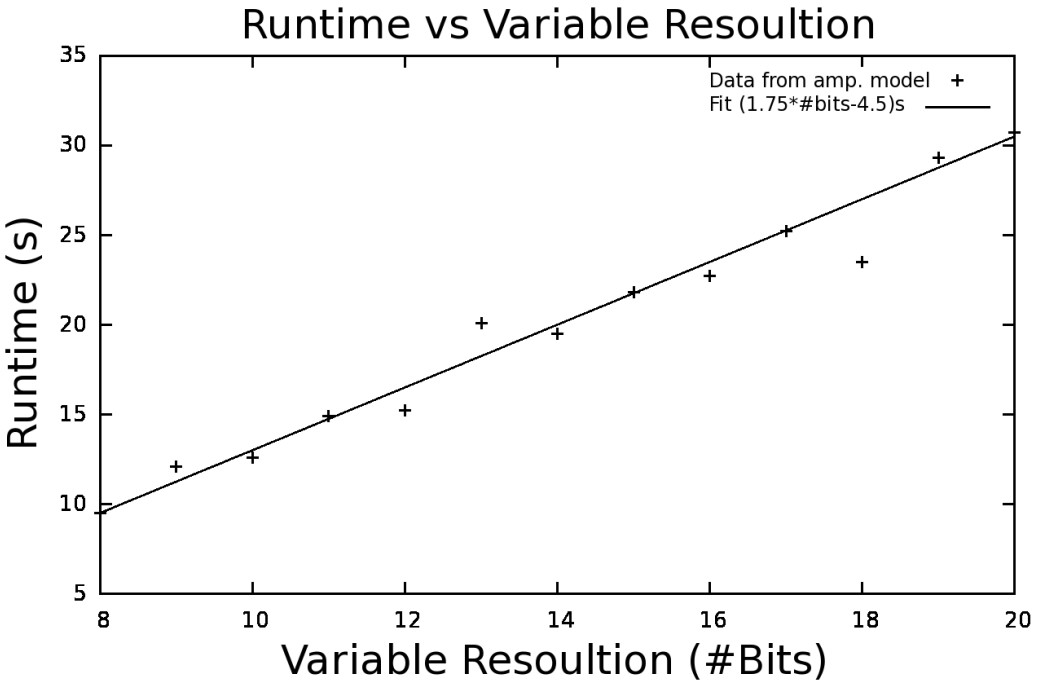


Figure 3.6: Time to solve for 4 bias voltages of circuit from Fig. 3.5 vs number of bits used to represent a voltage or current. Run-time scales linearly with resolution; scales as the log of representational precision

established conditions. As a consequence, a set of bounds discovered for the *pbias* signal implies that there is no working circuit, even with device variation, with a *pbias* outside of that range.

Given the higher transistor complexity of this amplifier, it is useful study how precision, as dictated by the number of bits used in the numeric relationships, impacts operational speed of the solver. Fig. 3.6 shows the time dependence of solution on number of bits used in the representation. The growth in computation time has an approximately linear relationship with the number of representation bits used. An approximate fit to the measured run-time VS variable size data is shown in Figure 3.6.

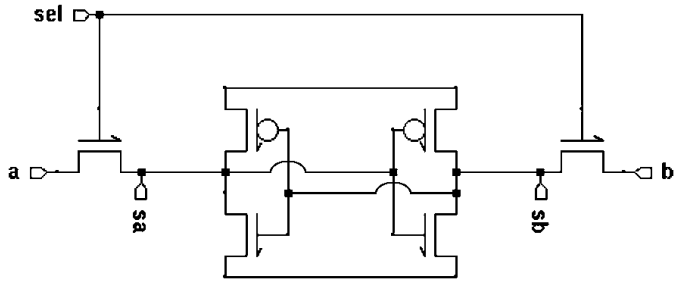


Figure 3.7: SRAM cell. This cell is used for the characterization exercise. For this exercise all NMOS are sized equally, and all PMOS are sized equally.

3.5.3 SRAM characterization

SRAM cells, due to the large number that are produced, are a common area of study for failure. Because of the multiple operating states of an SRAM cell there are steps to analysis, so that individual states can be isolated and studied. Additional constraints are added to the system to cut the space into cases. One case will be established to discover bounds on the metastable region, and another is constructed to find the bounds of the stable states.

3.5.3.1 Meta-stable region

The first operating point of the SRAM cell that we can discover is the region where the meta-stable point exists. The meta-stable region is the region where there is neither strong drive out of the region, nor any attraction to the region. If we assume that for a given cell there is reasonable transistor matching there is an easy condition for meta-stability: when the two storage voltages are equal (sa and sb in the schematic in Fig 3.7). The additional constraint is written formally as $V(sa) = V(sb)$.

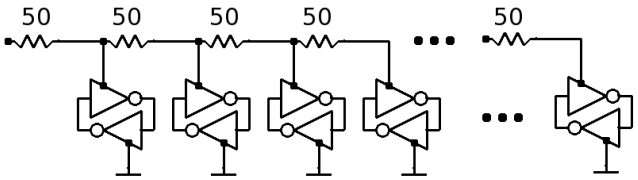


Figure 3.8: Schematic of SRAM array test. 10 SRAM cells used as part of this test

With the models described above the solver determines that this meta-stable state is bounded within the region of $.6v \leq sa, sb \leq .8v$

3.5.3.2 Stable state

The next operating mode of the cell to be considered is the behavior when the cell stores a value. In this case one of the voltage storage nodes is above the other. Due to the nature of the transistor models used the meta stable region can also satisfy this requirement, so the answer requires the exclusion of the meta stable region. Due to the symmetry of the problem the constraint $V(sa) > .8\,; V(sb) < .6$ is sufficient for this exclusion

The results of this run indicate that $1.4v \leq sa \leq 1.5v$ and $0v \leq sb \leq .1v$. This means that the high storage value can be as low as 1.4v and the low storage voltage can be as high as .1v given the model and other circuit assumptions

3.5.4 SRAM array

The SRAM array test consists of an array of SRAM cells connected by a power network that has resistive elements, and hence a voltage drop that is state dependent. The cells used are the same as the cells from Sec. 3.5.3; from that analysis the voltage required to force the

cells out of the meta stable region is known (not between .6V&.8V when the supply is 1.5V). The network topology used is shown in Fig. 3.8. The successive drops mean that each cell is operating at a different voltage than the other cells. This directly affects the behavior of adjacent cells. The array test seeks to verify the properties of an SRAM cell that is the 10th in the chain. The test thus has 20 NMOS transistors, 20 PMOS transistors and 10 resistors with non-trivial relation to the result.

The test exercises the limits of the ability of the solver. With the medium complexity transistor model the problem contains 75k Pseudo Boolean constraints. The run takes approximately 20 minutes to verify that, for the last cell, the voltage drop is significant enough that the final SRAM cell is incapable of reliably storing a bit (the model predicts that cells that will always store one value or another are possible outcomes). This is consistent with the single cell test since the voltage droop is approximated to be down to .9v in worst case, allowing the lower edge of meta-stable to crash into the low-stable state. Although a sizable use of time, this verification is fundamentally a proof. In current design practice, enormous amounts of computing resources are spent on Monte-Carlo validation of memory designs. There are stories of entire floors of some well-known silicon processor houses having audible noise increase as every unassigned CPU is tasked with such modeling on closing time.

3.6 Application to pulse circuits

Developing an asynchronous pulse logic circuit, like the development of any ordered logical construction, benefits from using a characterized cell set. The circuit's systemic description hinges on the knowledge of critical pulse sensitivity, maximum pulse rates, and characteristic

pulse widths. A key application of SAT-based circuit analysis is to determine some of the bounds on behavior for the triggering and resetting actions of a pulse gate.

Figure 3.9 is repeated from 2.5; this gate will serve as the model for analyzing behavior within the cell set. Steady-state analysis of this gate cannot give all of the critical timing behaviors, but it can give insight into the bounds of intended operation, and entry into a meta-stable, poor performing state. The analysis will break into to operations, one where transistor A is active, the reset operation of the circuit; an the other where transistor B is active, the triggering action of the circuit. For the analysis here the worst-case gate (a 2-of-3 consensus voter experiencing a disagreement on its inputs) will be assumed for the pull-down network.

3.6.1 Minimum input voltage

The minimum voltage for an incoming pulse is the minimum voltage guaranteed to not cause a meta-stable state within the gate. Any pulse that cannot attain this voltage cannot trigger the gate under any circumstances. Meta stability is determined by taking the version of the gate in its pull down (B-Active) format and determining the minimum voltage at the input that can fight the keeper circuit out of the meta-stable region.

The first step is defining the meta-stable region. This is the region of values at the critical node that have an indeterminate current drive. For this analysis, $\pm\frac{35\mu A}{\mu m}\cdot nmos\ width$ was the current used; it is roughly 5% of maximum drive, and one of the largest current magnitudes that correlates to a input voltage that does not explicitly dictate the sign of output current. The SAT analysis suggests the region of voltages between .55V and .85V (rounded to the nearest 50mV) at the critical node should be considered meta-stable for the gate. A pull-down network must pull the gate below .55v for proper (deterministic) operation. In practice gate timing will

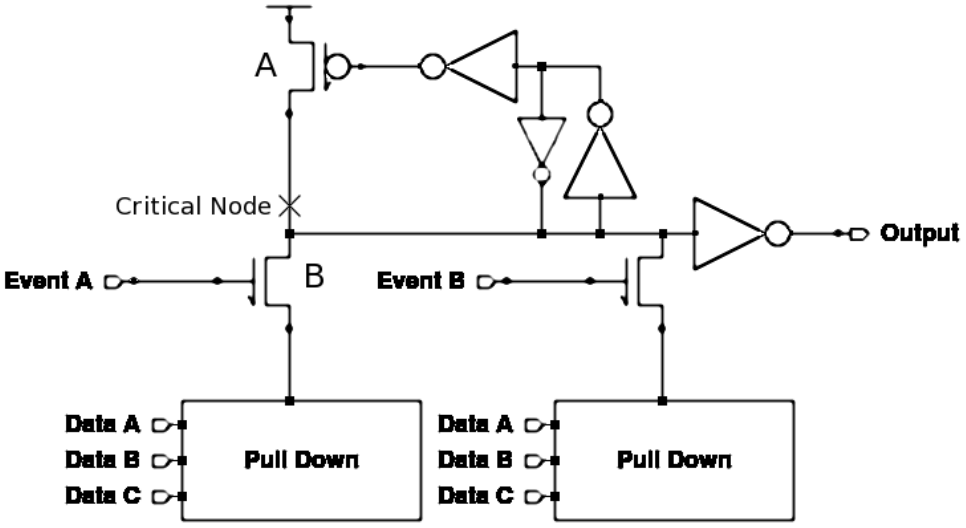


Figure 3.9: Drawing of a P-Gate. This gate is used to determine the characteristics of pulses within a gate cell set for construction of asynchronous pulse logic. By symmetry we will assume that events A and B will perform in the same manner given similar pull-down networks. Transistor A and Transistor B will be driven separately to analyze the circuit in reset and trigger operations

Table 3.2: Minimum pulse voltage to avoid meta-stability

	Keeper rel. strength	
Pull-down network	$\frac{1}{4}$	$\frac{1}{2}$
Just an Event	0.6V min.	0.8V min.
Event + 1 Data	0.6V min.	1.0V min.
Event + 2 Data	0.6V min.	1.1V min.
Event + 3 Data	0.6V min.	1.2V min.
Event + Vote(all in agreement)	0.6V min.	0.8V min.

be poor with a marginal pull-down; instead this internal voltage will indicate a threshold for detection in the internal node of the circuit.

The voltages used to describe cell behavior are the voltages presented to the front-end of the cell rather than the critical node voltage. Knowing the voltage needed to trigger the cell allows the derivation of the required pulse voltage to achieve this value. Using the critical node voltages from the 1st run; we use the solver to search for the minimum input voltage that will pull past the meta-stable threshold (.55v in this case). For a pull-down network that includes a single event and a 2-of-3 voter the resulting minimum event voltage is 1.1v in worst case, while it is 0.8V when all three inputs are in agreement. Table 3.2 shows a number of minimum pulse heights given pull-down networks. When the keeper strength is reduced; the minimum pull-down voltages for the tested networks all reduce to 0.6V. A small keeper is thus important if the gates are to be triggered easily and reliably.

Chapter 4

Implementation and Circuit Characterization

The development of a asynchronous pulse circuit requires a number of circuit components. As discussed in Chapter 2, there are a number of timing features of a cell that need to be characterized for the verification procedure. Here, specific decisions about wire size, current drive and transistor size are justified for radiation environments.

4.1 Optimal Wire Width and The Utility Metric

Full custom design gives a designer a large amount of freedom in wire width and spacing. Minimal wire delay tends to imply very wide wires – an intuitively impractical solution. Utility, the data rate per unit of wiring space used, is offered as an alternative metric. This metric maximizes the amount of data a given width of wire track can carry. As an added bonus

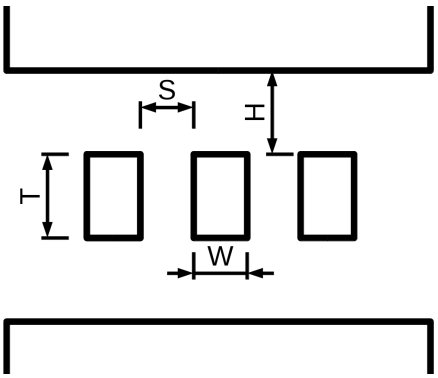


Figure 4.1: Critical Dimensions of On-Chip Wiring

most of the analysis is based on geometric arguments; process scaling does not change the basic trade-off.

Optimizing routing decisions based on this metric can lead to systems that can communicate upwards of 50% more data per second for a given width of routing track when compared to systems that use minimum wire size. Furthermore the optimal results presented are of practical scale, recommending trace widths and spacing slightly larger than conductive film thickness.

4.1.1 Wire Width and Spacing for Maximum Utility

Interconnect wires in an integrated circuit are a known limiting factor in the rate at which data can be moved on-die[30, 58]. Fine scaling of ICs has given small, high-resistance wires without a similar reduction in capacitance, leading to higher wire time-constants[30, 58]. In fact, wire bandwidth is known to be limited by the ratio of the wire size to its length even if resistive loss is not the dominant limiting factor [59]. In the case of custom ASICs, once the process is decided a designer has limited options for making larger cross-section wires; the film

thickness from which the wires are etched is set by the foundry, though usually a few options are made available.

Maximum utility of a portion of a metal wiring level occurs when the most data can be moved through that area. Given that the levels of metal wiring and dielectric are constant for a given process, the figure of merit for metal utility (in the context of data motion) is $\frac{Number\,of\,Bits}{Time \cdot Total\,Width}$. Finely dividing the available metal into a wide bus will both allow a large number of bits in parallel, and increase the total delay of each wire as capacitance of the structure increases due to the added inter-wire capacitance and resistance increases due to the loss of metal to the dielectric used to subdivide. Using the formulas from [60] to approximate wiring capacitance, and a resistance of $\frac{\rho \cdot length}{W \cdot T}$ we can estimate the maximum utility values for S and W given fixed T and H. Assuming the permittivity of the dielectric and the resistivity of the conductive film are independent of geometry, optimal values can be found independent of these constants.

4.1.2 Delay Based Metric

Link design, like many engineering problems, attempts to minimize the worst-case operating condition. In this case the total delay of the link in its worst-case is the cost function. Worst case delay occurs when adjacent wires are switching simultaneously and in the opposite direction causing the effective coupling capacitance to double. The total effective capacitance in this case is $4 \cdot C_{couple} + C_{af}$. The number of bits in a bus per unit width is $\frac{1}{W+S}$ and the time constant is $R \cdot C = \frac{\rho \cdot length}{W \cdot T} \cdot (4 \cdot C_{couple} + C_{af})$ giving the FOM in eq 4.1. Taking the case where $T = H$ the optimal values (solved numerically) are : $S = 1.273 \cdot T$ and $W = 1.844 \cdot T$; figure 4.2 shows the relative cost near the optimal point.

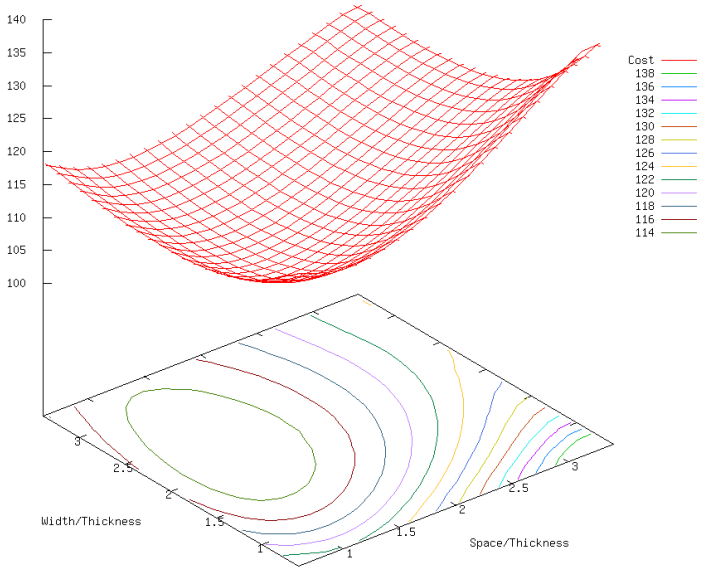


Figure 4.2: Relative cost (inverse utility) of different wire widths and spacing for the case where metal film thickness is equal to dielectric layer height

$$\frac{1}{\frac{\rho \cdot length}{W \cdot T} \cdot (4 \cdot C_{couple} + C_{af}) \cdot (W + S)} \tag{4.1}$$

4.1.3 Jitter Based Metric

Not all link technologies are limited by their total propagation delay. In a link it is common to use a strategy where the skew, or average time of flight, between the sending and receiving sides of a link is both known and compensated for. If a circuit exists that can predict when a new piece of data arrives, multiple pieces of data can be in transit between the transmitter and receiver of a circuit. The speed in this condition is limited by the capacity of the circuit to predict that time of flight. When considering the coupling capacitance alone as the source of this jitter, another metric for width and spacing can be developed. As in the total

delay metric, the worst case coupling sees $4 \cdot C_{couple}$ as the side coupling capacitance. In the best case delay time, none of this capacitance is apparent. The difference in these two cases is thus the $4 \cdot C_{couple}$ giving a cost function of

$$\frac{\rho \cdot length}{W \cdot T} \cdot 4 \cdot C_{couple} \cdot (W + S)$$

This Cost function does not have a minimum unlike the total delay cost function. This means that any jitter soultion must take the exact environment into account, so that power coupling noise, and the costs of adding extra buffers can properly factor into the wire spacing solution

4.1.4 Optimal Utility - Big Picture

The optimal utility metric is a useful way to determine appropriate resource allocation. Selecting wire sizes guided by utility will lead to solutions that have the narrowest wiring tracks for the relative required throughput desirable both in terms of cost of chip manufacture and the reduced routing distance wires orthogonal to the optimized tracks will require to bypass the optimized section.

4.2 Radiation Damage Modeling for Cell Characterization

Damage modeling is a critical part of properly characterizing cells for use in a radiation environment. The 130nm process used to construct the circuits discussed here has been well studied, as it is the same, or at least very similar, to the one studied in [20] and used to construct

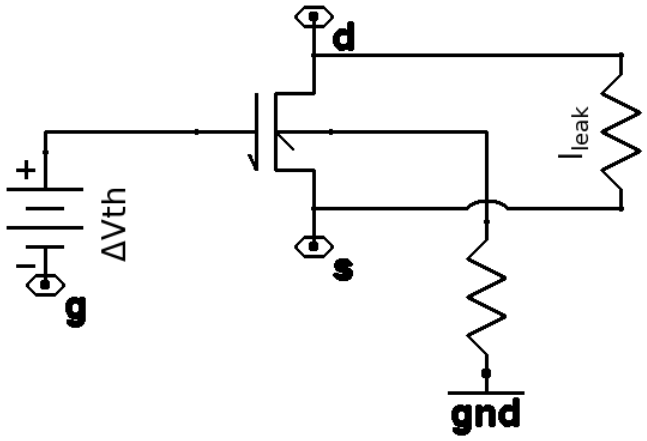


Figure 4.3: Electrical equivalent model for a radiation damaged FET. Key damage modeling includes the increased leakage primarily due to charge trapping on the sides of the FET, and voltage threshold shift due to damage to the main oxide layer.

circuits for projects at CERN. Creating correct worst-case timing models requires a means of incorporating damage.

Figure 4.3 Shows the electrical equivalent model used to adapt a foundry definition of a FET to the post-radiation damaged version. The key threshold shift and leakage channels are shown. The side leakage is modeled as a resistor instead of a current source, as there is evidence that lower V_{DS} leads to lower leakage, and a resistor is a reasonable model of this voltage dependence. Additionally shown in Figure 4.3 is the reduction in body conductivity which is expected to rise during radiation, and increase the transistor's sensitivity to substrate noise.

4.2.1 Gate Threshold Shift

The threshold voltage of MOSFET is a key parameter in FET models describes the voltage at which a charge inversion occurs in the channel and marks the onset of channel conduction. There are numerous methods of extracting this value from physical transistors, correlating to the different impacts the value can have on any given model[61]. Pulse asynchronous circuits are full-swing circuits, where event timing is critical; the effective shift in threshold is the one that helps determine decision voltages in the circuit. This means that the modeled threshold shift should make the overall FET model match roughly around $V_{DS} > \frac{V_{DD}}{2}$ and $V_{GS} \approx \frac{V_{DD}}{2}$, and consider a fit over a range, rather than a best match at a point.

Predicting behavior in an insufficiently studied process requires some reasonable and bounded guesses. One of the most important factors in using FETs in High-dose environments is the fact that there is a upper bound in damage to an oxide, and this upper bound is mediated by the oxide layer's thickness [62]. At the scales of readily available CMOS processes, field effect transistors can have a single worst-case-damage model, rather one that takes into account the extent of total environmental radiation. Additionally very thin oxides, on the scale of gate oxides, are much harder to damage than thick ones [11, 63] making core transistor channels relatively resilient. There still remains the issue of the effects of channel sides (STI processes) which will impact the core channel threshold and leakage characteristics[20, 64, 65]. From these charted impacts we can extrapolate the rough change in threshold voltage due to trapped charges. In the case of wide transistors in the $130nm$ process node, this value appears to be roughly $25mV$, and wide for this purpose appears to be roughly channel width 10 times greater than channel length.

4.2.2 Leakage channel

The parallel leakage is modeled with a resistor, as shown in figure 4.3. A current source would also be a logical choice for modeling this leakage, as it represents a parasitic FET channel that has been turned on by trapped charge. The choice of a resistor here is for pragmatic reasons; any model physical or non physical can be fitted to at least some region of operation, and in this case the resistor can be fit reasonably well. The resistor is taken to be a size such that the current through it closely follows the leakage current within the operating region of interest, the $V_{DS} > \frac{V_{DD}}{2}$ and $V_{GS} \approx \frac{V_{DD}}{2}$ case of approximate switching conditions for characterizing full-swing logic. The pragmatic value to the resistance is that under simulation resistors have stable behavior, unlike a dependent source, there is no error in fitting, barring a negative value, that can cause this leakage mechanism to appear as a power source. Additionally resistor models tend to be easy for most simulation platforms to handle mathematically.

The determination of the leakage channel value is fitted to account for the current that is not already accommodated in the V_{th} shift model (Section 4.2.1). The goal of the damaged FET model is to track worst-case conditions, even if maximum leakage occurs at a different dose than maximum threshold shift; the leakage resistor must model the additional current relative to the damaged shift value to create the damaged corner transistor model. This will be later reconciled by using Monte-carlo simulation that takes into account a radiation dose factor, and when these values reach their peak relative to total dose. To match the currents measured in the 130nm process [20, 66], a $300K\Omega$ resistor was used in the model. This size is somewhat too aggressive to match the actual process leakage, but lead to more conservative choices regarding compensating extra current.

4.3 Electronics for radiation test

Radiation testing was enabled by creating a number of specialized components that, while ill suited for production, aided in performing practical test. Two components in particular, the upset hardened current source, and the radiation monitoring ring oscillator, were key in conducting the first radiation characterization detailed in Chapter 6.1

4.3.1 Upset Hardened Current Source

Traditionally, analog circuits require bias generators and reference circuits to provide the nominal operating conditions for the circuit. Since reference circuits establish bias conditions for many circuits, faults at a bias generator become faults for a larger portion of the circuit.

Radiation induced upset is of particular concern for analog/mixed signal components for two reasons. First, unlike a digital system that can easily remove voltage noise that does not corrupt data, a continuous voltage, or small swing circuit can easily be corrupted by extra charge appearing at a node. Second, since the system uses continuous voltages it is likely that there are many constituent transistors that are not at their full on or full off configuration, leading to a limited drive strength to bring the system back toward correct operation.

The first circuit investigated to help with analog operations in an upset environment is the current source shown in Figure 4.4. This current source is closely related to a peaking current source, with a very simple cascode provided to the lowest current transistor. The goal of this added cascode transistor was to enable high current densities within the transistors, across the circuit. High current density enables a rapid recovery from upset. The high current density additionally reduces variation and sensitivity to radiation damage.

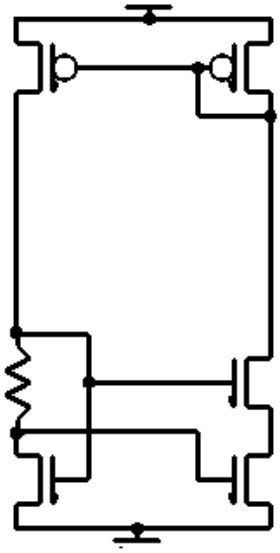


Figure 4.4: Upset hardened current source used for the first round of radiation test.

The relative sizing of transistors within the current source was optimized based on simulated performance. Monte Carlo simulation was used to capture the statistical distribution of the circuit behavior. This allows the projection of behavior over variation. A gradient decent optimizer was used to carry out this optimization. At each iteration the circuit was simulated with the given transistor sizes as well as an additional simulation for each transistor in the circuit. this additional simulation models the behavior with a small change in transistor size for the given transistor. The gradient with respect to size can be approximated by calculating the ratio between cost function and transistor size. The algorithm then adjusts the relative sizes of transistors and repeats.

Figure 4.5 shows the relative variation for simulated current in current sources over radiation. Transistors were grouped into ratios to better match the layout styles likely to be

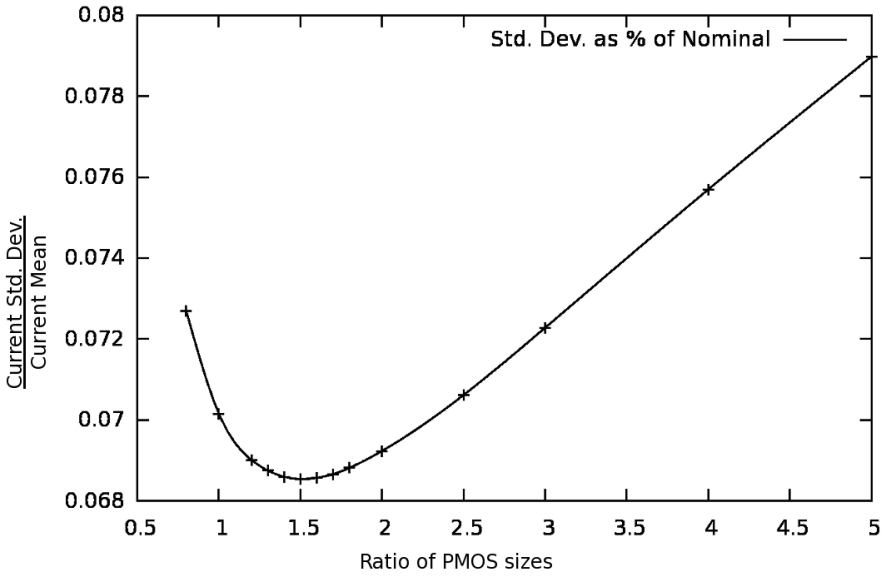
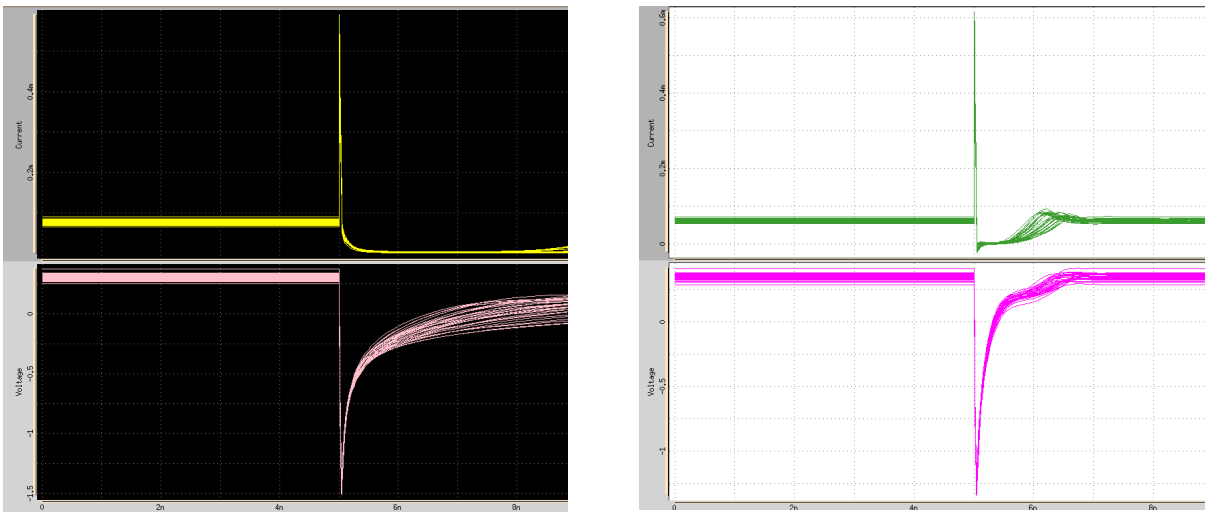


Figure 4.5: Current source variation vs relative sizing. This curve demonstrates that there is an ideal point in the relative sizing of two transistors. Many of the optimization operations were expressed in relative sizes of groups of transistors so that the resulting parameters matched designer expectations, and common needs for layout matching.

used in construction. In the case shown in figure 4.5, the curve is relatively shallow, but a 10% improvement in variation can be had by using the optimized version.

Upset in current sources is a more difficult problem, as the natural solution is to make as large and powerful of a circuit as possible thus overcoming any injected current. A practical design must also consider the trade-off in resources spent and advantage received. Thus a qualifier of recovery time for a fixed area allocation was used. Theoretically alternate qualifiers could be used such as $\frac{\tau_{upset} \times \sigma_{ref\ current}}{area \times ref\ current}$ would incorporate both variation and recovery qualifiers presented here. Figure 4.6 shows the difference between a circuit not optimized for upset, and one that was. In this case upset times were reduced by a factor of at least 5.



(a) Circuit not optimized for recovery time. Recovery time exceeds 5ns

(b) Circuit optimized for recovery time. Recovery takes between 1ns and 2ns

Figure 4.6: Simulations of bias circuit recovery from radiation induced upset. Altering the sizing of transistors in the bias circuit changes the upset induced behavior.

The circuit presented here does have a flaw in its construction: it is highly temperature sensitive. The optimization was carried out assuming a constant die temperature of $70°C$. In a cooled detector the structure would not function as expected. Since the circuits were used in practice in room temperature test, this error was never corrected.

4.3.2 Radiation monitoring ring oscillator

A custom built 27 stage ring oscillator was constructed and tested for the purpose of characterizing the behavior of logic in the 130nm process under radiation. A single stage of this oscillator is shown in Figure 4.7. The shown oscillator is designed such that the speed of the oscillator is determined by a radiation-compensated pull-down and a radiation sensitive pull-up. It is speculated, based on data from [20] that PMOS slow down is a greater challenge

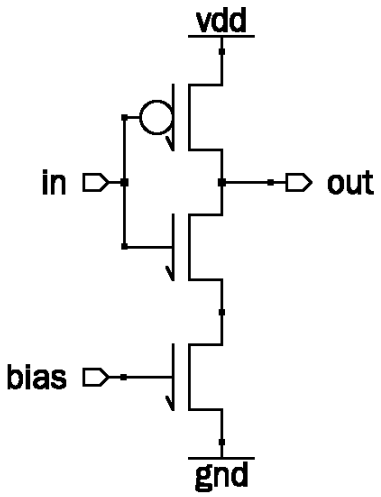


Figure 4.7: A single stage of the radiation test ring oscillator. Delay stage is compensated for drift in NMOS current, but not in PMOS current. Preliminary research determined that PMOS damage was more likely to cause slow circuit operation; this circuit allows the quantification of the slowdown.

for system timing, than NMOS speed-up. Since the structure of the oscillator is compensated on the NMOS side, testing will primarily indicate the slow-down due to degradation of PMOS transistors. Additionally the structure shown is sensitive to variation in voltage. This sensitivity allows more test operation, as the core voltage for the chip can be varied to test the oscillator. The downside of these sensitivities is that the oscillator cannot be reused in a non-test oriented application.

Chapter 5

Link Design

5.1 Architecture

Overall, the link is intended to behave like the familiar asynchronous FIFO (first-in-first-out) buffer. This style of design was chosen to simplify system integration and use as IP in larger designs for use in physics experiments. The design must survive very high levels of radiation, both SEU/SEE (transient soft upsets) and TID (aging and insulation charging as well as lattice damage and doping effects from induced traps). These effects are mitigated by a number of techniques including TMR (modular redundancy of I/O and state information), replication of function, temporal redundancy and device segmentation. As always, mitigation comes at a cost in power, area and system performance. Design choices were made to be as conservative as possible while still meeting design constraints. The high intended performance of the link forced many of these decisions.

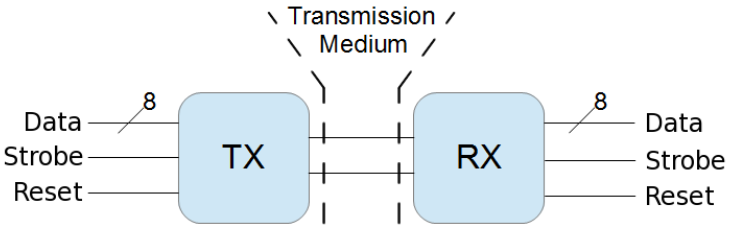


Figure 5.1: Interface of the transmitter and receiver. The interfaces on both sides have slow clock and data buses (all triplicated) as well as an asynchronous reset

5.1.1 System interface

On both the transmit and receive interfaces, the system presents a 8-bit data bus, a clock, and a reset line. All of these signals are presented in triplicate to meet the redundancy requirement. On the receive side, the data bus and signal dubbed clock are both outputs with the clock being the strobe recovered from the data stream. On the transmit side, the strobe and data bus are both inputs to the system again triplicated to meet the redundancy requirement. On both sides, the reset signal is an input and also triple redundant. The reset signal is used to put the transmitter and receiver into a known state at boot-up or in the event that a higher-level system catches an error. This system interface is described in figure 5.1.

5.1.2 Implementation of Triplication

Soft-errors are of a large concern for fast, small systems in radiation environments. In particular, for the 130nm process, single latches are known to be vulnerable to upset. Multiplicity is one of the few ways of storing information that reduces upset risk[67]. Because of this we incorporated triplication of information into the way our design works. This meets the multiple

redundancy requirements of our interface as well as hardening against upset. Figure 5.2 shows our majority logic protecting information at both ends of our system input and output of each of the serializer and deserializer blocks.

5.1.3 Clock and Reset Triplication

In this system all signals are triplicated including clock/strobe and the asynchronous reset. This means that voting on clock and asynchronous reset occurs within the system. Such voting may give many designers pause, due, in part, to concerns with adding logic to timing critical paths. Key to allowing both a voted high-speed data stream, and enabling triplicated, timing critical systems was the implementation of a glitch-free, fast voter. This voter consisted of a 2-2-2 AOI gate followed by an inverter wired to be a voter, as seen in Fig. 5.3. By using this same gate on all external signals, there is little issue of skew, since each data line, as well as each clock line, should see similar delay under normal operation. When one of the three lines disagrees with the other two, the worst case forward propagation is seen for the voter; simulation projected no more than 25ps of skew during a disagreement amongst the triplicated lines. This limited skew was achieved in part by the relative sizing between the AOI gate and the inverter.

5.2 Encoding

The link encoding is a two-wire encoding, where transits are allowed on only one wire at a time. A pulse on one wire is one combined data/timing symbol and a pulse on the other wire is another symbol. There is a '1' wire and a '0' wire; a pulse on the '1' wire indicates the arrival of a 1 bit; a pulse on the zero wire indicates the arrival of a 0 bit. This encoding makes the decoding state-machine very easy to construct, as well as making link signalling pre-emphasis an

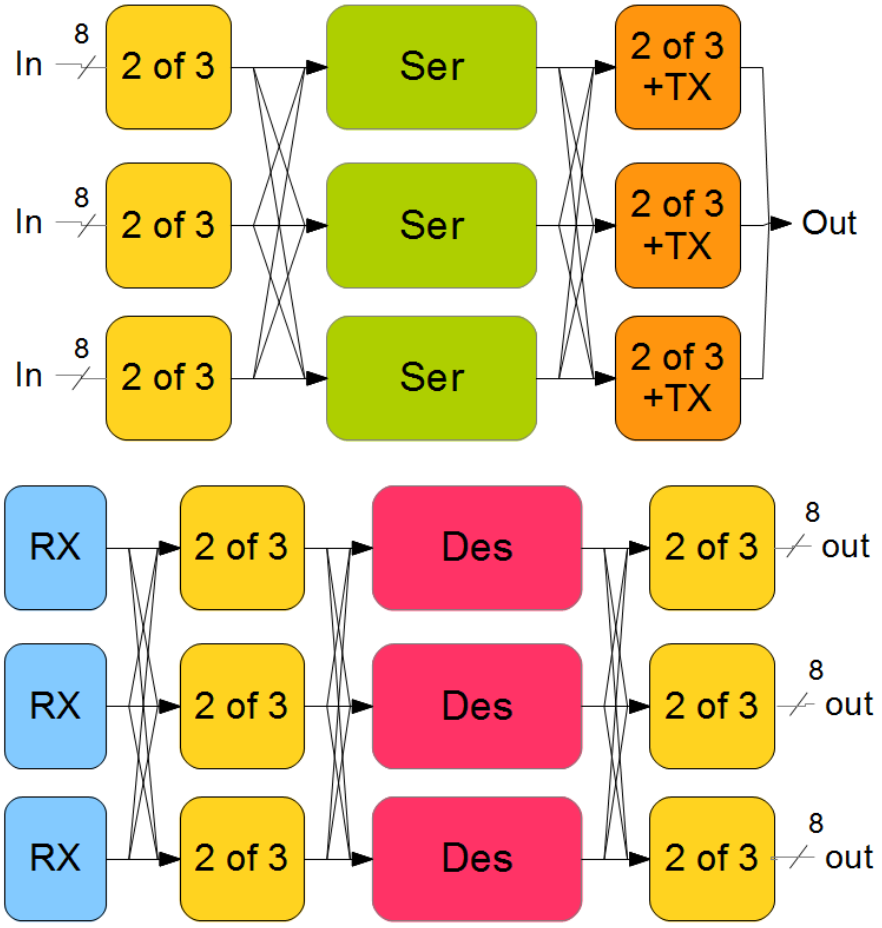


Figure 5.2: Triplicated transmit and receive units. Majority voters used at both the inputs and outputs of each of the Serializer and Deserializer so that all memory is protected.

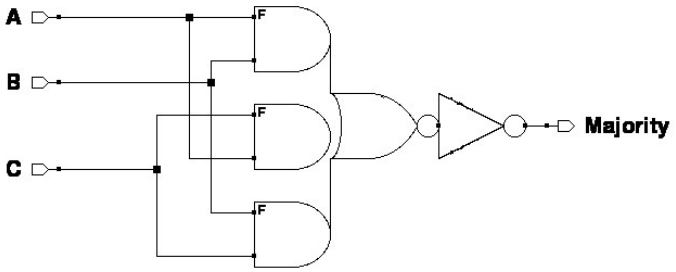
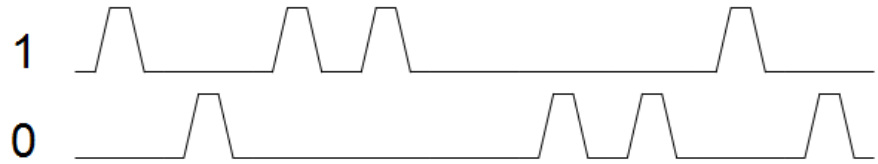
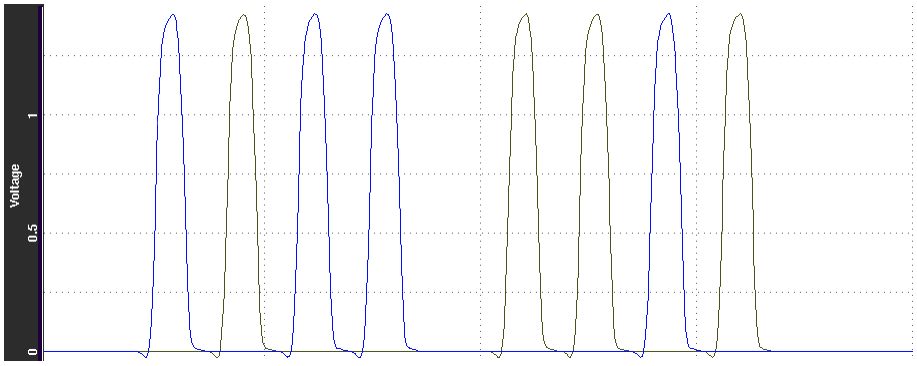


Figure 5.3: 2 of 3 voter topology used extensively in this implementation. This topology has a high level of predictability in terms of propagation delay, achieved by having control over the ratio between the first gate and second, lightly loading the signal with timing variance. Additionally this topology is not subject to glitching as there is no skew amongst signals internally. By using this block, voting on timing-critical signals is enabled.



(a) Idealized pulse stream showing 1-wire and 0-wire



(b) Simulated pulse stream; blue pluses are for the 1 line, gray for the 0 line

Figure 5.4: An example pulse-stream. Both 1 and 0 lines shown. These pulses encode 1011 0010 in two 4-bit nibbles. In the constructed circuit, the serializer sends 4 pulses per clock edge. This allows clock recovery to be a simple state-machine that toggles once for every 4 pulses.

easy task. Figure 5.4 shows the encoding. This encoding has been studied before for high-speed links; a skew compensation circuit was described in[28]. There was also an exploration of such signaling for fast buses in FPGAs in [7]. Both works were inspired by the use pulse surfing based pipelines described in[6]. This encoding has the advantage of easy clock recovery, similar to Data/Strobe encoding used in SpaceWire[68, 69]

5.2.1 Timing constraints for pulse encoding

The pulse system is created from the asynchronous self-resetting domino logic gates described earlier. [6, 4, 28]. While the transmit amplifier and receive amplifier are built with native pulse circuits, the whole system interfaces with normal logic, leaving the rest of the blocks to handle the interface between a pulsed system and one without.

The first set of constraints to consider are the practical ones of defining a pulsed interface. Each pulse must be both detectable and distinguishable from the next one. Detectability and distinguish-ability setup requirements on pulse width and pulse spacing. The minimum pulse space is set-up by the time it takes the detector to reset it self. While this delay is detector-dependent, the minimum detector delay can be roughly approximated as 3 inverter delays, as the fastest possible detector will need two inverters and a reset transistor in its feedback path. Pulse to pulse spacing that is less than this delay will result in reduced detectability. Specifically pulses that have less than this requisite quite time before arriving will have a portion of the pulse ignored due to a on-going a reset from the previous pulse. During the design phase, detectors with intrinsic reset times on the order of 80 picoseconds worst-case were realizable in the selected 130nm process.

The issue of pulse width is a more difficult one to specify, as larger pulse-heights translate into easier detectability, as well as wider pulse widths. If large pulse amplitudes are acceptable, other limitations come into consideration. The packaging and transmission medium thus also play a critical role in setting minimal acceptable pulse width. This pulse system targeted the QFN packaging system, since it is both inexpensive and high-performance. The draw-back of QFN, from a performance point of view, is that it uses bond-wires for signaling,

limiting the available bandwidth for signaling.[1] Within the confines of this package, rise and fall times faster than 40ps were difficult to achieve within a reasonable power budget and layout. Thus the transmitting pulse width was set to be above 80ps, with an amplitude set to allow for detection of an attenuated and noisy signal. In cases where package-limited speed was not required, longer pulse times could be used, with a voltage vs. detectability trade-off that results in lower operating voltage.

5.2.2 Transmission Medium Considerations

The pulse encoding system, when compared to a traditional DC-balanced differential encoding, has additional requirements on the transmission medium. Like true differential encoding, this pulse encoding requires two wires but the system does not naively operate on a differential pair. The system in present form is implemented as two wires that are both ground-referenced. There are a number of consequences to this signaling that are less of a concern with traditional differential encoding. The design terminates both lines to ground, which means that in operation all of the current is sourced at the transmitter and sunk at the receiver. This motion of current could conceivably cause an issue for sensitive power grids, additionally requiring a solid connection from ground on the transmitter to ground on the receiver. The advantage of ground termination is that without a voltage reference supply two important features can be implemented: first that there is no power dissipated in the termination and second that a very compact and simple termination scheme can be implemented on chip.

[1] The issue here is inductive effects on the bond wires which run about 1nH/mm. An alternative packaging strategy with much lower inductance and potentially higher performance is bump bonding. We chose the wire bond system to lower the cost and lead time of prototypes as well as the cost of test equipment.

The pulse signaling system can be used on common shielded twisted pair (STP) cables despite the differences between pulse encoding and traditional differential signaling. The adaptation for pulse encoding would require a termination scheme that matched both the single-ended and differential modes for the cable. With the existing designs that could be achieved by adding a single resistor between the two signal lines. This simple modification serves to help reduce noise on the line, since the differential mode in a shielded twisted pair line has lower impedance than twice the common mode, which is the case that the receiving amplifiers are designed for. By designing for the case where each wire was independently shielded, both applications could be targeted with the simple addition of a single resistor.

5.3 Link Latency

In the presented system, propagation delay is kept to a minimum. This reduction in delay is one consequence of the design decision to keep the link simple. The serializer and deserializer are designed to use a minimum of time to send and receive data. Both serializer and deserializer use half of a cycle of the word level clock to buffer incoming and out-going data respectively. The data is re-timed on the interfaces of the serializer and deserializer to buffer external timing concerns from internal timing. This extra buffering takes a full cycle, 1.6ns for buffers. The serialization and deserialization process is designed to take almost all of the word-level clock cycle at maximum rate. The first bit exits the serializer within 200ps of the start of the serialization procedure. Due to the design of the serializer the time for the first bit to exit is tied to the rate at which the serializer can operate. Even though this delay time is process and damage dependent, this margin is predicted to be held. This means that the

serializer adds .2ns. The deserializer consists of a multiplexer and a buffer that fills during the cycle. This architecture gives a few hundred picoseconds of propagation through the receive amplifiers. In specific stack has three levels of circuits that have a forward propagation limit of 120ps in a working 5Gbps system. Thus these stages take .36ns in worst case. The buffer takes a full cycle to fill by definition, thus a buffer meeting spec will add 1.6ns of delay. This gives a whole system delay of:

$$Time\ of\ Flight\ + 3.76ns$$

5.4 Design

Both the serializer and deserializer work on the concept of a cascade of cells that activate each-other in sequence. For short bursts, the serializer will operate as fast as the individual cells can fire, slowing down to match the data rate of the transmit clock. This slow-down occurs on each clock edge, giving the characteristic bunching of the transmitted pulses. The Deserializer takes advantage of these gaps, fitting the recovered clock edges into these spaces. While a clock is used for this example interface, an alternative asynchronous hand-shaking can be used at the ends for a full asynchronous FIFO interface. Given the design of the system, the maximum clock rate at transmit is set not only by the maximum rate that the transmitter can support but also the maximum rate the receiver can support.

5.4.1 Serializer

The serializer is constructed of cells as shown in Figure5.5. The serializer passes data until the rising edge of the enable signal, at which it creates a pulse corresponding to the bit

loaded into the cell. The cell creates a delayed copy of it's enable; This enable is sutiably delayed to create the create a correct spacing between pulses – that is the time for the pulse width of the serializer cell plus at least a minimum pulse-space worth of time. The relative delay between the delay circuits and the pulse-generating circuits is critical and care must be taken so that the delays of these two systems track with each other. This imposes layout matching constraints on the cell. The serializer action is thus dictated by this chain of cells and events. The serializer can function properly because the forward delays of the enable signal are laid out such that they will be longer than the cell takes to generate a forward going pulse. The pulse shape is derived form a sub-section of this delay line to further reinforce this relationship – that the generated pulses are appropriately sized and spaced. Figure 5.5 shows the pattern for the whole system two groups of four cells that have this forward propagating logic. One group of four operates on the rising edge of the clock, while the other group works with the falling. This helps the clock regeneration be more symmetric (since the original pulse pattern is similarly timed from the rising and falling edges of the clock).

5.4.2 Deserializer

The deserailizer has a cascade of cells, schematic in Fig. 5.7, similar to the serializer with the exception that instead of a delayed enable triggering a cell, the capturing of a bit triggers the next cell. This again creates a cell with a number of timing-critical details. The detection of capture time plus the enabling time of the cell must be at least as long as the pulse-width of an incoming pulse (to prevent double-capture) but must be no longer than the pulse width plus the pulse spacing (to prevent missing the next pulse). This narrow window in time is easily met, because the serializer system sits after a pulse-shaping stage. The system

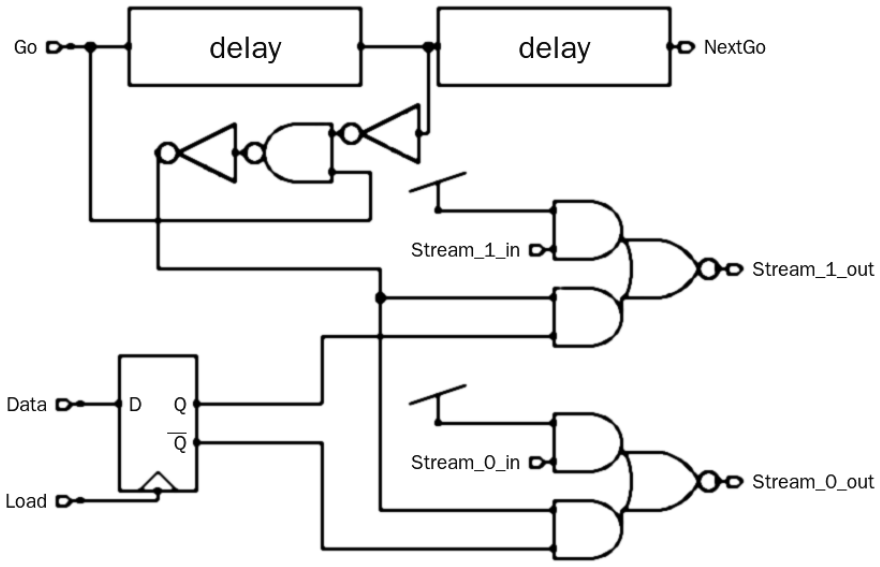


Figure 5.5: Serializer Single Cell. The cell allows data from previous stages until the enable signal from the previous cell allows it to transmit. The first cell in the chain is triggered by either the system clock, or a data valid signal in an asynchronous system.

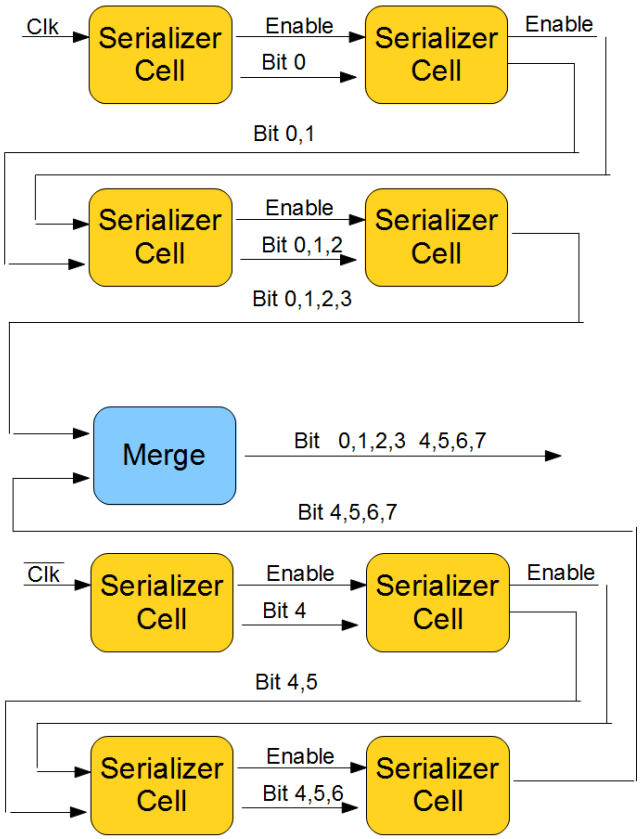


Figure 5.6: Serializer architecture; showing the two banks of 4-bit chains of serializer cells. The two banks operate off of opposite phases of the clock to evenly spread the pulses between clock phases. This compensates for the self-timed behavior of high-speed operation

guarantees that the pulse width is three gate delays, the loop between detection and the next detection stage is four gates long (as shown in Fig. 5.7). The deserializer's function thus requires that the gates in the deserializer track with the pulse-shaping gates in terms of behavior.

The clock recovery behavior of the deserializer comes from separating the deserializer into two equal-length banks. A toggle latch is used to track which bank is currently being filled by the bit-stream and which bank is static. The toggling of this latch generates a signal that can function as a clock signal for the purposes of latching data. This clock signal has all of the inherent jitter of the data system, as there is no PLL to average the jitter across multiple cycles. Though the resulting singal may not be appropriate for all possible clock applications, this is an extremely small state machine for clock recovery.

5.4.3 Transmitter and Receive amplifier

The transmit and recieve amplifiers are made out of self-resetting circuits as these circuits are very good at maintaining the characteristics of pulsed systems[**?**]. The voltage of the pulse at the reciever was selected so that a high-rate, low power reciever could be constructed. Large voltage swings allow a large difference in current between on and off states. The specification of signaling between ground and >450mV was selected for this implementation. Because the system is ground referenced, and a low-current idle was desired, a low-threshold nmos was selected for the input amplifier stage. with a small-bias current, a fast low power front-end amplifier is created. The ground referenced voltage also allows for a 0 dissipation transmitter without needing an additional termination reference source. The ~500 mV swing height was chosen for two reasons. First 500mV is sufficiently large to operate the receiver as a switch rather than a linear amplifier. Second the 1.5V power rail could be used to directly drive an

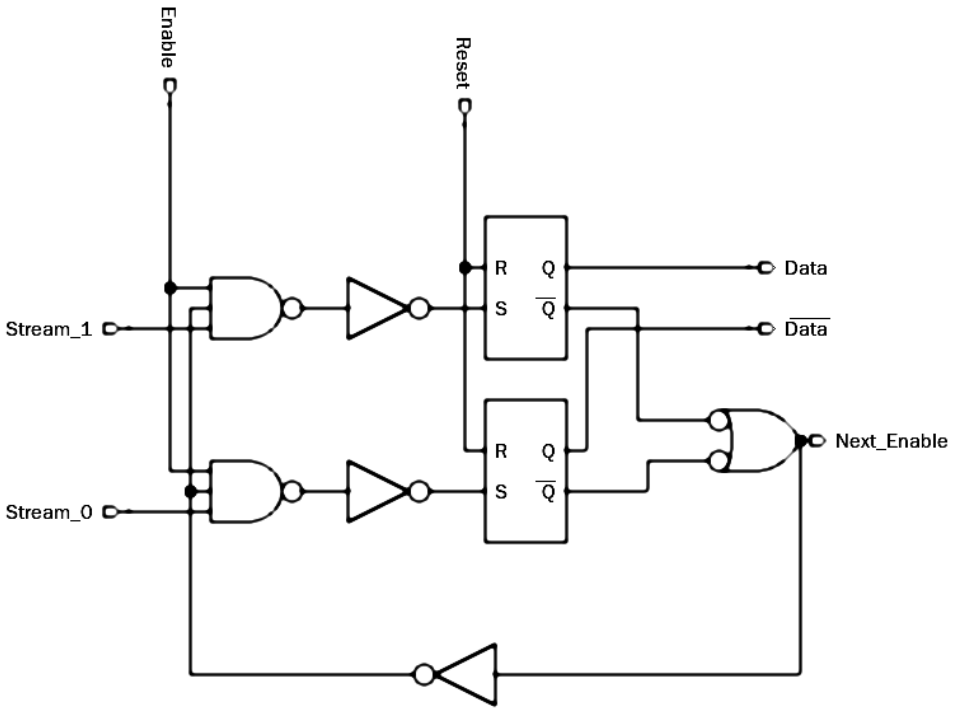


Figure 5.7: Deserializer cell schematic. The deserializer should have its first cell normally enabled, the enable out of the last cell is a data valid signal for the system. Alternately, in a clock recovery system, the receiver is broken into two equal banks. A toggle latch is used to generate the clock. One bank is enabled on each phase of the clock, with the bank done signal causing a toggle. This is the extremely small state machine for clock recovery that makes such a system tractable

output at 750mV through a matching network; allowing for both margin in the signal and a mild form of pre-emphasis. In short the transmitter and receiver are correctly sized CMOS inverters and have similar characteristics, including limited dissipation when not switching.

5.4.4 Driver segmentation

Most of the circuits in the link are built with triple redundancy for hardness against upset. The driver stage couldn't easily be made triple redundant, since there was a single pair of wires to be driven. Instead, the driver circuit was designed as a set of 12 segments to survive single event upsets/transients without failing to meet desired specification. Each segment of the drive circuit is connected resistively to the output, shown in the schematic in Fig. . Each segment takes triplicated input, and either drives a shaped pulse or not. In the event that there is an upset in any of the 12 driver slices 11 of the slices will have the correct behavior (either driving or terminating the line) and the upset slice will have incorrect behavior. In the case that the errant slice is falsely driving a driver with 1/12 of the rated current will drive into both the line and a terminator that is at 12/11 of the line's rated impedance resulting in an erroneous pulse 1/23 (~4.3%) of the full voltage – an acceptable amount of noise. In the case that the slice in error is falsely refusing to drive a pulse then 11/12 of the current is going into both the line and a terminator at 12 times the line's impedance in which case the output pulse has 11/13 (~85%) of the rated voltage – a worse but still acceptable amount of noise for the driver.

5.5 Layout

The layouts of these systems are very small ($100\mu m \times 200\mu m$) in addition to the required bond-pads. In addition to the described system a very fast random number generator

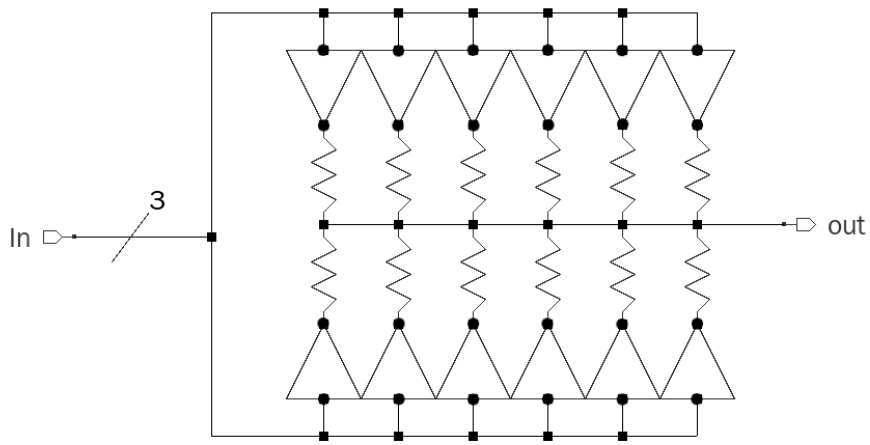


Figure 5.8: Segmented driver schematic. By splitting the driver into multiple segments and resistivly coupling those segments, the worst case of single event upset can be compactly analyzed. A 12-way segmented driver with matching resistors has a worst-case output at 85% of rated voltage in the event of an upset.

was created to feed test-patterns into the system. The number generator (also designed to be upset hardened, using similar techniques) occupies much of the extra size in the transmitter layout while the receiver has no extra attached hardware. The layout of the transmit system is shown in Figure 5.9. Key elements of the transmit system are highlighted such as (E) The random number generator (F) the logic handling the test-mode (D) the serializer (C) the 12-way segmented transmit amplifier (B) the ESD Clamps and (A) the bond pad. This should give reference to the relative sizes of all of the involved components. The entire system is designed to be as compact as feasible, to enable easy integration into other systems. Future work on the layout may see some further integration of the serializer into the pad area, making an even smaller layout. The layout of the receive block is shown in Fig. 5.10 and similarly consists of a (A) bond pad and (B,C) ESD protection, additionally (D) receiving amplifiers and (E) deserializer. The deserializer has no provision for special test included in the layout, and can also see easy integration due to its small size. Again future developments may see some of the logic block incorporated into the pad region, allowing for a much more dense system.

5.6 Predicted Performance

Performance estimates were generated using post-layout extraction and the models of expected radiation induced damage. The characterization relied on Monte-carlo simulation to verify that many different stages of damage and process variation were considered. This methodology renders the best available approximation of the system performance.

The whole link system (minus the test generator) uses on average 60mW (40mA @1.5V) when operating at 5Gb/s – the average maximum operating frequency was predicted to be

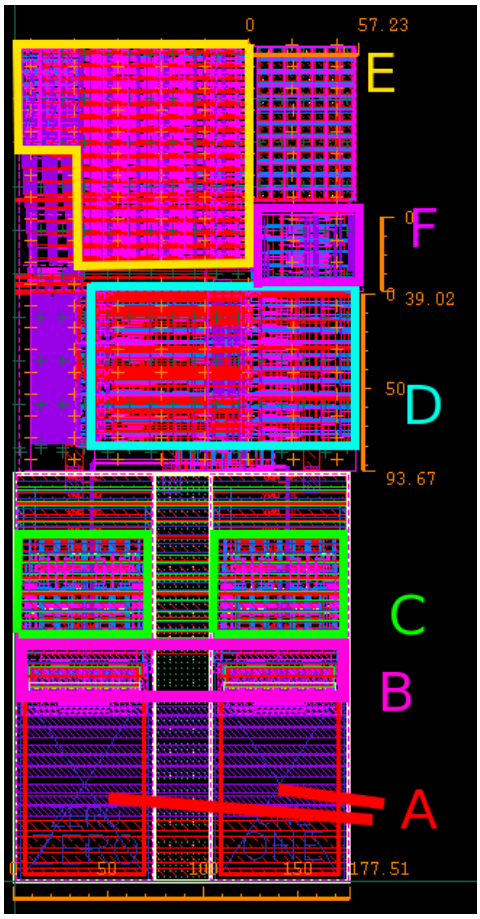


Figure 5.9: Transmitter layout. Labeled sections: A marks the bond-pads. B refers to the ESD protection diodes. C is the 12-way segmented drivers, allowing for high-speed and hardness against upset. D is the serializer. E is random number generator used for testing; it implements a 17bit linear feedback shift register in triplicated logic that can generate 6+Gb/s. F is a state machine to approximate an as-fast-as possible clock.

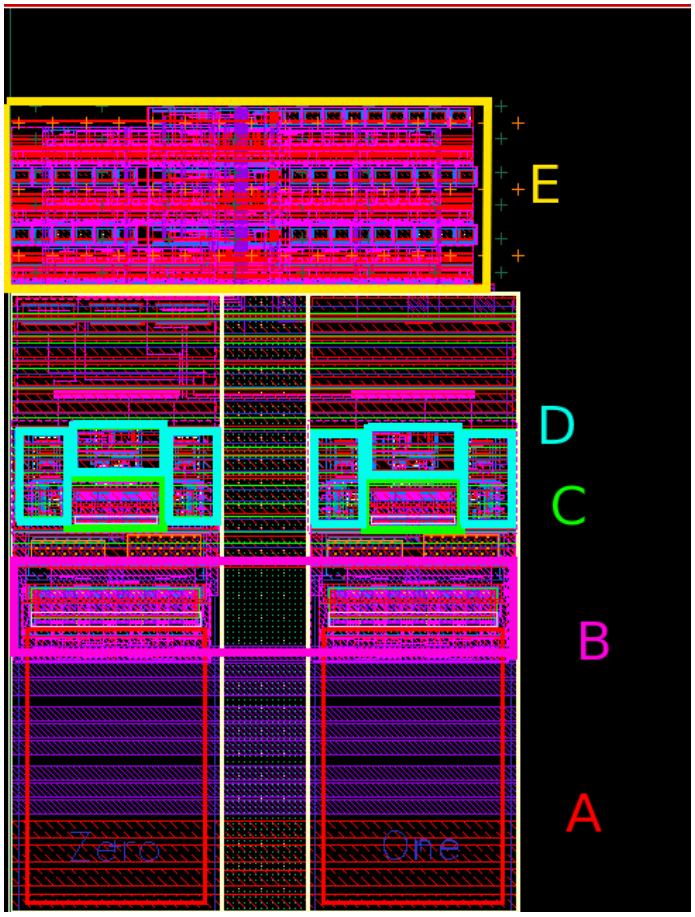


Figure 5.10: Receiver layout. A is the bond pads, B shows the HBM ESD diodes. C is the Termination network and CDM ESD protection. D highlights the triplicated receive amplifiers for each channel. E is the full deserializer, including clock recovery state-machine.

5.1Gb/s. Of the estimated dissipation 35 mW was estimated for the transmitter, 4.1mW was estimated for the receiver. The combined power estimate for the serializer, the deserializer and the random number generator is 46mW with 25mW taken by the random number generator and roughly 10mW each for the serializer and deserializer. This represents a large amount of power savings compared to other efforts at serial link technology in 130nm, notably the serializer and deserializer in CERN's GBT consume 330mW and 450mW respectively[**?**], and run at a similar rate.

Chapter 6

Test and Data

6.1 Radiation Testing and Characterization

Limited radiation testing was preformed to confirm the general design approach of this work as well as to quantify expected damage. For the test a number of candidate structures for high-speed serial links were generated and fabricated in the 130nm process node. Chief amongst these structures were I/O driver and receiver pairs, an upset hardened current reference, and a digitally controlled ring oscillator. The design of these components is described in Chapter 4.3. The ring oscillator was specifically structured to assist in gathering test data on chip sensitivity to radiation.

The current source and ring oscillator were tested in a x-ray radiation machine housed at CERN. This machine produced a well-calibrated stream of 40keV X-rays and a week's worth of machine time allowed multiple samples to be given multiple MGy of total dose. The highest tested radiation dose was 4.5MGy – above the expected life-time damage for the highest-

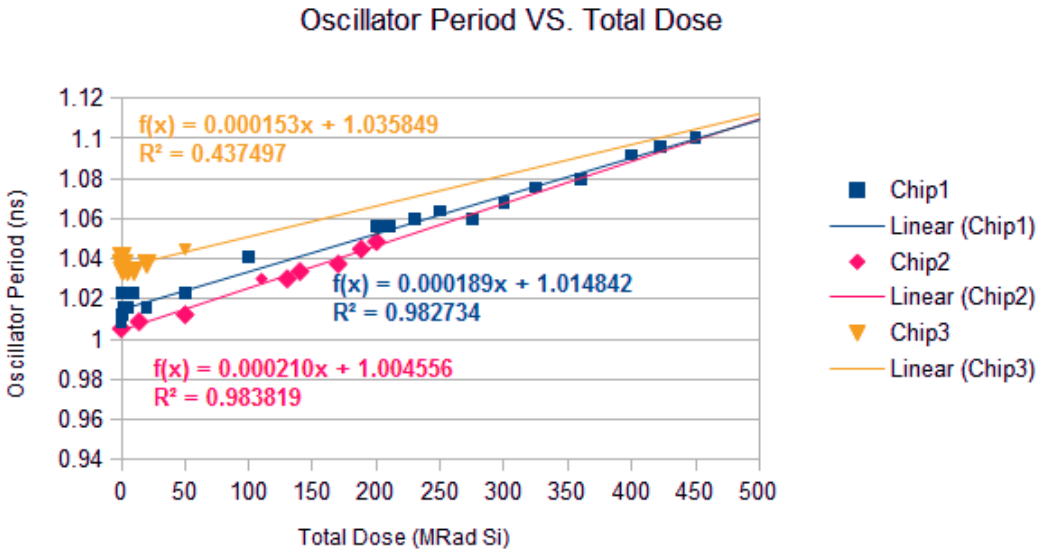


Figure 6.1: Radiation sensitive oscillator period VS total dose

radiation parts of the high luminosity LHC, a forthcoming upgrade to the current LHC configuration.

Figure 6.1 shows the observed rate of slow-down for the radiation sensitive oscillator in the presence of high total doses of X-rays. The total slow-down of the oscillator in this case is 10% over the expected life-time total dose for an integrated circuit mounted in the highest radiation areas of the LHC. This value give promise to the notion that designs can survive the radiation environment, and that an appropriate design margin can be constructed. An interesting feature of Figure 6.1 is that over the total dose of radiation, the measured periods of the three different oscillators converged. This suggests that at least some of the manufacturing variability is being made more consistent by the radiation bath. While this might mean that devices are more consistently damaged, it is a good sign for the overall margining process, as manufacturing variance is causing a converging rather than diverging total device variance.

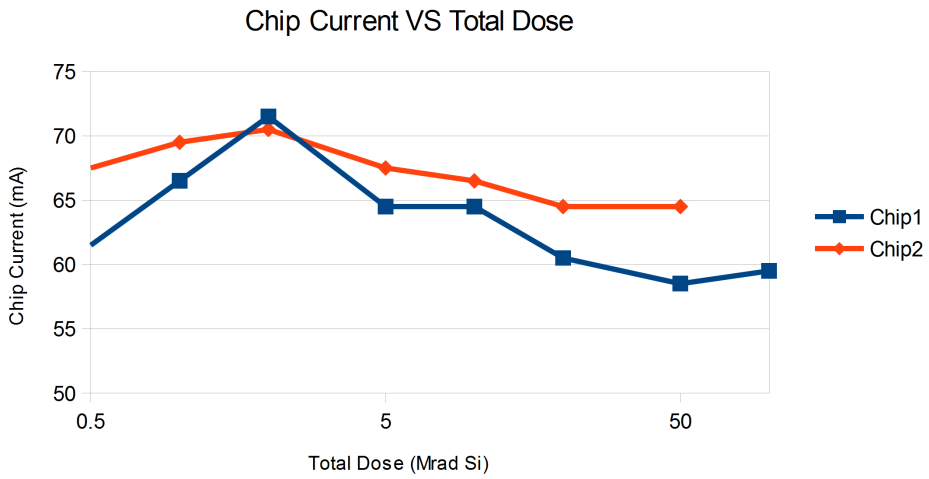


Figure 6.2: Leakage current vs Total Dose for two test chips. The results here suggest that the measurements of transistor leakage in [20] apply to the design styles used to construct the test chip.

Figure 6.2 shows the observed system current during radiation test. This current shows an early peak at 3Mrad (30kGy) indicating a peak leakage current for the system. This test verifies the applicability of the leakage models from [20] to the design styles utilized in this test chip, and carried throughout the work. The peak in leakage represents less than 15% extra current flowing, and contributes to simple design marginalizing since there is a clear worst case in damage.

6.2 5Gbps Link Testing

Before going deep into the analysis of the 5Gbps link it is important to note that a bug was discovered in the physical implementation of the link. Currently, this bug has been localized to the function that copies data from the random number generator into the serializer.

The timing verification of the block that handles this task was computed by hand rather than by computer, under the assumption that the timing was simple enough to not warrant exhaustive study. This decision is not only a mistake in retrospect, but also serves as impetus for using automated verification – a human being, this author, saw a handful of working test-cases and assumed that the hand calculations were correct. A computer given this task could be more exhaustive in its search, more strict in the application of appropriate timing margins, and prevent this outcome.

Even with this bug, the asynchronous serializer and random number generator were capable of creating a pseudo random pattern that could be observed at the output of the transmitter logic. This output did, indeed, consist of a set of very narrow pulses at high rate. Figure 6.3 shows a sample stream operated in the chip's slow (4Gbps) mode for clarity. It is important to note that the apparent variance in height amongst the displayed pulses is due to a beat pattern with the oscilloscope's sampling frequency of 40GS/s - a pulse is characterized by only 10 data points in this display, making it unlikely that all points will land sufficiently close to the peaks of each pulse to not witness this effect. Section deals more exhaustively with the characterization of the pulse

Despite the high-jitter nature of the serializer, and the added noise from the sampling of the scope, the circuit behavior can be demonstrated to be very stable. Figure 6.4 is an equivalent to an eye diagram for the pulsed signal signal; sampling is triggered by the rising edge of a single pulse and allowed to continue for 3ns before the system is set to re-trigger on the next incoming pulse. This allows a view similar to the one a pulse circuit would take – the timing is relative to a triggering function based on one pulse. Pulse heights in this diagram are very consistent, confirming that the variability from figure 6.3 on the following page is due

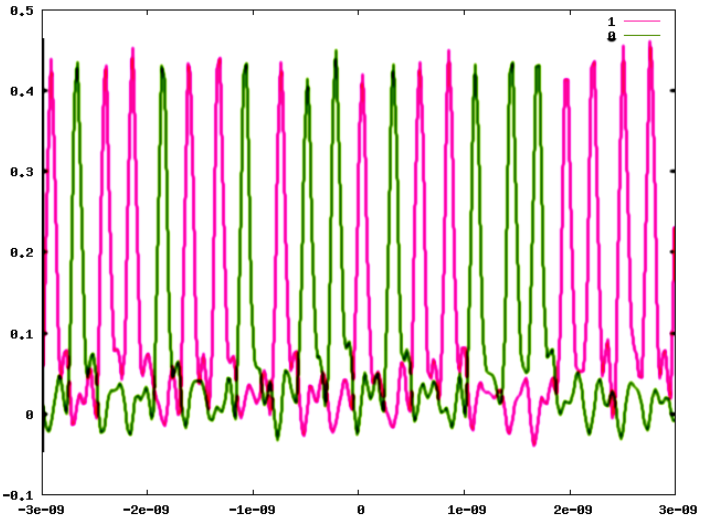


Figure 6.3: Measured output stream from a test chip containing the first generation of the asynchronous pulse link

to a limitation in the oscilloscope's sampling ability. It is also very clear that the period for a 8-pulse packet is very stable despite the high-jitter design of the serializer used to create the pulse stream. Even for cases of high pulse position uncertainty the pulses are individually well defined as is demonstrated by the small handful of stable pulse shapes relative to the triggering pulse.

The stability of the 8-pulse packet begs another investigation: the relative arrival times of pulses within the packet. The timing within the serializer is taken from a series of non-calibrated delays. Packet to packet stability indicates that this 8-bit delay is stable on its own[1]. Not only is the delay of an 8 bit packet stable, but also the arrival times of individual bits originating form the serializer are highly predictable. This can be demonstrated by taking

[1]The source of this stability has not yet been decisively ascertained, but the symmetrical nature of the circuit design, the existence of frequent voting in the timing chain, and the constant current draw of the circuit as a whole (in its test mode) can all contribute to the predictability of timing with out being unique to the presented technology. Unique to the presented circuits is the intrinsic sensitivity to a limited set of transition cases – in effect creating a narrow-band filter around the operating frequency when connected in a self-timed loop

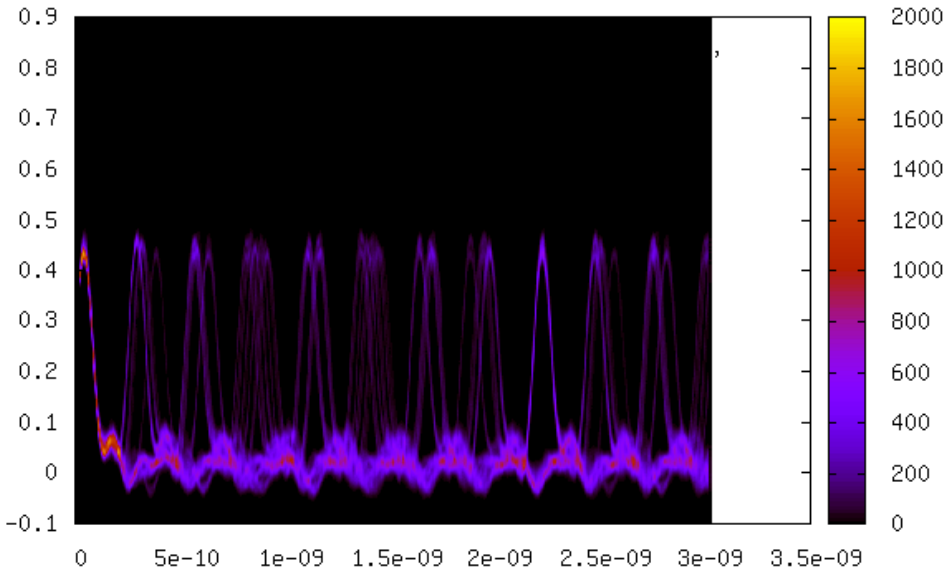


Figure 6.4: Pulse Eye Diagram for the 4Gbps mode of the pulse link. Important features of the diagram include: the high stability in 8-bit periodicity, the confirmation that variance in pulse height is due to a sampling beat pattern, an estimation of the noise in a pulse system – the uncertainty or fuzz demonstrated at the bottom of the diagram.

a measurement of the positions of a pulse relative to the start of a word. When measuring from the bench, a word-start is demarcated by the first pulse after longest pause between pulses in a stream of 8[2]. Figure 6.5 contains the results for a nominal run of the serializer at its fastest settings. The data shows time since the last pulse of the previous word. All pulses, not just the first pulse, show up at very predictable times. It can be seen that pairs of pulses tend to be close to each-other in time, despite the electronics being explicitly designed to operate on groups no smaller than four. This non-linear effect was predicted in simulation and similar multi-period pulse phenomena are known in other pulsed systems, such as neurons [70], and pulse lasers[71, 72].

6.2.1 Output Pulse Characterization

Characterization of the output driver took place using an Aglient (now Keysight) DSA90804A Infiniium 8GHz Oscilloscope. The cabling between the scope and the PCB pigtail had a measured roll-off 3db frequency of 7GHz (measured on higher speed equipment). The pigtail and PCB are not characterized directly, but are calculated to have limited loss (in part due to small size) and matching within 10% (based on manufacturer specification). This is all to say that there is a strong limit to the accuracy of reported numbers, due to the short pulse duration and high frequency of operation. These measurements do, on the other hand, give a good idea of what a receiver circuit would see at the end of a practical cable. Table 6.1 describes the test-cabling set-up including the package parasitics.

Pulse peak voltage is one of the critical measures of the pulse. The peak voltage can be directly measured off of a oscilloscope waveform, but the voltage measurement is complicated by

[2] This is a feature of the serializer created to allow for the deserializer to latch the current word and reset.

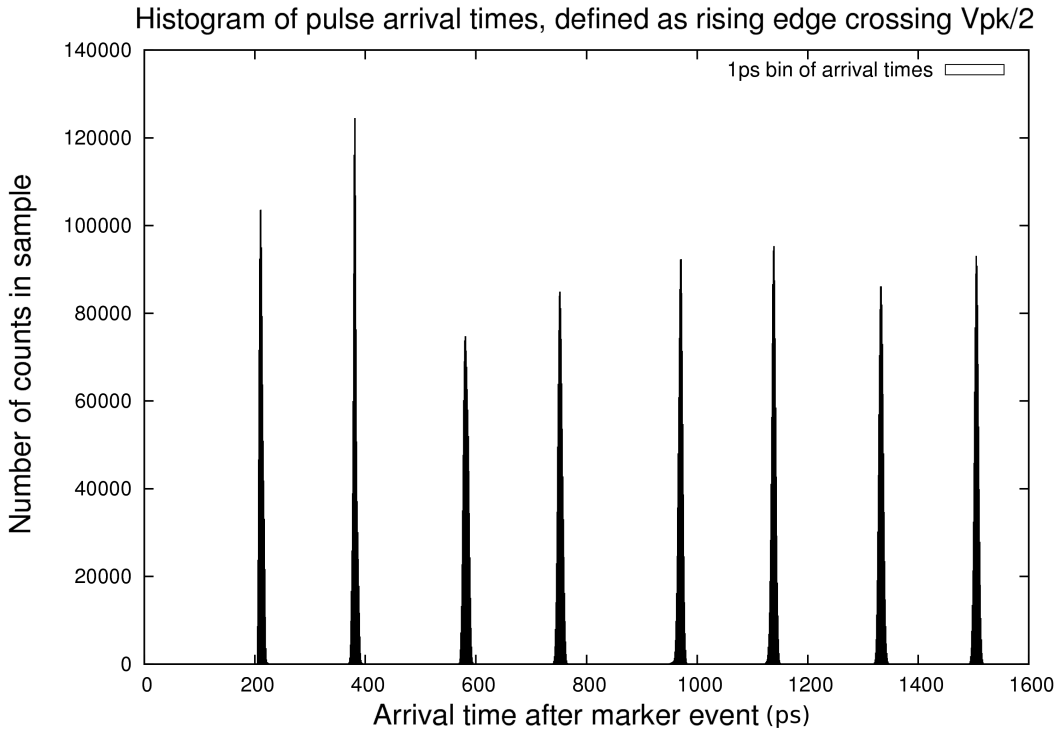


Figure 6.5: Arrival time histogram for a full 8-bit word. Arrival times were determined by rising edge crossing Vpk/2 (250mV) as outlined in section 6.2.1. The histogram shows that the relative positions of pulses within the stream are very stable (vary by only a few ps). Successive pulse periods are not exactly regular, the histogram for these can be found in Fig. 6.8

Table 6.1: Test bench cable configuration in order from test chip to oscilloscope

	Length	Impedance	Loss	Notes
Package	1.5mm	$63 - 73\Omega$	1Ω	Bond-wire and copper landing based on models in[73, 74, 75]
PCB Trace	1.5cm	$50 - 60\Omega$	$.1\Omega$	Microstrip 15 mil wide, 10mil FR4 (4.0-4.4 ϵ_r, 8mil-11mil thickness)
Pigtail	10cm	$50 \pm 2\Omega$	$.04\Omega$	1.37mm coaxial cable UFL to SMA
Cable	1m	$50 \pm 2\Omega$	$.1\Omega$	RG-174, SMA-SMA

not only the loss in transmission but also the limited sampling frequency of the oscilloscope. The oscilloscope samples at 40Gs/s, a 100ps wide pulse will be characterized by only about 5 samples (4 time periods). Assuming ideal sampling and a linear rise time this would result in $\pm 6.25\%$ in the measurement (an accurate measure to a loss of 12.5%). Figure 6.6 shows the measured distribution of measured peaks form the pulse system running at peak rate. Data collection ran for 1.28ms rendering just over 6.8 million pulses counted. This distribution is nearly a Gaussian distribution, though slightly denser on the lower-valued side of the distribution. The distribution has an average value of 499.6mV and a sample standard deviation of 18.7mV.

Pulse width is the next critical measure of a pulse. Using the same data set as was used to generate figure6.6 is again studied to determine pulse width and spacing as seen from a potential receiver. A threshold value of 250mV, half the pulse height, is selected as a threshold value to measure pulse width. Because sampling points at exactly this value are rare, linear interpolation between two sequential points is used to determine the crossing point for the rising

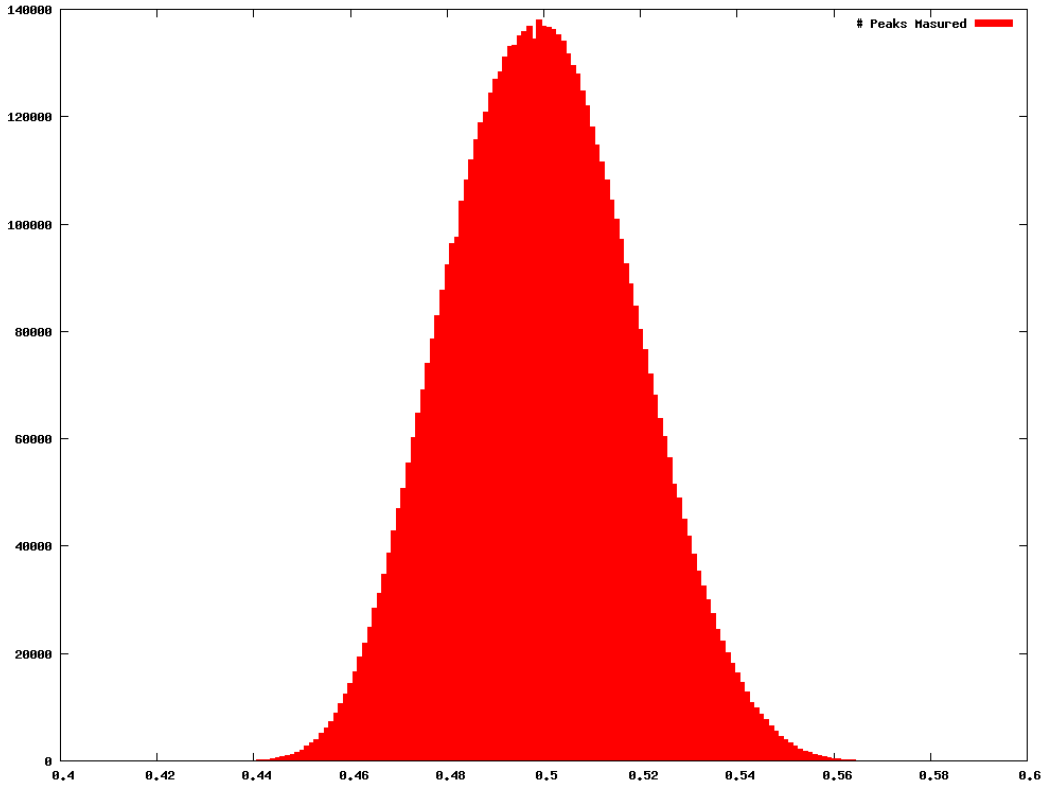


Figure 6.6: Distribution of measured voltage peaks. 51.25Ms data set (1.28ms of run time), 6.8 million pulses counted

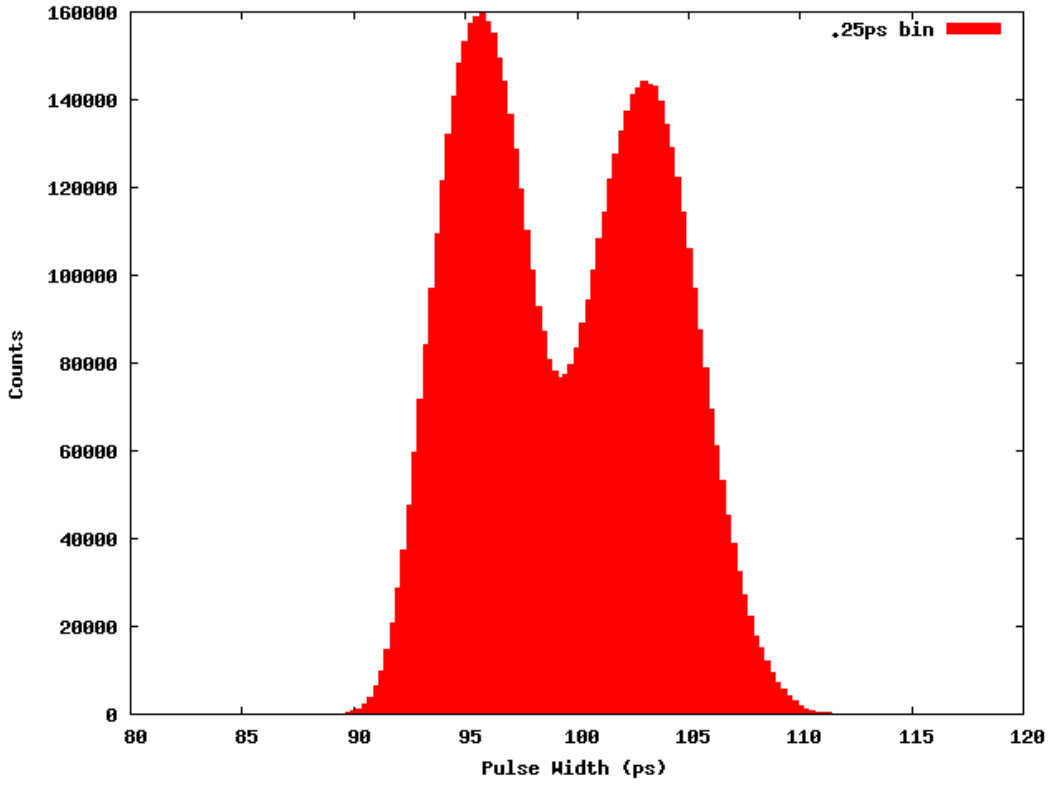


Figure 6.7: Measured pulse widths. Same data set as for the pulse heights of Figure6.6

and falling edge of each pulse. The bi-modal distribution of pulse-widths are shown in Figure 6.7. This bi-modal distribution is present in both 1 and 0 channels.

Spacing between pulses can be more variable than pulse widths. Considering spacing between only adjacent pulses (1 followed by a 1 or 0 followed by a 0), the distribution of spaces meeting this criterion is shown in figure 6.8.

6.2.2 Pulse Timing Accuracy

The timing stability of pulses and the matching detection system warrants closer inspection due it its high stability. Figure 6.9 is generated by revisiting the data from Figure 6.5,

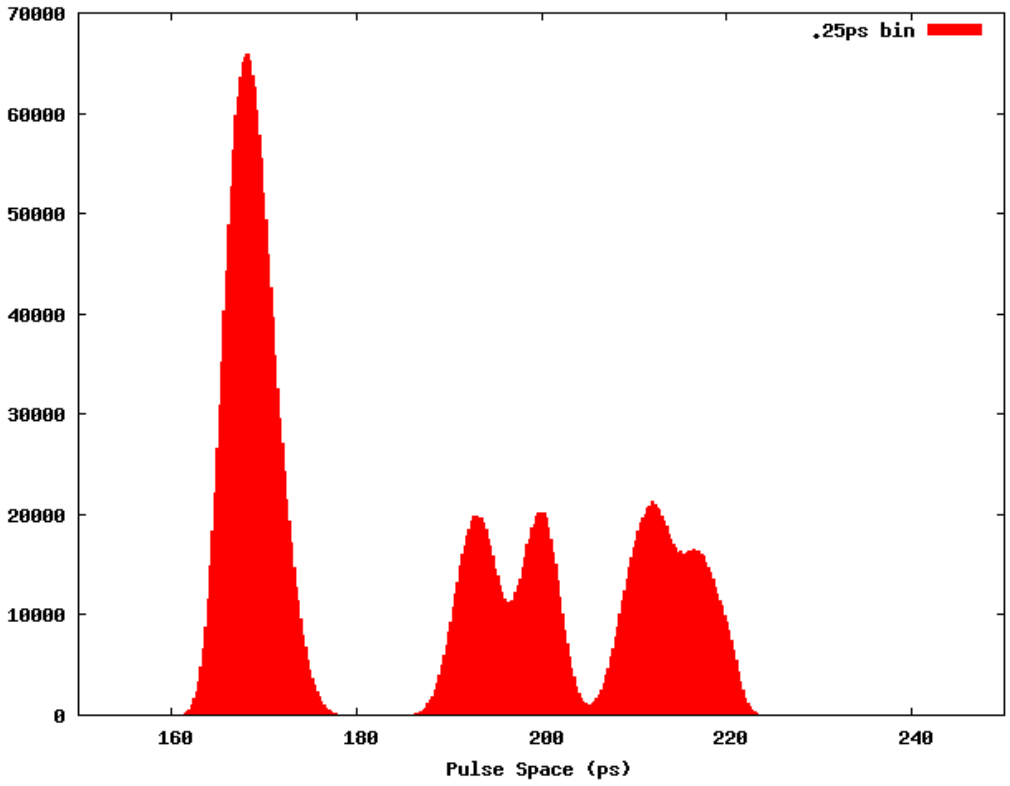


Figure 6.8: Pulse to pulse timing. Time measured rising edge to rising edge, marked when the rising edge crosses .25V.

Table 6.2: Count data for the fifth pulse from the 8-pulse histograms, Figure 6.5 and6.9

Time(ps)	Count	Time(ps)	Count	Time(ps)	Count	Time(ps)	Count
945	0	955	353	965	35605	975	38606
946	6	956	459	966	47960	976	24519
947	11	957	439	967	60781	977	14010
948	39	958	737	968	74283	978	6966
949	64	959	1350	969	85426	979	3132
950	91	960	2790	970	92165	980	1221
951	118	961	5399	971	92240	981	363
952	201	962	9720	972	84525	982	96
953	251	963	16096	973	71326	983	6
954	302	964	24586	974	54656	984	0

and plotting the data with a log scale. The log-scale plot clearly shows that in a 6+ million point data set, no pulses are seen more than 25ps from the corresponding peak arrival time when using the word-long pulse identification method. In this data set the 5th pulse has the largest spread, the actual statistics shown in Table 6.2. This data can be fit with a Gaussian curve with a $R^2 \approx .998$; the mean is $970.21ps$ and the standard deviation is $3.66ps$. Since there is a noticeable skew to the early side of the data, an artificial data set which mirrored the data points about $970.5ps$ was created and fitted. This fit has an $R^2 \approx .9997$ and $\sigma \approx 3.98ps$, meaning only 1 in 10^{15} pulses will be more than $32ps$ earlier than the nominal case – a very tight jitter bound for no active control of timing.

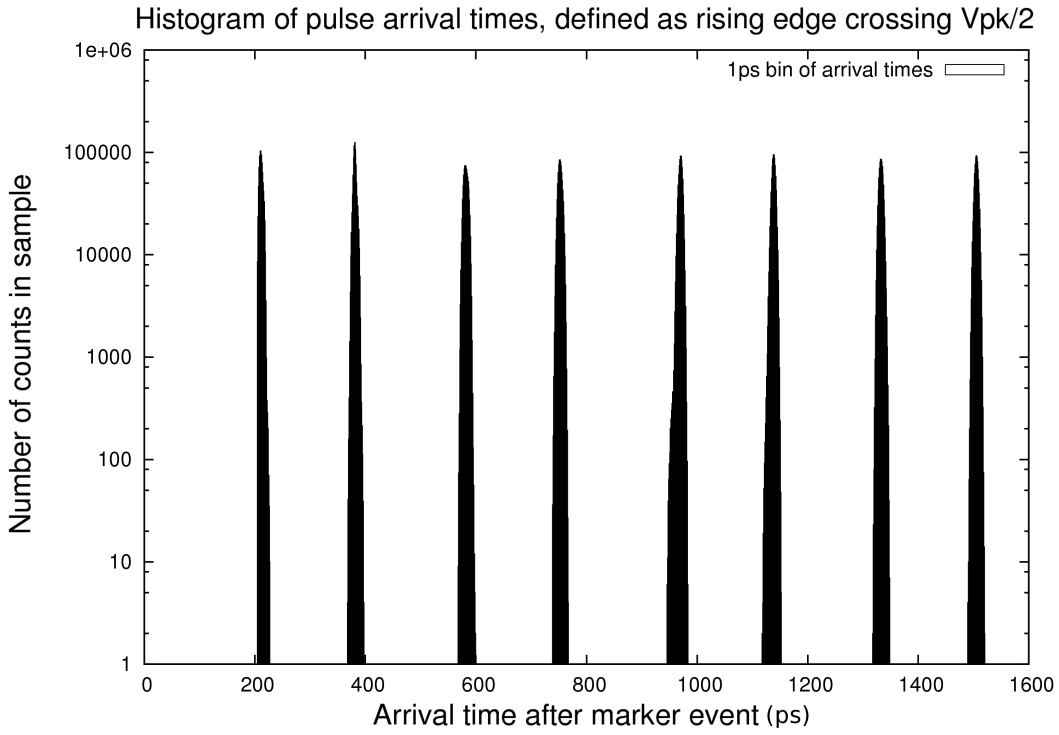


Figure 6.9: Log scale arrival time plot. This plot shows the same data as Figure 6.5 with a log scale in Y, showing that across millions of data points, no counts are seen more than +/- 20ps from nominal

6.2.3 Power

Some limited power measurements can be made to confirm if the presented serial link is meeting its power requirements. The key factor hindering power measurement is the presence of other structures on the chip besides the link. The most power hungry of these other structures were a set of 4 HSTL I/O pads, part of an experiment to see if the radiation hardening techniques of the pulse link were applicable to other interfaces. Two of these pads had parallel termination that was always active when the link was running, causing, by design, 30mA of shunt current between VDD and GND. This current projection is only accurate to 15% according to device specifications. This is on top of the the design has a power-hungry random number generator for which the power figures cannot be separated form the link.

Idle current measurement of the test chip yielded 29mA for the sample used for all of the traces, and between 26mA and 29mA for 5 samples measured at 1.5v. For the traced sample, full-speed transmit is measured at 68mA (67mA-70mA for 4 samples). Combined transmit and receive current is measured at 76mA with transmit and receive in a loop-back test. By subtracting the idle current (power consumed by other devices on the test chip) the full link current is measured at 47mA including the current for the random number generator. This gives an operating power of 71mW, less than the projected 85mW for the random number generator and transceiver combined. If this power savings was spread proportionately amongst the test structures and the transceiver, the actual operating transceiver power is estimated to be 50.1mW while running at 5.3GBps.

Chapter 7

Conclusion

This dissertation presented a number of different facets to the problem of systematic construction of high-speed circuitry. In general, the approach was to remove high-speed clocking from the system in an effort to make timing comparisons tractable. The Asynchronous logic style remaining is feed-forward in nature, requiring more designer effort than a delay insensitive approach. The freedom to the designer is restricted, by design policies, to keep the difficulty of checking and analyzing the work low. A number of timing comparisons are required for correct operation; as well as the strong typing of signal classes such that only one Event can occur at a time, separating correct latch operation from event sequencing.

7.1 Future Work

The present state of the art in feed-forward asynchronous pulse logic is an incomplete picture. The methods developed in Chapter 2 allow a designer to specify a sub-set of possible asynchronous pulse circuits. The method does not well capture approaches that modulate the

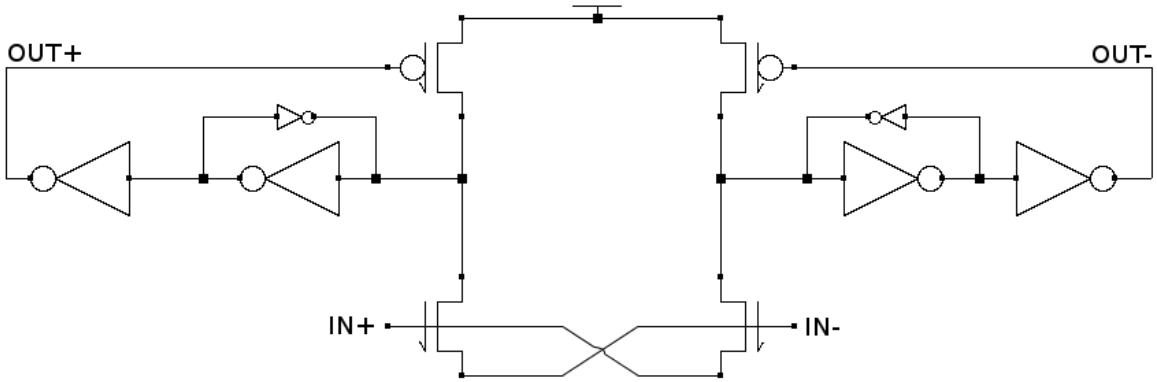


Figure 7.1: Basic Pulse Arbiter and amplifier

feed-back network of a pulse gate. Figure 7.1 for example is an arbiter that can be used to re-enforce pulse to pulse spacing across two wires. This circuit is useful in reducing uncertainty relative timing and hence can limit the worst-case timing of a two pulse system to the jitter of a single stage. Take, for example, the case presented in Chapter 2 Table 2.2, where the key limit on pulse stream performance was uncertainty in arrival order. If the circuit of Figure 7.1 could be used in the circuit, a solution put forward in [28], the pulse signaling rate would become 2 times the pulse width regardless of line length, assuming that the arbitration circuit was part of repeater stages often enough that order certainty was grater than the maximum period.

The problem with incorporating an arbiter, such as the one in Figure 7.1, into the analysis is that a correct specification requires describing behavior that involves operation, under limited circumstances, on two pulses at the same time. This violates the one-at-a time rule required for constraining the number of checks in formal verification. In addition to this, there are two other more technical issues with incorporating the circuit: the feed-back network is

modulated by a signal besides the trigger, and the forward propagation time is input-dependent. These three reasons make it difficult to directly insert this gate into the paradigm with out admitting systems that are not, strictly speaking, correct by construction. A theory that can decisively handle local arbitration such as that of Figure 7.1, and differentiate it from a failing circuit would be useful.

Many of the methods presented still require designer intervention through the process. For example the tool that takes a linguistic description and produces a dependency graph does not directly compute timing; it can merely produce a listing of dependent arcs for another tool or a designer to sum and compare to the design goals.

All of the formal analysis presented should ease the process of design automation; but synthesis was not attempted in this work. The extension of these concepts to a synthesis method would be a useful endeavor.

Bibliography

[1] Vinod Narayanan, Barbara A Chappell, and Bruce M Fleischer. Static timing analysis for self resetting circuits. In *Proceedings of the 1996 IEEE/ACM international conference on Computer-aided design*, pages 119–126. IEEE Computer Society, 1997.

[2] B.T. Nguyen, M.D. Papermaster, G.N. Pham, T.K. Ta, and W.B. van der Hoeven. Pipelined clock distribution for self resetting cmos circuits, June 9 1998. US Patent 5,764,083.

[3] Charles E Molnar, Ian W Jones, William S Coates, Jon K Lexau, Scott M Fairbanks, and Ivan E Sutherland. Two fifo ring performance experiments. *Proceedings of the IEEE*, 87(2):297–307, 1999.

[4] I. Sutherland and S. Fairbanks. Gasp: a minimal fifo control. In *Asynchronus Circuits and Systems, 2001. ASYNC 2001. Seventh International Symposium on*, pages 46–53, 2001.

[5] Mika Nyström and Alain J Martin. *Asynchronous pulse logic*. Springer, 2002.

[6] Brian D Winters and Mark R Greenstreet. A negative-overhead, self-timed pipeline. In *Asynchronous Circuits and Systems, 2002. Proceedings. Eighth International Symposium on*, pages 37–46. IEEE, 2002.

[7] Paul Teehan, Guy GF Lemieux, and Mark R Greenstreet. Towards reliable 5gbps wave-pipelined and 3gbps surfing interconnect in 65nm fpgas. In *Proceedings of the ACM/SIGDA international symposium on Field programmable gate arrays*, pages 43–52. ACM, 2009.

[8] Kaizad Mistry, C Allen, C Auth, B Beattie, D Bergstrom, M Bost, M Brazier, M Buehler, A Cappellani, R Chau, et al. A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 cu interconnect layers, 193nm dry patterning, and 100% pb-free packaging. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 247–250. IEEE, 2007.

[9] L. Bollinger, S. Harris, C. Hibdon, and C. Muehlhause. Neutron absorption and scattering by hafnium. *Phys. Rev.*, 92:1527–1531, Dec 1953.

[10] D Wilmore and Peter Edward Hodgson. The calculation of neutron cross-sections from optical potentials. *Nuclear Physics*, 55:673–694, 1964.

[11] Nelson S. Saks, M.G. Ancona, and J.A. Modolo. Generation of interface states by ionizing radiation in very thin mos oxides. *Nuclear Science, IEEE Transactions on*, 33(6):1185–1190, Dec 1986.

[12] Robert C Baumann. Radiation-induced soft errors in advanced semiconductor technologies. *Device and Materials Reliability, IEEE Transactions on*, 5(3):305–316, 2005.

[13] M Allenspach, JR Brews, I Mouret, RD Schrimpf, and KF Galloway. Evaluation of segr threshold in power mosfets. *IEEE Transactions on Nuclear Science (Institute of Electrical and Electronics Engineers);(United States)*, 41(CONF-940726–), 1994.

[14] Jeffrey L Titus and C Frank Wheatley. Experimental studies of single-event gate rupture and burnout in vertical power mosfet's. 1996.

[15] Ronald R Troutman. *Latchup in CMOS technology: the problem and its cure*, volume 13. Springer, 1986.

[16] Paul E Dodd and Lloyd W Massengill. Basic mechanisms and modeling of single-event upset in digital microelectronics. *Nuclear Science, IEEE Transactions on*, 50(3):583–602, 2003.

[17] F Ahmadov. Irradiation tests and expected performance of readout electronics of the atlas hadronic endcap calorimeter for the hl-lhc. *Journal of Instrumentation*, 9(01):C01028, 2014.

[18] HJ Barnaby. Total-ionizing-dose effects in modern cmos technologies. *Nuclear Science, IEEE Transactions on*, 53(6):3103–3121, 2006.

[19] S. Diez, M. Ullan, A A Grillo, J. Kierstead, W. Kononenko, F. Martinez-McKinney, F. M. Newcomer, S. Rescia, M. Ruat, H. F W Sadrozinski, A Seiden, E. Spencer, H. Spieler, and M. Wilder. Radiation hardness evaluation of a 130 nm sige bicmos technology for the atlas electronics upgrade. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pages 587–593, Oct 2010.

[20] Federico Faccio and Giovanni Cervelli. Radiation-induced edge effects in deep submicron cmos transistors. *Nuclear Science, IEEE Transactions on*, 52(6):2413–2420, 2005.

[21] Merritt Miller and Forrest Brewer. Formal verification of analog circuit parameters across variation utilizing sat. In *Proceedings of the Conference on Design, Automation and Test in Europe*, pages 1442–1447. EDA Consortium, 2013.

[22] Intel. Intel®Xeon®Processor E7 v2 Family. `http://ark.intel.com/products/family/78584/Intel-Xeon-Processor-E7-v2-Family?q=xeon`. Accessed: 2014-11-30.

[23] Atmel. AT32UC3A3 32-bit AVR Microcontroller Summary. `http://www.atmel.com/Images/32072s.pdf`. Accessed: 2014-11-30.

[24] Gunok Jung, V.A. Sundarajan, and G.E. Sobelman. A robust self-resetting cmos 32-bit parallel adder. In *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, volume 1, pages I–473–I–476 vol.1, 2002.

[25] Ayoob E Dooply and Kenneth Y Yun. Optimal clocking and enhanced testability for high-performance self-resetting domino pipelines. In *Advanced Research in VLSI, 1999. Proceedings. 20th Anniversary Conference on*, pages 200–214. IEEE, 1999.

[26] T. Säntti and J. Isoaho. Modified srcmos cell for high-throughput wave-pipelined arithmetic units. In *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, volume 4, pages 194–197 vol. 4, May 2001.

[27] Mark R Greenstreet and Jihong Ren. Surfing interconnect. In *Asynchronous Circuits and Systems, 2006. 12th IEEE International Symposium on*, pages 9–pp. IEEE, 2006.

[28] Merritt Miller, Greg Hoover, and Forrest Brewer. Pulse-mode link for robust, high speed communications. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pages 3073–3077. IEEE, 2008.

[29] Ivan E. Sutherland. Micropipelines. *Communications of the ACM*, 32(6):720–738, 1989.

[30] Ron Ho, Kenneth W Mai, and Mark A Horowitz. The future of wires. *Proceedings of the IEEE*, 89(4):490–504, 2001.

[31] H Buman Bakoglu. Circuits, interconnections, and packaging for vlsi. 1990.

[32] M-JE Lee, William J Dally, and Patrick Chiang. Low-power area-efficient high-speed i/o circuit techniques. *Solid-State Circuits, IEEE Journal of*, 35(11):1591–1599, 2000.

[33] D Breuer and K Petermann. Comparison of nrz-and rz-modulation format for 40-gb/s tdm standard-fiber systems. *Photonics Technology Letters, IEEE*, 9(3):398–400, 1997.

[34] James F Buckwalter, Mounir Meghelli, Daniel J Friedman, and Ali Hajimiri. Phase and amplitude pre-emphasis techniques for low-power serial links. *Solid-State Circuits, IEEE Journal of*, 41(6):1391–1399, 2006.

[35] Anirudh Devgan and Chandramouli Kashyap. Block-based static timing analysis with uncertainty. In *Proceedings of the 2003 IEEE/ACM international conference on Computer-aided design*, page 607. IEEE Computer Society, 2003.

[36] Edsger W Dijkstra. Guarded commands, nondeterminacy and formal derivation of programs. *Communications of the ACM*, 18(8):453–457, 1975.

[37] Charles Antony Richard Hoare. Communicating sequential processes. *Communications of the ACM*, 21(8):666–677, 1978.

[38] Mohamed H. Zaki, Sofiï¿œne Tahar, and Guy Bois. Formal verification of analog and mixed signal designs: A survey. *Microelectronics Journal*, 39(12):1395 – 1404, 2008.

[39] G.G.E. Gielen and R.A. Rutenbar. Computer-aided design of analog and mixed-signal integrated circuits. *Proceedings of the IEEE*, 88(12):1825 –1854, dec. 2000.

[40] L. Hedrich and E. Barke. A formal approach to verification of linear analog circuits wth parameter tolerances. In *Proceedings of the conference on Design, automation and test in Europe*, DATE '98, pages 649–655, Washington, DC, USA, 1998. IEEE Computer Society.

[41] W. Daems, G. Gielen, and W. Sansen. Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 22(5):517 – 534, may 2003.

[42] Amith Singhee and Rob A. Rutenbar. From finance to flip flops: A study of fast quasi-monte carlo methods from computational finance applied to statistical circuit analysis. In *Quality Electronic Design, 2007. ISQED '07. 8th International Symposium on*, pages 685 –692, march 2007.

[43] K.J. Antreich, H.E. Graeb, and C.U. Wieser. Circuit analysis and optimization driven by worst-case distances. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 13(1):57 –71, jan 1994.

[44] R.P. Kurshan and K.L. McMillan. Analysis of digital circuits through symbolic reduction. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 10(11):1356 –1371, nov 1991.

[45] W. Hartong, L. Hedrich, and E. Barke. Model checking algorithms for analog verification. In *Design Automation Conference, 2002. Proceedings. 39th*, pages 542 – 547, 2002.

[46] S. Little, N. Seegmiller, D. Walter, C. Myers, and T. Yoneda. Verification of analog/mixed-signal circuits using labeled hybrid petri nets. In *Computer-Aided Design, 2006. ICCAD '06. IEEE/ACM International Conference on*, pages 275 –282, nov. 2006.

[47] Chao Yan, F. Ouchet, L. Fesquet, and K. Morin-Allory. Formal verification of c-element circuits. In *Asynchronous Circuits and Systems (ASYNC), 2011 17th IEEE International Symposium on*, pages 55 –64, april 2011.

[48] Chao Yan and M.R. Greenstreet. Verifying an arbiter circuit. In *Formal Methods in Computer-Aided Design, 2008. FMCAD '08*, pages 1 –9, nov. 2008.

[49] S.R. Nassif. Modeling and analysis of manufacturing variations. In *Custom Integrated Circuits, 2001, IEEE Conference on.*, pages 223 –228, 2001.

[50] S.K. Tiwary, A. Gupta, J.R. Phillips, C. Pinello, and R. Zlatanovici. First steps towards sat-based formal analog verification. In *Computer-Aided Design - Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on*, pages 1 –8, nov. 2009.

[51] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, STOC '71, pages 151–158, New York, NY, USA, 1971. ACM.

[52] N. Eén and N. Sörensson. Translating pseudo-boolean constraints into sat. *Journal on Satisfiability, Boolean Modeling and Computation*, 2(3-4):1–25, 2006.

[53] O. Bailleux, Y. Boufkhad, O. Roussel, et al. A translation of pseudo boolean constraints to sat. *Journal on Satisfiability, Boolean Modeling and Computation*, 2:191–200, 2006.

[54] Michael Garland and Paul S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, pages 209–216, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[55] A.D. Kalvin and R.H. Taylor. Superfaces: polygonal mesh simplification with bounded error. *Computer Graphics and Applications, IEEE*, 16(3):64 –77, may 1996.

[56] S. Gupta, B. H. Krogh, and R. A. Rutenbar. Towards formal verification of analog designs. In *Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design*, ICCAD '04, pages 210–217, Washington, DC, USA, 2004. IEEE Computer Society.

[57] Min Chen, Wei Zhao, Frank Liu, and Yu Cao. Fast statistical circuit analysis with finite-point based transistor model. In *Proceedings of the conference on Design, automation and test in Europe*, DATE '07, pages 1391–1396, San Jose, CA, USA, 2007. EDA Consortium.

[58] HB Bakoglu and James D Meindl. Optimal interconnection circuits for vlsi. *Electron Devices, IEEE Transactions on*, 32(5):903–909, 1985.

[59] David A. B. Miller and Haldun M. Ozaktas. Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture. *Journal of parallel and distributed computing*, 41(1):42–52, 1997.

[60] Shyh-Chyi Wong, Gwo-Yann Lee, and Dye-Jyun Ma. Modeling of interconnect capacitance, delay, and crosstalk in vlsi. *Semiconductor Manufacturing, IEEE Transactions on*, 13(1):108–111, Feb 2000.

[61] Adelmo Ortiz-Conde, FJ Garcıa Sánchez, Juin J Liou, Antonio Cerdeira, Magali Estrada, and Y Yue. A review of recent mosfet threshold voltage extraction methods. *Microelectronics Reliability*, 42(4):583–596, 2002.

[62] HE Boesch, FB McLean, JM Benedetto, JM McGarrity, and WE Bailey. Saturation of threshold voltage shift in mosfet's at high total dose. *Nuclear Science, IEEE Transactions on*, 33(6):1191–1197, 1986.

[63] J.M. Benedetto, H.E. Boesch, F.B. McLean, and J. P. Mize. Hole removal in thin-gate mosfets by tunneling. *Nuclear Science, IEEE Transactions on*, 32(6):3916–3920, Dec 1985.

[64] Marek Turowski, Ashok Raman, and RD Schrimpf. Nonuniform total-dose-induced charge distribution in shallow-trench isolation oxides. *Nuclear Science, IEEE Transactions on*, 51(6):3166–3171, 2004.

[65] James R Schwank, Marty R Shaneyfelt, Daniel M Fleetwood, James A Felix, Paul E Dodd, Philippe Paillet, and Véronique Ferlet-Cavrois. Radiation effects in mos oxides. *Nuclear Science, IEEE Transactions on*, 55(4):1833–1853, 2008.

[66] L. Gonella, F. Faccio, M. Silvestri, S. Gerardin, D. Pantano, V. Re, M. Manghisoni, L. Ratti, and A. Ranieri. Total ionizing dose effects in 130-nm commercial CMOS technologies for HEP experiments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 582(3):750–754, 2007. VERTEX 2006 Proceedings of the 15th International Workshop on Vertex Detectors.

[67] C. Hafer, M. Lahey, H. Gardner, D. Harris, A. Jordan, T. Farris, and M. Johnson. Radiation hardness characterization of a 130nm technology. In *Radiation Effects Data Workshop, 2007 IEEE*, volume 0, pages 123–130, July 2007.

[68] IEEE. Ieee standard for heterogeneous interconnect (hic) (low-cost, low-latency scalable serial interconnect for parallel system construction). *IEEE Std 1355-1995*, pages i–, 1996.

[69] S.M. Parkes and P. Armbruster. Spacewire: a spacecraft onboard network for real-time communications. In *Real Time Conference, 2005. 14th IEEE-NPSS*, pages 6–10, June 2005.

[70] William M Lewis Jr. Phase locking, period-doubling bifurcations, and irregular dynamics in periodically stimulated cardiac cells. *Oecologia (Berlin)*, 19:75, 1975.

[71] G Sucha, SR Bolton, S Weiss, and DS Chemla. Period doubling and quasi-periodicity in additive-pulse mode-locked lasers. *Optics letters*, 20(17):1794–1796, 1995.

[72] Nail Akhmediev, JM Soto-Crespo, and G Town. Pulsating solitons, chaotic solitons, period doubling, and pulse coexistence in mode-locked lasers: complex ginzburg-landau equation approach. *Physical Review E*, 63(5):056602, 2001.

[73] Nansen Chen, Kevin Chiang, TD Her, Yeong-Lin Lai, and Chichyang Chen. Electrical characterization and structure investigation of quad flat non-lead package for rfic applications. *Solid-State Electronics*, 47(2):315–322, 2003.

[74] Yeong-Lin Lai and Cheng-Yu Ho. Electrical modeling of quad flat no-lead packages for high-frequency ic applications. In *TENCON 2004. 2004 IEEE Region 10 Conference*, volume 500, pages 344–347. IEEE, 2004.

[75] Frank Mortan and Lance Wright. Quad Flatpack No-Lead Logic Packages. *Texas Instruments Application Report* `http://www.ti.com/lit/an/scba017d/scba017d.pdf`, February 2004.

[76] E.S. Ochotta, R.A. Rutenbar, and L.R. Carley. Synthesis of high-performance analog circuits in astrx/oblx. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 15(3):273 –294, mar 1996.

[77] Tom Verhoeff. Delay-insensitive codes - an overview. *Distributed Computing*, 3(1):1–8, 1988.

[78] Jordi Cortadella, A Kondratyev, L Lavagno, and C Sotiriou. A concurrent model for de-synchronization.

[79] Paulo Moreira. GBT Project: Present & Future, Mar 2014. ACES 2014 - Forth Common ATLAS CMS Electronics Workshop for LHC `https://indico.cern.ch/event/287628/session/1/contribution/12/material/slides/1.pdf`.

[80] Paulo Moreira. The Radiation Hard GBTX Link Interface Chip, Jul 2014. TDC - PLL meeting `http://indico.cern.ch/event/323782/contribution/1/material/slides/1.pdf`.

[81] Chenming Hu. Gate oxide scaling limits and projection. In *International Electron Devices Meeting*, pages 319–322, 1996.

[82] T. Heijmen, D. Giot, and P. Roche. Factors that impact the critical charge of memory elements. In *On-Line Testing Symposium, 2006. IOLTS 2006. 12th IEEE International*, pages 6 pp.–, 2006.

[83] Paulo Moreira. GBT Project Status, Mar 2011. ACES 2011 - Common ATLAS CMS Electronics Workshop for SLHC .

[84] B.M. Haugerud, S. Venkataraman, A.K. Sutton, A.P.G. Prakash, J.D. Cressler, Guofu Niu, P.W. Marshall, and A.J. Joseph. The impact of substrate bias on proton damage in 130 nm cmos technology. In *Radiation Effects Data Workshop, 2005. IEEE*, pages 117–121, July 2005.

[85] Hugh J. Barnaby, Michael Mclain, and Ivan Sanchez Esqueda. Total-ionizing-dose effects on isolation oxides in modern CMOS technologies. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 261(1-2):1142 – 1145, 2007.

[86] Robert E Lyons and Wouter Vanderkulk. The use of triple-modular redundancy to improve computer reliability. *IBM Journal of Research and Development*, 6(2):200–209, 1962.

[87] Y Boulghassoul, LW Massengill, AL Sternberg, BL Bhuva, and WT Holman. Towards set mitigation in rf digital plls: From error characterization to radiation hardening considerations. *Nuclear Science, IEEE Transactions on*, 53(4):2047–2053, 2006.

[88] Compaq, Hewlett-Packard, Intel, Lucent, Microsoft, NEC, and Phillips. *Universal Serial Bus Specification Revision 2.0.* USB Implementers Forum, April 2000.

[89] Armin Tajalli, Paul Muller, Mojtaba Atarodi, and Yusuf Leblebici. A multichannel 3.5 mw/gbps/channel gated oscillator based cdr in a 0.18μm digital cmos technology. In *Solid-State Circuits Conference, 2005. ESSCIRC 2005. Proceedings of the 31st European*, pages 193–196. IEEE, 2005.

[90] Albert X. Widmer and Peter A. Franaszek. A dc-balanced, partitioned-block, 8b/10b transmission code. *IBM Journal of research and development*, 27(5):440–451, 1983.

[91] CMS Collaboration, S Chatrchyan, et al. The cms experiment at the cern lhc. *Jinst*, 3(08):S08004, 2008.

[92] Dan Lei Yan, M Kumarasamy Raja, and Aruna B Ajjikuttira. A gated-oscillator based burst-mode clock and data recovery (cdr) circuit. In *Radio-Frequency Integration Technology, 2007. RFIT 007. IEEE International Workshop on*, pages 90–93. IEEE, 2007.

[93] Masafumi Nogawa, Kazuyoshi Nishimura, Shunji Kimura, Tomoaki Yoshida, Tomoaki Kawamura, Minoru Togashi, Kiyomi Kumozaki, and Yusuke Ohtomo. A 10 gb/s burst-mode cdr ic in 0.13 mm cmos. In *IEEE International Solid-State Circuits Conference*, 2005.

[94] Alan Sill. {CDF} run {II} silicon tracking projects. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 447(1–2):1 – 8, 2000.

[95] J.R. Carter et. al. The silicon microstrip sensors of the {ATLAS} semiconductor tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 578(1):98 – 118, 2007.

[96] Frank Hartmann. The CMS all-silicon tracker – strategies to ensure a high quality and radiation hard silicon detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 478(1–2):285 – 287, 2002. Proceedings of the ninth Int.Conf. on Instrumentation.

[97] D Braga, G Hall, L Jones, P Murray, M Pesaresi, M Prydderch, and M Raymond. Cbc2: a microstrip readout asic with coincidence logic for trigger primitives at hl-lhc. *Journal of Instrumentation*, 7(10):C10003, 2012.

[98] M.J. French, L.L. Jones, Q. Morrissey, A. Neviani, R. Turchetta, J. Fulcher, G. Hall, E. Noah, M. Raymond, G. Cervelli, P. Moreira, and G. Marseguerra. Design and results from the apv25, a deep sub-micron {CMOS} front-end chip for the {CMS} tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 466(2):359 – 365, 2001. 4th Int. Symp. on Development and Application of Semiconductor Tracking Detectors.

[99] R. Hentschke, F. Marques, F. Lima, L. Carro, A Susin, and R. Reis. Analyzing area and performance penalty of protecting different digital modules with hamming code and triple modular redundancy. In *Integrated Circuits and Systems Design, 2002. Proceedings. 15th Symposium on*, pages 95–100, 2002.

[100] S.M. Parkes and P. Armbruster. Spacewire: a spacecraft onboard network for real-time communications. In *Real Time Conference, 2005. 14th IEEE-NPSS*, pages 6–10, June 2005.

[101] F Lemeilleur, M Glaser, EHM Heijne, P Jarron, and E Occelli. Neutron-induced radiation damage in silicon detectors. *Nuclear Science, IEEE Transactions on*, 39(4):551–557, 1992.

[102] HJ Barnaby, SK Smith, RD Schrimpf, DM Fleetwood, and RL Pease. Analytical model for proton radiation effects in bipolar devices. *Nuclear Science, IEEE Transactions on*, 49(6):2643–2649, 2002.

[103] O. Hauck and S.A. Huss. Asynchronous wave pipelines for high throughput datapaths. In *Electronics, Circuits and Systems, 1998 IEEE International Conference on*, volume 1, pages 283–286 vol.1, 1998.

[104] ITRS. The International Technology Roadmap for Semiconductors 2013 Interconnect, 2013.

[105] W.C. Elmore. The transient response of damped linear networks with particular regard to wideband amplifiers. *Journal of Applied Physics*, 19(1):55–63, Jan 1948.