MEASUREMENT AND REGULATION OF ON-CHIP POWER SUPPLY NOISE

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

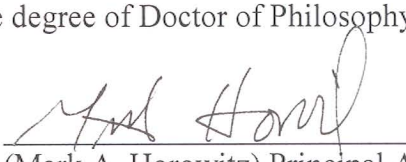
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Elad Alon

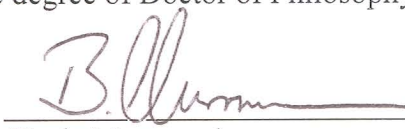
December 2006

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.



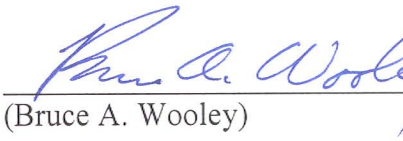
(Mark A. Horowitz) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.



(Boris Murmann)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.



(Bruce A. Wooley)

Approved for the University Committee on Graduate Studies

Abstract

The rules of constant-field scaling have driven CMOS devices to operate at ever-lower supply voltages while maintaining constant or even increasing power densities. This scaling has caused drastic reductions in the required impedance of the supply distribution network, reaching ~1 mΩ in today's 1 V, 100 A microprocessors, and because of this the integrity of the on-die supply voltage has become a critical issue. This thesis presents circuits and techniques to measure the supply noise as it is seen on the die, and to efficiently regulate the power supply to actively improve its quality.

Despite the increase in concern over supply noise, directly measuring the noise in order to experimentally quantify its impact on circuits can be very challenging. To overcome this challenge, we extend previous sub-sampling methods to measure autocorrelation, allowing the noise spectrum to be extracted using only two low-rate samplers. Calibrating the measurement circuits allows the samplers to be implemented with a sampling switch and a simple high-resolution VCO-based ADC, and we present measured results using these samplers that confirm the cyclostationary nature of supply noise. We also highlight some of the challenges that arise from implementing the samplers in aggressively scaled technologies, and describe a technique that employs averaging of many dithered, low-resolution samples to eliminate the difficult to scale sampling switch entirely.

Integrated regulators have found widespread adoption in isolating the supplies of sensitive analog or mixed-signal circuits from externally-generated noise. To understand the tradeoffs between noise rejection and total power consumption inherent to such regulators, we describe an analysis showing that feedback gain-bandwidth (and hence power) must increase quadratically to improve regulator noise rejection. We next show that in applications where the regulator's principal goal is isolation, we can make use of knowledge of the regulator's load to improve the effective gain-bandwidth of the feedback path – making the required gain-bandwidth of the amplifier increase only linearly with noise rejection.

Despite their use in analog or mixed-signal applications, the high power overheads of traditional regulators (both series and shunt) have precluded their successful adoption in regulating the supply of energy-efficient digital circuits. We therefore present a push-pull shunt regulator that makes use of a secondary, higher-than-nominal power supply, comparator-based feedback, and a switched-source follower output stage to minimize the regulator's static power dissipation. Using these techniques, a push-pull shunt regulator implemented in a 65 nm SOI technology was able to simultaneously reduce the supply's effective impedance by ~30% and the chip's total power dissipation by ~1%.

Acknowledgments

The past nine years I have spent at Stanford have been an incredible journey – mostly because of the amazing people who have been a part of it.

First and foremost, I would like to thank my advisor, Prof. Mark Horowitz. I couldn't have found a better advisor anywhere in the world, and I can only aspire to one day come close to the example he has set as a mentor, researcher, and teacher. I can not count the number of times that Mark's ability to decompose any problem or idea into its most fundamental form has left me utterly amazed, and I am infinitely grateful to have had the opportunity to learn so much from him.

I would also like to thank some of the other excellent faculty members at Stanford who have shaped my experience. I would like to thank Prof. Boris Murmann for many interesting discussions, for serving on my orals committee, and for his feedback on this thesis. I am grateful to Prof. Bruce Wooley for his teaching, his support, and for taking the time to read this thesis. I am especially thankful to Prof. Thomas Lee for the incredible teaching that sparked my interest in analog design, and for his encouragement and advice. I also wish to thank Prof. Stephen Boyd and Prof. Claire Tomlin not only for their excellent lectures on convex optimization and control theory, but for their great enthusiasm and support. Finally, I would like to thank Prof. Joseph Kahn for an exciting collaboration on multi-mode fiber optical communications.

The current and former students that I had the pleasure of interacting with were no less important in creating a vibrant and energetic atmosphere. First, I would like to thank Vladimir Stojanović – not only for the many successful projects we worked on

together and for his invaluable help and advice, but also for his passion for research and commitment to excellence. I am truly lucky to have found such a great friend, and will always proudly wear my "destabilizer" badge. I am indebted to Stefanos Sidiropoulos and Dean Liu for their teaching and support while I was still an undergraduate, and to Valentin Abramzon, Bita Nezamfar, Hae-Chang Lee, Ken Mai, Jaeha Kim, Eric Kim, Dinesh Patil, Amir Amirkhany, Xiling Shen, Sam Palermo, Ron Ho, and Jim Weaver for their friendship and many enjoyable discussions and collaborations.

I was very fortunate to collaborate with Rambus multiple times, and I would like to thank all of my friends and colleagues there for making it an incredible place to work. In particular, I would like to thank Fari Assaderaghi, Ken Chang, Jared Zerbe, and Carl Werner for their support and belief in me, and Andrew Ho, Sudhakar Pamarti, and Fred Chen for their friendship and hours of fun and fascinating discussions.

I would also like to thank many people from IBM, HP, AMD, and Intel for giving me the wonderful opportunity to work with and learn from them – especially Sam Naffziger, Sanjay Sethi, Kevin Nowka, Doug Josephson, and Greg Taylor.

I wish to thank MARCO C2S2 and AMD (through the CIS FMA program) for enabling this research with their financial support.

I also owe many thanks to Teresa Lynn, Ann Guerra, and Penny Chumley for being excellent admins and resolving countless issues. I would especially like to thank Ann Guerra for making EE313 run much more smoothly than I could have ever hoped for.

My warmest thanks go to Andrew Ho, Michelle Lee Ho, Viresh Hingarh, and Christina Ip for their support, great friendship, and many wonderful memories.

Words cannot express my gratitude to my girlfriend Meredith for her constant love, support, and understanding – without which I would have never completed this journey. I can only hope to provide her the same as she continues down what I have no doubt will be the path to the extremely successful completion of her own graduate studies.

Finally, I would like to express my enormous thanks to my mother Ruth, my father Zvi, my brother Eyal, my grandparents, and the rest of my family. No matter how far away they may be physically, they are never far from my heart and mind. Without their boundless love and belief in me, I would never have had the strength and courage to pursue my dreams, and for that I dedicate this thesis to them.

Table of Contents

List of Figures

Chapter 1

Introduction

As Dennard explained in [1], by reducing the supply voltage along with the critical dimension of CMOS transistors, a chip scaled from one technology generation to the next could integrate roughly twice the transistors for the same die area, run them at a higher frequency, and still maintain roughly constant total power. However, even if scaling had followed the ideal constant-field rules, constant power at reduced supply voltage necessarily leads to increased current consumption. Therefore, to maintain a fixed percentage budget of variation on the supply voltage, the combination of lower voltage and higher current requires the supply network impedance to drop with the scaling factor squared. This requirement on the supply network translates not only into maintaining a low resistance, but also into low dynamic impedance (i.e., low series inductance, large parallel capacitance) to mitigate the impact of the (potentially large) transient currents inherent in the operation of CMOS digital gates.

As shown in Figure 1.1, the scaling of supply impedance for high-performance microprocessors actually followed a trend even worse than that predicted by constant-field theory – mainly because higher performance circuit designs led to higher operating frequencies, and hence actually increased the power density significantly [2]. This has led to today's high-performance microprocessors running off of roughly 1 V power supplies with 100 A of current, making their required impedance ~1 mΩ.

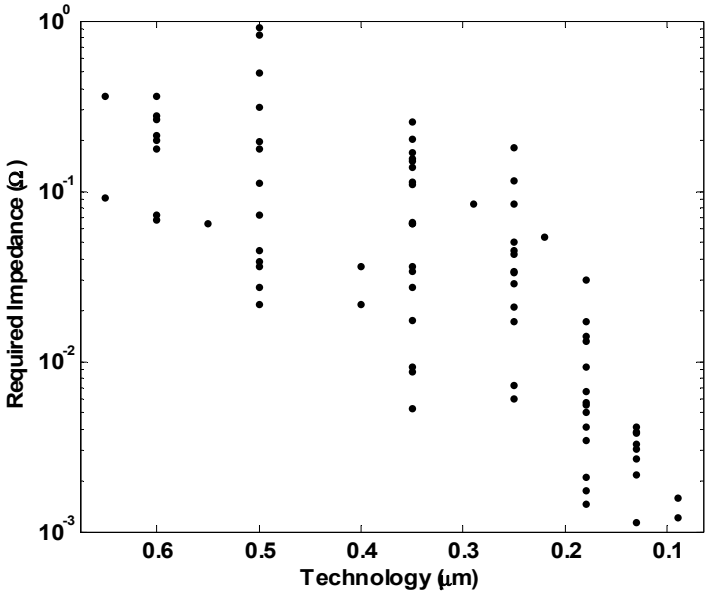


Figure 1.1: Required impedance for a 10% noise budget on high-performance microprocessors vs. technology node.

To evaluate how this impedance scaled relative to interconnect and packaging technology, many researchers in both industry and academia used various predictive models and examined the scaling of the supply voltage's overall integrity [3,4,5,6,7,8]. The essential outcome of all of these studies was that meeting the impedance requirement (both static and dynamic) would become increasingly difficult and costly – necessitating optimization of the supply distribution network all the way from the printed-circuit board, through the package, and onto the die [9]. Improved supply impedance was arguably one of the main drivers of the adoption of flip-chip packaging technology [10], and attaining mΩ levels of supply impedance has forced designers to dedicate an ever-larger percentage of the on-chip metal resources purely to power distribution [11,12].

For these reasons, it is clear that the integrity of the on-chip supply voltage has become a critical concern in modern designs – even for chips containing purely digital circuits. Therefore, in this thesis we set out to develop techniques to enable full measurement and characterization of noise on the supply voltage using only simple hardware, and to develop integrated, efficient regulation circuits to actively improve the integrity of the power supply.

1.1 Organization

In the first part of this thesis, we describe our work on circuits and techniques for measurement of supply variations as they are seen on the die. Chapter 2 begins by motivating the need for such measurements, and explaining the limitations of previous, mostly sub-sampling-based measurement schemes. To overcome these limitations, we then describe how to extend these techniques to measure autocorrelation, allowing the noise spectrum to be extracted using only two low-rate samplers. Calibrating the measurement circuits allows the samplers to be implemented with a sampling switch and a simple, compact, and high-resolution VCO-based ADC.

Chapter 3 continues the discussion of supply noise measurement by highlighting some of the challenges associated with integrating even this relatively simple scheme into chips manufactured in aggressively scaled, digital-performance optimized technologies. The majority of the issues are traced to the sampling switch itself, so after describing modifications to the sampler that can mitigate these issues, we conclude the chapter by explaining a technique that eliminates the need for a sampling switch entirely by averaging many low-resolution samples.

Having discussed our approach to supply noise measurement, the next two chapters describe our work on building integrated regulation circuitry to actively counter variations in the supply voltage. Since modern chips are severely power or thermally-constrained, the issue of a regulator's efficiency takes a central role in the discussion, and in Chapter 4 we focus on tradeoffs in the control design of CMOS linear regulators when the power consumption of the feedback circuitry is limited. As part of this chapter, we also show how to make use of knowledge of the behavior of the load circuitry to significantly improve the efficiency of regulators whose principal goal is isolation of the supply of sensitive (typically analog or mixed-signal) circuits.

In Chapter 5, we focus on applying integrated regulation to the power supplies of high-performance, energy-efficient digital chips. We first show that the energy costs of traditional, single-supply topologies are simply too large to successfully apply such regulators to power-limited digital chips. We therefore next describe the design of a

push-pull shunt regulator that makes use of a secondary, higher-than-nominal power supply to overcome this limitation, and demonstrate that this approach can simultaneously reduce the chip's power consumption and the variations on its supply.

Finally, in Chapter 6 we summarize some of the key insights we gained into the challenges associated with measuring and regulating on-chip supply noise. In addition, we examine potential future research directions building upon this thesis to solve challenges in the measurement of more general on-chip signals, and to improve the efficiency of generating multiple, local supply voltages on the die.

Chapter 2

Supply Noise Measurement

As scaling caused power supply integrity to become a significant issue, increasingly sophisticated models of the supply network and chip current consumption were developed [13,14,15], leading to CAD tools [16,17] whose goal is to give chip designers the ability to find potential supply integrity issues in simulation. The latest versions of these tools can not only calculate the IR drop on the supply network at various locations on the die, but can run dynamic simulations to predict the impact of switching currents on the supply voltage.

Despite their sophistication, these simulations of the noise on the power supply are inherently based on models. Hence, verification of the tools' results by actual measurements of noise as it is seen on the die is a key step to signing off on the supply integrity of the design – sparking significant interest in techniques to perform such on-chip supply noise measurements [18,19,20,21].

The principal challenge in developing schemes for supply noise measurement is that the current variations that generate the noise can have extremely high bandwidths (limited roughly only by the delay of a digital gate). Therefore, capturing a full time-domain snap-shot of the supply voltage waveform would require a measurement bandwidth of over 10-20 GHz with an effective resolution of ~6-8 bits. While ADCs

developed for real-time oscilloscopes have attained this level of performance [22], these converters occupy an entire chip and dissipate ~9 W.

To avoid the excessive overhead of such a high-speed converter, most supply noise measurement circuits have been designed to make use of some form of sub-sampling (as typically used in equivalent-time oscilloscopes) – typically either by explicitly taking analog samples of the noise [18,21], or by making use of variable threshold comparators [19,20]. Making the supply waveform periodic (and hence measurable through equivalent-time sampling) typically requires the chip to operate under controlled testing conditions in a repetitive manner, and hence this type of measurement unfortunately may not capture supply noise behavior during normal chip operation.

If the samplers used in the measurement circuit do not rely on averaging for accuracy, further characterization of the supply variations can be accomplished by using the sampler to collect data about the distribution of the supply noise at each point within a given cycle (like an oscilloscope with "infinite persistence" turned on). While this measurement can provide a large amount of information about the behavior of the supply noise, the dynamics of the non-repetitive variations inherent in typical chip operation would remain uncharacterized. This is a limitation because both the distribution and frequency spectrum of the supply variations must be known in order to characterize the effect of supply noise on circuits – especially on sensitive analog or mixed signal circuits.

To overcome this limitation, in this chapter we extend the previous measurement techniques by treating supply noise as a random process, and use its statistical properties – in particular, its autocorrelation [23] – to measure the noise spectrum.[1] As we will describe, the attractiveness of this technique lies in the fact that autocorrelation can be measured using only two samplers with precise sampling instants, but low sampling rates. We further extend this technique to measure the dynamics of a repetitively time-varying, or cyclostationary [24], noise process by recording the sampling instants in addition to the distance between the samples. This additional information captures the periodic structure of the noise, and better characterizes its behavior.

[1] The noise measurement techniques and circuits described in this chapter and the next were developed in collaboration with Vladimir Stojanović, Valentin Abramzon, and Bita Nezamfar.

2.1 Random Supply Noise and Autocorrelation

While supply noise may actually be deterministic in nature, the number of state variables that would need to be tracked to precisely calculate it is enormous. Therefore, instead of finding its exact behavior, supply noise is usually modeled as a random process, and can be characterized by its autocorrelation or its frequency spectrum. This spectrum is one of the properties necessary to calculate the effects of the noise on sensitive circuits. In addition, the shape of the spectrum can give insights into the source of the noise.

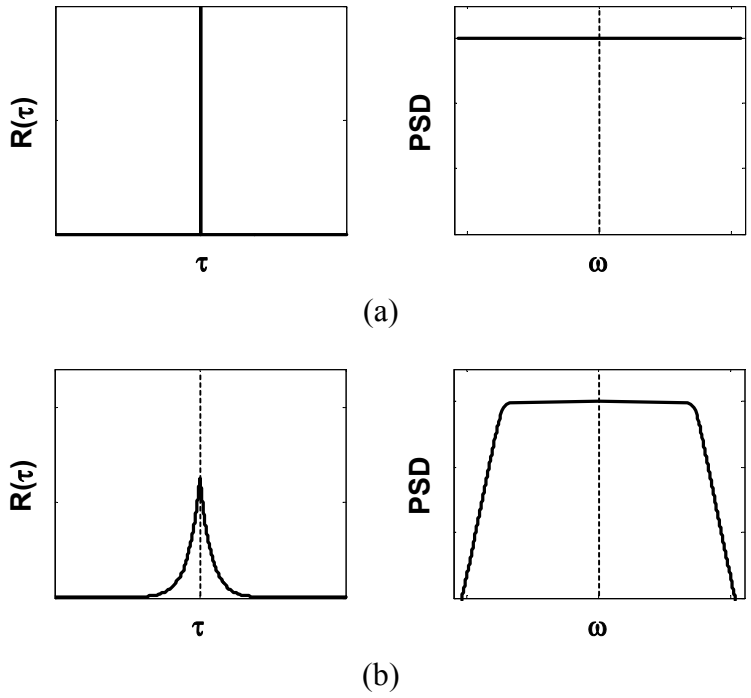


Figure 2.1: Autocorrelation and power spectral density for a) white noise, and b) low-pass filtered white noise.

For a time-invariant (also known as stationary) random process, the autocorrelation (R) is defined as

$$R(\tau) = E[v_{noise}(t+\tau/2)\cdot v_{noise}(t-\tau/2)] \quad (2.1)$$

where τ represents the separation of the two noise samples in time and E[] returns the expected value of the random process. The Fourier transform of the autocorrelation gives the power spectral density (PSD) of the noise process. For example, as shown in Figure

2.1a, a white, zero mean noise process has a spectral density that is flat and infinitely wide – correspondingly, its autocorrelation is an impulse whose magnitude is set by the variance (σ^2) of the noise process. Intuitively, if a noise process contains high frequency components, it will change within a short period of time, and therefore samples of the noise that are far apart in time will not be correlated with each other.

As an additional example of the relationship between autocorrelation and noise spectrum, consider a low-pass filtered white noise process. The filter spreads each noise impulse in time, and thus closely spaced samples of the noise are highly correlated with each other. As shown in Figure 2.1b, this causes the autocorrelation to become broader, matching the shape of the impulse response of the filter.

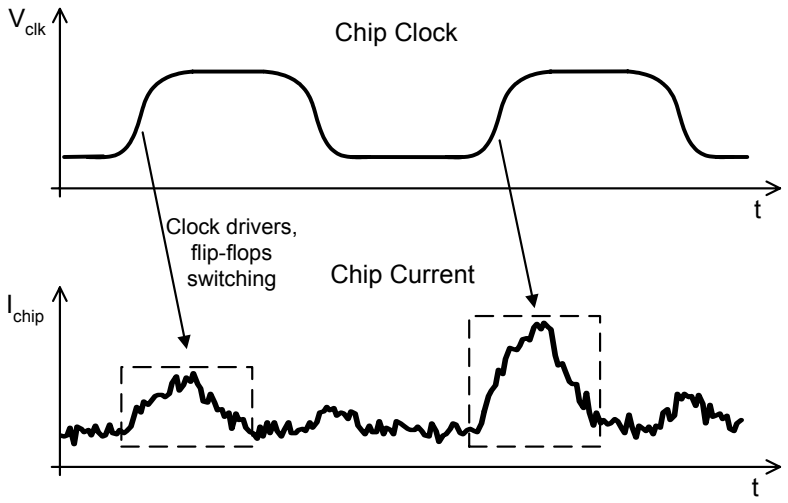


Figure 2.2: Example of the clock modulating the chip current consumption waveform.

Thus far, the discussion of autocorrelation has been limited to stationary noise processes that do not vary with time. Most chips run synchronously to one or more clocks, and these clocks can modulate the occurrence of noise events. For example, as shown in Figure 2.2, switching events may be more likely to occur at the beginning of the clock cycle than at the middle or end since all of the clock drivers and flip-flops toggle near the rising edge of the clock. Because of this modulation, it is unlikely that the properties of supply noise will be independent of time. However, since this modulation is repetitive, at a particular point in time relative to the cycle the properties of the noise

should be the same. This is known as a cyclostationary process, and can be characterized by measuring the autocorrelation (and hence PSD) of the noise at each time point in the cycle. In other words, the autocorrelation R becomes a function of both the time separation τ and the time within the noise cycle at which the samples were taken.

The most important characteristic of autocorrelation for the measurement system is that it is an average statistical property of the noise, and therefore we do not need to know the exact behavior of the noise at all times in order to extract its frequency content. The Nyquist frequency of the measurement is set by the minimum time spacing between the samples – not by the repetition rate of the sampling – which greatly reduces the requirements on the throughput of the sampling circuits. A further benefit of measuring autocorrelation[2] is that any uncorrelated, additive noise that is independent between the samplers (e.g. thermal noise) can be eliminated because the expected value of the product of two uncorrelated noise terms is zero.

2.2 Measurement Circuits

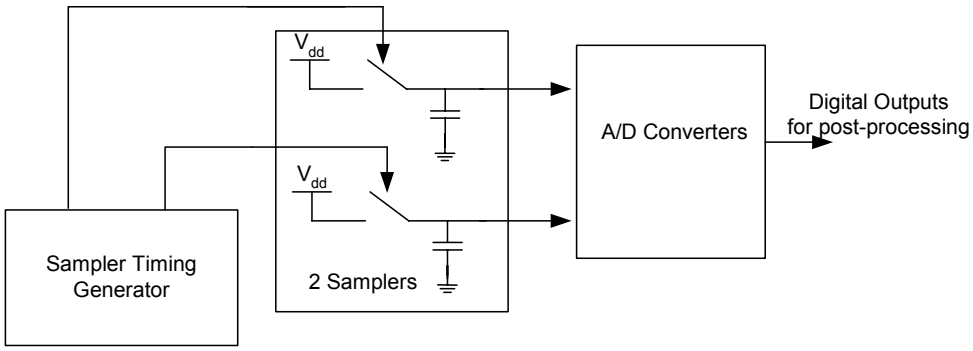


Figure 2.3: Supply noise measurement system block diagram.

A block diagram of the noise measurement system developed based on the concept of measuring autocorrelation [25] is shown in Figure 2.3. As will be described in Section 2.2.3 of this chapter, the timing generation controlling the sampling instants can either be performed off-chip with external equipment, or with internal circuitry. As we will

[2] In order to reduce the sensitivity of the measured results to residual offset errors in the samplers, it is often desirable to measure autocovariance, $C(\tau) = E[(v_{noise}(t+\tau/2) - E[v_{noise}(t+\tau/2)])\cdot(v_{noise}(t-\tau/2) - E[v_{noise}(t-\tau/2)])]$.

describe in the calibration section, in both cases, on-chip hardware is included in the measurement system to allow calibration of mismatch between the two sampling signals.

To avoid additional noise that would be coupled into the measurement system from communicating analog quantities off the chip, A/D conversion is performed internally. Since all the digital samples are post-processed using measured calibration curves, the linearity and offset requirements of the on-chip converters are not stringent. This allows a simple and compact design that can be integrated onto many different parts of the die.

2.2.1 Sampling Switch

The sampling switch is the only circuit in the measurement system that must be on the die in order to measure supply noise relative to on-chip V_{ss} – it is also the only circuit whose bandwidth must be high enough to avoid filtering high frequency noise content. The bandwidth of the sampling switch should meet or exceed the bandwidth of the highest bandwidth circuits on the chip so that the measurement will accurately capture the supply noise behavior as it is seen by these circuits. For example, in many digital designs the highest bandwidth circuits are fanout-of-four (FO4) inverters, and therefore the sampling switch would need a bandwidth above that set by the rise time of these inverters. While previous designs such as [18] and [26] used NMOS switches, when sampling V_{dd} it is more straightforward to obtain the required bandwidth using PMOS switches.

To avoid undesirable filtering of the measurements, during hold mode the sampled voltage should be as independent of the current value of V_{dd} as possible. The most direct way to achieve this goal is to make use of a separate, slightly higher than nominal power supply for the samplers (e.g., V_{ddQ} = 1.3 V, whereas V_{dd} = 1 V). This was the approach taken in [25] and these circuits and results will be described in this chapter; however, in the next chapter we will describe methods to avoid this separate supply.

Despite the conceptual simplicity of using a separate supply for isolation, care must be taken in the design and use of this supply. Specifically, since any noise on V_{ddQ} (e.g. due to ground-bounce) will couple onto the sample node through the sampling switch parasitic capacitances, V_{ddQ} must be heavily bypassed to on-chip V_{ss}.

2.2.2 VCO Converter

Analog-to-digital conversion of the sampled voltage could be performed in a variety of ways, but in the application of supply noise measurement, it is especially desirable to perform this conversion without the need for an explicit voltage reference (since this reference would need to be carefully distributed around the chip). Therefore, we achieved the A/D conversion by simply using the sampled voltage to set the frequency of a VCO (which is typically an inverter-based ring oscillator with a controlled supply [27]), and then digitally estimating this frequency by counting the number of clock edges the oscillator outputs over a specified window of time, as shown in Figure 2.4 [25,28,29].

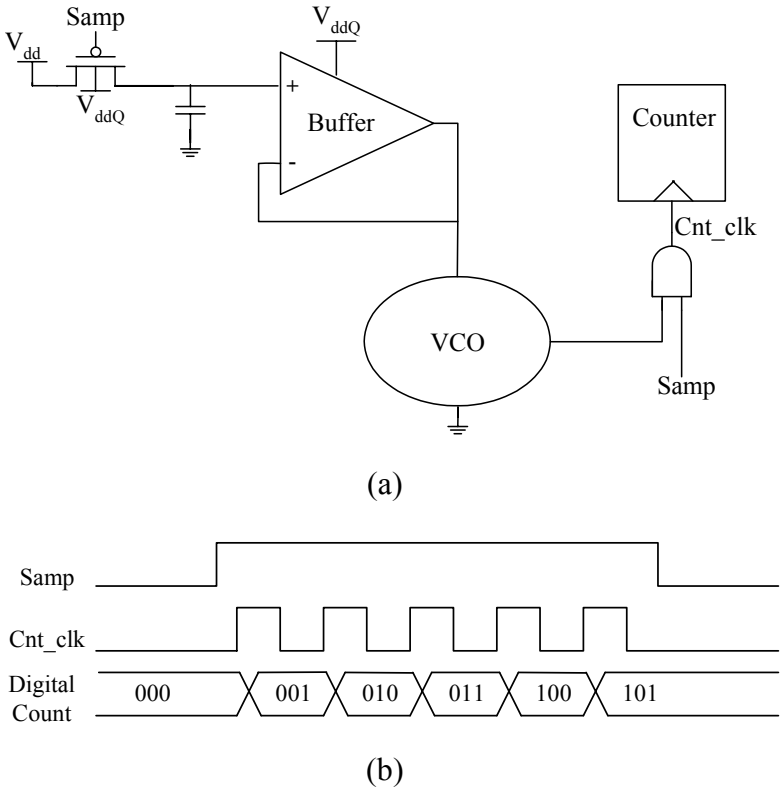


Figure 2.4: a) VCO converter schematic. b) Example waveforms.

Despite the fact that the oscillator is essentially free-running, the quantity that is being digitally measured is the average VCO frequency over the conversion window. Therefore, this scheme is insensitive to high frequency jitter (induced either by thermal

noise or noise from V_{ddQ}) because the counting process averages out cycle-to-cycle variations in the oscillator's period.

Unlike high frequency jitter, the averaging does not attenuate noise that is on the same time scale as the counting window. Fortunately, for noise frequencies too low to be averaged by the counter, on- and off-chip V_{ss} tend track each other very closely (since with a 1 μs conversion window, at these frequencies the impedance of the package inductance is negligible). Therefore, even if the on-chip bypass capacitance does not fully couple V_{ss} to V_{ddQ} at these frequencies, the external regulator and bypass capacitors will. Experimental results presented in the next section of this chapter verify that V_{ddQ} remains well-coupled to V_{ss} at these frequencies.

The buffer and VCO used in the converter were implemented with a voltage regulator and pseudo-differential regulated supply CMOS buffers [30]. Since V_{ddQ} must be quiet for the system to perform accurate measurements, the supply rejection requirements of the regulator are not very strict.

The resolution achieved by the VCO converter is set by the minimum change in voltage necessary to measure a difference of one count over the conversion window;

$$1\ \text{LSB} = 1/(T_{win}K_{vco}) \tag{2.2}$$

where K_{vco} (in Hz/V) is the VCO gain and T_{win} is the conversion time. One of the advantages of the VCO-based approach is its flexibility in achievable resolution; as long as leakage in the sampling switch is negligible,[3] the resolution of the converter can be improved by increasing the conversion time.

Since the phase of the oscillator is random with respect to the start of the counting window, the oscillator effectively adds a uniformly distributed noise with a magnitude of 1 LSB peak-to-peak to the output of the converter. This noise can actually be helpful

[3] Particularly in modern technologies, negligible leakage can be very difficult to achieve. Chapter 3 will describe techniques to both reduce the effective leakage of the sampling switch, as well as avoid the need for a sample and hold entirely.

because it creates dither in the output of the converter, allowing the resolution of the converter to increase as the measurements are externally averaged together.

In order to maximize the resolution for a given conversion time, K_{vco} should be as large as possible. Hence, if implemented as a ring, the operating frequency of the VCO should also be high. While ring oscillators can easily achieve 6-8 FO4 cycle times, a counter using a large adder with a latency of less than that could consume significant area and power. Fortunately, the counter's latency does not need to be set by the operating frequency of the VCO – only its throughput must be high enough that no counts are lost. Therefore, we used a simple toggle counter to minimize the required hardware.

If the VCO is implemented as a ring oscillator, the resolution can be further increased by counting the edges out of all phases of the VCO instead of only a single phase. By tracking all phases in the oscillator, the effective VCO cycle time can be reduced to the delay of a single stage of the VCO, which can be 1 FO4 or less. This technique may be particularly desirable in technologies where leakage in the sampling switch limits the resolution of the overall converter (as described further in Chapter 3).

One of the drawbacks of the VCO-based converter is that the voltage to frequency gain of the oscillator may fluctuate with temperature. Depending upon the VCO implementation, temperature variations can cause offset and gain errors in the measurement. However, with additional circuitry such as a local thermometer – which could be implemented by another VCO converter sampling a relatively constant voltage (e.g. a divided version of V_{ddQ}) – the temperature dependence of the oscillator could be measured during calibration and cancelled in post-processing.

2.2.3 Timing Generation

To measure repetitive waveforms, time-dependent distributions (i.e., infinite persistence on an oscilloscope), or noise spectra, the sampler timing pulses need to be placed with relatively fine resolution within a time basis that spans the longest events of interest. If package pins and metal tracks on the chip are available, the sampler timing signals can be generated in a relatively straightforward manner by external equipment (such as a pulse

generator) and then routed to the on-die measurement circuitry. For many implementations of the measurement circuitry (such as [25] and [31]), this approach was taken. However, in some applications it is necessary to generate the sampler timing signals using only internal circuitry, so we next describe an implementation of this timing circuitry (which was used in [32]).

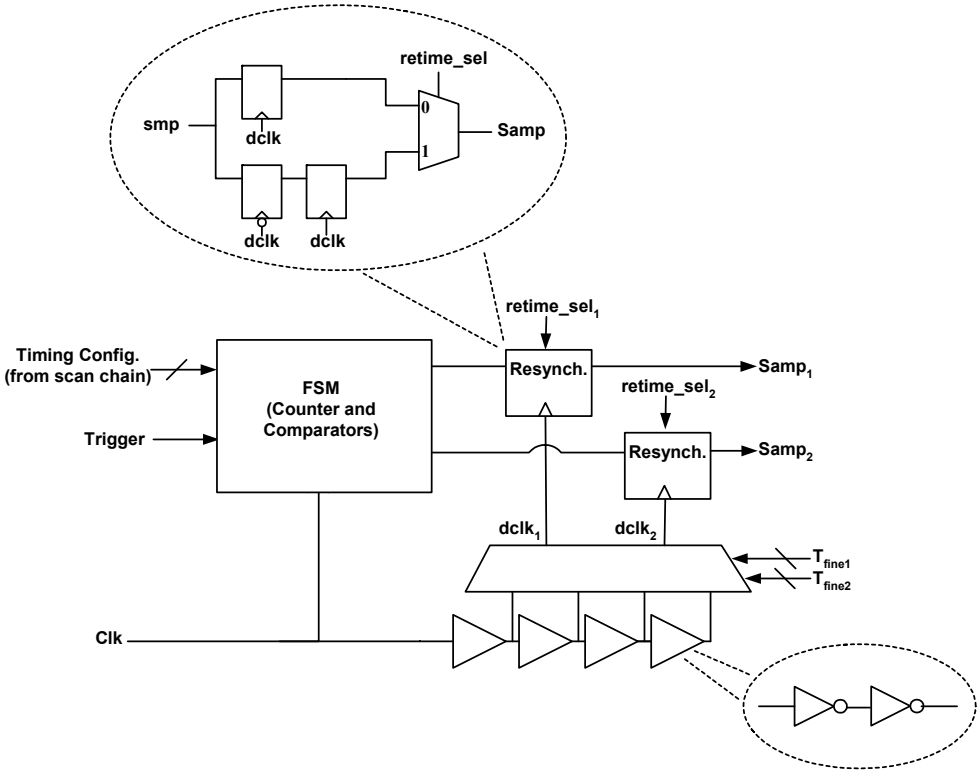


Figure 2.5: Coarse/fine architecture for sampler timing generation.

To meet the requirements of fine resolution within a wide time basis span with on-chip circuitry, we employed a coarse/fine architecture (Figure 2.5) where the timing signals are placed with clock cycle granularity by a scan-chain controlled state machine, and then adjusted with finer steps by a low-fanout inverter-based delay line. To avoid any dependency on the duty cycle of the clock, the state machine used only rising edges and the delay line was designed to cover at least one full clock cycle.

Instead of using two separate delay lines for the two sampling signals, the clock was fed through a single delay line and the outputs of the state machine were resynchronized by the selected delayed clocks ($dclk_{1,2}$) in order to minimize mismatches between the

paths of the two sampling signals. For small delay settings, the smp signals from the state machine were retimed by the rising edge of dclk. As the delay of a dclk approaches a full cycle, the smp signal may not meet the setup or hold time of the rising-edge triggered resynchronization flip-flop – leading to indeterminism in the signal's timing. Therefore, for these delays smp is first captured by the falling edge of dclk before it is retimed by the rising edge. The setting where this occurs can be found through calibration with on-chip flip-flop phase detectors that determine the early/late relationship between $Samp_1$ and $Samp_2$. These detectors are in any case necessary to calibrate the delay line by measuring the number of delay stages that span a known clock cycle, and as described in the next section, would also be required to calibrate routing mismatch in an implementation that generates the sampling signals externally.

Since the fine delay is generated by a set of inverters that run off of V_{core}, the timing of the sampling signals will be affected by the noise that the circuits are measuring. While this does cause filtering of the true supply signal, as long as the delay sensitivity of the inverters is similar to that of the other circuits on the chip (mostly logic gates), the measurement circuits will see the noise with the same (supply-noise jittered) timing reference that the rest of the circuits do.

2.3 Measurement Results

The proposed supply noise measurement circuits were implemented on several chips in a variety of technologies [25,31,32]. The most extensive characterization work was done on the first implementation of these circuits [25], and hence this section will focus on the results from this chip. This first chip with the noise measurement circuits was built in a 0.13 μm CMOS process, and included 4 high-speed links and a central ASIC that controls the links [33]. The measurement circuits on this chip had the capability to measure both the digital supply (V_{dd}) and the supply dedicated to the analog blocks such as the PLL and phase interpolators (V_{ddA}). In addition, noise generators were integrated to allow testing and validation of the measurement system.

2.3.1 Measurement System Calibration

Calibration is a key element of the measurement system since it enables the realization of a high degree of accuracy with relatively simple (and non-ideal) circuits. The first step in the calibration is to characterize the digital code versus input voltage curves of the two VCO-based converters. To perform this calibration, the rest of the chip is placed in a power-down state and the supply voltage is swept across the range of interest using an external power supply and the average digital code at each voltage is recorded. The measured VCO count versus voltage curves of the two samplers with a 1 μs conversion window are shown in Figure 2.6. At a 1 V supply, K_{vco} is roughly 2.6 GHz/V, corresponding to a nominal LSB of 385 μV.

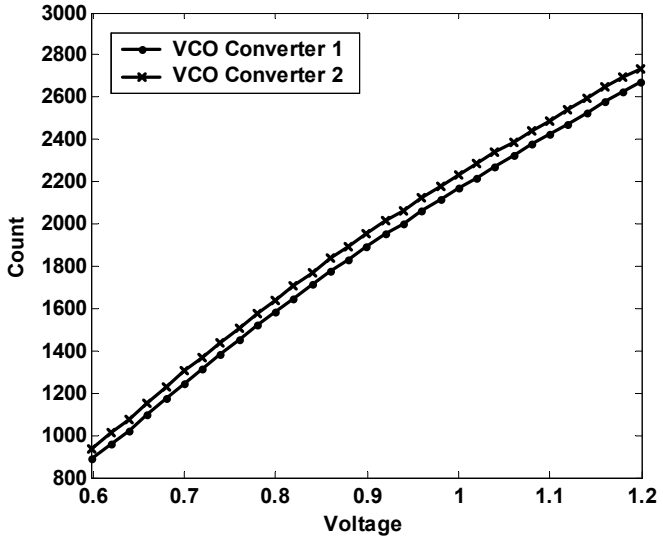
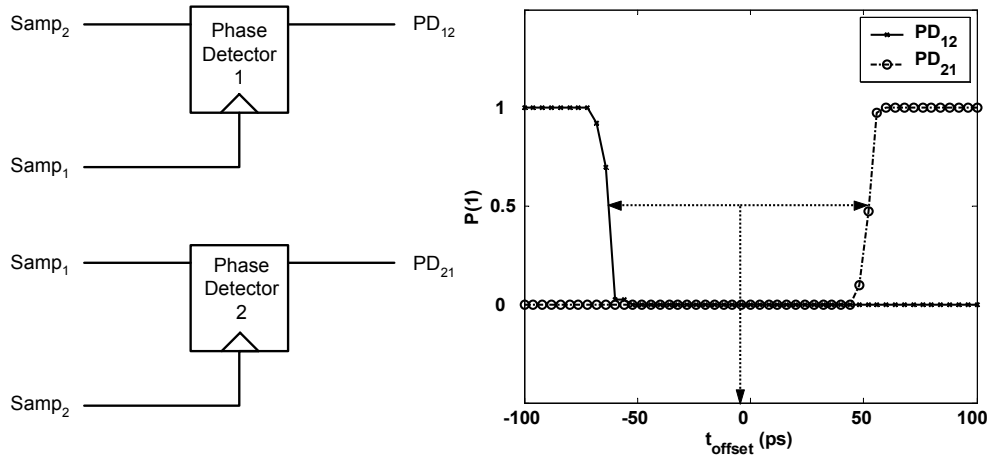


Figure 2.6: VCO calibration curves with a 1 μs conversion window.

In addition to voltage calibration of the A/D converter, timing calibration is necessary to characterize any skew between the two sampling signals caused by mismatches in routing from the pulse generator to the measurement circuits. To simplify the calibration procedure, two flip-flop phase detectors, shown in Figure 2.7a, are included as part of the measurement circuit.



(a) (b)

Figure 2.7: a) On-chip flip-flop phase detectors for timing calibration. b) Measured probability of each phase detector outputting a 1 versus timing offset between $Samp_1$ and $Samp_2$. Negative t_{offset} indicates $Samp_1$ is skewed to arrive after $Samp_2$.

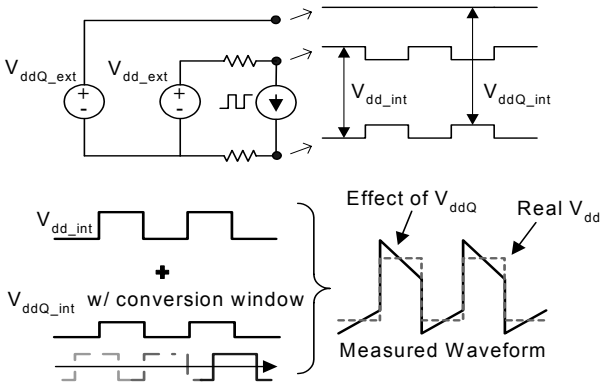
To measure the timing skew, the external pulse generator is used to sweep the timing offset between the two sampling signals and the outputs of the phase detectors are collected multiple times at each setting. The result of this measurement, shown in Figure 2.7b, is the percentage of time that each phase detector outputs a 1 (i.e. for phase detector 1, $Samp_1$ arrives after $Samp_2$) at each timing offset. If the setup times of the two flip-flops are identical, the timing skew can be calculated by finding the center point between the two 50% probability points.

Clearly, in using this timing calibration method any mismatch in the setup times of the two phase detectors will result in an error in the estimated skew. Since the setup time of the flip-flops is generally on the order of the minimum step-size in τ, and the mismatch between setup times should be only on the order of 10-20%, this timing error will not cause a large error in the measured autocorrelation. In addition, although rarely necessary, this calibration error can be removed by making use of the fact that the autocorrelation will be maximized when the two sampling signals are exactly aligned.[4]

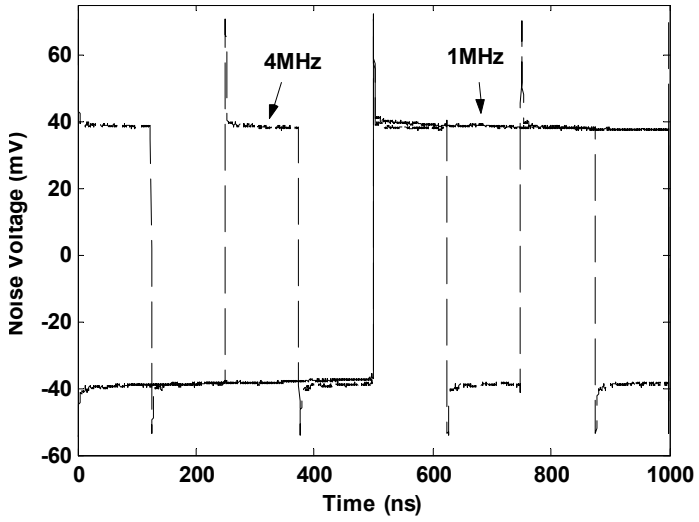
[4] Because of the variance associated with estimating autocorrelation using only a finite number of samples, this method will also result in calibration errors; however these errors can be minimized with a large number of samples.

2.3.2 Measurement System Validation and Characterization

Since noise on V_{ddQ} directly affects the sampled node, the main accuracy concern for this measurement system is the degree to which V_{ddQ} remains coupled to on-chip V_{ss}. In order to characterize this effect, we used the noise generator to inject square wave current onto the supply grid while the rest of the chip was inactive.



(a)



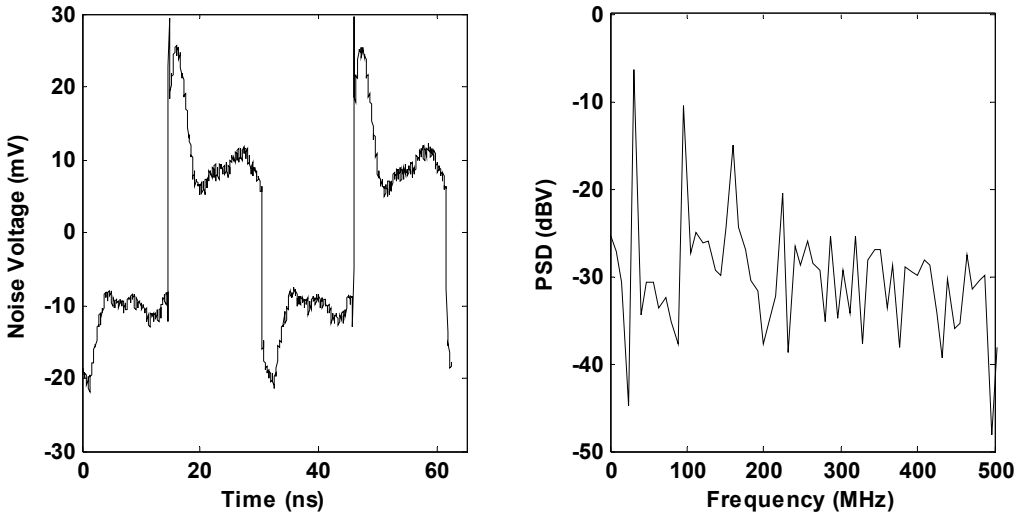
(b)

Figure 2.8: a) Potential effect of uncoupled V_{ss} noise through V_{ddQ} on measured waveform. b) Measured noise injected on V_{ddA} (mean subtracted) at 1 MHz and 4 MHz with a 500 ns conversion window.

The injected current causes the chip's power supplies to collapse towards each other. As shown in Figure 2.8a, if V_{ddQ} is not perfectly coupled to on-chip V_{ss}, it will effectively

move in the same direction as V_{dd} because of the change in V_{ss}. Since the VCO integrates V_{ddQ} noise over the conversion window, if the conversion window is the same width as the injected pulse, this uncoupled ground-bounce would cause the measured waveform to be somewhat triangular instead of square. The height of this triangle relative to the amplitude of the square provides a worst-case bound on the relative error of the measurements due to shifts in V_{ss} causing V_{ddQ} noise. With a 500 ns conversion window, the measured 1 MHz and 4 MHz waveforms of Figure 2.8b exhibit only negligible differences in magnitude and shape, showing that uncoupled ground-bounce on V_{ddQ} is minimal for this chip.

Despite the fact that these generated waveforms are deterministic, their spectral densities can be characterized using autocorrelation to validate the measurement procedure. A 32 MHz square wave was created on V_{dd} and the measured waveform and PSD are shown in Figure 2.9. The PSD exhibits the odd harmonics associated with the square-wave waveform.



(a) (b)

Figure 2.9: Measured a) waveform and b) PSD of injected 32 MHz square wave on V_{dd}. The PSD has been scaled by the Nyquist frequency (10 GHz) to calculate the power and shown with the unit (dBV) instead of the usual (dB V^2/Hz). For PSDs in dBV, the average level in dB corresponds roughly to the noise σ.

Finally, to complete the characterization of the samplers, the noise floor of the system was measured in two different ways. First, using a single sampler with the chip powered down, 100k samples were taken and the voltage distribution around the mean was recorded; the measured σ was 800 μV. Second, both samplers were used to measure the noise floor in the frequency domain. The measured noise floor of the PSD (overlaid in Figure 2.12) was 300 μV. The reduction from the single sampler is due to the averaging of the noise sources that are uncorrelated between the two samplers.

2.3.3 Measured Supply Noise

With the characterization of the samplers complete, the chip was activated with the links running at 4 Gb/s in serial loop-back on 2^{31} PRBS data. Before proceeding to measure autocorrelation, we used a single sampler as a sub-sampled oscilloscope to collect the distribution of the supply noise at each point in time (Figure 2.10). This measurement can provide a large amount of information about the behavior of the power supplies. For example, the noise on V_{dd} exhibits much larger random variations (and hence a larger peak-to-peak variation) than the noise on V_{ddA}. In fact, the large majority of the noise on V_{ddA} is due to a deterministic, repetitive noise waveform; this waveform can be extracted by taking the mean of the supply noise voltages at each point in time (Figure 2.11).

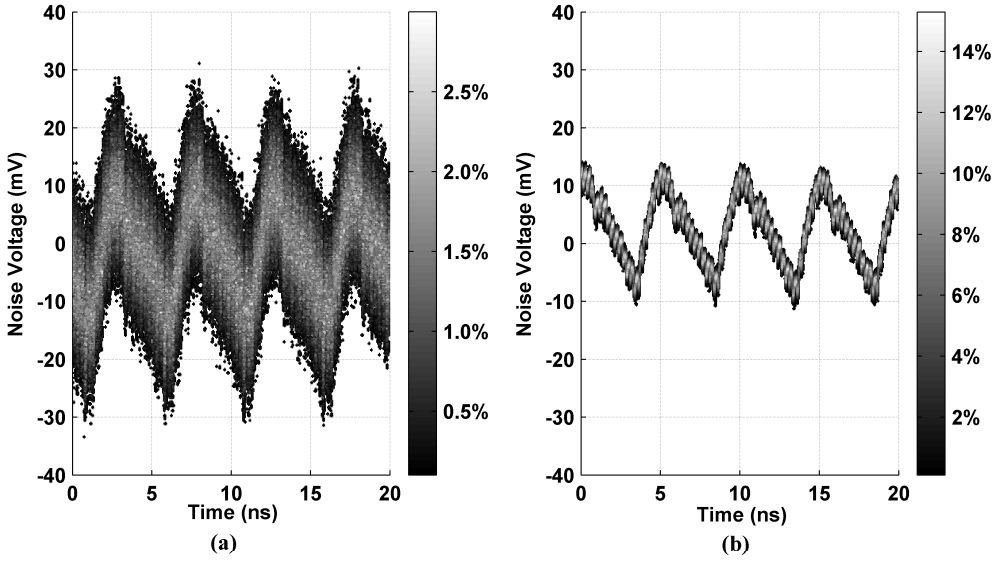
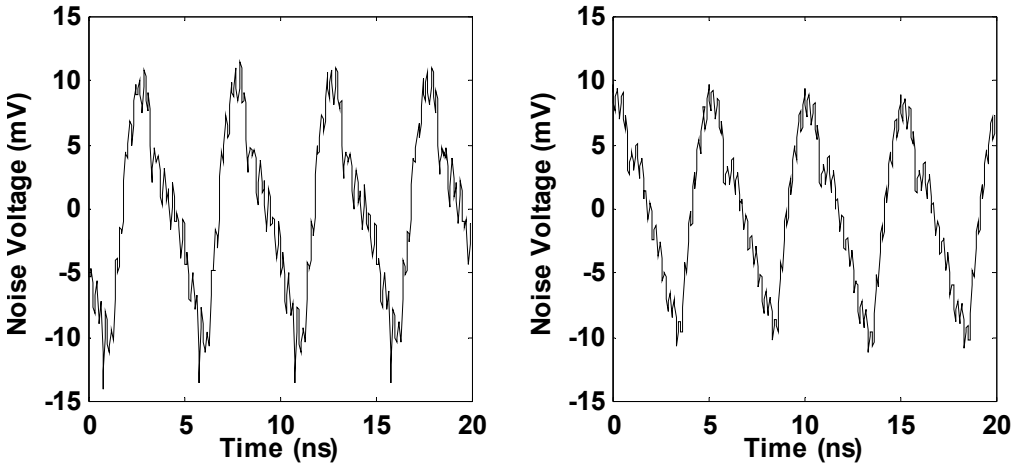
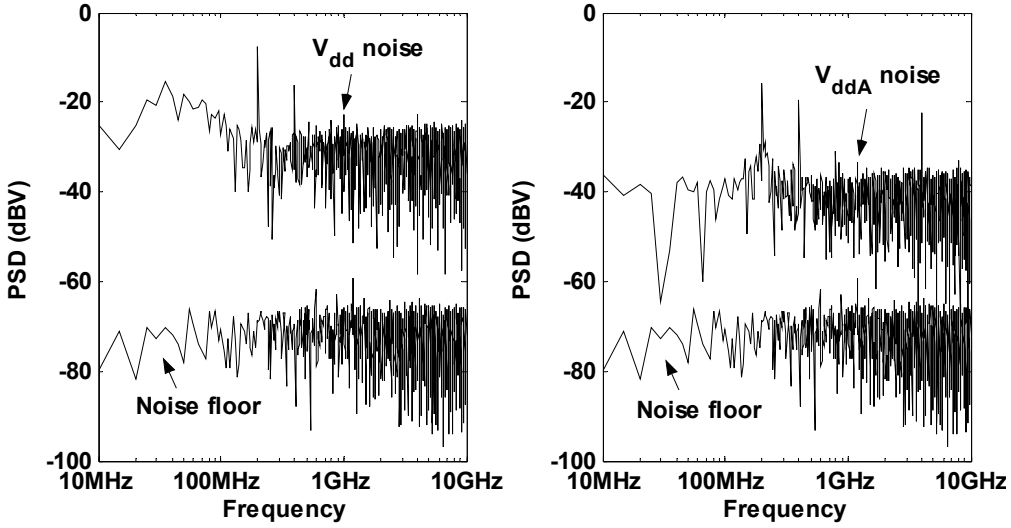


Figure 2.10: Measured noise distributions on a) V_{dd} and b) V_{ddA}.



(a) (b)

Figure 2.11: Measured deterministic noise on a) V_{dd} and b) V_{ddA}.



(a) (b)

Figure 2.12: Measured PSD with noise floor for a) V_{dd} and b) V_{ddA}.

To characterize the dynamics of both this deterministic waveform and the random supply noise, we measured the stationarized [34], or time-averaged PSD with the chip operating under the same conditions (Figure 2.12). As can be seen by the peaks in the spectral densities of Figure 2.12, the most dominant component of the repetitive waveforms appears at 200 MHz. Since the clock frequency of the ASIC core is 200 MHz for this data rate, and the core has its own supply voltage but a shared V_{ss}, this noise is

likely due to ground bounce that is not fully decoupled to the supplies, explaining why the waveforms on V_{dd} and V_{ddA} look very similar. An additional noise component exists at 400 MHz, which is the frequency of the reference clock that is fed to the PLLs in the links and drives some of the digital logic. Finally, the noise waveform also has a component at 4 GHz, likely due to transitions in the clock buffers that drive both polarities of the 2 GHz main clock, and due to modulation of the tail currents in the differential pairs used in the transmitter, VCO, clock buffers, and phase interpolators.

Both the distribution measurements from Figure 2.10 and the spectral densities from Figure 2.12 indicate that a large portion of the supply noise is due to these deterministic variations. Except for the frequency components of the deterministic signal, the PSD of V_{ddA} is nearly flat, implying that the noise is essentially white in nature. The PSD of V_{dd} is also relatively flat; however it exhibits some low frequency peaking in the 20-100 MHz range. Since the voltage waveform shown in Figure 2.9a created by injecting a square wave current onto V_{dd} exhibits damped sinusoidal ripple, this peaking is most likely due to the transfer function of the supply distribution network. In contrast, the square wave current response of V_{ddA} exhibits only overshoot on the transitions and no sinusoidal ripple, matching the relatively flat PSD. The difference in the impedance of the two distribution networks is intentional; V_{ddA} has a more resistive distribution network because it supplies less power than V_{dd} and because the additional series resistance is beneficial in isolating V_{ddA} from external noise.

For circuits that respond to noise equally at all points in time, the stationary PSD gives a complete description of the supply noise. However, the random noise is actually cyclostationary in nature. In the time-averaged spectral densities of Figure 2.12, the noise appears white because its time-varying behavior has been averaged by the measurement procedure, which for the stationary measurement samples the supply voltage uniformly throughout the noise cycle.

For cyclostationary noise processes, both the distribution and the dynamics of the noise can vary within the cycle. The cycle at which the characteristics of the noise repeat is not necessarily the clock cycle of the chip, especially if the chip has multiple clocks or

may exhibit somewhat repetitive modes of operation (e.g. a code loop running on a processor). Since the noise cycle may not be known *a priori*, the most straightforward method to detect the noise cycle is a guess-and-check approach. In many chips all of the clocks and/or events are harmonically related to each other, and therefore the measurement can be taken assuming that the noise cycle is some integer multiple of the slowest clock cycle in the system. If the integer multiple that was chosen is large enough that the assumed noise cycle encompasses two or more true noise cycles, the measured characteristics of the supply noise will repeat within the assumed cycle.

On this chip the cyclostationary behavior was not very pronounced, and therefore, instead of presenting the cyclostationary spectral densities at each time in the cycle, we measured the PSD of the random noise at two different times in order to observe examples of noise variation. Additionally, in order to make the cyclostationary behavior more apparent, we decreased the operating frequency of the links to 2 Gb/s. At this reduced rate, the majority of the logic should complete before the beginning of the next clock cycle, causing a period of relative calm on the supply.

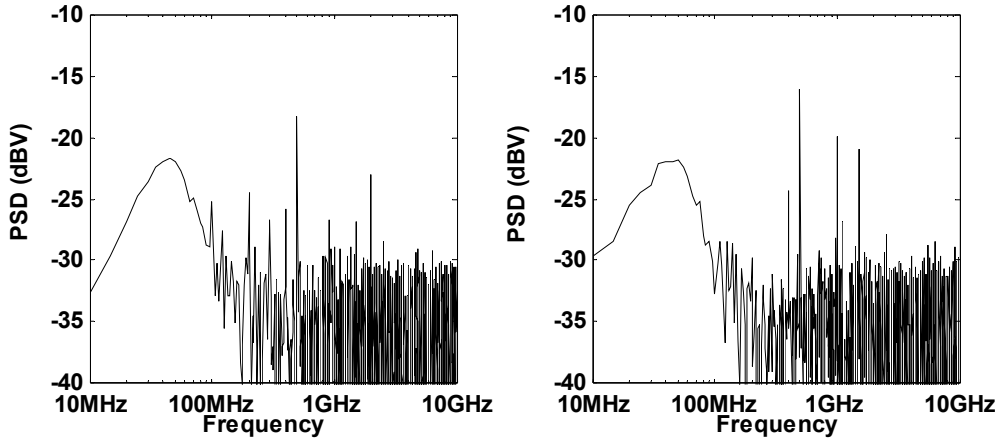


Figure 2.13: Measured PSDs of V_{dd} noise at two different times in the noise period with the links operating at 2 Gb/s.

Although the total power in the PSD shown on the right of Figure 2.13 is only slightly higher than that of the PSD on the left, their dynamics are clearly different - the PSD on the right has a strong component at 1 GHz that is not present in the other PSD.

Since a large portion of the transceiver runs off of a 1 GHz clock at this data rate, these two spectral densities correspond to times of relatively low and relatively high switching activity in the link circuits.

2.4 Summary

In this chapter we have presented a measurement system for characterization of power supply noise using two low-rate samplers to enable autocorrelation (and hence noise spectrum) measurement. The system made use of a VCO-based A/D converter and calibration procedure in order to achieve high-resolution measurements with relatively simple circuitry.

The system was used to measure supply noise on a high-speed link chip, and a deterministic signal of ~20 mV peak-to-peak was detected on both the digital and analog supplies. In addition to this deterministic signal, the non-deterministic portion of the noise was shown to exhibit cyclostationary behavior, and example time-indexed spectra showing the cyclically varying properties of this noise were presented.

By integrating such supply noise measurement systems in multiple locations on their chips, designers can use verified noise distributions and spectral densities to characterize both the supply grid and the impact of supply noise on their circuits. As the issue of supply integrity is most acute in chips built in advanced technologies, performing this type of validation on these chips is of great interest. However, because of issues caused by leakage in these advanced, digital-optimized transistors, many of the analog circuits which are part of the measurement system (in particular, the sample and hold) do not scale very well to these technologies. Therefore, in the next chapter, we present improvements to the sampling circuitry that mitigate the impact of source-drain leakage, as well as a measurement technique that eliminates the need for a sample and hold (and hence separate supply) by averaging many dithered, low-resolution samples.

Chapter 3

Supply Noise Measurement in Modern Technologies

While the measurement system we described in the previous chapter is flexible enough to collect noise distributions, sub-sample periodic waveforms, and extract the spectrum of supply noise, it requires high-bandwidth, low-leakage sample and hold (S/H) circuits as well as analog buffers. Unfortunately, these analog circuits – particularly the S/H – can be very difficult to scale to modern, performance-optimized CMOS technologies. These issues are of particular importance when a separate, clean power supply and/or alternative, low-leakage (i.e., thicker gate oxide and higher threshold voltage) transistors are not available.

Our first opportunity to implement the supply noise measurement system in an advanced technology was on a 90nm dual-core Itanium microprocessor code-named Montecito [32]. While a separate power supply was available for this design, for cost reasons the process did not include low-leakage devices. Therefore, in the first part of this chapter, we present improvements to the S/H circuit design that significantly reduce the impact of source-drain leakage on the performance of the sampler, as well as measurements taken from the processor using this sampler design.

Although the circuit techniques used in the Itanium sampler were successful in limiting the impact of source-drain leakage, they increase the overall analog complexity of the design and rely on the existence of a separate power supply – both of which make integrating the noise measurement system onto a broad variety of chips more difficult. As we describe in the next section, modifications can be made to the sample and hold circuitry to eliminate the need for a separate power supply – but these modification further increase the complexity of the design and require additional calibration procedures to guarantee accurate final measurements.

Therefore, in the last section of this chapter we describe a measurement system that eliminates both the analog complexity of the S/H and the overhead of a separate power supply by averaging many dithered, low-resolution samples. After describing one implementation of such a system, we conclude with measurement results comparing its performance with a S/H-based system, and comments on the tradeoffs between the two approaches.

3.1 Improvements for Itanium Measurement System

As described in the previous chapter, the VCO-based ADC that is part of the measurement system uses the supply voltage sample to set the frequency of the VCO, and then counts clock pulses over a fixed time window to estimate this frequency. The resolution of this ADC is set by the Hz/V gain of the VCO and by the width of the conversion window; to measure supply noise with mV-level resolution the conversion window is typically 100's of ns wide.

Due to the unavailability of high threshold, thick oxide devices, the sampling switch had to be implemented using standard, high-leakage devices. To avoid undesirable filtering of the measurements, during hold mode the sampled voltage should be as independent of the supply voltage as possible; with standard devices the leakage time constants were often significantly shorter than the conversion window, making this isolation very difficult to achieve.

To mitigate the source-drain leakage of the sampling switch and hence improve its isolation, the sample and hold was modified from the previous design as shown in Figure 3.1. The PMOS switch was split into two, and during hold mode the unity gain buffer forces V_{mid} to track V_{samp}. The voltage across the switch connected to C_s is hence nearly zero (limited by buffer offset), and the switch's source-drain leakage is essentially eliminated.

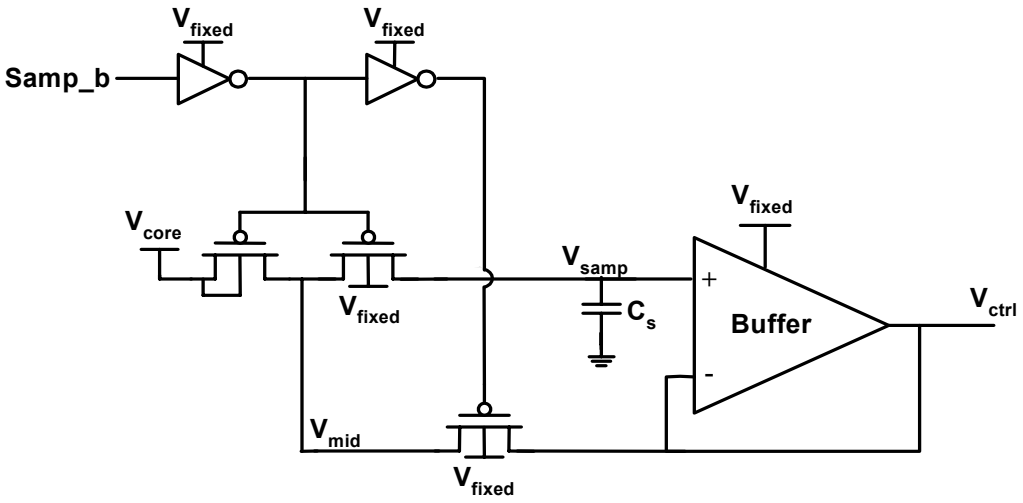


Figure 3.1: Sample and hold with source-drain leakage cancellation. The sampler circuits were powered off of a separate, relatively quiet supply V_{fixed} that was in any case required for the clock generation circuitry.

With their 1-2 nm thick gate oxides, modern high-performance transistors often exhibit significant gate leakage; at high supply voltages, the gate leakage current can even be larger than the source-drain leakage [35]. Fortunately, because of the use of an independent supply V_{fixed}, the gate leakage of the sampling switch is of less concern than source-drain leakage. This is because the effect of gate leakage on the sampled voltage is independent of future values of V_{core}, and hence it does not lead to filtering of the measurements. However, to maximize the ADC's resolution it is still desirable to keep the net leakage current low. While the gate leakage of the sampling switch could be eliminated in a manner similar to the source-drain leakage, this step was not taken because the gate leakage of the buffer opposes the gate leakage of the switch (V_{fixed} is raised above V_{core} when taking measurements to keep all of the devices in the buffer in

saturation), and hence canceling the switch gate leakage can often result in a lower leakage time constant.[5]

3.1.1 Itanium Measurements

Figure 3.2 shows the distribution and spectrum obtained from measuring the noise with both cores operating at 1.4 GHz with a nominal supply of 1.05 V; the cores were set to toggle between a power virus and low activity. As shown by the sharp dips in the distribution which repeat every clock cycle and the corresponding pulse train in the noise spectrum, a significant amount of noise is generated by clock-related activity that repeats every cycle. This is most likely because the flip-flops and many of the logic gates toggle at or near the rising edge of the clock, leading to a current profile with repetitive pulses whose magnitudes are modulated by the number and size of the transitioning gates.

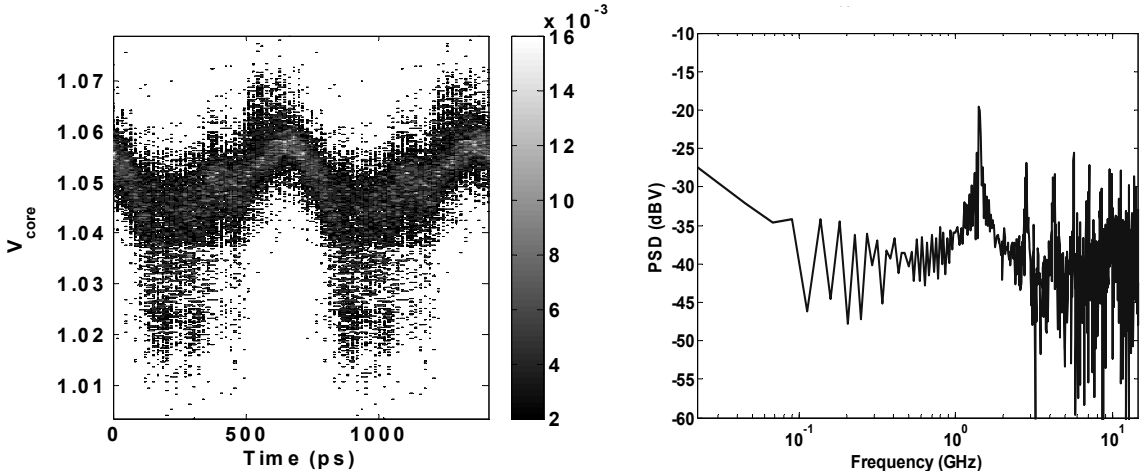


Figure 3.2: Measured supply voltage distribution and spectrum with both cores toggling between high and low activity; the clock frequency is 1.4 GHz and V_{core} is 1.05 V.

The spectrum also shows an increase in the noise density at the lower frequencies (less than ~50 MHz) that is likely due to the resonance of the power distribution network. Since this lower frequency noise persists for many clock cycles, in a standard digital

[5] The magnitudes of the two gate currents (from the sampling switch and from the buffer) are sensitive to sizing and process variations, and hence canceling the switch gate leakage can sometimes be beneficial. To cover both cases the cancellation needs to be programmable, but the potential benefit in resolution is generally not worth this complexity.

system the noise directly impacts performance because the critical paths must meet timing at the lowest voltage.

Montecito implemented an adaptive frequency control system that adjusts the processor's operating frequency based on the current supply voltage [36,37]. One of the advantages of this adaptive frequency control is that the chip adjusts itself to these low-frequency variations and hence on average achieves higher performance. In this case, the measurements show ~70 mV of peak-to-peak noise, indicating that along with its other advantages, dynamic frequency management can improve performance by ~5% even in relatively quiet conditions [36].

3.2 Samplers without an Independent Power Supply

The main motivation for using a separate power supply for the sample and hold circuit is to isolate the sample node from the supply that is being measured while the sample is being converted. However, the use of a separate, higher than nominal power supply also solved the practical issue of providing headroom for the buffer to drive the sampled voltage onto the control node of the VCO. To solve the practical issue of buffer headroom without the use of a separate supply, the sampler circuit can be modified as shown in Figure 3.3.

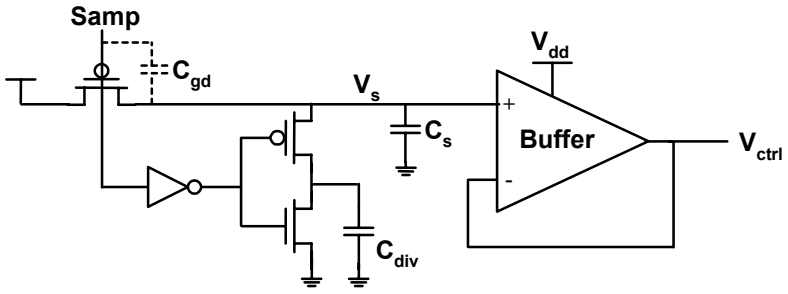


Figure 3.3: Sample and hold powered by the same supply that is being measured. For simplicity, the circuit modifications for source-drain leakage cancellation (from Figure 3.1) are not shown here, but in high-leakage technologies may need to be applied to this sample and hold as well.

In this sampler, capacitive charge sharing is used to divide down the sample voltage and guarantee that it never exceeds the current supply voltage. Clearly, dividing the

sampled voltage (reducing its magnitude) degrades the per-sample signal-to-noise ratio (SNR) of the measurement system. In addition, the effective resolution of the VCO-based ADC will be degraded, since a larger change in the original supply voltage would be required to change the frequency count by one step.

Fortunately, the degree of division is typically not high. This is because the division ratio is set only by the requirement that after sampling the maximum possible voltage, the buffer must have enough headroom to drive the sample voltage even if the supply later reaches its minimum value. For example, if the supply voltage ranges from 0.95 V to 1.05 V and the buffer headroom is 150 mV, the sampled voltage can be set to ¾ of the original supply voltage.

While dividing the sample voltage solves the headroom issue, the more important issue of sample node isolation must still be handled. As mentioned in the previous chapter, even if source-drain leakage is negligible, the voltage at the gate and body of the sampling switch will couple into the sample node through the sampling switch's gate-to-drain[6] and body-to-drain capacitances. During conversion, both the gate and the body of the sampling switch are connected to V_{dd} (the supply being measured), causing the sample node voltage to be set not only by the value of the supply voltage at the sampling instant, but also by the current value of the supply voltage. In other words, the sample voltage V_s can be written as:

$$V_s(t) = V_{dd}(t_{samp}) + k_f \cdot \left(V_{dd}(t) - V_{dd}(t_{samp})\right) \tag{3.1}$$

where $(V_{dd}(t)-V_{dd}(t_{samp}))$ is the feedthrough term, and k_f is the feedthrough coefficient from V_{dd} to V_s (i.e., $k_f \approx (C_{gd}+C_{bd})/(C_{gd}+C_{bd}+C_s+C_{div})$).

As in the previous sampler designs, the VCO-based ADC adds the average value of the feedthrough over the conversion window to the voltage sample. However, unlike the previous designs where the additional supply could be strongly coupled to V_{ss}, the

[6] Since the sampling switch is off during conversion, the gate-to-drain capacitance in this case is set mostly by the overlap and fringing fields from the gate to drain – which can be ~20% or more of the total gate capacitance in modern technologies [38].

feedthrough signal can not be decoupled from the behavior of the supply being measured. The impact of this is that instead of taking instantaneous samples of the supply voltage, the sampler essentially measures a filtered version of the supply noise signal. The transfer function of the filter the sampler applies to the measurement can be expressed as:

$$H_{samp}(s) = (1 - k_f) + k_f \cdot H_{window}(s) \tag{3.2}$$

where $H_{window}(s)$ is the transfer function of a rectangular window of width T_{win}, which has the form of a sinc [39].

Since the form of the filter that the sampler applies to the measured supply noise is known, an unfiltered measurement can be recovered in post-processing by applying the inverse transfer function to measured repetitive waveforms or noise spectra.[7] While it is conceptually simple to apply this inverse filter, the correction can be simplified even further if only high frequency (i.e., $f >> 1/T_{win}$) components of the supply noise are of interest. This is because at high frequencies, the conversion window averages out all of the supply variations, and $H_{samp}(s)$ is then very well approximated by $(1-k_f)$. In fact, this approximation is exact for frequency components that are harmonics of $1/T_{win}$. Therefore, the correction can simply be achieved by multiplying the measured (high-frequency) spectrum by $1/(1-k_f)$.

3.2.1 Feedthrough Calibration

While correcting for the effect of feedthrough on the sampler is mathematically straightforward, the feedthrough coefficient k_f is strongly dependent upon several process parameters such as gate oxide thickness, dopant concentration, and metal line width and spacing. Therefore, to guarantee accurate final results (especially if multiple samplers are being used to compare noise at various locations on the die), it is highly desirable to perform a calibration that directly measures the level of feedthrough in each sampler.

[7] While repetitive waveforms and noise spectra can be recovered in this manner, "infinite persistence" measurements taken via sub-sampling can not. This is because in the case, the inverse filter would need to be applied to the full time domain waveform, which would not be available from a sub-sampled measurement.

To perform this calibration, an AC signal of known (or externally measurable) magnitude must be induced onto the power supply and then measured using the samplers. Since harmonics of $1/T_{win}$ are attenuated by exactly $(1-k_f)$, to simplify the calibration it is desirable to induce noise on the supply only at $1/T_{win}$ or its harmonics. One of the most straightforward ways to achieve this is to create a square-wave whose frequency is $1/T_{win}$ (or a harmonic of $1/T_{win}$) on the power supply, in which case k_f can be calculated by

$$k_f = 1 - A_{sampler} / A_{square} \tag{3.3}$$

where $A_{sampler}$ is the magnitude measured by the sampler, and A_{square} is the known magnitude of the square-wave.

Conceptually, this calibration procedure is quite straightforward, but performing this procedure in real systems often leads to several practical difficulties. Specifically, while on-chip circuitry to internally generate voltage waveforms of known magnitude on the power supply could be built, such a design would likely require either several calibration steps (for example, as described in [37]), or multiple package pins to allow external probing of the on-chip voltage waveforms. In addition, these circuits would consume significant area (relative to the samplers) since they would need to sink large currents (in a high-performance processor, 10's of amps) in order to generate reasonable voltage magnitudes on the low-impedance supply network.

To avoid these on-chip overheads, the calibration waveform could be generated (and its magnitude measured) off-chip with external equipment and/or power transistors on the PCB. However, this approach is not without its difficulties. The calibration square wave's frequency should be above $1/T_{win}$, but to guarantee that the on-chip waveform closely matches the off-chip waveform, the frequency should also be well below the bandwidth of the package distribution network. Typically, this leads to a square wave[8] in the low MHz range. While power transistors and/or external equipment that can switch

[8] Although a sine wave may seem preferable to avoid transients in the voltage waveform (e.g. ringing due to the supply inductance), any clipping in the sine wave would negate this advantage. Instead, the transient issue can simply be avoided by calculating the square wave magnitude based on the height of the flat region of the waveform (the square wave should be low enough in frequency for the transients to die out).

10's of amps at this frequency are available, large decoupling capacitors on the PCB may lower the impedance of the supply network on the board even below the on-chip impedance – requiring even larger currents to achieve reasonable voltage amplitudes.

With careful design and planning, all of the difficulties associated with performing the AC calibration can be overcome, but a sampler design that avoids the need for this type of dynamic calibration would significantly reduce the barriers to adopting the measurement system. Since the coupling issues which forced the need for calibration arise from the fact that the gate of the PMOS switch is tied to V_{dd} while A/D conversion is being performed, one initially tempting alternative is to use to an NMOS-based sampling switch where the gate is tied to V_{ss} during conversion.

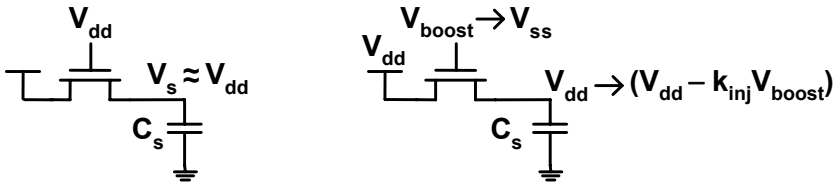
3.2.2 NMOS-Based Sample and Hold

The most immediate issue with using an NMOS sample and hold to measure the supply voltage is that if the gate of the sampling switch is only driven to V_{dd} (as shown in Figure 3.4a), the switch will be cut-off and the sampler's bandwidth will be extremely low. To improve the overdrive of the NMOS switch, a boosted supply could be generated with a charge pump, or a more sophisticated (and complicated) bootstrapping scheme such as [40] could be used, but in both cases charge would be injected onto the sample node during the switch turn-off. Since the NMOS gate voltage (and hence the injected charge) will depend upon previous values of the supply voltage, as shown in Figure 3.4b, the sampler would still effectively measure a filtered version of the supply noise.[9]

To improve the overdrive of the NMOS sampling transistor without the complexities of a boosted supply or bootstrapping, the supply voltage could be divided at the input of the sampling switch using a resistive voltage divider – as shown in Figure 3.5a. As with the charge sharing sampler, the larger the voltage division, the worse the effective voltage resolution of the ADC and the SNR of each measurement will be. These concerns are

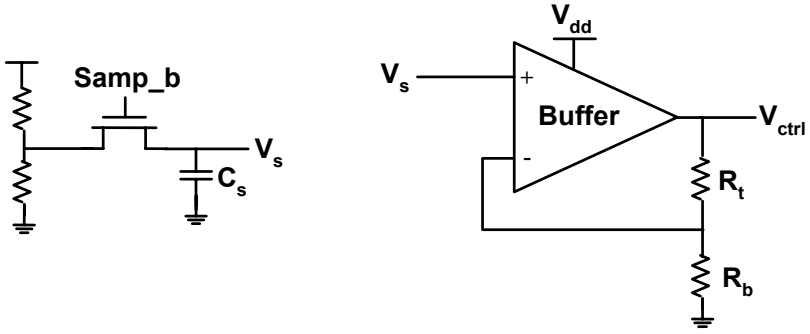
[9] This filtering can not be solved by adding a charge-injection cancellation switch [41], because the gate of the cancellation switch would be tied to the boosted voltage during A/D conversion. Low-pass filtering the output of a charge pump that generates a boosted supply to drive the sampler's gate does not solve this issue either – the charge pump inherently samples the measured supply, and hence aliases some high frequency portions of the supply noise into the bandwidth of the low-pass filter.

more relevant in the NMOS sampler, because to achieve a reasonable overdrive on the sampling switch, the input to the sampler typically needs to be divided by a factor of two to three. In addition, because charge injection during the switch turn-off further reduces the sample node voltage, the effective division ratio is typically even higher than that set by the resistive divider – often leading to division by a roughly a factor of 5.



(a) (b)

Figure 3.4: a) Terminal voltages for an NMOS S/H. b) Terminal voltages for an NMOS S/H with a boosted gate voltage before and after the sampling switch is turned off.



(a) (b)

Figure 3.5: a) NMOS sampling switch with resistively divided input. b) Buffer with voltage multiplication to recover ADC resolution lost from dividing the input supply.

While the SNR of the measurement with respect to kT/C noise can only be recovered by increasing the sampler capacitance C_s (and hence power consumption) or increasing the degree of averaging, some of the voltage resolution can be recovered by using the buffer amplifier to introduce a multiplication factor between the sampling voltage and the VCO control voltage – as shown in Figure 3.5b. The buffer's allowable multiplication factor is once again set by the need to maintain buffer headroom at the minimum supply voltage after sampling the maximum supply voltage:

$$A_{buffer} = \left(V_{dd,min} - V_{od}\right) \Big/ \left(A_{div} \cdot V_{dd,max}\right) \tag{3.4}$$

where A_{buffer} is the gain of the buffer (~$1+R_t/R_b$), V_{od} is the buffer headroom, and A_{div} is the ratio between V_{dd} and V_s after sampling (i.e., including the switch's charge injection).

Despite the reduction in per-sample SNR, this configuration does initially look promising since it does not appear to include any direct coupling paths between V_s and V_{dd} while A/D conversion is being performed. However, because of the voltage division, in order for the buffer amplifier to function properly it must accept input voltages in the vicinity of 200-300 mV. To avoid subthreshold operation, this low input voltage essentially rules out the use of a buffer amplifier with an NMOS-based differential pair as the input stage of the amplifier. Therefore, the buffer amplifier must use a PMOS-based input stage – such as the level-shifting source follower shown in Figure 3.6.

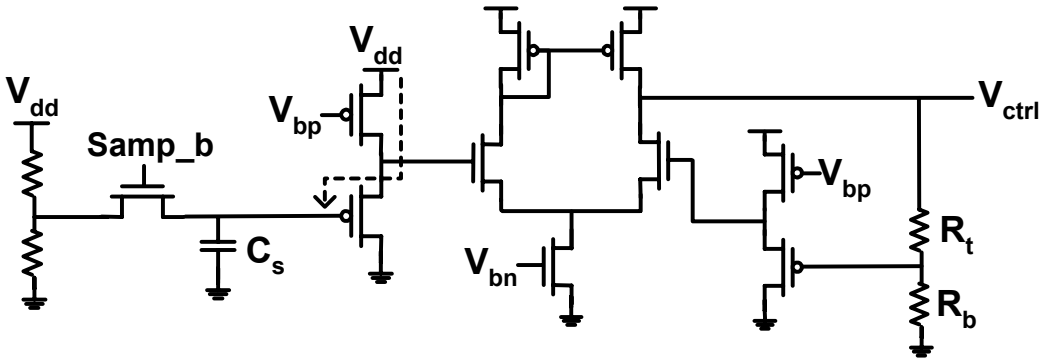


Figure 3.6: Example transistor-level implementation of an NMOS-based S/H and buffer amplifier. To keep the NMOS differential pair in saturation, the buffer uses PMOS source followers to level-shift the 200-300 mV input voltage from the S/H.

Unfortunately, a PMOS-based input stage will couple the sample node to V_{dd} through the C_{gs} of the PMOS device – as shown by the dashed arrow in Figure 3.6. While the magnitude of the direct coupling to the sample node is typically lower than with a PMOS-based sampler, the impact of this coupling on the effective filtering of the sampler must be evaluated relative to the divided supply voltage. For example, if the magnitude of the coupling from V_{dd} to V_s during conversion is 2% and A_{div} is 0.2, the effective feedthrough would be 10%. Therefore, despite the initial promise of an NMOS-

based sampler in eliminating the coupling issues, this design also typically requires dynamic calibration to guarantee accurate measurements.

In both the NMOS and PMOS-based samplers, leakage cancellation may need to be employed to achieve acceptable leakage time constants – although the NMOS sampler does suffer less from source-drain leakage since the V_{gs} of the switch is negative during conversion. Whether leakage cancellation is employed or not, eliminating the separate power supply increases the analog complexity of the sampler design, and leads to the requirement for dynamic calibration procedures to ensure accurate measurements. While these issues can be overcome to realize an accurate, high-resolution noise measurement system, these designs are very sensitive to the detailed transistor parameters. Hence, such designs require significant verification and optimization – making porting much more challenging. Since nearly all of the issues can be traced back to the need for a S/H, in the next section we describe a measurement system that eliminates the S/H altogether by averaging many dithered, low-resolution samples.

3.3 Averaging-Based System

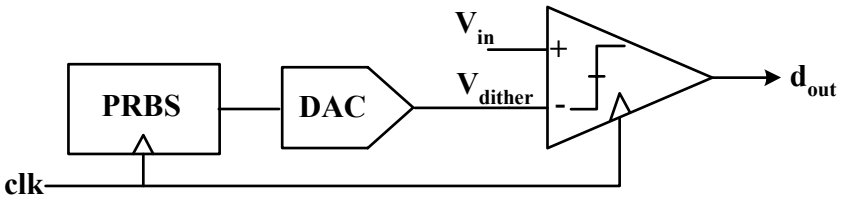


Figure 3.7: Use of a DAC and pseudo-random bit sequence to generate dither and increase the resolution of a single-bit ADC (i.e., a single comparator).

It is well known that averaging multiple low-resolution (i.e., coarsely quantized) samples yields a higher resolution measurement if the samples have an appropriate dither signal added to them [42]. To make use of this concept, in many ADCs the dither signal (which should be zero-mean) would have to be artificially created. To maintain reasonable complexity, this dither is typically created by using a digital-to-analog converter (DAC) and a pseudo-random code (as shown in the single-bit ADC example of Figure 3.7).

However, as one may expect, the resolution of the DAC will limit the achievable resolution of the measurement (even after averaging) [43].

Fortunately, as we mentioned in the previous chapter, in a VCO-based converter this resolution-enhancing dither is inherent and truly uniformly distributed – avoiding the resolution issues of artificially generated dither. A VCO-based converter's dither stems from the fact that the VCO is oscillating asynchronously with the rest of the system, which means that the initial phase of the VCO at the beginning of each conversion window is random and uniformly distributed between 0 and 1 period.[10]

At the end of the conversion window, a deterministic phase shift proportional to the sampled voltage V_{dd} of $K_{VCO} \cdot T_{WIN} \cdot V_{dd}$ is added to the random initial phase. The counter quantizes the resulting total VCO phase with a quantization step of 1 period, resulting in

$$\varphi_Q = \text{floor}[\varphi] = \text{floor}[\varphi_V + \varphi_N] = \varphi_V + e_Q \tag{3.5}$$

where φ is the total VCO phase shift, φ_V is the deterministic phase shift due to the sampled voltage, φ_N is the random initial phase, and e_Q is quantization noise. Since φ_N is independent of the measured voltage and spans an entire cycle, the quantization noise is uncorrelated with the sample voltage and has zero mean (in other words, $E[e_Q] = 0$). Thus, averaging the quantized measurements φ_Q yields the average of φ_V, and hence the average voltage sample:

$$E[V_{dd}] = E[\varphi_Q]/(K_{VCO} \cdot T_{WIN}) = E[\varphi_V]/(K_{VCO} \cdot T_{WIN}) \tag{3.6}$$

where E[] denotes the expectation operation.

Of course, in order to perform averaging the input signal must be periodic so that the same voltage can be measured multiple times; such periodic waveforms can be measured with low-resolution samplers by sub-sampling, but any non-periodic variations will be removed by the averaging many samples together. Directly related to this is that with

[10] For clarity, all phases will be measured in units of VCO periods instead of degrees or radians.

low-resolution sampling, time-dependent distribution measurements (e.g., left of Figure 3.2) can no longer be taken by simply collecting many samples and then plotting their histogram, because each sample will be corrupted by the coarse quantization and dither.[11]

Although time-dependent distributions can no longer be directly measured with a low-resolution VCO-based converter, the dynamics of the non-repetitive noise can still be recovered by measuring autocorrelation:

$$R(\tau) = E[V(t-\tau/2)V(t+\tau/2)] = E[\varphi_V(t-\tau/2)\varphi_V(t+\tau/2)]/(K_{VCO}\cdot T_{WIN})^2 \tag{3.7}$$

Just like repetitive waveforms, averaging individually dithered measurements also increases the resolution of the measured autocorrelation. This is because the two samplers measuring $\varphi_V(t-\tau/2)$ and $\varphi_V(t+\tau/2)$ have uncorrelated zero-mean quantization noise. Therefore, the mean of the quantization noise of the product of the samples remains zero, and

$$R(\tau) = E[\varphi_Q(t-\tau/2)\varphi_Q(t+\tau/2)]/(K_{VCO}\cdot T_{WIN})^2 \tag{3.8}$$

Thus, averaging allows high-resolution measurements of both voltage and autocorrelation to be obtained from low-resolution samples. We can make use of this technique to dispose of the S/H and buffer amplifier (and their associated issues) altogether.

As shown in Figure 3.8, if the conversion window in a VCO-based system is short enough that the input voltage stays roughly constant during the entire window, then no explicit S/H is required. The VCO (i.e. ring oscillator) runs directly off of V_{dd} such that its instantaneous frequency tracks the instantaneous supply voltage. Just as with a high-resolution VCO-based ADC, this frequency is then estimated by counting the number of edges that occur within a given window – but in this case the window is significantly shorter than a single period.

[11] However, with a controllable DAC (either to generate the dither signal, or to sweep the effective thresholds of the ADC), time-dependent distributions can be measured to within the resolution of the DAC.

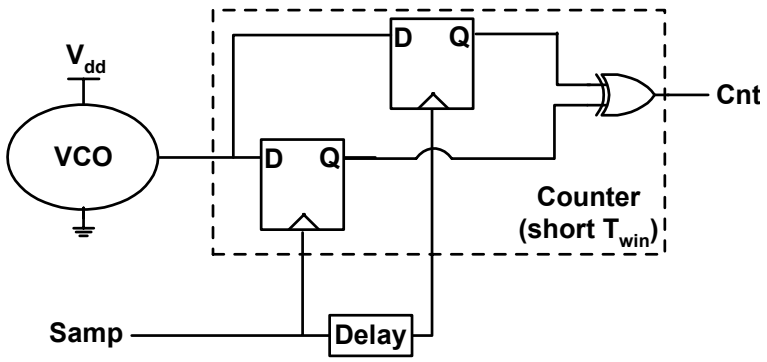


Figure 3.8: Low-resolution VCO measurement circuit.

While the counter could be implemented as described in the previous chapter (i.e., with an AND gate and a toggle flip-flop), since the window is significantly shorter than a VCO period, in this case it is desirable to detect both rising and falling edges of the VCO. As compared to detecting only rising edges, this has the effect of increasing the K_{vco} by a factor of 2. To detect any transitions in the VCO output, the Samp signal latches in the current phase state of the VCO and then the phase state of the VCO after a short, roughly supply-independent delay. If the VCO output has transitioned, the outputs of the two latches will differ, and hence XOR'ing the two outputs indicates the "count" of interest.

This counter essentially acts as a single-bit quantizer of $\varphi = \varphi_N + \varphi_V$, where φ_V is the phase accumulated by the VCO over a short delay, and φ_N is the VCO's random initial phase. When several 1-bit quantized outputs φ_Q are averaged together, the random initial phase is removed, leaving only φ_V (which is proportional to the supply voltage V_{dd}). Similarly, the average of the product of the outputs from two such circuits triggered τ apart in time will be proportional to $R(\tau)$.

As previously stated, in this circuit the supply voltage must remain roughly constant during the conversion window. This is because the VCO integrates the supply voltage during the delay between the latches (T_{win}), acting as a running average filter of the supply voltage. More formally, the bandwidth of the measurement circuit is inversely proportional to T_{win}, and can be approximated by

$$f_{-3dB} \approx 0.44/T_{win} \tag{3.9}$$

Clearly, decreasing the delay yields higher bandwidth, but since the resolution of the converter is proportional to $1/(K_{vco}T_{win})$, this comes at the cost of lower measurement sensitivity. Therefore, the larger the required measurement bandwidth, the larger the number of averages N needed to achieve a desired overall measurement SNR. The worst-case SNR for a sinusoidal signal of amplitude A_{sin} can be lower-bounded by

$$SNR_V = 2N \cdot A_{sin}{}^2 (K_{vco} \cdot T_{win})^2 \tag{3.10}$$

for sub-sampled voltage waveform measurement, and

$$SNR_R = \tfrac{1}{2} N \cdot A_{sin}{}^4 (K_{vco} \cdot T_{win})^4 \tag{3.11}$$

for autocorrelation measurement.

As an example, for a VCO with an effective K_{vco} of 10 GHz/V (i.e., detecting both edges of a 5 GHz/V VCO), measurement bandwidth of 10 GHz ($T_{win} \approx 44$ ps), and a sinusoidal input with amplitude $A_{sin} = 50$ mV, $N \approx 10^7$ averages per time point would be required to achieve an SNR_V of 40 dB (~6.6 effective bits of resolution), and $N \approx 8.5 \cdot 10^{10}$ for an SNR_R of 40 dB. Clearly, this extremely large number of samples is the biggest disadvantage of this type of system, but several steps can be taken to mitigate this issue.

First, since a CMOS ring oscillator will be constructed out of several (typically, 3-5) stages, the effective K_{vco} of the converter can be increased by looking for transitions on all of the oscillator's phases. This simply involves latching and checking for transitions the output from each stage (instead of only one), and increases the effective K_{vco} by the number of stages in the VCO. Since SNR_V and SNR_R are quadratically and quartically related to K_{vco}, counting all phases has a major impact on the required number of averages. Specifically, for the same system described in the previous example, but counting all phases from a 5-stage VCO, the number of samples would be reduced to $N \approx 4 \cdot 10^5$ and $N \approx 10^8$ averages per time point for 40 dB of SNR_V and SNR_R respectively.

While significantly reduced by counting all phases, the number of samples required is still very significant. However, the major practical limiter of measurement time is often the latency of the PC controller interface. Therefore, the impact of the need for a large number of samples on measurement time can be mitigated by performing on-chip averaging of many samples (for each externally initiated scan) using a simple counter and a state machine. Without further additional hardware, the time required to measure low-frequency repetitive waveforms will be limited by the cycle of the waveform itself, but this is rarely an issue for the frequencies at which an on-chip measurement provides more information than an off-chip measurement (~10's of MHz).

3.3.1 Averaging-Based System Implementation and Measurements

To verify the feasibility of an averaging-based system and to compare its performance to that of a S/H-based system, both systems were integrated in a 90 nm SOI process as part of a test chip for characterization of the parallel interface described by K. Chang *et al.* in [44]. In addition to his collaboration on developing the averaging-based system, Valentin Abramzon performed the measurements and characterization of these circuits, and described the technique and the results in [31].

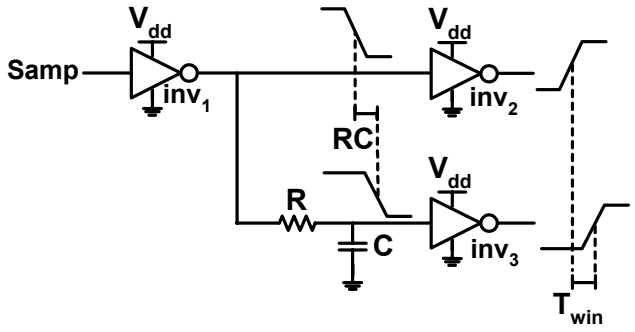


Figure 3.9: Supply-independent delay generator.

The S/H-based system used charge sharing (Figure 3.3) to operate off of the supply that was being measured, and also made use of a thick-oxide, high-threshold sampling transistor to minimize leakage. The averaging-based system used two samplers similar to Figure 3.8, but with all 5 stages of the VCO checked for transitions. The roughly supply-independent T_{win} was generated using an RC delay (Figure 3.9). The RC time constant

was set to be less than the rise time of inv_1, so that the edge rates at the inputs of inv_2 and inv_3 (and thus their delays) were nearly identical. Thus, the delay between the outputs of inv_2 and inv_3 to first order depended only on the RC product. However, this circuit works best for generating small values of T_{win}, yielding very high bandwidth but low sensitivity.

Both the S/H- and averaging-based circuits were triggered by the same pair of externally generated clocks and calibrated by varying the DC supply voltage of the chip with as much circuitry as possible turned off to minimize the DC voltage drop across the supply network. Figure 3.10 shows the measured calibration results. All of the curves are roughly linear for $V_{dd} \geq 0.9$ V, but below this the S/H-based circuits show a pronounced non-linearity – likely due to voltage range limitations in the analog buffer.

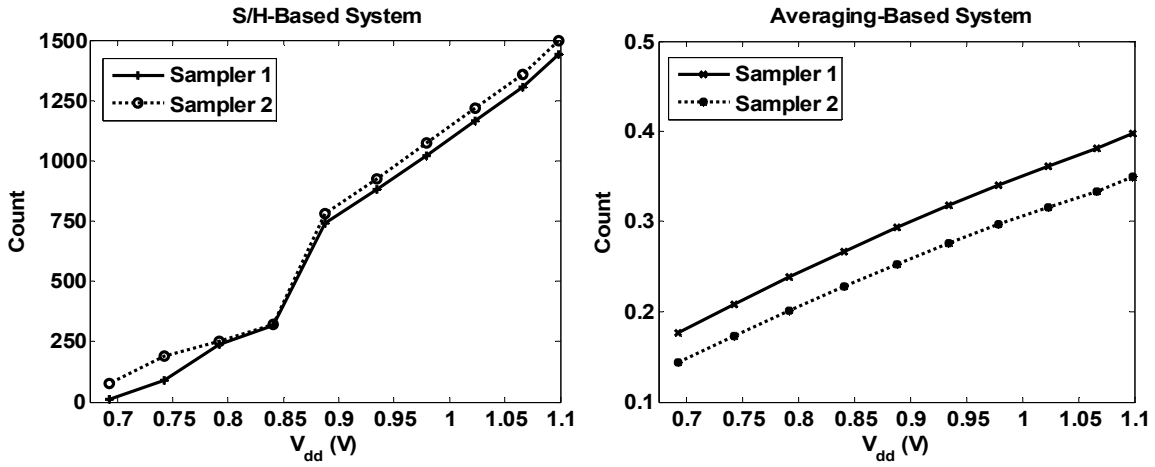


Figure 3.10: Measured calibration curves for $T_{win} = 1$ μs for the S/H-based system and $T_{win} \approx 10$ ps for the averaging-based system.

To verify the functionality of the noise measurement circuits, a 1 MHz square wave was induced on the chip's supply with an off-chip noise generator. Since 1 MHz is low enough in frequency for the chip package to be essentially transparent and for the supply grid to be excited uniformly across the entire die, the measurement circuits should record the same waveform as an off-chip probe regardless of their exact locations on the die. Once the feedthrough in the S/H-based system was calibrated (using this 1 MHz square wave) and corrected (k_f was measured to be ~0.25), the waveform measured with the averaging-based system coincided (within ~2%) with the S/H-based measurement and with the waveform measured externally to the chip – as shown in Figure 3.11.

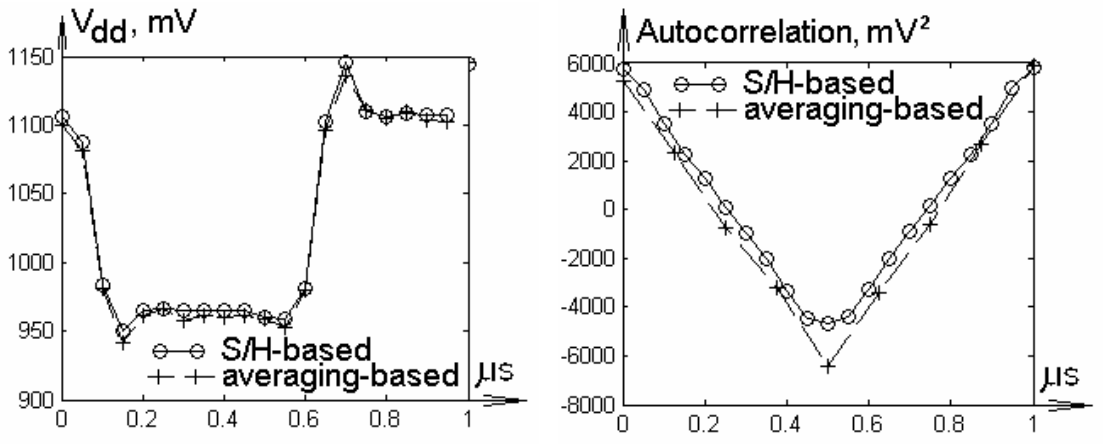


Figure 3.11: S/H-based and averaging-based measurements with S/H feedthrough cancelled in post-processing.

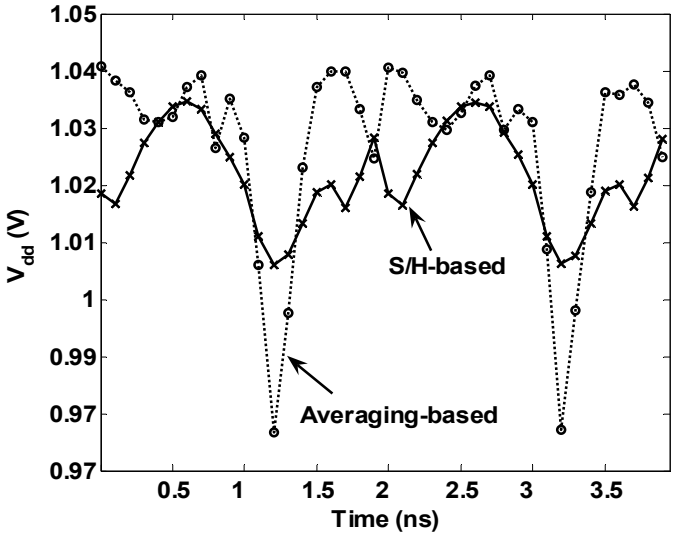


Figure 3.12: 500 MHz measured noise waveforms.

After verifying the operation of the measurement circuits with a known external waveform, we used them to measure the actual supply noise induced by the on-die circuits (Figure 3.12). On this chip, a periodic noise synchronous to the 500 MHz system clock appeared on the power supply, allowing us to simply sub-sample this noise waveform. As opposed to the 1 MHz noise, the clock-related noise contained frequency components high enough to be distributed non-uniformly across the supply network. Specifically, the averaging-based circuit was located right next to strong noise sources and measured much higher noise amplitude than the S/H-based circuit located further

away and next to a large (~25 nF) on-die bypass capacitor. As shown in Figure 3.12, this bypass capacitor along with the series supply rail resistance of ~10 mΩ attenuates the noise by approximately a factor of 2.

3.4 Summary

Modern CMOS processes are typically optimized to achieve very high digital switching speeds energy-efficiently, leading to low threshold voltages [45] and thin, leaky gate oxides. In addition, since they add complexity to the manufacturing process (and hence increase cost), many of these technologies do not include alternative, low-leakage transistor designs – making the design of supply noise measurement samplers that must hold voltages on a capacitor for 100's of nanoseconds (in order to achieve high resolution) much more challenging.

Fortunately, even with leaky, high-performance transistors, re-using the analog buffer to nearly eliminate the voltage across the sampling switch can significantly reduce the level of source-drain leakage, and we presented measured results using this technique from a 90 nm dual-core Itanium microprocessor. This design made use of a separate, quiet power supply for the sampler circuitry, but in many cases such a supply is unavailable or would simply incur too much overhead. Therefore, we next presented a PMOS-based sampler that uses charge-sharing to divide the sample voltage and allow the sampler to operate off of the supply being measured. While this eliminates the need for a separate supply, coupling from the supply being measured (through the switch parasitics) to the sample node leads to filtering of the noise measurements and the need for dynamic calibration to eliminate this filtering. Even with NMOS-based samplers where the sampling switch's gate is not tied to the measured supply during hold-mode, coupling through the input stage of the buffer still leads to the need for dynamic calibration.

To avoid the complications associated with building a high-resolution S/H-based ADC in these modern technologies, we next described how to reconstruct high-resolution measurements of both autocorrelation and periodic noise waveforms by averaging many dithered, low-resolution samples from a converter that does not require a S/H. The VCO-

based implementation presented here confirms the feasibility of this approach and is significantly simpler and more robust than the S/H-based systems, but requires many averages due to its steep tradeoffs between bandwidth and per-sample SNR. Despite this drawback, the averaging-based approach is a promising avenue for achieving very robust and scalable noise measurement systems.[12]

[12] An implementation with controllable dither would require significantly less averaging than the VCO-based approach, and would also enable time-dependent distribution measurements. Hence, in Chapter 6 we will briefly describe potential future measurement systems building upon the averaging-based approach.

Chapter 4

Feedback Amplifier Design for Linear Regulators

Having described in the previous two chapters circuits and techniques to enable efficient measurement of the characteristics of on-die supply noise, in this chapter and the next we describe the design of integrated regulation circuitry to actively counter supply noise. Because modern high-performance CMOS chips are severely power or thermally-constrained and further technology scaling will likely only accelerate this trend [2], the issue of power efficiency (i.e., tradeoffs between a regulator's performance and its power consumption, and the impact of regulation on the power consumption of the entire chip) will take a central role throughout the discussion.

It is because of the need for power efficiency that in this thesis we will focus on integrated linear regulators (vs. switching converters) to improve the integrity of the supply network. As shown by Lee *et al.* in [46], the efficiency of fully integrated switching converters is severely limited by the resistance of the on-chip inductor. Even if the inductor's resistance is reduced (e.g., with magnetic materials [86] or package inductors [90]), the high frequency output impedance of a switching converter is limited by its switching frequency, and high conversion efficiency leads to a relatively low switching frequency compared to the bandwidth that would be required to counter broad-band load current noise.

This first chapter on regulation focuses on tradeoffs that arise in the control design of CMOS linear regulators when the power consumption of the feedback amplifier is limited. Specifically, we first describe how traditional compensation schemes that reduce the feedback amplifier's bandwidth compromise the regulator's noise performance. Since an unstable regulator would exhibit poor noise performance, to elucidate design strategies that avoid this compromise, we next analyze regulator performance as a function of the amplifier's gain and bandwidth – leading to design equations that maximize rejection for a given gain-bandwidth (GBW). This analysis highlights the nature of the tradeoff between amplifier gain-bandwidth (and hence power dissipation) and noise rejection, and shows that this tradeoff stems from the first-order high-frequency roll-off of the supply network's impedance. We then examine the implications of these tradeoffs on the regulator's topology – specifically, the advantages of a source-follower output stage over a common-source – and finally show how knowledge of the load can be used to significantly improve the nature of the GBW vs. rejection tradeoff.

4.1 Compensation and Noise Performance

Since they are more straightforward to analyze, in this section we will use a typical series regulator (Figure 4.1) to highlight the degradation in noise performance caused by classic compensation schemes. Although they will not be explicitly described here, shunt regulators exhibit very similar issues with regard to compensation and noise performance.

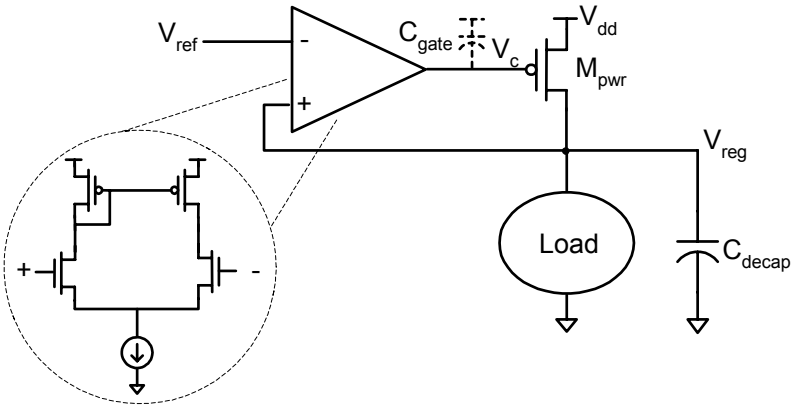


Figure 4.1: Typical low-dropout series regulator implementation.

The regulator's common-source PMOS output device M_{pwr} uses a low overdrive to maintain a low dropout voltage, leading to a large output device. The gate capacitance of this large transistor loads the output of the amplifier, creating a pole that limits the amplifier's bandwidth. This pole is in addition to the pole formed at V_{reg} by the large decoupling capacitance used to suppress the impact of load current variations and supply noise.[13] Therefore, the regulator will contain two (typically closely spaced) poles in its feedback loop, requiring compensation to achieve stability [47,48].

To minimize the size of the additional capacitance necessary to achieve stability, analog designers often compensate feedback loops at the highest impedance node. In a linear regulator, the output of the amplifier is often the highest impedance node, and designs such as [49] perform compensation by increasing the capacitive loading on the amplifier. Unfortunately, compensating the regulator by reducing the natural bandwidth of the amplifier has a very negative impact on the regulator's performance. To clarify this statement, we will qualitatively consider two regulators whose transfer functions from V_{ref} to V_{reg} are identical; one whose dominant pole is at V_{reg}, and one whose dominant pole is at V_c. In both cases, we will assume that the non-dominant pole is at a frequency much higher than the closed-loop regulator bandwidth such that the feedback loop is stable.

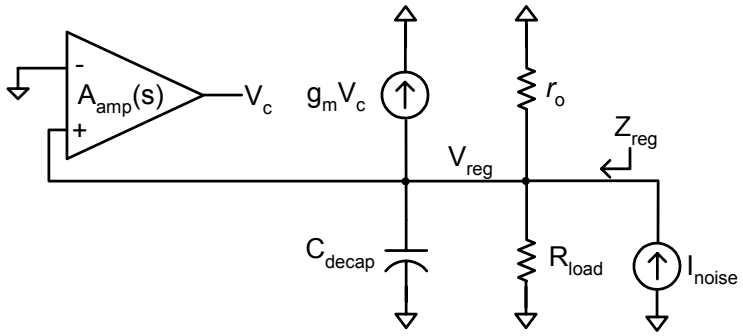


Figure 4.2: Simplified small signal model for a series regulator's output impedance – i.e., its voltage response to load current variations (I_{noise}).

[13] Unlike external decoupling capacitors, with proper layout on-chip decoupling capacitors can achieve very low equivalent series resistance (ESR). As will be described further in the next section, low ESR is highly desirable to minimize high frequency noise on V_{reg}, and since this thesis focuses on integrated regulators, we will assume negligible ESR in our analysis.

To calculate the response to variations in load current of these two regulators, we will use the simplified small signal model shown in Figure 4.2; g_m and r_o model the transconductance and output resistance of M_{pwr}, R_{load} is the load's linearized output resistance, and Z_{reg} is the effective impedance at the regulated supply V_{reg}. To simplify the description, this model assumes that there is no coupling from V_{reg} to V_c (through the C_{gd} of M_{pwr}) and hence that $A_{amp}(s)$ is simply a first-order low-pass filter with a DC gain of A_a, but this simplification does not alter the insights gained from the example.

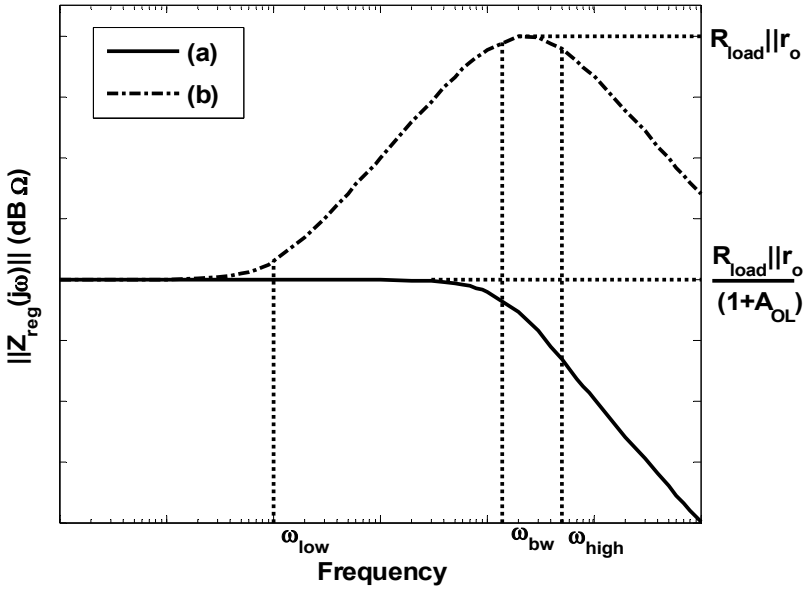


Figure 4.3: Example output impedance for a regulator with a) $\omega_o = \omega_{high}$, $\omega_a = \omega_{low}$, and b) $\omega_o = \omega_{low}$, $\omega_a = \omega_{high}$. $A_{OL} = g_m \cdot R_{load}||r_o \cdot A_a$ is the open loop gain of the regulator.

Using the small signal model of Figure 4.2, the output impedances for these two regulators are plotted in Figure 4.3.[14] For the regulator whose dominant pole is at V_{reg}, as the frequency of the load current variations passes $1/(R_{load}||r_o \cdot C_{decap})$, the decreasing impedance of the decoupling capacitance begins to reduce the open loop impedance at V_{reg}. However, due to this reduction in impedance, the gain of the output device – and hence the gain of the regulator feedback loop – also begins to drop; these two effects initially cancel each other such that there is no net change in Z_{reg}. Past its closed loop bandwidth (ω_{bw}) the regulator can no longer actively attenuate supply noise, and therefore

[14] As we will describe in more detail in the next section, from the standpoint of the impact of the feedback loop on a linear regulator's noise performance, output impedance and supply sensitivity are identical.

the output impedance is set entirely by the output RC network. At ω_{bw}, the impedance attenuation (relative to $R_{load}||r_o$) of C_{decap} approaches the open loop gain, and therefore past this frequency Z_{reg} falls with the first order roll-off of C_{decap}'s impedance.

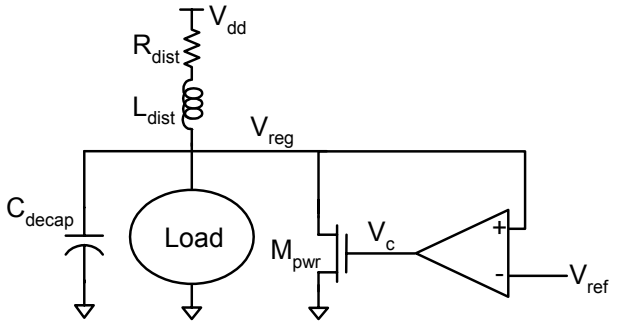
The output impedance of the regulator for which the amplifier is the dominant pole behaves quite differently. As soon as the current noise frequency surpasses the amplifier bandwidth, the amplifier gain begins to drop and the output impedance increases. Past the closed loop bandwidth of the regulator, the impedance flattens out at $R_{load}||r_o$ – in other words, the peak impedance is worse than it was with a high bandwidth amplifier by roughly the open loop gain. Only once the bandwidth of the output RC network has been passed will the impedance begin to drop from this peak value.

Clearly, a more careful stabilization approach than simply reducing the amplifier bandwidth must be taken to achieve good dynamic noise rejection. To develop such an approach, we first note that if a regulator is optimized for high supply noise or load current variation rejection, such a regulator will have to be stable – after all, an unstable (or marginally stable) regulator could actually amplify such variations. Therefore, in the next section we analyze regulator performance as a function of the amplifier's design parameters, leading to design equations for the amplifier that intrinsically achieve stability by maximizing the regulator's noise rejection at a given amplifier GBW.

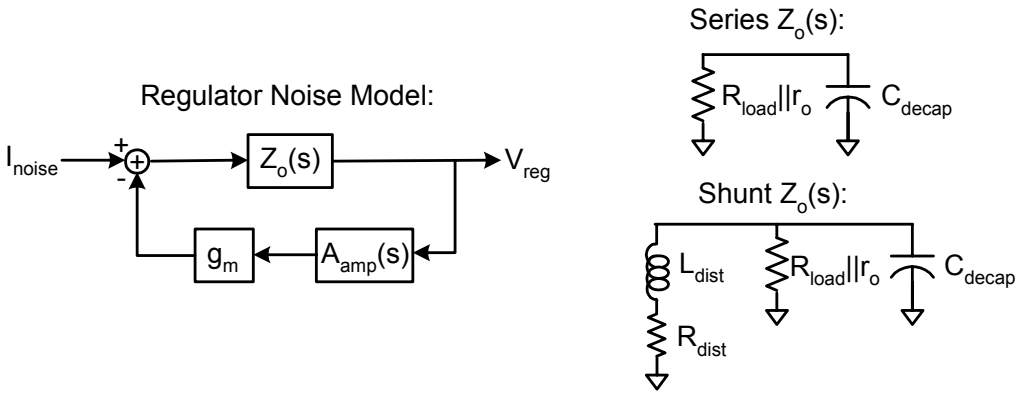
4.2 Regulator Feedback Amplifier Analysis

To analyze the noise performance of both series and shunt regulators (whose typical implementation is shown in Figure 4.4a), we will use the small-signal model shown in Figure 4.4b. In this model, $Z_o(s)$ represents the intrinsic impedance of the regulated supply node (i.e., the impedance without any regulator feedback). We have drawn the regulator's model this way because it directly leads to the regulated impedance $Z_{reg}(s) = V_{reg}(s)/I_{noise}(s)$. In addition, since supply sensitivity can be found by setting $I_{noise}(s) = V_{dd}(s)/r_o$, the model shows that the impact of the regulator's feedback on the PSRR of a

series regulator is identical to its impact on impedance.[15] While this model continues to assume that undesired coupling of the power device's gate V_c to V_{reg} or V_{dd} is negligible, including this coupling would not change the nature of the tradeoffs shown by the model.



(a)



(b)

Figure 4.4: a) Typical implementation of a shunt regulator, highlighting the most relevant elements of the supply network impedance. b) Generalized small-signal model for noise at regulator output – $Z_o(s)$ is the open loop impedance of the regulated supply.

Using a first-order feedback amplifier with a DC gain of A_a and bandwidth of ω_a, the regulator's impedance is:

$$Z_{reg}(s) = \frac{Z_o(s)\left(1+s/\omega_a\right)}{\left(1+s/\omega_a\right)+g_m A_a Z_o(s)}. \tag{4.1}$$

[15] $I_{noise}(s) = V_{dd}(s)/r_o$ ignores any direct coupling from V_{dd} to V_{reg} (e.g., through the body-to-drain capacitance of M_{pwr}), but this typically only causes the regulator's PSRR to approach a constant value at very high frequency, and has no impact on the behavior of the feedback loop.

As previously discussed, the amplifier's limited bandwidth and the first-order roll-off of the decoupling capacitor's impedance make regulators intrinsically second-order (or higher) feedback systems. The amplifier's limited bandwidth also causes a zero in Z_{reg}, which is the reason that compensating the regulator by simply reducing the amplifier's bandwidth sacrifices dynamic noise performance.

In fact, because of the presence of this zero at ω_a and that the device's g_m is effectively multiplied by A_a, Equation (4.1) makes it clear that good dynamic noise rejection requires the GBW of the amplifier to be maximized. However, since amplifier gain can always be traded for bandwidth (e.g., through local negative feedback[16]), what is not as immediately apparent is how to optimally allocate the available GBW. Although both excessive bandwidth (with low gain) and excessive low-frequency gain (with all of the gain disappearing before the impedance of C_{decap} drops below R_{load}) are clearly wasteful, a strategy for balancing the two parameters is not initially as obvious. Therefore, we next develop an analysis that leads to the allocation between amplifier gain and bandwidth that maximizes the regulator's performance with two different types of noise excitation: worst-case sinusoidal, and random, white noise.

Since it is a lower-order system, we will only describe in detail here the analysis of a series regulator. However, particularly for a worst-case sinusoidal excitation, the isolating inductance L_{dist} (mostly due to the package and external distribution network) makes the results for shunt regulators essentially identical.[17]

For a series regulator, $Z_o(s) = R_{load}||r_o/(1+s/\omega_o)$, where the output pole ω_o is $1/(R_{load}||r_o \cdot C_{decap})$, and we define the output device's gain as $A_o = g_m \cdot (R_{load}||r_o)$. To simplify the expressions we will normalize the amplifier's bandwidth so that $\omega_a = \kappa \cdot \omega_o$. We can then define the amplifier's normalized gain-bandwidth as $A_{GBW} = A_a \cdot \kappa$. Finally,

[16] Compensating the amplifier with an RC circuit as described by Lee in [50] can achieve a similar effect in terms of trading between amplifier gain and bandwidth, but the capacitance required to achieve this is often too large to be practical.

[17] For worst-case sinusoidal noise current, with typical supply network parameters the difference in optimized impedance between a series and a shunt (using Equation (4.3) for both regulators) is less than 10% across a broad range of amplifier GBW. For random, white noise, the shunt regulator's allocation would result in higher bandwidth. However, the impedance vs. allocation curve is very shallow near the optimum; when both regulators use Equation (4.5), the difference in impedances is typically less than 5%.

for sinusoidal noise, we define the regulator's minimum feedback-contributed load rejection as $LR_{min} = (R_{load}||r_o)/||Z_{reg,max}||$ (such that the R_{load}-referred rejection ratio is $(1+R_{load}/r_o)\cdot LR_{min}$).

Using these definitions and Equation (4.1), it can be shown that for a regulator that is not overdamped, LR_{min} is

$$LR_{min} = \sqrt{1-\kappa\left[\kappa+2A_oA_a-2\sqrt{A_oA_a\left(2\kappa+2+A_oA_a\right)}\right]} \tag{4.2}$$

To find the allocation between gain and bandwidth that maximizes LR_{min}, we make use of the fact that $A_a = A_{GBW}/\kappa$ and solve for κ the equation given by setting $d(LR_{min})/d\kappa = 0$. While this procedure is conceptually simple, the results are significantly simplified when the regulator has high open loop gain (i.e., $A_aA_o >> 1$). Under this approximation, the load rejection is maximized when

$$\kappa = \sqrt{\tfrac{3}{2}A_{GBW}A_o} \qquad A_a = \sqrt{\tfrac{2}{3}A_{GBW}/A_o} \tag{4.3}$$

While the exact allocation of Equation (4.3) is difficult to arrive at through means other than this mathematical analysis, the fact that both gain and bandwidth scale with the square-root of GBW has a simple, intuitive explanation that provides additional insight into this result. Specifically, it has been previously shown that to minimize a supply network's or regulator's peak-to-peak response to load current steps, the impedance should exhibit no peaking [51,52,53] – this concept is also known as voltage positioning.

As shown in Figure 4.5 – which uses the allocation of Equation (4.3) – because of the first-order roll-off of $Z_o(s)$, maintaining this flat behavior while reducing the regulator's impedance requires both the gain and the bandwidth of the amplifier to be increased – explaining the form of Equation (4.3). The need to increase both gain and bandwidth to reduce impedance also shows that the gain-bandwidth of the amplifier must scale with desired noise rejection squared; in the specific case of sinusoidal noise:

$$GBW \approx \frac{2 \cdot LR_{min}^{\ 2}}{A_o} \cdot \omega_o \tag{4.4}$$

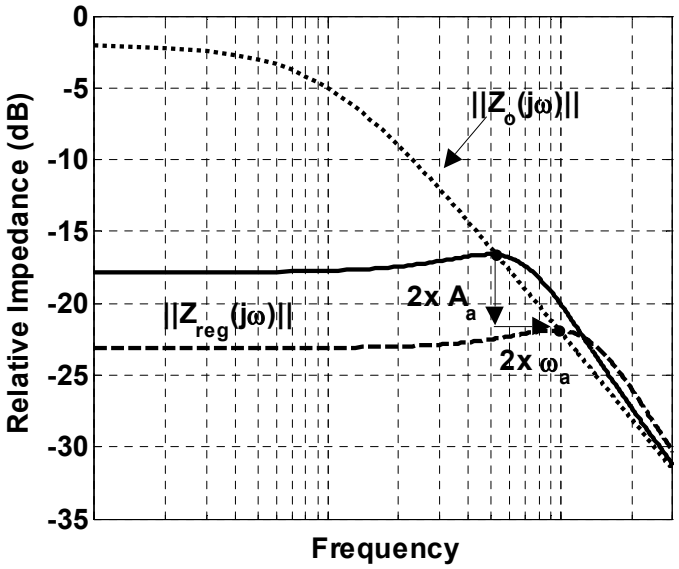


Figure 4.5: Magnitude plot of $Z_{reg}(s)$ showing that both amplifier gain and bandwidth must be increased by 2x to decrease impedance by 2x.

Several other insights can be gained by examining the results of the gain-bandwidth allocation. The first point to notice is that because of the loop gain A_o contributed through M_{pwr}, most of A_{GBW} will be allocated to bandwidth. For example, with an A_{GBW} of 10 (GBW of ~1 GHz), an A_o of 6 will make $A_a \approx 1$. Since straightforward amplifier implementations usually have high gain but low bandwidth, this makes applying local negative feedback to the amplifier to reduce its gain and increase its bandwidth a particularly attractive solution.

In addition, as A_{GBW} is increased, the allocation given by Equation (4.3) will cause the regulator's damping ζ to approach a constant value of ~0.61 – leaving a small amount of peaking in the regulator's response (as seen in Figure 4.5). This slight discrepancy from traditional voltage positioning results from the lack of capacitor ESR, and from the fact that the response to sinusoidal current (instead of square current steps) was optimized. However, this does not mean that the ESR of C_{decap} should be intentionally

increased, since this would significantly degrade high-frequency noise sensitivity while minimally reducing the peak-to-peak ripple from a current step.

While load currents are unlikely to display truly sinusoidal behavior, maximizing LR_{min} nearly minimizes the sensitivity of the regulator to a variety of deterministic or repetitive current variations. Interestingly, optimizing the response of the regulator to random, white load current noise (i.e., maximizing $\sigma_{Inoise}/\sigma_{Vreg}$)[18] leads to a very similar allocation of:

$$\kappa = \sqrt{A_{GBW}A_o} \qquad A_a = \sqrt{A_{GBW}/A_o} \tag{4.5}$$

Intuitively, the slight reduction in bandwidth is easily explained by the fact that for white, random load current, the noise energy is not concentrated at a single frequency, and hence additional peaking is actually desirable to minimize the total voltage noise energy. Correspondingly, as A_{GBW} is increased the regulator's ζ approaches a value of ~0.5.

4.3 Implications on Regulator Topology

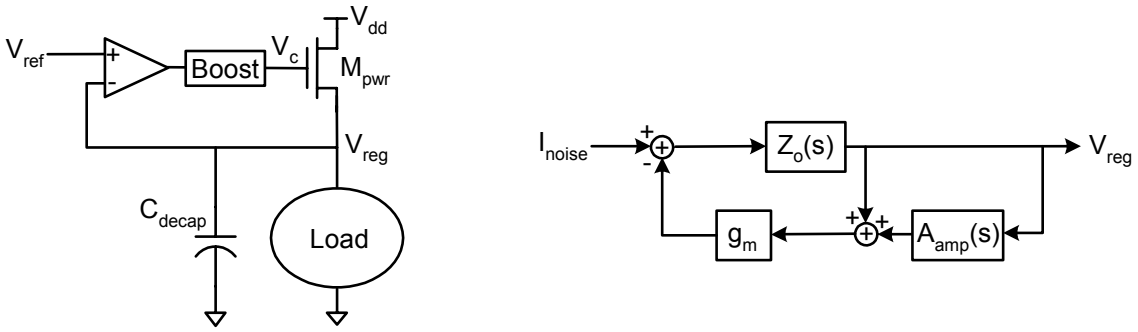
These tradeoffs between amplifier gain-bandwidth and desired noise rejection stem from the 1st order nature of $Z_o(s)$ at high frequencies, and therefore from a load current noise rejection standpoint this tradeoff exists in essentially all regulator designs. However, as we will explain in the next three subsections, the results of the feedback amplifier analysis have direct implications on the topology of the regulator. Specifically, the regulator can make use of a source-follower output device to significantly improve the effective GBW of the feedback. In addition, in some applications, the load is very well known and the effect of supply noise on the regulated output can be sensed separately

[18] This analysis is carried out in a similar manner to that for the worst-case sinusoid, where $\sigma^2_{Inoise}/\sigma^2_{Vreg}$ is computed from $\int_0^\infty \left\| Z_{reg}(f) \right\|^2 df$. Useful formulas for this type of integration of up to 7th order linear systems can be found in [54].

from the output itself, and this fact can be used to change the form of the GBW vs. power supply noise rejection tradeoff from quadratic to linear. The use of such replica feedback is then extended to applications where the regulator must also achieve high-bandwidth tracking of the reference input, and experimental results highlighting the improvements from this design are presented.

4.3.1 Source-Follower Output Stage

The analysis of the previous section showed that for optimal rejection characteristics the feedback amplifier's gain will be modest – often 1 or less. In these cases, the amplifier is essentially acting only as a power-consuming wire routing the error signal onto M_{pwr}'s gate, while using a source follower (SF) power device (Figure 4.6) applies the feedback signal to M_{pwr} through a high-bandwidth wire that has no power dissipation.



(a) (b)

Figure 4.6: a) Series regulator with source follower (SF) power device. b) Noise model including intrinsic negative feedback of the SF.

While for series regulators the use of an SF requires additional complexity for gate boosting circuitry (or alternatively, a separate power supply for the amplifier) to maintain low dropout (as shown in Figure 4.6a), the intrinsic feedback of the SF increases the rejection over a common source (CS) design by A_o ($LR_{min,SF} \approx LR_{min,CS}+A_o$). Hence, if the regulator is targeting an LR_{min} less than A_o, the power-consuming amplifier can be removed from the feedback path entirely.

To truly achieve an LR_{min} of A_o when the amplifier has been removed requires the gate of the SF to be very well isolated from its source – but in this case this can be

achieved by simply adding capacitance to V_c. This was demonstrated by den Besten in [55], where V_c was set through a replica-biased charge-pump. As shown by Hazucha in [53], the intrinsic feedback of the SF gave the design from [55] significantly better figure of merit than all of the other regulators included in the comparison.

An SF-based regulator can have advantages over a CS-based design even when the regulator is targeting an LR_{min} greater than A_o (and hence a feedback amplifier is required). Without adding any intentional capacitance to V_c (to maintain high amplifier bandwidth), the ratio between the output device's C_{gs} and the other parasitics at V_c ensures that independent of the action of the amplifier, the gate overdrive of M_{pwr} will be modulated by 40-50% of the noise on V_{reg}. With the quadratic tradeoff between rejection and GBW, for $LR_{min}<\sim 3 \cdot A_o$ even this partial feedback through M_{pwr}'s g_m cuts the SF regulator's required A_{GBW} by over 40%.

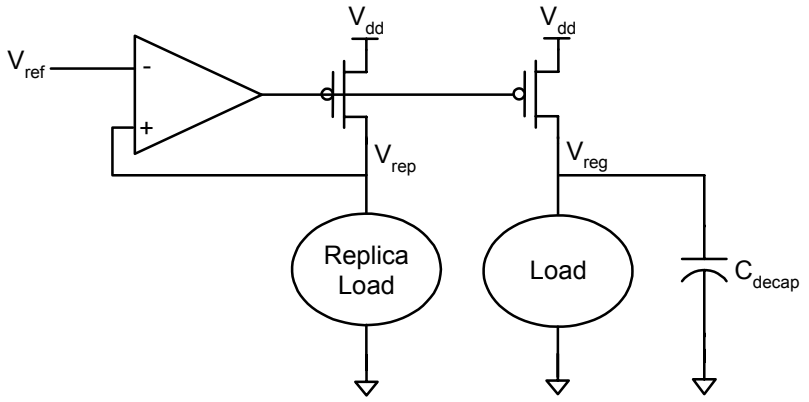
4.3.2 Replica Feedback

While the primary role of many regulators is to provide a low output impedance (while perhaps translating voltage levels) to minimize supply voltage variations, in some situations most of the noise is not caused by current draw variations in the load circuits themselves, but rather by other circuits on the die (e.g. an analog/mixed-signal component integrated as part of a larger digital chip [27]). In these applications, the chief task of the regulator is to reject the externally generated noise on the power supply.

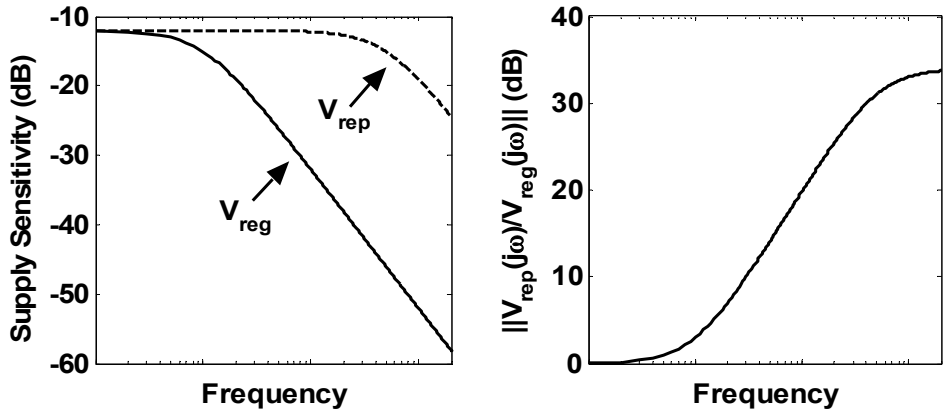
For a regulator whose primary task is to reduce the supply impedance seen by the load, the error signal created by the noise is only available at the regulator's output. However, the effect of externally generated noise on the regulator's output can be sensed separately – especially when the behavior of the load with respect to supply variations is well-understood and can easily be mimicked (i.e. a replica of the load can easily be built).

We can make use of this fact to drastically improve upon the feedback amplifier gain-bandwidth vs. rejection tradeoff. Specifically, consider the replica-biased regulator [55] shown in Figure 4.7a, where the replica supply node is intentionally left without any decoupling capacitance. By sensing the impact of V_{dd} noise on the replica supply (V_{rep}),

as long as the replica load's static supply sensitivity matches that of the real load, the error signal applied to the amplifier will essentially be an unfiltered version of the noise at the output. Therefore, as shown in Figure 4.7b, as the supply noise frequency surpasses the cutoff of the output RC filter, the magnitude of the error signal (relative to the noise on V_{reg}) will automatically increase.[19]



(a)



(b)

Figure 4.7: a) Replica-biased regulator. b) Open loop supply sensitivities (i.e., $||V_x(j\omega)/V_{dd}(j\omega)||$) of V_{rep} and V_{reg}, highlighting the increase in coupling to V_{rep} at high frequencies.

From the standpoint of the regulator's noise performance, this increase in the magnitude of the error signal is indistinguishable from increased amplifier gain.

[19] The parasitic capacitors at V_{rep} will also eventually filter the coupling from V_{dd} and limit the increase in the magnitude of the error signal. Therefore, as described in Appendix A, the replica load should be designed to minimize the effective capacitance it places on V_{rep}.

Therefore, as shown in Figure 4.8, even increasing only the bandwidth of the amplifier causes the effective gain to increase by the same amount. This means that in order to increase the supply rejection of the regulator, only the amplifier's bandwidth must be increased (as opposed to both its gain and its bandwidth). In other words, the required amplifier GBW increases only linearly with worst-case power supply rejection:

$$\mathrm{GBW} \approx \frac{\mathrm{PSR}_{\min}}{\mathrm{A}_{\mathrm{o}}} \cdot \omega_{\mathrm{o}}, \tag{4.6}$$

where $\mathrm{PSR}_{\min}$ has been defined as the minimum feedback contributed power supply rejection (such that the total worst-case PSRR is $(1+r_{\mathrm{o}}/\mathrm{R}_{\mathrm{load}})\cdot\mathrm{PSR}_{\min}$).[20]

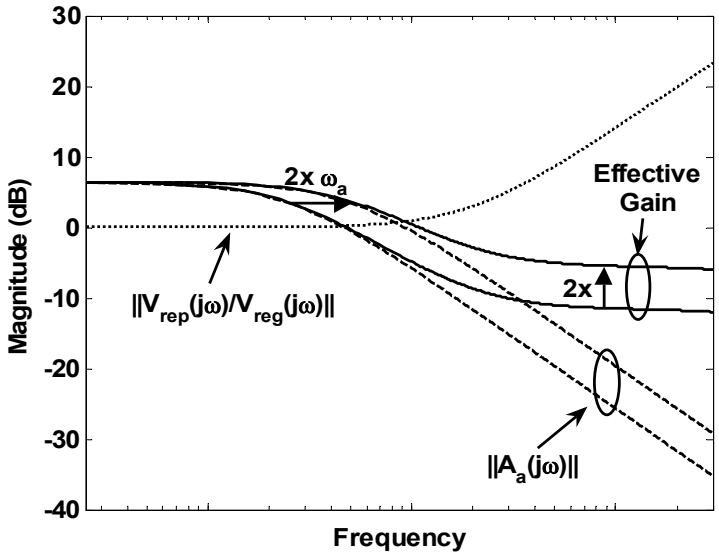


Figure 4.8: Plot of effective amplifier gain in a replica-biased regulator, showing that because of the increasing magnitude of the error signal on $\mathrm{V}_{\mathrm{rep}}$ relative to $\mathrm{V}_{\mathrm{reg}}$, increasing the amplifier's bandwidth by 2x increases the effective gain by 2x.

The supply rejection advantages of a replica-biased regulator make this topology extremely well-suited to applications where a constant, quiet supply voltage is desired (such as generating the supply voltage for an LC oscillator or CML gates [56]) –

[20] Equation (4.6) is derived assuming that the amplifier's bandwidth ω_a is less than or equal to the bandwidth of the output network ω_o. If $\omega_a > \omega_o$, the worst-case supply rejection will actually be at low frequency, because the effective feedback gain will start to increase at ω_o. However, in this situation the gain of the amplifier can simply be increased at the expense of its bandwidth, making Equation (4.6) valid.

especially since as long as the shapes of the replica and actual I-V curves match, the replica load current can be a fraction of the actual load current. However, in a replica-biased design, the lack of feedback from the true output means that the tracking bandwidth of the regulator will be limited by the bandwidth of the output RC network (ω_o). Hence, a replica-biased design may not be suitable in applications where the regulated supply must track the reference input at a reasonably high rate.

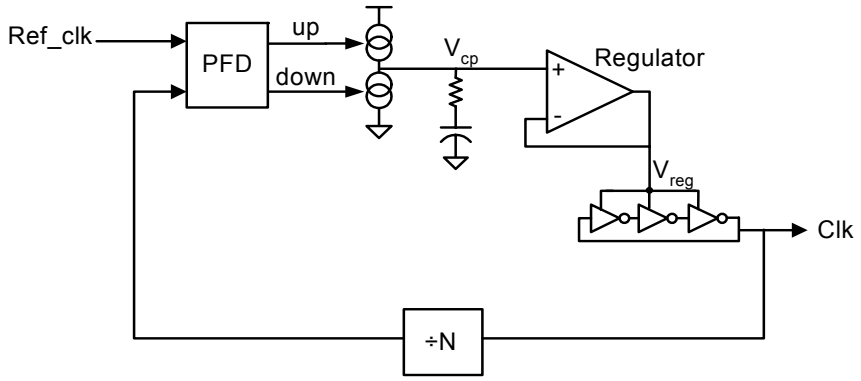


Figure 4.9: Supply regulated PLL block diagram.

One such application that demands high tracking bandwidth from the regulator is the supply-regulated phase-locked loop (PLL) proposed by von Kaenel in [27] and extended for adaptive bandwidth by Sidiropoulos in [57] (Figure 4.9).[21] The supply-regulated topology has found widespread use in modern technologies [30,49,57,60,61] because of its relaxed headroom requirements and simple VCO buffer design, but relies heavily upon regulator supply rejection to achieve low jitter. The fact that the regulator is in the forward path of the PLL precludes the use of a replica-biased topology, but the supply rejection advantages of replica feedback would significantly improve the jitter performance of the PLL. Therefore, in the next section we describe the regulator topology we proposed in [62] that makes use of a replica as part of a local feedback loop around the amplifier. With the appropriate choice of gain for this local negative

[21] It is interesting to note that the self-biased topology proposed by Maneatis in [58] and then improved in [59] essentially makes use of a replica-biased regulator to generate the (shared) tail voltage for the differential oscillator. It is exactly due to the tracking bandwidth requirement that additional decoupling capacitance is not added from this tail node (V_{tail}) to the supply (V_{dd}), compromising the dynamic supply rejection of these designs.

feedback, the regulator can meet the tracking bandwidth requirement while utilizing the replica feedback for improved supply rejection.

4.3.3 Replica Compensated Regulator

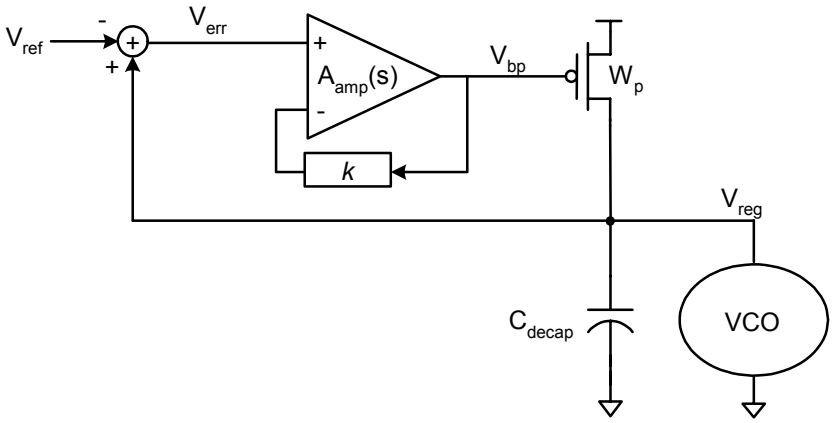


Figure 4.10: Application of a local negative feedback loop to the regulator's amplifier.

In order to arrive at the proposed structure, we will start with the application of a local negative feedback loop on the regulator's amplifier, as shown in Figure 4.10. As previously described, this additional negative feedback allows the amplifier's forward gain to be traded for bandwidth, and changes the regulator's transfer function to

$$\frac{V_{reg}(s)}{V_{in}(s)} = \frac{A_{amp\_eff}(s)A_o(s)}{1 + A_{amp\_eff}(s)A_o(s)},\ A_{amp\_eff}(s) = \frac{A_{amp}(s)}{1 + kA_{amp}(s)} = \frac{A_{a\_eff}}{1 + s/\omega_{a\_eff}}. \tag{4.7}$$

Since in a supply-regulated PLL the structure of the load is easy to mimic (as shown in Appendix A), we can make use of replica feedback (while still allowing the desired tracking bandwidth to be obtained with the appropriate choice of k) to improve the regulator's supply rejection by modifying the local negative feedback such that it is taken through a replica load, as conceptually illustrated in Figure 4.11.

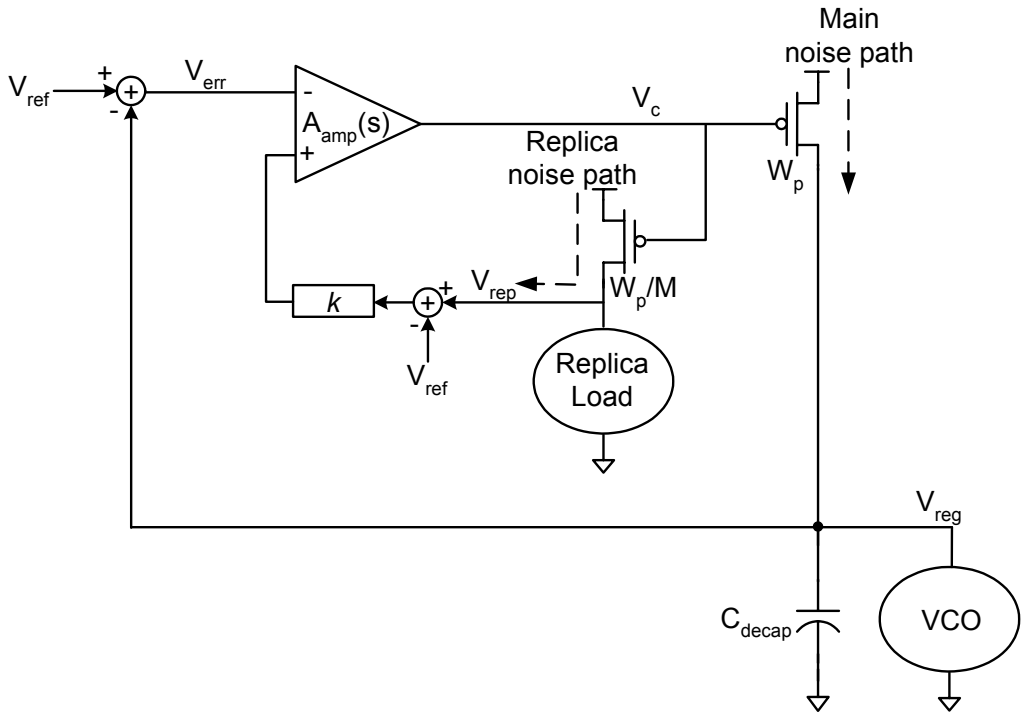
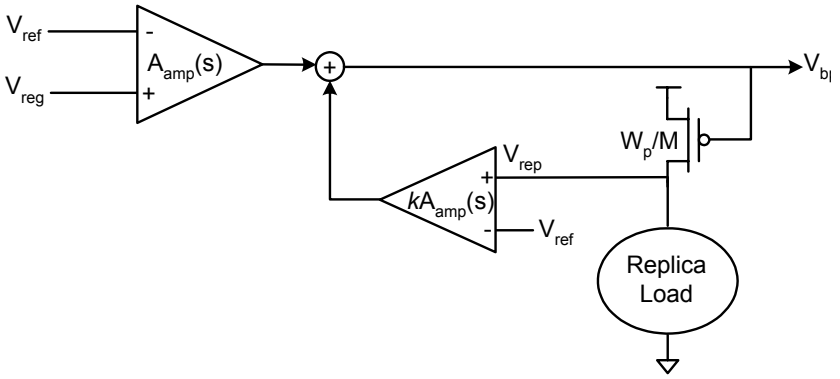


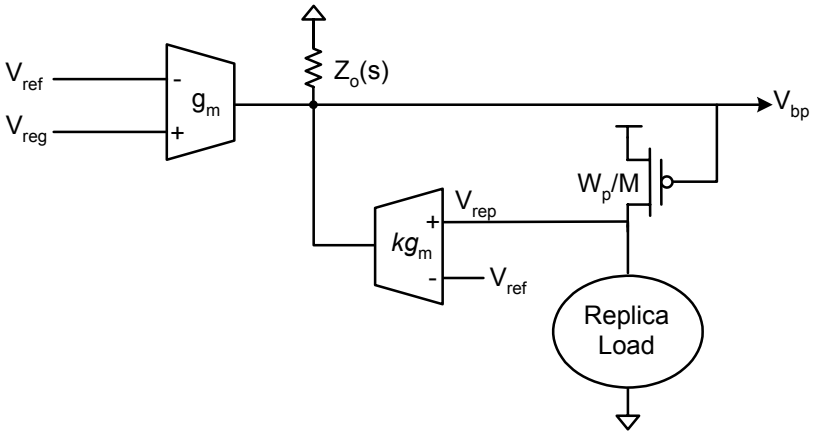
Figure 4.11: Addition of a replica to the amplifier's local negative feedback to improve the regulator's supply rejection. Note that to maintain the correct DC voltage at V_{reg}, the feedback from the replica supply node is also referenced to V_{ref}.

In order to simplify the implementation of the regulator, we can restructure the amplifier and its local feedback loop by separating the feedforward and feedback gain paths as shown in Figure 4.12a. Since the addition between the feedback from the replica and the feedback from the actual output occurs at the output of their respective amplifiers, as shown in Figure 4.12b the summation could easily be implemented in the current domain by shunting together the outputs of the transconductance stages that implement the amplifiers.

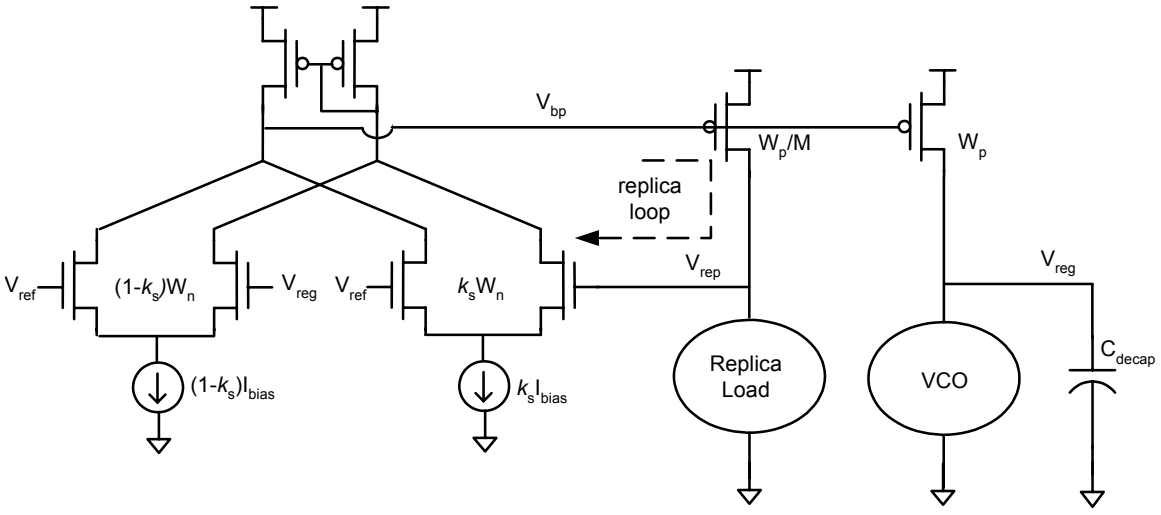
Notice that in Figure 4.12b, the total g_m of the two amplifiers is proportional to $1+k$ – i.e., the total g_m increases with k. From a practical standpoint, for a given total feedback amplifier current, the total g_m would be allocated between the two amplifiers (rather than increased with k). Therefore, we implemented the final replica compensated regulator as shown in Figure 4.12c, where two differential pairs share a single current-mirror load, and the (re-scaled) feedback gain $k_s = k/(1+k)$ is set by the current and device width allocated to each pair.



(a)



(b)



(c)

Figure 4.12: a) Separation of the amplifier feedforward and feedback paths. b) Feedforward and feedback amplifiers implemented using transconductance stages with their outputs summed in the current domain. c) Circuit implementation of the replica compensated regulator with two differential pairs sharing a single current-mirror load.

4.3.3.1 Replica Compensated Regulator Bandwidth

As with a standard regulator to which local negative feedback is applied to the amplifier, as the local feedback gain k_s is increased, the open-loop gain from V_{ref} to V_{reg} drops and the tracking bandwidth decreases. To arrive at design equations that estimate the regulator's closed loop bandwidth, we will briefly examine the reference to output transfer function of the regulator, and its behavior for different values of k_s.

Defining $A_{o\_rep}(s)$ as the transfer function of the replica output stage, the closed loop transfer function of the regulator is

$$\frac{V_{reg}(s)}{V_{ref}(s)} = \frac{A_{amp}(s)A_o(s)}{1 + A_{amp}(s)\left((1-k_s)A_o(s) + k_s A_{o\_rep}(s)\right)}. \tag{4.8}$$

To make the local feedback applied to the amplifier more apparent, this transfer function can be manipulated to match the form of Equation (4.7), with

$$A_{amp\_eff}(s) = \frac{A_{amp}(s)}{1 + k_s A_{amp}(s)\left(A_{o\_rep}(s) - A_o(s)\right)}. \tag{4.9}$$

Interestingly, the $A_{o\_rep}(s)–A_o(s)$ term shows that the local feedback is only applied to the amplifier at frequencies where the gain of the replica output stage is large compared to that of the actual output stage. This means that it is only at frequencies between ω_o and ω_{o_rep} (the pole due to the parasitic capacitances at V_{rep}) that the feedback takes effect and extends the amplifier's bandwidth, thus stabilizing the regulator. Therefore, ω_{o_rep} should be at a high enough frequency that the local feedback maintains the extended amplifier bandwidth until well beyond the open-loop cross-over frequency.

Figure 4.13 shows the root locus of the regulator transfer function as k_s is swept from zero to one with example values for the gains and poles. Initially, as the effective bandwidth of the amplifier is increased with k_s, the output pole becomes more dominant and the phase margin of the regulator improves. Once ω_{a_eff} has been increased enough that the overall regulator becomes overdamped, the closed loop dominant poles of the

regulator (ω_{bw} and ω_{2nd_order}) meet on the real axis and split. In this region, the regulator bandwidth is approximately set by the output pole,

$$\omega_{bw} \approx \frac{(1+A_a A_o)}{(1+k_s A_a A_o)}\omega_o \approx \frac{1}{k_s}\omega_o. \tag{4.10}$$

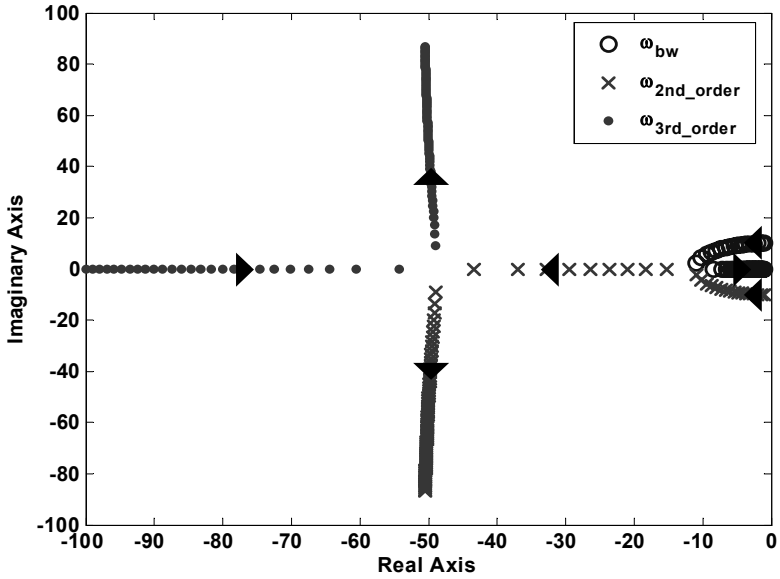


Figure 4.13: Replica compensated regulator root locus vs. k_s with $A_a = A_o = 10$, $\omega_o = \omega_a$, and $\omega_{rep} = 100\omega_o$.

Since at this point the output pole is already completely dominant, further increases in amplifier bandwidth do not affect the stability of the loop; it is at this point that as k_s is increased ω_{bw} begins to decrease because of the reduction in effective amplifier gain. Of course, once all of the gain has been allocated to the replica loop (i.e. $k_s = 1$), the regulator becomes replica-biased, with its bandwidth set by ω_o and the other two poles (ω_{2nd_order} and ω_{3rd_order}) at the closed loop poles of the replica loop. As shown by the pole locations when $k_s = 1$ in the example of Figure 4.13, the replica loop may itself be underdamped. Therefore, in order to avoid any peaking in the response of the regulator, k_s should be within the range that makes all three poles of the regulator purely real and negative – which is also the region in which Equation (4.10) can be used to easily estimate the bandwidth of the regulator.

4.3.3.2 Replica Compensated Regulator Supply Rejection

While increasing the local feedback gain of the replica compensated regulator decreases its tracking bandwidth, the additional effective gain provided by the replica feedback improves the regulator's supply rejection as k_s is increased. In order to quantify this tradeoff, in this section we will briefly examine the supply rejection properties of the regulator. The analysis will show that because of the opposing effects of k_s on bandwidth and rejection, this topology simply trades tracking bandwidth linearly for worst-case rejection, and that because of the use of the replica, the regulator maintains a linear tradeoff between required amplifier gain-bandwidth and rejection.

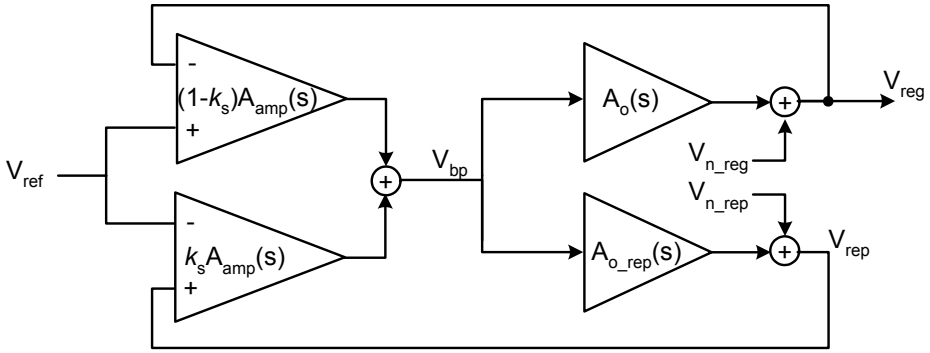


Figure 4.14: Noise transfer functions model of the replica compensated regulator.

In order to analyze the supply rejection properties of the regulator, we will derive the noise sensitivity transfer functions using the model shown in Figure 4.14. Defining the voltage noise terms as V_{n_reg} and V_{n_rep},[22] the transfer functions are

$$\frac{V_{reg}(s)}{V_{n\_reg}(s)} = \frac{1 + k_s A_{amp}(s) A_{o\_rep}(s)}{1 + A_{amp}(s)\left((1-k_s)A_o(s) + k_s A_{o\_rep}(s)\right)}, \tag{4.11}$$

and

$$\frac{V_{reg}(s)}{V_{n\_rep}(s)} = \frac{-k_s A_{amp}(s) A_o(s)}{1 + A_{amp}(s)\left((1-k_s)A_o(s) + k_s A_{o\_rep}(s)\right)}. \tag{4.12}$$

[22] We have separated the noise sources impacting V_{reg} and V_{rep} in this way in order to simplify the analysis of replica load mismatch presented in Appendix A.

For supply noise, V_{n_reg} and V_{n_rep} will simply be filtered (by the true output and replica output RC networks, respectively) versions of the noise on V_{dd}. Hence, if the static sensitivity and DC gain of the replica output stage match those of the actual load, the supply noise transfer function is

$$\frac{V_{reg}(s)}{V_{dd}(s)} = \frac{S_{Vdd}(s)}{1 + A_{reg}(s)} = \frac{S_{Vdd}(s)}{1 + A_{amp}(s)\left((1-k_s)A_o(s) + k_s A_{o\_rep}(s)\right)}, \tag{4.13}$$

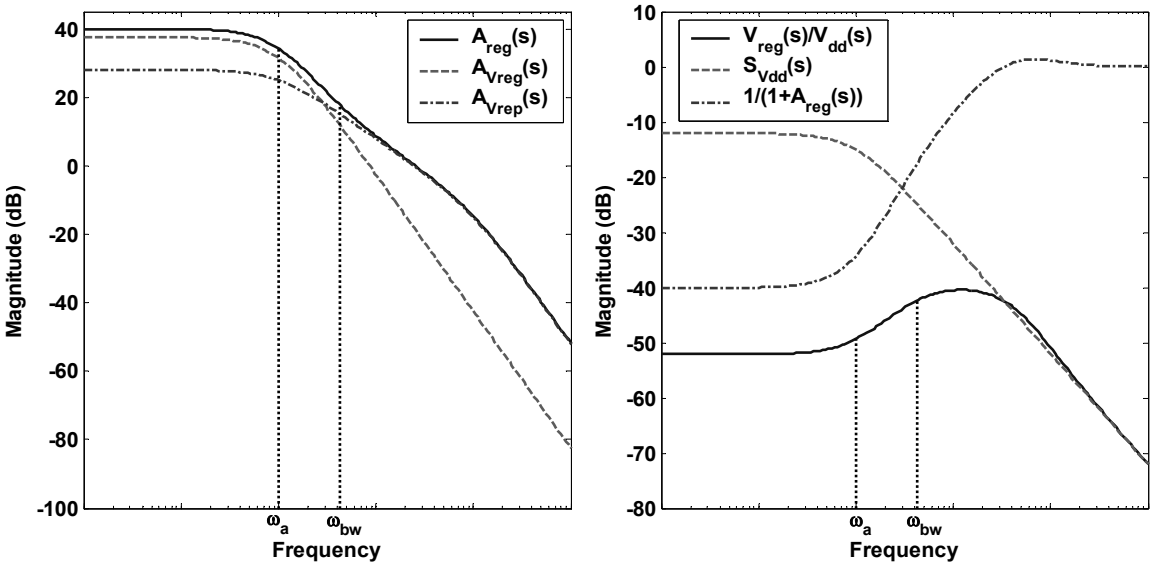
where $S_{Vdd}(s)$ is the transfer function of the RC filter from V_{dd} to V_{reg}.

As mentioned in the previous subsection, in order to avoid any peaking in its response the regulator should be designed such that all three of its poles are purely real and negative. If the regulator is designed in this manner, it is straightforward to predict its sensitivity to supply noise. Since the dominant pole of the regulator must be set at a frequency 5 to 10 times higher than the PLL bandwidth in order to maintain the PLL's stability, in analyzing supply rejection we will make the assumption that the closed-loop bandwidth of the regulator is higher than the open-loop bandwidth of the amplifier.

Since the local feedback is applied to the amplifier only at frequencies where $A_{o\_rep}(s)$ differs from $A_o(s)$, at low frequencies the rejection of the regulator is identical to that of a typical regulator with the same amplifier gain. Just as it did for the traditional regulator, once the supply noise frequency passes the open-loop bandwidth of the amplifier (ω_a), the amplifier's gain drops and the regulator will have less total gain $A_{reg}(s)$ to combat the noise, leading to a zero at ω_a (Figure 4.15).

When the frequency of the supply noise passes ω_o, the higher bandwidth of the RC filter at V_{rep} increases the effective gain of the replica path relative to the gain of the main path. Therefore, as shown in Figure 4.15a, there will be a frequency at which $A_{reg}(s)$ is dominated by the gain through the replica path – this frequency is the closed-loop bandwidth of the regulator, ω_{bw}. As shown in Figure 4.15b, once the supply noise

frequency passes ω_{bw}, the supply sensitivity of the regulator reaches its maximum value before it begins to roll off again due to the attenuation of the output RC filter.



(a) (b)

Figure 4.15: Replica compensated regulator a) open-loop gains and b) supply sensitivity components with $A_a = A_o = 10$, $\omega_o = \omega_a$, $\omega_{rep} = 100\omega_o$, $k_s = 0.25$, and $S_{Vdd} = R_{load}/(R_{load}+r_o) = 0.25$. $A_{reg}(s)$ is the total gain the regulator applies against supply noise, $A_{Vreg}(s)$ is the gain the regulator feedback applies to noise sensed on V_{reg}, and $A_{Vrep}(s)$ is the gain applied by the replica path to noise sensed on V_{rep}.

Knowing that the feedback-contributed rejection of the regulator is $(1+A_aA_o)$, and that at ω_{bw} the amplifier gain has dropped by ω_a/ω_{bw}, the minimum feedback-contributed supply rejection can be simply approximated by

$$PSR_{min} \approx \frac{\omega_a}{\omega_{bw}} \cdot (1 + A_a A_o) \approx \frac{\omega_a}{\omega_o} \cdot (1 + k_s A_a A_o) \tag{4.14}$$

The form of the tradeoff between regulator bandwidth and worst-case supply rejection is now clear – increasing bandwidth linearly decreases the minimum rejection. Equivalently, increasing k_s improves the regulator's supply rejection but linearly decreases its bandwidth. Since these mechanisms vary in the same manner with the feedback gain, the product of minimum supply rejection (PSR_{min}) and regulator bandwidth is independent of k_s, and is given by

$$\mathrm{PSR}_{\min} \cdot \omega_{bw} \approx \mathrm{GBW} \cdot \mathrm{A}_o \tag{4.15}$$

Therefore, in a supply-regulated PLL, maximizing supply rejection requires that k_s be chosen such that the regulator achieves the lowest possible bandwidth that maintains PLL stability. As can be seen from Equation (4.15), at this fixed tracking bandwidth, increasing the gain-bandwidth of the amplifier will linearly increase $\mathrm{PSR}_{\min}$ – once again highlighting the benefit of the use of a replica for the local feedback.

4.3.3.3 Replica Compensated Regulator Experimental Results

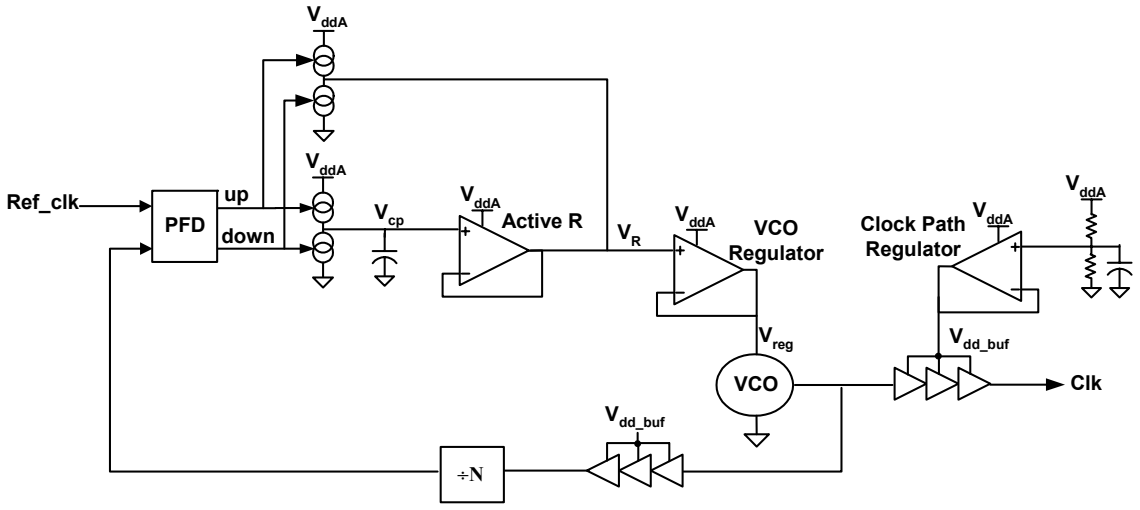
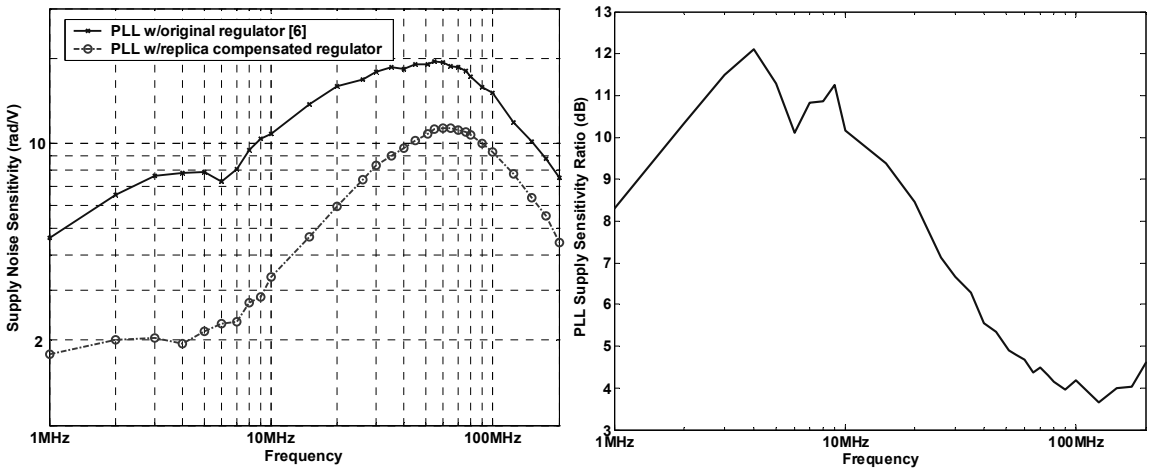


Figure 4.16: PLL and clock distribution architecture of the test chip for characterization of the parallel interface described in [44]. The stabilizing resistor for the PLL's loop filter is created by the output resistance of the unity gain buffer labeled "Active R". Nominally, $V_{ddA} = 1.5$ V, and $V_{dd\_buf} \approx .73 \cdot V_{ddA}$.

To experimentally demonstrate its efficacy, the replica compensated regulator was implemented in a 90 nm SOI process as part of one of the PLLs on a test chip for characterization of the parallel interface described by K. Chang *et al.* in [44]. Each transmit or receive byte-wide link had a multiply-by-5 PLL to generate its high-speed clock; the PLL architecture (Figure 4.16) was based upon a previous design [30].

For the purpose of comparison, two transmit byte-wide parallel links were fabricated; with all other components identical (including the decoupling capacitance for the VCO supplies), the PLL for one of the links used a VCO regulator based on [57] and

[30], and the VCO regulator in the other link was replica compensated. For the purpose of comparison, the replica compensated regulator was designed to consume roughly the same amount of power as the original regulator.



(a) (b)

Figure 4.17: a) Measured supply noise sensitivity versus frequency for the PLLs with the original regulator design and with the replica compensated regulator. b) Sensitivity of the original PLL divided by sensitivity of the PLL with the replica compensated regulator.

While the test chip did not include the capability to directly measure the regulator output voltage (and hence the regulator supply noise sensitivity), it did include supply noise generators and measurement circuits similar to those described in the previous two chapters. Therefore, we used the generators to inject sinusoidal noise onto the power supply, measured the resulting jitter by sending a clock pattern through the transmitter, and normalized this jitter by the measured supply noise magnitude to obtain the supply noise sensitivity for the two PLLs. The results of this measurement for both PLLs operating at 2.5 GHz are shown in Figure 4.17a.

The measured data clearly shows that the PLL with the replica compensated regulator has lower supply sensitivity. To better isolate the sensitivity differences due to the regulators from the filtering that the PLL itself applies to the noise (as described by Mansuri *et al.* in [63]), Figure 4.17b shows the measured sensitivity of the original PLL divided by the sensitivity of the PLL with the replica compensated regulator. Because of the additive supply noise paths through the active resistor and clock buffers (shown in Figure 4.16), the ratio shown in Figure 4.17b will not directly match the ratio of regulator

sensitivities – but it does provide a lower bound on the improvement of the replica compensated regulator over the previous design.

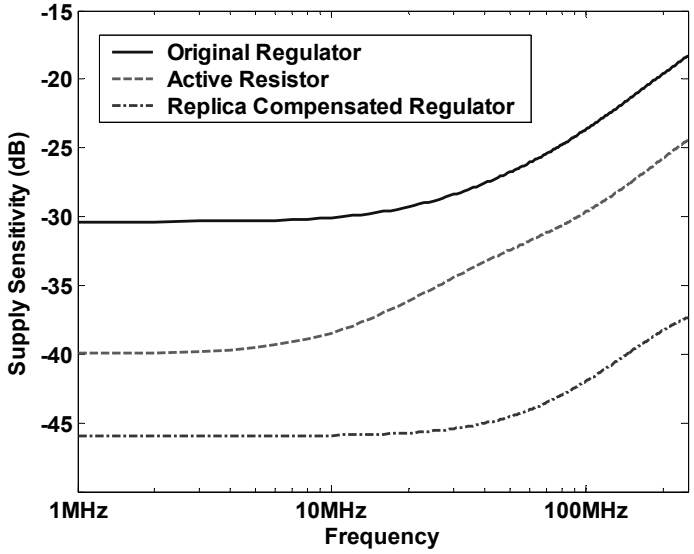


Figure 4.18: Simulated supply noise sensitivities for the original regulator, active resistor, and replica compensated regulator.

The measurement shows that the additional gain stage of the replica compensated regulator improves the low frequency noise sensitivity[23] by at least ~10 dB. In fact, the simulation results of Figure 4.18 show that even at higher frequencies, the replica compensated regulator's sensitivity has been reduced enough that the active resistor becomes the dominant contributor of noise coupling from V_{ddA} to the VCO supply.[24]

The fact that the active resistor is the dominant source of sensitivity to supply noise in the PLL with the replica compensated regulator unfortunately makes it difficult to isolate the replica compensated regulator's sensitivity; however, the sensitivity ratio does

[23] Because of the integration from the charge pump/loop filter, it is normally expected that the supply sensitivity of an overdamped PLL rises at 20dB/dec at low frequencies (i.e. below ~10% of the PLL bandwidth). The sensitivity of the original PLL displays this behavior, but the low frequency sensitivity of the PLL with the replica compensated regulator is essentially flat – leading to the low frequency slope in the sensitivity ratio. The floor in the noise sensitivity of the PLL with the replica compensated regulator is most likely due to mismatch between the clock buffers on the forward path and those in the feedback path, and hence it is unlikely that the low frequency slope in the sensitivity ratio is due to the behaviors of the two regulators.

[24] Note that the reason for the increase in sensitivity of the active resistor and original regulator that begins at 10-20MHz is most likely the reduction in effective output impedance of body-contacted SOI devices at frequencies where the resistive body contact is no longer effective [44]. Due to differences in sizing and topology, the active resistor displays this behavior at a lower frequency than the original regulator, leading to the reduction in measured sensitivity ratio that begins at ~10MHz shown in Figure 4.17b.

provide a lower bound on the relative improvement of the replica compensated design. Thus, the measured ratio shows that with roughly the same power consumption, the replica compensated regulator achieves a minimum of ~4 dB higher supply rejection than the previous design. Furthermore, the typical corner simulations from Figure 4.18 indicate that the regulator in isolation achieves an improvement of greater than 15 dB in supply noise rejection.

4.4 Summary

In this chapter, we began to explore the issue of efficient regulator design with an examination of the control design of CMOS linear regulators whose feedback amplifiers have limited power consumption, and hence limited gain-bandwidth. The first concern of many regulator designers is to guarantee the stability of the feedback loop – which is intrinsically second-order (or higher) because of the amplifier's limited bandwidth and the first-order roll-off of the supply network's high-frequency impedance. However, we showed that traditional compensation schemes that reduce the natural bandwidth of the amplifier significantly degrade the dynamic noise performance of the regulator.

To point to better design strategies, we developed an analysis of regulator performance that shows how to allocate between the feedback amplifier's gain and bandwidth in order to maximize noise rejection, naturally leading to stable designs. The analysis showed that because of the gain contributed by the output device, the amplifier should typically be designed with a high bandwidth and relatively low gain. This allocation leads to relatively flat output impedance characteristics – matching well with the concept of voltage-positioning [51].

In order to increase the regulator's noise rejection, both the amplifier's gain and its bandwidth must be increased to maintain flat output impedance, making the GBW required of the amplifier scale quadratically with the desired worst-case rejection. This tradeoff arises from the first-order reduction in the impedance of the supply network's decoupling capacitors, and therefore from the standpoint of output impedance is inherent to essentially all regulator designs.

The implications of these results on the topology of integrated regulators were then examined, starting by showing that the intrinsic negative feedback provided by a source-follower output device can significantly improve the regulator's performance – especially in high power-efficiency designs. We then examined how to make use of the fact that in some applications where externally generated supply noise is more critical than output impedance, the effect of supply noise on the regulated output can be sensed separately from the load itself. By sensing supply noise through a replica whose supply is not filtered by additional decoupling capacitance, the effective gain of the feedback amplifier essentially increases with frequency – making the required amplifier GBW increase only linearly with power supply rejection. Finally, we presented a replica compensated regulator that employs local amplifier feedback through a replica to extend the benefits of replica feedback to regulators which must maintain high tracking bandwidths.

One of the most challenging designs from the standpoint of supply noise issues is a modern microprocessor, where the vast majority of the power consumption is due to digital logic gates. While in this chapter we showed that a source-follower output stage can significantly improve the noise performance of a regulator, as we will show in the next chapter, the power overheads of typical regulator topologies whose goal is to reduce the supply impedance are simply too large to enable their adoption in such power-limited digital designs. Therefore, in the next chapter, we develop a push-pull shunt topology that can simultaneously reduce both the supply impedance and the overall power consumption of digital chips.

Chapter 5

Push-Pull Shunt Regulation for Digital Circuits

As we saw in several of the examples from the previous chapter, on-chip regulation has found extensive use in isolating the power supplies of sensitive analog or mixed-signal circuits from externally generated noise. While there has been some work exploring active noise reduction for digital chips [64,65], in today's power-limited environment and given the robustness of digital circuits, on-die regulation of digital supplies will only be adopted if it effectively decreases the total power dissipation of the chip.

While building a regulator that must spend power to reduce the effective supply network impedance may at first seem contradictory to the goal of reducing the entire chip's power dissipation, variations in the supply voltage (caused by variations in the load current) are themselves a cause of additional power dissipation. As we will describe in more detail in the first section of this chapter, this additional power dissipation is due to the fact that in synchronous digital systems, performance (or operating frequency) is set by the minimum average supply voltage over a clock cycle; noise-induced drops in the minimum supply voltage necessitate an increase in the nominal voltage to maintain the same performance. Therefore, if a regulator consumes less power than the power saved by the reduction in supply voltage enabled by lower noise, it can reduce the overall chip power consumption.

Achieving such efficiency requires optimizing the regulator to use minimal static power (both in the output stage and in the feedback circuits). As we will show in further detail in the second section of this chapter, single-supply linear regulators intrinsically burn significant static power in the output stage, making them unsuitable for this application. Therefore, in the rest of this chapter, we describe the design and implementation of a push-pull shunt regulator that makes use of a second, higher-than-nominal supply voltage to enable the regulator to transiently deliver energy to the load, allowing us to achieve the goal of reducing the chip's power consumption.

5.1 Efficiency of an Unregulated Digital Chip

Since the main barrier to the adoption of regulation in digital applications is the need to maintain chip power consumption constant (or even lower it), we must first evaluate the additional power dissipation caused by supply variations. We will calculate this excess power by assuming that the digital circuits' performance – and hence the minimum supply voltage – are kept constant. Therefore, as the noise on the supply increases, the nominal supply voltage must increase. An additional factor to consider is that since the dynamic current (and hence current variations) of a digital circuit is proportional to $C_{sw} \cdot f_{clk} \cdot V_{dd}$ (where C_{sw} is effective switched capacitance per cycle and f_{clk} is the operating frequency), the effective noise current increases along with the nominal supply voltage.

In order to maintain a certain minimum voltage $V_{dd,min}$ under zero-mean load current variations proportional to a constant k_n times the average load current (and whose positive variations are equal in magnitude to the negative variations), it can be shown that

$$\left(1 - \tfrac{1}{2} k_n / LR_{dist}\right) V_{dd,nom} = V_{dd,min} \tag{5.1}$$

where $V_{dd,nom}$ is the nominal supply voltage and LR_{dist} is defined as the load rejection of the supply network (e.g.., for a worst-case sinusoid, $R_{load}/||Z_{o,max}||$). With these definitions, the relative noise on the supply voltage (i.e., $\Delta V_{dd,p2p}/V_{dd}$) is exactly k_n/LR_{dist}.

With the simplifying assumption that the chip's power consumption is purely dynamic (i.e., no leakage current), the power consumption of the chip is proportional to $V_{dd,nom}^2$. Hence, the efficiency of the chip (relative to the case with no noise) is:

$$\eta = \left(1 - \tfrac{1}{2}\left(k_n / LR_{dist}\right)\right)^2 \tag{5.2}$$

5.2 Efficiency Limitations of Single-Supply Regulators

5.2.1 Series Regulator

We will begin the discussion by examining the conceptualized series regulator shown in Figure 5.1. We have drawn the output stage as a variable resistor (instead of with the transistor that implements this variable resistor) because as we will show shortly, the major limitation on the efficiency of the regulator is essentially independent of the actual implementation, and is inherent to the series topology itself.

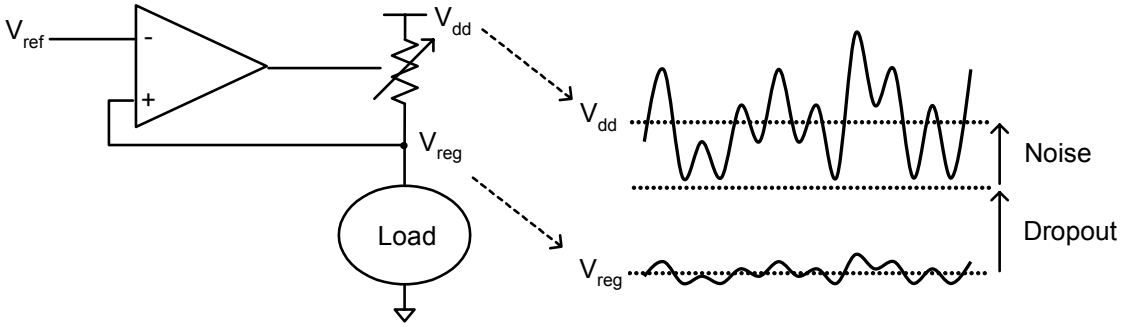


Figure 5.1: Series regulator and example waveforms on input supply (V_{dd}) and regulated supply (V_{reg}), highlighting the need to margin the dropout voltage of the regulator to include the variations on V_{dd}.

For a series regulator to maintain an output impedance that is decoupled from (and hence can be lower than) the impedance of the input supply V_{dd}, the input supply voltage must remain above the regulated output V_{reg}. The minimum allowable voltage drop from V_{dd} to V_{reg} is typically referred to as the dropout voltage. This dropout voltage must be

maintained not only in a DC sense, but as shown also shown in Figure 5.1, taking into account the expected dynamic variations on V_{dd}.

Unfortunately, since the series regulator does not significantly alter (relative to an unregulated chip) the impedance of the input supply or the load current variations, the margin required on V_{dd} to maintain a certain minimum V_{reg} due to load current variations will be just as large as in the unregulated case. Even worse, to keep the same performance, V_{dd} in the regulated system has also been increased by the regulator's dropout, making it clear that a series regulator can not achieve the goal of reducing the net power consumption of the chip.[25]

Intuitively, the fact that series regulators reduce the overall efficiency of a digital chip is not very surprising. The noise on the power supply is set by the supply impedance and by the variations in the load current, and although the regulator uses negative feedback to reduce the voltage noise seen by the load, the series regulator by definition adds series resistance to the impedance of the supply network.

Although they are not suited for digital supply regulation, the series resistance of a series regulator is exactly what allows it to isolate the regulated supply from externally generated noise on the input supply. As mentioned in the previous chapter, this isolation makes the series topology very suitable for regulating the supply of sensitive analog or mixed-signal components, and in Appendix B we describe an analysis framework for optimizing the overall efficiency (at a given supply rejection) of such regulators.

5.2.2 Shunt Regulator

Since the underlying cause of noise on the power supply of digital circuits is variation in the load current itself, a shunt regulator that directly reduces these current variations is much more suitable to this application than a series regulator. Figure 5.2 shows such a

[25] Since the nominal voltage seen by the digital circuits will be lower with the series regulator than in the unregulated case, the magnitude of the noise current in the regulated chip will be smaller than in the unregulated chip. However, for the expected magnitude of noise on the supply, this reduction will be small (typically, less than 10%) – only for extremely large noise magnitude (~50% variation on the supply) does this effect become large enough for the series regulated chip's power consumption to become comparable or less than that of an unregulated chip.

shunt regulator, where once again we have depicted the output stage with a conceptual current source instead of the actual transistor in order to highlight the fact that the efficiency limitation is inherent to the regulator's topology.

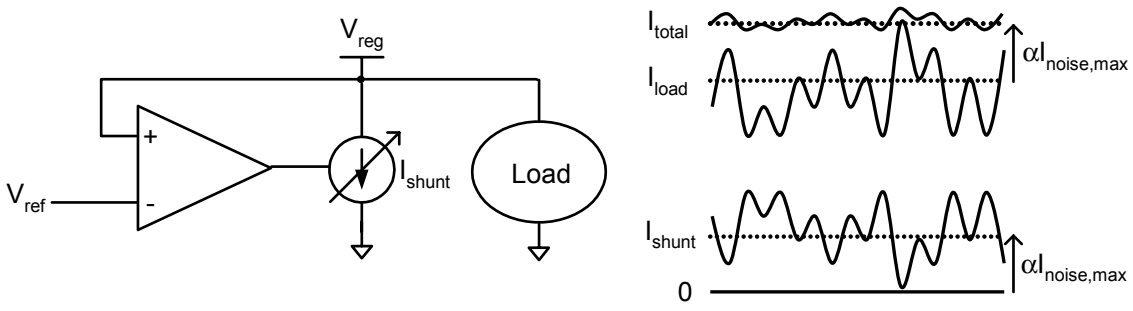


Figure 5.2: Shunt regulator and example load current (I_{load}) and shunt current (I_{shunt}) waveforms, highlighting the need for the shunt current source to statically draw current proportional to the maximum noise current ($I_{noise,max}$).

As also shown in Figure 5.2, the instantaneous load current draw can be larger or smaller than the average current – in other words, the current variations are bidirectional. However, without an additional power supply or energy storage element, the shunt regulator's current source must be unidirectional, and can only pull current out of the power supply. Therefore, the shunt regulator must statically pull a current that is proportional to the worst-case excess current, so that it can then dynamically reduce its current when the load demands excess current.

The need to statically dissipate the worst-case noise current significantly limits the efficiency of a chip with a shunt regulator (relative to an unregulated chip). Specifically, even with the optimistic assumptions that the regulator's feedback amplifier has infinite bandwidth and dissipates no power whatsoever, the efficiency of the chip will be

$$\eta = \frac{\left(1 - \frac{1}{2}\frac{k_n}{LR_{dist} \cdot LR_{reg}}\right)^2}{1 + k_{Idiv}\left(1 - \frac{1}{LR_{reg}}\right)k_n} \tag{5.3}$$

where LR_{reg} is the additional rejection contributed by the shunt regulator, k_{Idiv} accounts for the percentage of current the shunt regulator must handle for a specific noise excitation,[26] and the average I_{shunt}/I_{load} is $k_{Idiv} \cdot (1-LR_{reg}^{-1}) \cdot k_n$. As shown in the example of Figure 5.3, the required static output current makes the overall efficiency of a chip with a shunt regulator worse than that of an unregulated chip for any reasonable level of noise.

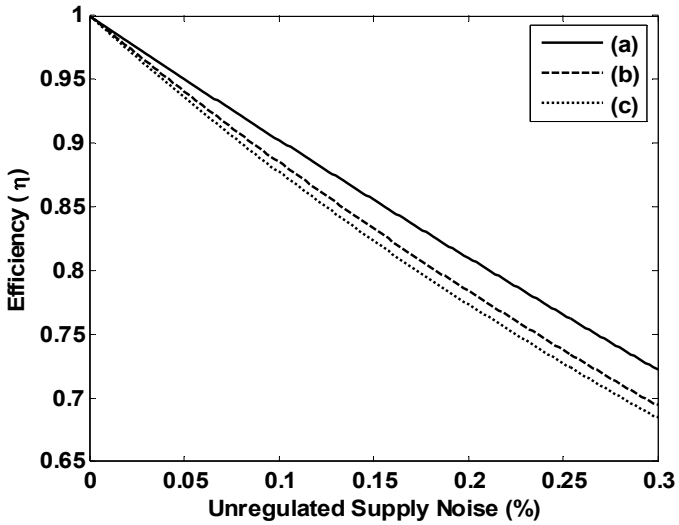


Figure 5.3: Digital chip efficiency vs. unregulated supply noise for a) an unregulated supply, b) a shunt regulated supply with $LR_{reg} = 1.5$, and c) a shunt regulated supply with $LR_{reg} = 2$. In all cases, $k_{Idiv} = \frac{1}{3}$ and $LR_{dist} = 5$.

The fundamental cause of the shunt regulator's limited efficiency is that it cannot deliver energy when the load transiently demands more power – all the regulator can do is burn excess static power in the output stage and then modulate this power dissipation.

5.3 Push-Pull Shunt Regulator

To enable the regulator to transiently deliver energy without the static overhead of typical single-supply topologies, we can introduce a second, higher-than-nominal supply voltage and build a push-pull shunt regulator (Figure 5.4). As we will describe next, the

[26] The percentage of the noise current that the regulator's output stage must handle will vary with the spectral content of the noise, and must be calculated for a given noise current excitation. With the assumption of an ideal amplifier, for sinusoidal noise current $k_{Idiv}=1$. For broad-band (white), random (but bounded) current noise, k_{Idiv} can be calculated using $\int_0^\infty \left\| g_m A_a(f) Z_{reg}(f) \right\|^2 df$, and is typically $\sim \frac{1}{3}$.

use of a second supply allows the dissipation of the regulator's output stage to be tied mostly to the average deviation of the noise current, giving the regulator the potential to improve upon the efficiency of an unregulated chip.

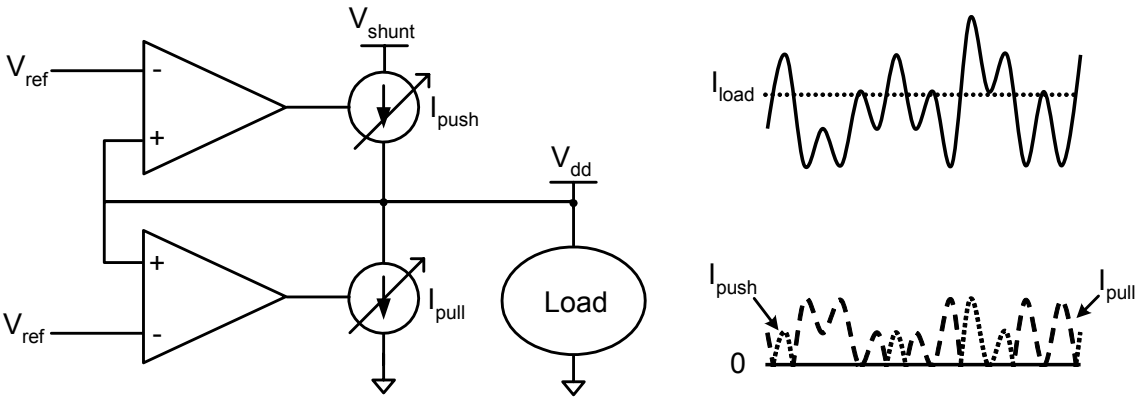


Figure 5.4: Push-pull shunt regulator topology.

With the use of a second power supply, the nominal output current of the regulator can be set to zero, and the appropriate current source turned on only when necessary. Therefore, as also shown in Figure 5.4, the average current through each of the regulator's output current sources is set by the average of one side of the noise current.

Clearly, if the peaks of the noise current are significantly larger than the average current deviation, the average current flowing through a push-pull shunt will be significantly lower than that of a single-supply shunt. If we continue to assume ideal (i.e., infinite bandwidth, zero power) feedback amplifiers, and for simplicity assume that the second power supply $V_{shunt} = 2*V_{dd,nom}$, the efficiency with a push-pull regulator is

$$\eta = \frac{\left(1 - \frac{1}{2}\frac{k_n}{LR_{dist} \cdot LR_{reg}}\right)^2}{1 + \left(\frac{k_{Idiv}}{k_{pk/avg}}\right)\left(1 - \frac{1}{LR_{reg}}\right)k_n} \tag{5.4}$$

where $k_{pk/avg}$ is simply the peak current deviation divided by the mean current deviation. As shown in Figure 5.5, even at a relatively moderate $k_{pk/avg}$ of 3, eliminating the

unnecessary static power finally enables the regulated chip to achieve higher efficiency than an unregulated design.

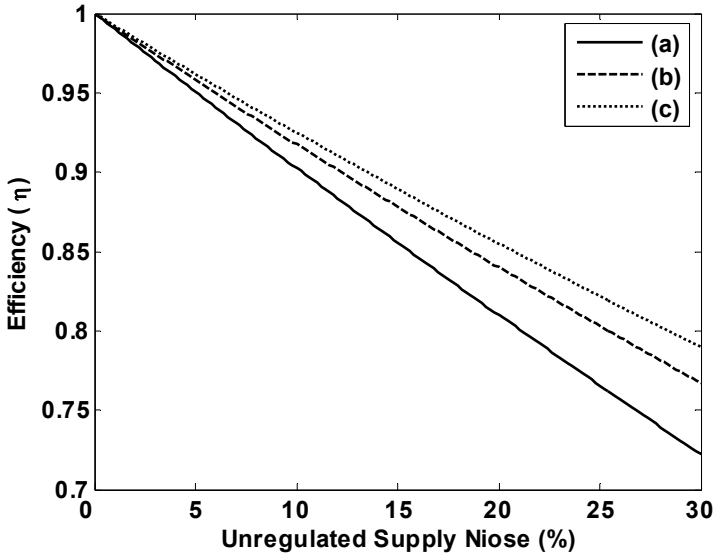


Figure 5.5: Digital chip efficiency vs. unregulated supply noise for a) an unregulated supply, b) a push-pull shunt regulated Supply with LR_{reg} = 1.5, and c) a push-pull shunt regulated supply with LR_{reg} = 2. In all cases, $k_{pk/avg}$ = 3, k_{Idiv} = ⅓ and LR_{dist} = 5.

This push-pull topology is similar to the "active clamps" developed for board-level power distribution applications [66,67,68], and in this section we build upon those works to design an integrated regulator specifically targeted to increase the efficiency of digital chips. Since the regulator must make use of a second power supply, we will first examine some of the important considerations in the design of the shunt supply network. We will next look at the regulator's implementation in more detail, specifically the steps that were taken to truly minimize the static power consumption of the design. Finally, we will present measured results from a test-chip that demonstrate the regulator's ability to both reduce noise and improve a digital chip's overall efficiency.

5.3.1 Shunt Supply Network Design

Since modern chips already dedicate a significant percentage of the available pad, pin, and metal resources to the supply distribution network, integrating a push-pull shunt regulator requires allocating these resources between the main power supply (V_{dd}) and

the additional shunt supply (V_{shunt}). Clearly, taking resources away from the main power supply will increase its loss (since its series resistance will increase), but allocating too few resources to the shunt supply would make it too lossy to be effective.

Intuitively, the resources for the two supplies should be allocated according to their relative contributions to the total loss, and this is exactly the allocation found from mathematically minimizing the total resistive losses of the two supplies:

$$p_{Vshunt} = \frac{I_{push,rms}}{I_{push,rms} + I_{load,rms}}, \qquad p_{Vdd} = \frac{I_{load,rms}}{I_{push,rms} + I_{load,rms}}, \tag{5.5}$$

where p_{Vshunt} is the percentage of the total resources dedicated to the shunt supply, and p_{Vdd} is the percentage dedicated to the main power supply. As long as the regulator delivers only transient currents with reasonable peak-to-average ratio, this allocation results in low p_{Vshunt} (typically less than 5%), and hence the impact on the resistive losses of the main supply will be relatively minor.

In addition to properly allocating resources to minimize the resistive losses of the supply network, steps must be taken to ensure that I_{push} returns only through the on-chip path (highlighted in Figure 5.6). Specifically, if V_{shunt} is supplied by an external voltage source, using the impedance model of Figure 5.6 it can be shown that the return current will flow almost entirely through the on-chip path when

$$\|Z_{Vshunt\_dist}(j\omega)\| \gg 1/(\omega C_{Vshunt}), \; \|Z_{Vshunt\_dist}(j\omega)\| \gg \|Z_{Vss\_dist}(j\omega)\| \tag{5.6}$$

Achieving these conditions at the frequencies of interest requires that part of V_{dd}'s decoupling capacitance be used for V_{shunt}. Of course, removing decoupling capacitance from V_{dd} increases the noise on the main supply. Therefore, to keep the capacitance reallocated to V_{shunt} low (~10% of the total available), it is typically necessary to intentionally increase $\|Z_{Vshunt\_dist}(j\omega)\|$ as well. To avoid sacrificing the DC resistive losses of the shunt supply, this increase in $\|Z_{Vshunt\_dist}(j\omega)\|$ is best achieved by increasing the shunt supply's series inductance.

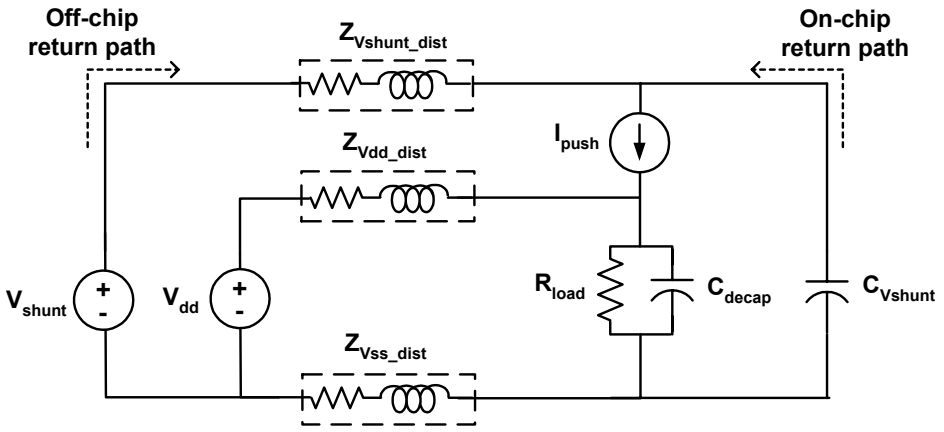


Figure 5.6: Simplified model of the supply impedances in a push-pull shunt regulated system, highlighting the on-chip and off-chip return paths for I_{push}.

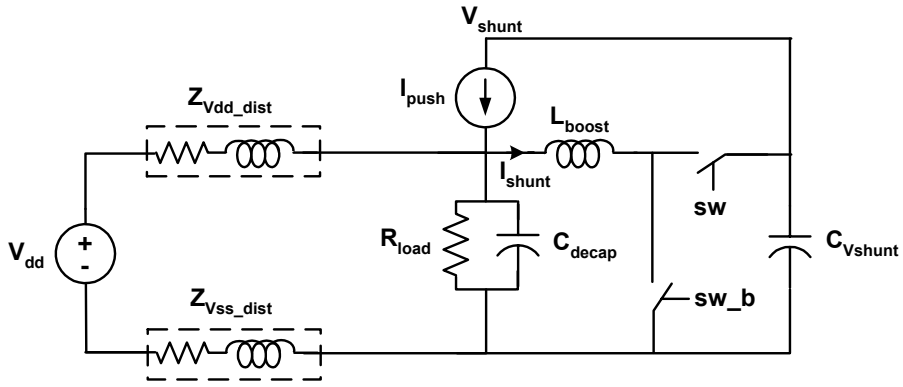


Figure 5.7: Generation of V_{shunt} using an on-chip boost converter.

One interesting alternative to an external power supply is to generate V_{shunt} directly from V_{dd} using an on-chip boost converter or charge pump, as shown in Figure 5.7. This approach can be advantageous because the package supply distribution resources can be dedicated entirely to the main supply network if the boost inductor L_{boost} (or flying capacitor for a charge pump) is integrated onto the die. Even if L_{boost} is not on-die but rather integrated into the package, V_{shunt} would no longer require any package pins.

Interestingly, even though I_{push}'s return path will intrinsically remain on-chip with a fully integrated boost converter,[27] because any current that flows from V_{shunt} actually

[27] If L_{boost} is in the package, care must be taken to minimize the parasitic capacitance of the package traces to either of the package power planes (V_{dd} or V_{ss}).

originates from on-chip V_{dd} (note I_{shunt} in Figure 5.7), care must be taken in the design of the converter to ensure that it does not interfere with the operation of the regulator. Specifically, it can be shown that for I_{shunt} to be a small fraction of I_{push} at the frequencies of interest, the series input impedance of the converter must be significantly higher than the impedance in parallel with the converter's output (i.e., $1/(j\omega C_{Vshunt})$). Much as was the case for an externally generated V_{shunt}, this requires decoupling capacitance for V_{shunt} and may require increasing the effective series inductance of the converter.

Even though the average current through V_{shunt} should be low to attain high overall efficiency, the transient currents are set by the full load current variations. In fact, the relative magnitude of these transient currents will inherently be higher than the percentage of resources dedicated to the shunt supply (since the resources should be allocated based on average regulator current). Hence, the impedance of the shunt supply will be relatively high, and the large transient currents make V_{shunt} very noisy. Therefore, to guarantee that the regulator's push-side output device does not fall out of saturation, appropriate margin (typically 200-300 mV) must be added to V_{shunt}'s nominal value.

5.3.2 Regulator Implementation

In Figure 5.4 we presented a conceptual schematic of the push-pull shunt regulator topology to highlight its potential for the regulation of digital chips, but the actual implementation of the regulator can have significant impact on the success of the design. To achieve the goal of improving the efficiency of digital chips, the regulator's implementation must be tailored to truly ensure minimal static power consumption – both in the output stage and in the feedback circuitry. As shown in Figure 5.8, robustly minimizing static power has led us to use local reference voltage generation, comparator-based feedback, and a switched source-follower output stage, and in this subsection we will describe each of these optimizations in more detail.

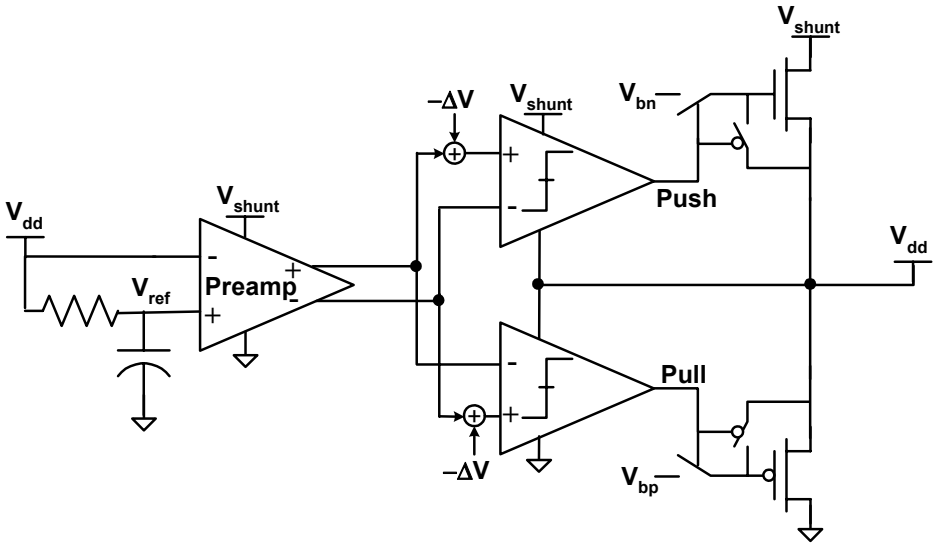


Figure 5.8: Push-pull shunt regulator implementation with local reference generation, comparator-based feedback, and switched source-follower output stage to minimize static power consumption.

5.3.2.1 Local Reference Generation

If the regulator were implemented directly from Figure 5.4, even relatively small offsets between the reference voltage V_{ref} and the average value of V_{dd} would cause the regulator to statically push or pull significant output current, greatly degrading the regulator's efficiency. Fortunately, it is relatively straightforward to avoid this issue by generating the regulator's reference by locally RC filtering the regulated supply voltage [66,68] – as shown in Figure 5.8. A local filter for each regulator location (vs. a single global filter) is necessary because IR drop will vary spatially within the die.

Clearly, generating the reference in this manner will cause the regulator to ignore any variations well below the bandwidth of the filter (which is typically set in the low-MHz range). This behavior is intentional, since it forces the regulator to operate only on voltage transients; given the low series resistance of the main supply network, attempting to counter slowly varying IR drops with the regulator would be ineffective and a waste of power. In fact, constructing the regulator such that it operates only on voltage transients can be especially important in modern chips since functional units or even the entire chip

may be shut down or placed into lower performance modes by the operating system for extended periods of time [69].

5.3.2.2 Comparator-Based Feedback

To ensure that the regulator achieves its efficiency target, the regulator's output stages should not conduct any current when the supply is quiet. This rules out linear (or "class-A" [48]) operation of the regulator's power devices, necessitating instead a mode of operation where the quiescent current of the output stage (relative to the maximum current) is set by the power device's bias voltage, IV characteristics, and gate voltage swing.

In a regulator with purely linear voltage-mode feedback amplifiers, the power device gate voltage swing is not an independent design parameter – it is set by the magnitude of the supply voltage noise and by the gain of the amplifier, which in and of itself is tied to the effective transconductance of the output stage (as derived in the previous chapter). Therefore, with such a design it is not straightforward to guarantee gate voltage swings large enough to mitigate any errors (systematic or random) in the biasing of the output device, potentially leading to large static output current and poor efficiency. The active clamp described by Wu in [66] avoided some of these issues by feeding the output of a transconductance feedback amplifier into a cascade of multiplying current mirrors (essentially building an amplifier that is non-linear in voltage, but linear in overall transconductance), but in this application the degradation in feedback bandwidth caused by these current mirrors is highly undesirable.

In order to completely eliminate the push-pull regulator's static output power in a robust manner, we employed comparator-based feedback as shown in Figure 5.8. To avoid unnecessary (and highly inefficient) limit cycles that can potentially arise in such a comparator-based system [70], the thresholds of the comparators are offset to create a dead-band.[28] In addition, the offsets of both comparators and the preamp are tunable to

[28] In addition to limit cycle considerations, it is important to note that it can beneficial from an efficiency standpoint to tune the dead-band along with the magnitude of the supply noise. In essence, since digital performance is set by the minimum voltage, the dead-band can be used to reduce the amount of current the regulator consumes without significantly impacting the minimum voltage. Although the exact relationship

compensate for mismatch and zero-center the dead-band. While the use of comparators increases the feedback gain required over a linear scheme (since the comparator outputs must swing full-rail nearly independent of the input voltage), it allows the use of CMOS inverters as gain stages in the comparator and as buffers to drive the output stages. As discussed in [71], the efficiency of CMOS buffers in modern processes allows the power consumption of this approach to be the same or less than a comparable design with class-A feedback amplifiers.

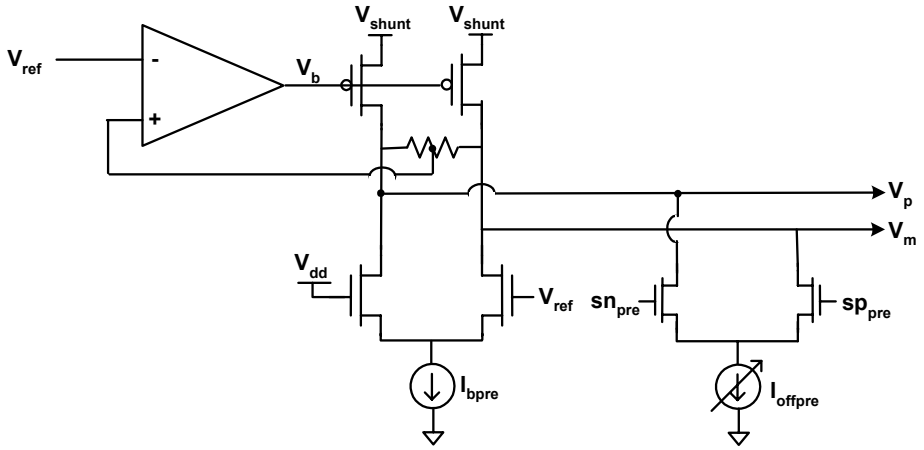
Even though the use of comparators in the feedback path makes the regulator a non-linear system, the intuition and design rules developed in the previous chapter from the analysis of linear-feedback regulators are still applicable to the push-pull regulator. Because of the high-frequency filtering provided by the decoupling capacitance, the feedback will act to make the average output current of the regulator track the noise current within the loop bandwidth. Intuitively, this is very similar to the behavior of quantizer outputs from ΣΔ A/D converters, which are often analyzed by linearizing the quantizer into an equivalent gain [72]. More formally, describing function techniques [70,73] (which rely on filtering of the non-linearity for their accuracy) can be used to quasi-linearize the comparator. With this linearization, the noise response of the regulator is qualitatively the same as that of a linear-feedback regulator, but with the effective g_m of the feedback set by the magnitude of the regulator's current relative to the noise current – allowing the results from the previous chapter related to the requirements on the feedback loop to be extended to the comparator-based design.

The circuit implementations of the preamp and comparator are shown in Figure 5.9. To implement the tunable offset and dead-band, digitally programmable (5-bit with a thermometer decoder) current sources are tied to the outputs of the pre-amp and comparator. This implementation was chosen for its simplicity, although other potentially more efficient schemes such as a pair of skewed-sizing differential pairs with programmable tail currents [74] could be adopted. Common-mode feedback is used to

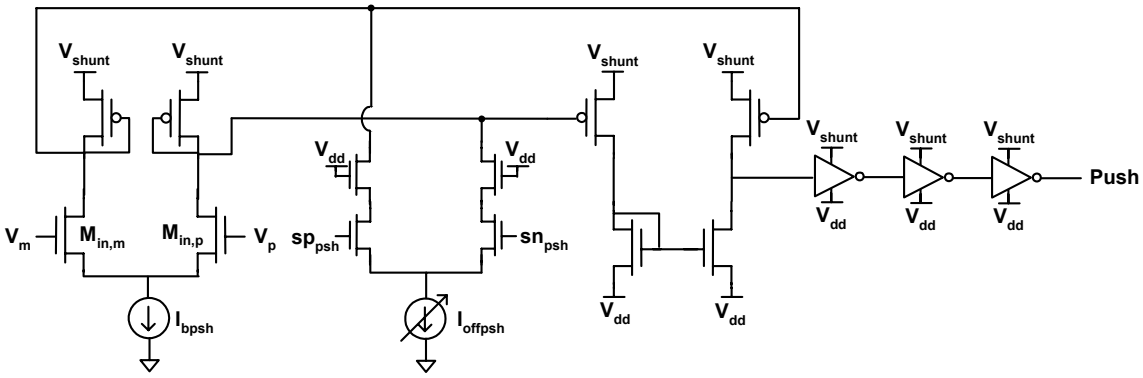
---

between the dead-band and noise magnitudes is strongly dependent upon the current noise distribution (and may not be straightforward to calculate analytically), a good general rule is to maintain the dead-band constant relative to the σ of the noise. Section D.2 of Appendix D provides further analysis and discussion of this issue.

set the preamp's common mode to V_{ref} (the average value of V_{dd}) in order to allow both the push side and pull side comparators to easily accept the single preamp output (as opposed to two preamps with potentially different offsets, gains, and bandwidths). Finally, the comparator design used a first current-mirroring stage in order to simplify the task of level-translating from an input centered at V_{ref} to a digital output swinging between V_{dd} and V_{shunt} (for the push side – the pull side swings between V_{dd} and V_{ss}).



(a)

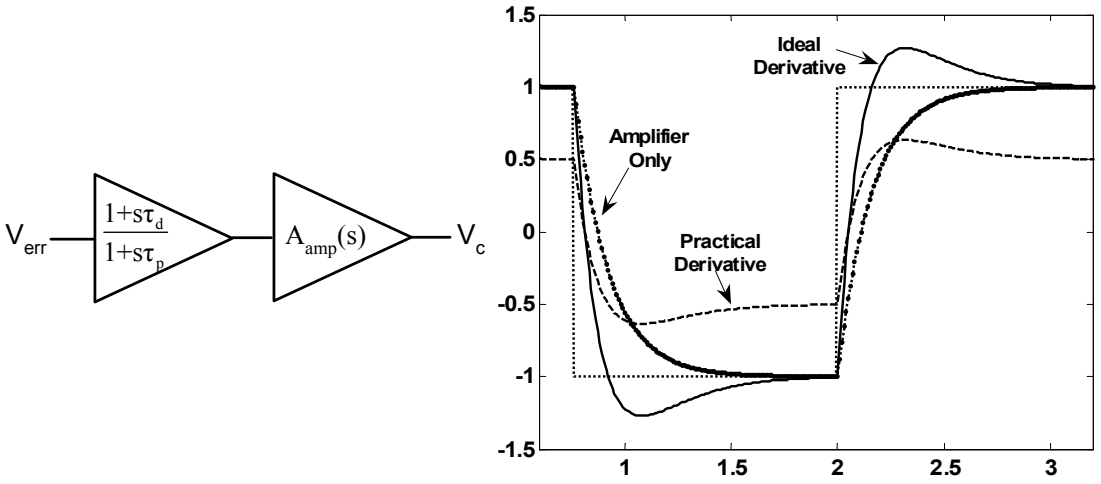


(b)

Figure 5.9: a) Preamp and b) push side comparator implementations, showing programmable current sources to tune individual offsets. The pull side comparator implementation is complementary (i.e., PMOS replacing NMOS, power and ground connections swapped) to the push side, but with the second mirror stage operating between V_{dd} and V_{ss} (as opposed to between V_{shunt} and V_{dd} for the push side).

In addition to robustly eliminating static output current, the use of comparator-based feedback also provides the opportunity to exploit the non-linear nature of the control loop

to reduce the effective feedback delay. In order to illustrate how this could be accomplished, we will begin by considering the use of an additive derivative term at the input of a linear feedback amplifier. Such a configuration is shown in Figure 5.10a, where $s\tau_d$ is the additive derivative term; a pole has been included as part of the input filter since any real implementation of such a derivative will inherently have finite maximum high-frequency gain.



(a) (b)

Figure 5.10: a) Additive derivative term at the input of a linear feedback amplifier. b) Example output waveforms for an amplifier in isolation, an amplifier with an ideal derivative filter, and an amplifier with a practical (DC gain reduced) derivative filter.

By appropriately choosing the derivative time constant τ_d relative to the amplifier's bandwidth (as described in Appendix C), some of the negative phase shift from the amplifier's limited bandwidth can be canceled (over a band of interest) by the positive phase shift of the derivative term. As shown in Figure 5.10b, this reduced phase shift translates into a reduction of the delay of the feedback action. Unfortunately, despite the fact that there are many ways to implement such a peaking filter (e.g., a resistive divider with a feedforward capacitor, negative feedback through a low-pass filter [75,76], etc.), without area-consuming peaking inductors [48] these implementations will all actually reduce the low-frequency gain of the feedback signal (rather than increase the high-frequency gain). Hence, as also shown in Figure 5.10b, although the delay of the feedback is improved, with a practical derivative filter the gain of the loop is also

decreased – limiting the potential performance improvement of this scheme in a purely linear loop.

Unlike the linear system, in a comparator-based feedback loop, as long as the comparator outputs continue to swing full-rail, it is only the shape of the transfer function in front of the comparator that sets the behavior of the system. In other words, the excess gain of the comparators compensates for the reduction of low-frequency gain caused by the derivative implementation. Therefore, with a properly chosen derivative time constant (typically, $1/\tau_d$ set to roughly half of the feedback bandwidth) the dominant impact of this scheme is to reduce the effective delay of the feedback action, improving the overall efficiency of the regulator.

The use of a $1+s\tau_d$ feedback filter with comparator-based control is very similar to the sliding mode controllers that have found use in DC-to-DC converters [77,78,79]. As shown in Appendix D, the region over which the push-pull shunt regulator slides along the surface defined by the feedback filter is set by the magnitude of the regulator current. However, if the load current noise is larger than the regulator current, it can force the system outside of the sliding region. To maximize the overall efficiency of the chip, the impedance reduction provided by the regulator is often only roughly a factor of 2. Hence, the regulator's current is typically not very large relative to the noise current, and therefore the regulator rarely operates within the sliding region.

In order to implement the derivative filter for the push-pull shunt regulator, we modified the input stage of the preamp as shown in Figure 5.11. The derivative behavior is accomplished by subtracting an RC-filtered version of V_{dd} (V_{hf}) from V_{dd}. Hence, the effective input to each segment whose reference is connected to V_{hf} is $s\tau_p/(1+\tau_p)$, where τ_p is the time-constant of the RC filter (which is significantly smaller than that of the filter generating V_{ref}). Therefore, ignoring the gain provided by the preamp, the transfer function implemented by this circuit is

$$H_d(s) = (1-k_d)\frac{1+s\tau_p/(1-k_d)}{1+s\tau_p}, \tag{5.7}$$

where k_d is the number of segments tied to V_{hf} divided by the total number of segments.

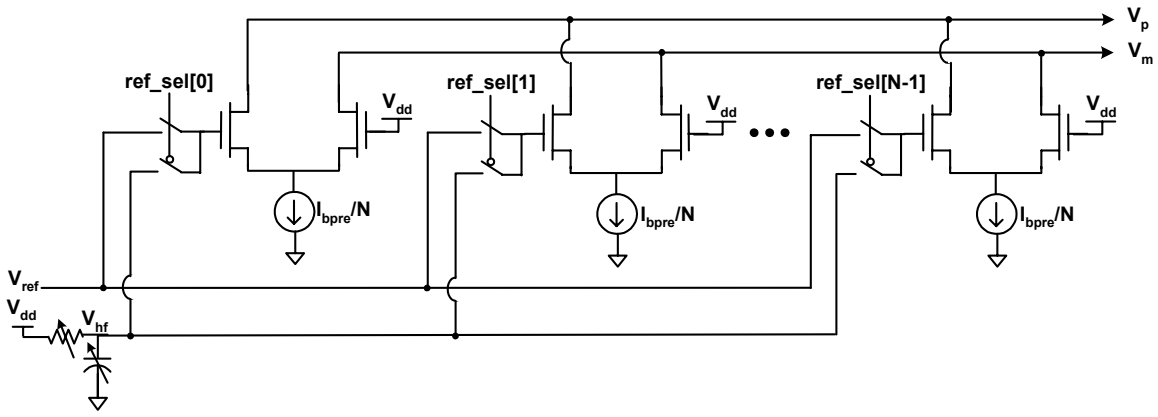


Figure 5.11: Preamp input stage modified to implement the derivative filter, where the filter time constants are programmable.

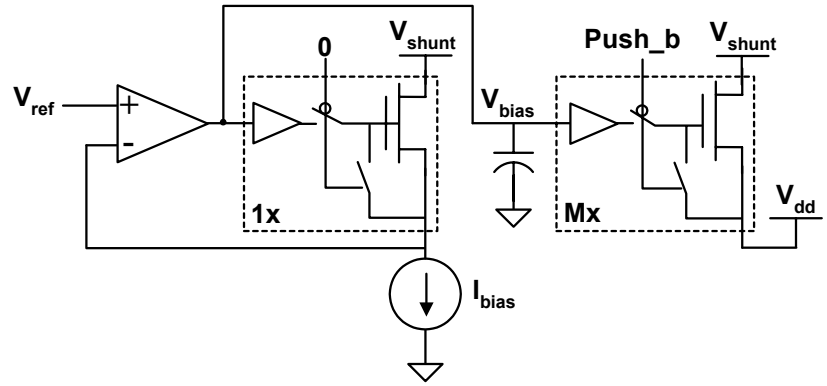
Applying the positive phase shift of the filter at the appropriate frequency requires the derivative time constant (which in this implementation is $\tau_p/(1-k_d)$) to be set as a relative fraction of the effective bandwidth of the feedback path. In addition, the bandwidth of the filter generating V_{hf} should roughly match the feedback bandwidth. This is because if the filter bandwidth is significantly greater than the feedback bandwidth, k_d would need to be increased to set τ_d, further lowering the DC feedback gain and requiring additional gain margin from the comparators. Alternatively, if τ_p is too large, it will reduce the positive phase shift of the filter at the feedback bandwidth, which is where the positive phase has the greatest impact on the effective delay. Since the bandwidth of the feedback path is affected by various transistor parameters which can vary independently of the resistance (typically polysilicon) and capacitance (either metal-to-metal or gate) which set τ_p, V_{hf}'s RC filter was made programmable in this design.[29]

5.3.2.3 Switched Source-Follower Output Stage

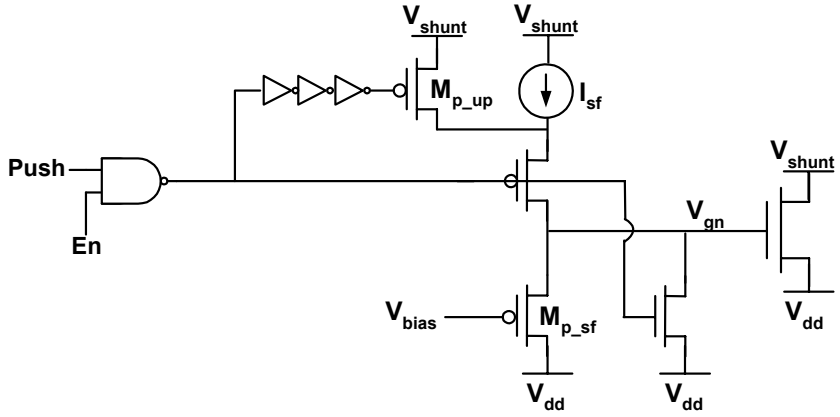
Although the push-pull regulator is non-linear and based on comparators, it benefits from a source-follower output stage (over a common-source) for reasons similar to those

[29] Programmability of k_d was also included for the purpose of testing and fine-tuning of ω_d.

described in the previous chapter.[30] The two principal goals in the design of such an output stage for a push-pull regulator are achieving low turn-on delay, and maintaining predictable output current with a noisy V_{shunt}.



(a)



(b)

Figure 5.12: a) Replica-bias generation for a source-follower output driver. b) Output driver implementation, highlighting the switched source-follower buffer isolating V_{bias} from V_{gn}. This figure shows the push side – the pull side implementation is once again complementary.

To achieve these goals, the regulator employs a replica-biasing scheme where a bias voltage is driven onto the gate of the power device (through a buffer), as shown in Figure 5.12a. For efficiency, the power consumption of the amplifier in the replica-bias loop

[30] As with a series regulator, the use of a source-follower output stage for the push side requires a higher voltage (vs. a common-source) at the output device's gate (to accommodate the output device's V_{th}). For simplicity, in the implementation described here V_{shunt} was raised by V_{th}; the regulator's efficiency could be further improved by separating the supply of the feedback path from that of the output stage

should be minimized, making its output impedance relatively high. This means that directly switching V_{bias} onto the gate of the power device (which is nominally discharged to V_{dd}) will create disturbances on V_{bias}. While adding capacitance can reduce the disturbance, this exacerbates the dependence of I_{push} on the history of the comparator outputs and would lead to undesirable filtering of the output current. Hence, it is advantageous to include a buffer to isolate V_{gn} from the high-impedance V_{bias}.

This additional buffer is implemented as another source follower, as shown in Figure 5.12b. To eliminate its static power, the buffer is turned on only when current is being pushed; since the buffer current flows into V_{dd} (or out of V_{dd} for the pull side) and only flows when the main output current does, I_{sf} can simply be treated as a portion of the regulator's total output current. However, waiting for I_{sf} to charge V_{gn} up to its final value can cause significant delay, so during the turn-on transition M_{p_up} is left on to provide a large transient current. For simplicity, in this implementation the width and magnitude of the pre-emphasis current pulse were not well-controlled, and hence there will be some transient imprecision in the output current – preferably an overshoot. Fortunately, as long as the transient is short its effects will be minor, while the improvement in the feedback delay directly impacts the regulator's performance.

To enable the magnitude of the regulator's output current to be tuned along with the magnitude of the current noise (and to enable the regulator to be disabled), several programmability features were added to the driver shown in Figure 5.12. Specifically, both the magnitude of I_{bias} and the ratio of real output current to replica current (M in Figure 5.12a) were set by scan-controlled digital bits. To ensure monotonicity, the digital bits controlling the current ratio were thermometer decoded and the output driver was split into an array of 15 identical segments (with each segment shown in Figure 5.12b).

5.3.3 Experimental Results

In collaboration with AMD, this push-pull shunt regulator was designed and taped-out in a 65 nm SOI test-chip used for technology bring-up. To facilitate the characterization of the regulator (as described below), the test-chip included noise generators and

performance monitoring circuitry. The die area of the regulator test-site[31] (excluding scan controls) was roughly 400 μm by 400 μm, and the supply distribution network was designed using essentially the same methodology (but with the reallocation of resources for V_{shunt} as described in Section 5.3.1) as the rest of the microprocessor test-chip.

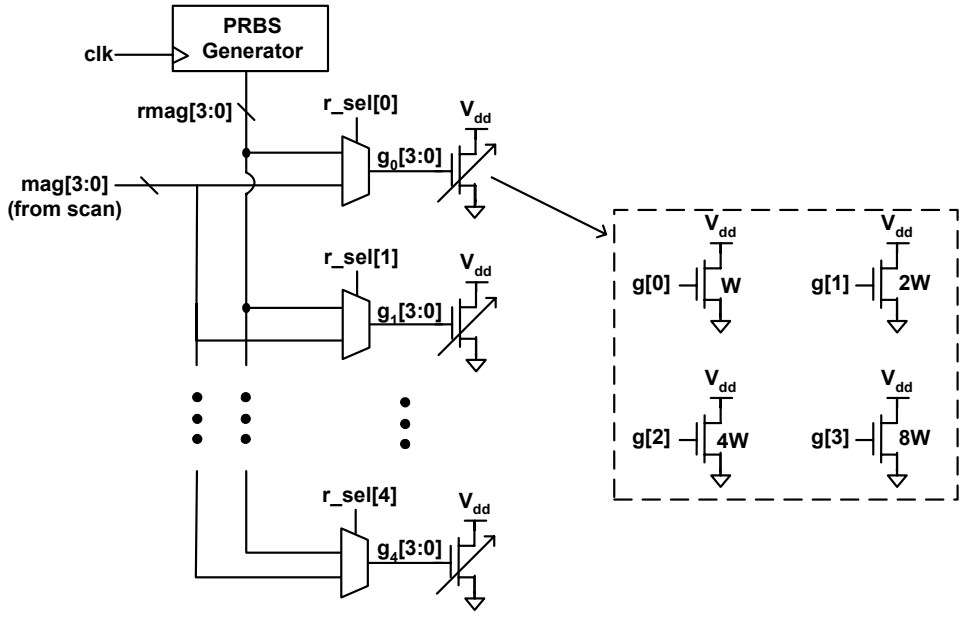


Figure 5.13: On-chip programmable broadband current noise generator.

To characterize the regulator at different noise levels and with broadband current noise, we used the programmable on-chip noise generator shown in Figure 5.13. As shown in the figure, there are 5 noise generator banks, where each bank consists of a binary-weighted, 4-bit array of NMOS devices tied between V_{dd} and V_{ss}. Each individual bank can have its magnitude set by a static control signal mag[3:0] (such that the bank draws only DC current), or by the lower 4 bits of a 2^{15}-1 PRBS generator (such that the bank draws a pseudo-random amount of current each clock cycle). The current noise generated by the PRBS is essentially white up to roughly half of the PRBS generator's clock frequency, and the relative magnitude of the current noise is set by the number of

[31] The impact of high-frequency current noise on the supply network is relatively localized because of the resistance of the on-chip supply grid, and therefore a full-chip implementation of the push-pull regulator would likely have many regulator sites distributed around the die. The results presented here are from a single regulator site in isolation, but these results can be used to project the impact of many such sites on an entire chip. An important topic for further future investigation would be to examine the density of regulator sites required to avoid undesirable interactions between the sites (leading e.g. to oscillations).

banks whose inputs are controlled by the PRBS generator (i.e., the relative current noise can range from roughly 0 to 5).

Since the goal of the regulator is to reduce power consumption at a fixed performance, to ensure that performance remained constant as the noise and regulation settings were varied during testing, we also integrated the performance monitor circuit shown in Figure 5.14. The monitor operates by generating a signal that transitions every clock cycle (cdiv), and then checking on every cycle that this signal was correctly captured by a flip-flop after N stages of fanout-of-four inverters. If on any cycle an error is detected (i.e., dval ≠ tval), the err signal will remain high until the circuit is externally reset. In this manner, the circuit will correctly indicate whether the performance set by the frequency of clk can be maintained at the worst-case supply voltage.

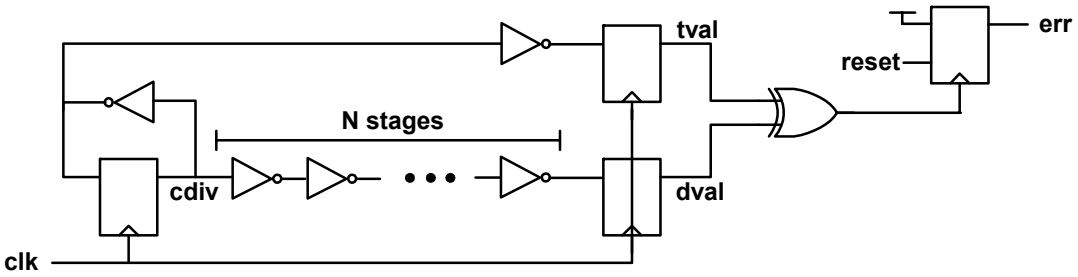


Figure 5.14: On-chip performance monitor circuit.

The measurement procedure for characterizing the regulator's impact on supply noise and efficiency[32] began with picking a supply voltage $V_{dd,set}$, and with the regulator disabled, setting the noise generator to statically draw roughly half of its maximum current (i.e., no noise). The performance monitor was then used to determine the maximum frequency of operation. Leaving this frequency constant, measurements were then taken at 4 magnitudes of current noise.

At each level of current noise, three different measurements were taken. First, since the noise generator banks and individual devices do not match perfectly, the supply voltage was set to $V_{dd,set}$ with the regulator still disabled, and the average power

[32] Since they are relatively straightforward, we will not describe here the measurement procedures used to calibrate the offsets of the preamp and comparators.

consumption of the chip (at the current noise setting) was recorded. This measurement provides the baseline power against which to evaluate the chip's efficiency (for that level of noise). Next, still leaving the regulator off, the supply voltage was raised to the minimum level at which the performance monitor was error free ($V_{dd,unreg}$), and this voltage and the chip's power consumption were once again recorded. Finally, the regulator was turned on, and its parameters (e.g., the magnitude of the output currents) were tuned along with the supply voltage (calling the final voltage $V_{dd,reg}$) to achieve minimum power with no errors from the performance monitor. The peak-to-peak voltage noise corresponding to each noise generator setting was simply calculated with

$$\text{p2p\_noise}_{\{unreg,reg\}} = 2\left(V_{dd,\{unreg,reg\}} - V_{dd,set}\right). \tag{5.8}$$

Figures 5.15 and 5.16 show the results measured from this chip using this procedure with $V_{dd,set}$ = 1.188 V and V_{shunt} = 2.14 V. At this $V_{dd,set}$ the maximum frequency indicated by the performance monitor was 1.67 GHz, and the power consumption was ~105 mW; the static power of the regulator's biasing and feedback circuitry was less than ~1.5 mW. Based on the measured supply current and calculated peak-to-peak voltage noise, the effective (broadband) unregulated supply impedance was ~869 mΩ. For a 12 mm by 12 mm chip operating at the same power density (i.e., 94.5 W total chip power), this would correspond to an impedance of ~966 μΩ.

In the range of expected supply noise (~8% peak-to-peak and above), the regulator successfully reduces the chip's power by up to ~1% while reducing the noise by ~30%.[33] Based on these results and simulations, in a production 65 nm technology with higher transistor f_t, we anticipate the regulator to achieve a ~50% reduction in noise with a ~4% improvement in total power.

[33] The external power supply used to generate V_{dd} for the test chip had a voltage resolution of ~5 mV. Hence, the measurements had to be snapped to this relatively coarse grid. This coarse snapping is the most likely cause of the low noise reduction at low unregulated noise (since the unregulated supply droops are only ~30mV in this case) and of the non-monotonicity of the noise reduction curve.

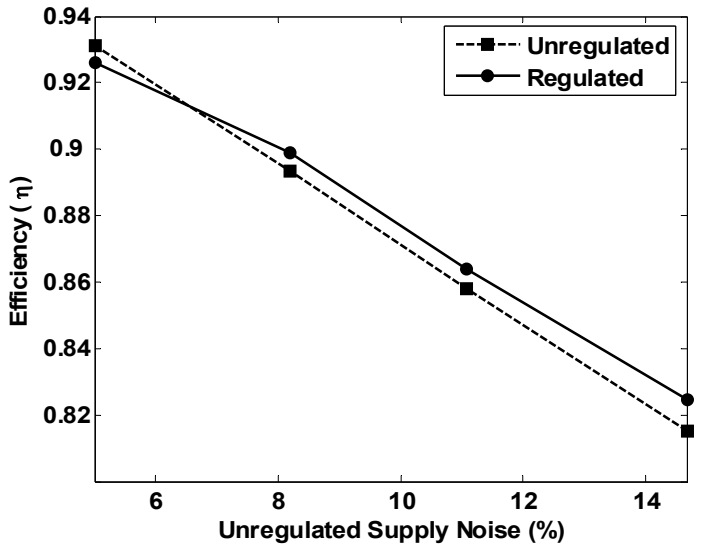


Figure 5.15: Measured efficiencies of unregulated and push-pull shunt regulated systems vs. magnitude of the supply noise without regulation.

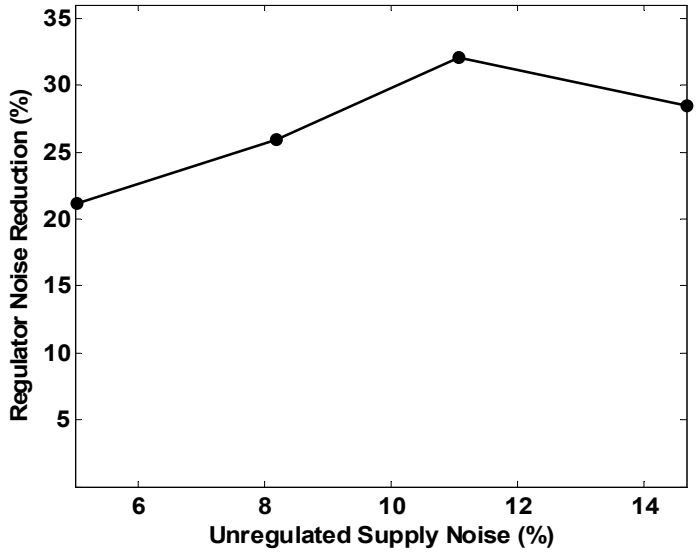


Figure 5.16: Measured noise reduction provided by the push-pull shunt regulator.

5.4 Summary

In order for on-chip regulation of digital power supplies to be effective in modern, power-limited chips, the use of regulation must decrease (or at the worst not adversely impact) the total power dissipation of the chip. Since variations on the supply voltage

force the nominal supply to be increased, reducing supply noise can actually lower the power dissipation of the chip. However, single-supply regulator topologies intrinsically cannot achieve a net reduction in power because of the significant static power they dissipate in their output stages – either because of the required drop-out voltage in a series regulator or the static current draw in a shunt design.

To overcome this limitation, we presented in this chapter the design of a push-pull shunt regulator that makes use of a second power supply to allow the regulator to transiently deliver power – making its power dissipation tied mostly to the average noise current. In order to realize the ideal capability of this topology to reduce the total chip power dissipation, we described some of the key issues in the design of the additional power supply and the regulator itself.

To enable the use of this second power supply for the regulator, resources from the main power supply must be allocated to the additional supply, and we derived a strategy for this resource allocation that minimizes the total loss. In addition, we showed that keeping the regulator's return current flowing through the correct on-chip path requires careful design of the shunt supply's impedance – including the allocation of decoupling capacitance to store energy for this supply.

The design of the regulator itself must be optimized to minimize all sources of static power consumption – both from the output stage and from the feedback path, which can be particularly challenging given the high bandwidth required to maintain a voltage positioned response. The use of a local RC filter-based reference guarantees that the regulator will not inefficiently attempt to correct for quasi-static shifts in the supply voltage. Comparator-based feedback with a dead-band robustly eliminates quiescent current from the output devices, and an input peaking filter exploits the excess gain of the comparators to reduce the effective feedback delay without increasing the feedback circuitry's power consumption. Finally, the regulator's switched source-follower output stage made use of replica biasing and transient current peaking to achieve low turn-on delay with relatively predictable output current.

Measurement results from a fabricated regulator using these techniques confirm that the design can achieve efficiency high enough to simultaneously improve the power dissipation of a digital chip and reduce the noise on its power supply. While issues related to distributing many regulator sites on a die and adapting each of these regulators individually need further exploration, the results presented here show that the push-pull shunt is a promising approach for the regulation of energy-efficient digital circuits.

Chapter 6

Conclusions

As technology scaling and the demand for ever-increasing performance drove CMOS supply voltages down and power consumption up, the impedance required of the supply distribution network dropped dramatically. Since power consumption is no longer rising and supply voltage scaling has drastically slowed down in modern technologies [2], the required supply impedance is fortunately no longer significantly decreasing. However, with the required impedance having already reached roughly 1 mΩ in today's high-performance microprocessors, the integrity of the supply network will remain a significant concern for the foreseeable future.

Given the difficulty of designing such a low impedance supply distribution network, to guarantee the performance or even functionality of modern chips, there is a need to characterize supply noise as it is seen by the circuits on the die during normal operation of the chip. Even though the bandwidth of supply noise can be extremely high and the noise may not be repetitive, by exploiting the fact that supply noise can be fully characterized by extracting only its average properties, this type of measurement can be performed using only compact and simple circuits.

Indeed, it has long been known that if a waveform is repetitive, a single, relatively simple sampler circuit could trace out the waveform with excellent resolution by

averaging together many voltage samples in a sub-sampled manner. In this thesis, we showed that even if the noise is not repetive and exhibits statistical behavior, as long as the properties of the noise do not vary with time (or simply vary repetitively), we need only average together the outputs of two low-rate samplers to find the autocorrelation of the noise and hence recover its spectrum.

Even though the types of sampling circuits used in these measurement systems (which often store analog voltages on capacitors) can be difficult to implement in modern technologies with leaky transistors, the fact that only average properties of the noise need to be measured can once again be exploited to tackle this issue. Specifically, since dynamic errors which are independent of the measured voltage (such as quantization noise when appropriate dither is added) can be eliminated by averaging a sufficient number of samples, low-resolution samplers without any analog storage nodes (and hence no leakage issues) can robustly implement high-resolution measurements of supply noise.

Having created circuits and techniques to characterize the noise on a chip's power supply, we next explored the use of regulation to actively improve the quality of the power supply. Since modern chips are severely power or thermally-limited, regulation wil only be adopted if it is efficient enough that the power a regulator spends in reducing supply noise is less than the power recovered from reducing the margin required for the underlying circuits to maintain a certain level of performance. Regulators that meet this criterion can in fact be built, and therefore circuits that employ regulation can achieve both lower supply noise and lower total power consumption than circuits that do not.

The difficulty of achieving these benefits from regulation is strongly tied to the nature of the circuits being regulated. In sensitive analog or mixed-signal circuits (like PLLs), it may be impossible to achieve the required performance (e.g., magnitude of output jitter) with the expected level of unregulated supply noise. Thus, any regulator that achieves enough noise reduction to make the performance target feasible would clearly be more efficient than an unregulated design. Of course, once regulation is applied the principal concern is to minimize the power needed to achieve the required performance. Since in most chips the power consumption of these types of blocks is

much smaller than the power of the rest of the (mostly digital) chip, a series regulator that isolates the supply of such sensitive circuits from externally-generated noise results in the lowest total power (despite the relatively low efficiency of the series regulator).

Unlike analog or mixed-signal circuits where performance specifications are usually impacted mostly by the magnitude of the supply voltage variations, in synchronous digital circuits it is the minimum supply voltage (over a clock cycle) that sets performance (i.e., operating frequency). This makes achieving the benefits of regulation significantly more challenging, because in this case the nominal value of the supply can simply be raised to maintain the required performance. However, by applying techniques to minimize static power consumption – including the use of a second supply so that the regulator spends power only to counter voltage transients – a push-pull shunt regulator that spends less power reducing noise than the power recovered from operating the digital circuits at a lower nominal supply can be built. This fact not only enables such regulators to simultaneously improve both the effective supply impedance and the energy-efficiency of today's digital chips, but as we will describe in the next section, may further improve the efficiency of future designs with many local supply voltage domains.

6.1 Future Work

In addition to supply noise, there are many other on-chip signals (such as low-jitter clocks or the inputs of high-speed receivers) with high bandwidths and statistical behavior whose analog properties may be of interest. This has driven significant work on integrated measurement and characterization systems to capture the behavior of such on-chip signals [18,26,80,81,82], where many of these systems are based solely on the use of traditional sub-sampling.

As described in this thesis, by extending the averaging performed by a measurement system already based on sub-sampling to include autocorrelation (which only requires a single additional sampler), the system can then measure the input signal's spectrum – including the behavior of any statistical components of the signal. To enable measurement of a broad variety of high-bandwidth signals distributed throughout the

chip, a compact, low input capacitance implementation of each of these samplers is highly desirable. Hence, it may be particularly attractive to once again employ averaging to enable the samplers to be implemented simply as clocked comparators, each with an independent DAC for dither generation (autocorrelation measurement) or threshold shifting (distribution measurement). Averaging can be leveraged even further by applying it to these DACs (for example, by using over-sampled techniques such as ΣΔ modulation) in order to allow them to achieve sufficient resolution while remaining compact and simple to implement. An additional benefit of this DAC-based system is that it would allow the required number of averages to be minimized by adjusting the magnitude of the dither along with the expected magnitude of the input signal.

Returning to the issue of supply integrity, although supply voltage scaling has slowed down, it is clear that future chips will require many independently-controlled local supply voltage domains in order to counter the increasing variability of scaled transistors [2,83,84]. The need for independent supply voltages will lead to heavily split package and board power planes, on-chip series regulators (which may simply be variable-width sleep transistors [85]), or perhaps even integrated switching converters (if, e.g., magnetic materials are added to CMOS processes [86]). No matter which of these approaches is taken, the impedance of the local supply is likely to be negatively impacted, since even in the best case implementation using an integrated switching converter, high efficiency would dictate relatively low switching frequency (as compared to what would be required to counter broad-band current noise).

Given this trend towards local supplies and the benefits of regulation for digital circuits demonstrated in this thesis, a promising avenue of further research is to explore the combined design of a parallel regulator (like the push-pull shunt) with the power conversion element that sets the local supply voltage in these future chips. By using the parallel regulator to control the supply's impedance (perhaps adaptively since the conversion element's output impedance may vary with the output voltage), unlike a system without the parallel regulator, the power conversion element can be optimized principally to minimize its loss. Thus, in these future systems as well, a chip employing

regulation could simultaneously achieve both lower supply noise and lower total power consumption than a chip without regulation.

Appendix A

Replica Loads and Mismatch Analysis

The replica compensated regulator described in Chapter 4 relies on a well-matched, low-overhead replica of the VCO to improve the regulator's performance, and therefore in this appendix we describe potential replica load structures, as well as the impact of mismatch between the replica load and the actual VCO.

A.1 Replica Loads

The choice of replica load depends on the specific VCO topology used in the PLL, and may involve tradeoffs between I-V curve tracking, switching noise generation, and parasitic capacitance. One of the most straightforward options for a replica load is a scaled copy of the main VCO, as shown in Figure A.1a. There are some disadvantages to this approach however; the parasitic capacitance at the supply node of the replica will be relatively large, which can limit the performance of the replica compensated regulator. In addition, the scaled VCO will generate switching noise that due to mismatch may not be aligned in frequency with the real VCO. As shown in Figure A.1b, this issue can be alleviated by using a delay line fed by the VCO clock instead of a scaled VCO, but this replica load will still have high parasitic capacitance.

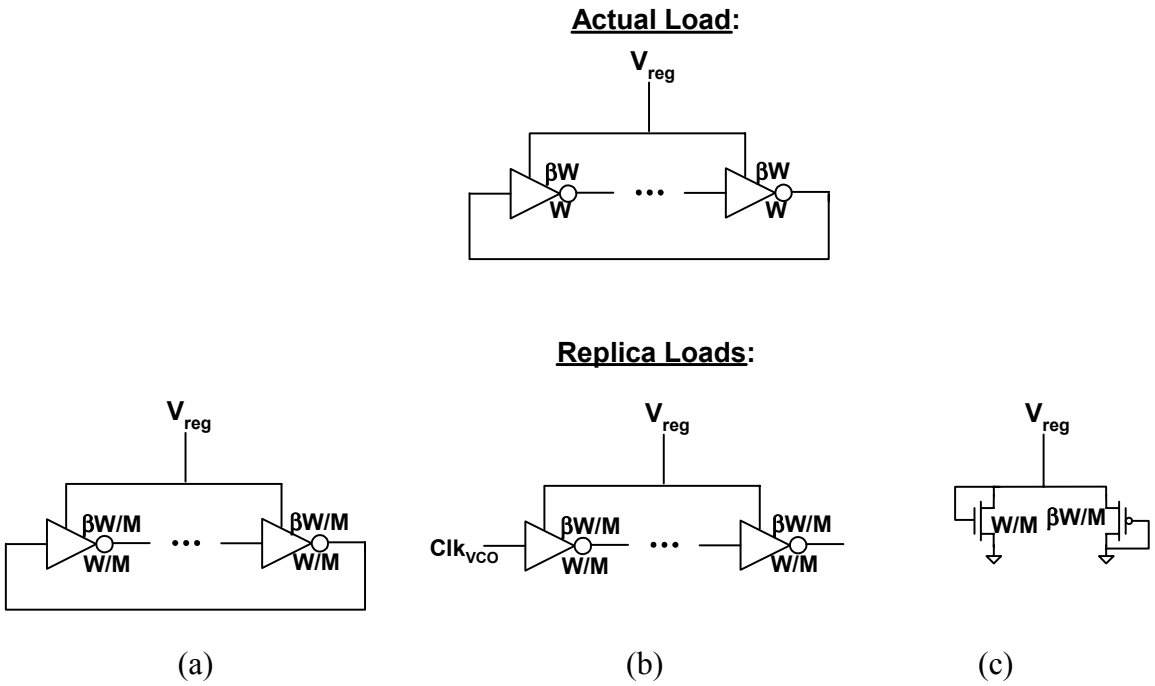


Figure A.1: Replica load options: (a) Scaled copy of the actual VCO. (b) Delay line with the same number of stages as the VCO fed by the VCO clock. (c) Diode-connected devices.

The most desirable replica load is one that does not generate switching noise as well as has a low parasitic capacitance. To find a replica with these characteristics, consider the current drawn by a CMOS inverter-based ring oscillator. Ignoring crowbar and leakage currents:

$$I_{vco} = n_{stages} C_{stage} V_{reg} f_{osc} \tag{A.1}$$

where n_{stages} is the number of stages in the ring and C_{stage} is the effective capacitance driven by each inverter in the oscillator. If the capacitance seen by each inverter is roughly the same for both the rising and falling transitions and ignoring slope effects, the period of the VCO ($T_{osc} = 1/f_{osc}$) is

$$T_{osc} = \frac{n_{stages} C_{stage} V_{reg}}{2 I_{on\_p}} + \frac{n_{stages} C_{stage} V_{reg}}{2 I_{on\_n}} = \frac{n_{stages} C_{stage} V_{reg}}{2 \cdot (I_{on\_p} \| I_{on\_n})} \tag{A.2}$$

where I_{on_p} and I_{on_n} are the saturated drain currents of the PMOS and NMOS transistors in the inverters, respectively. Combining Equation (A.1) and Equation (A.2), the current drawn by the VCO is simply twice the parallel combination of the two on-currents:

$$I_{vco} = 2\left(I_{on\_p} \| I_{on\_n}\right) \tag{A.3}$$

Therefore, if the inverters in the oscillator are sized such that the on currents for both rise and fall transitions are roughly equal (which as shown by A. Hajimiri *et. al.* is desirable for phase noise considerations [87]), the VCO current will equal I_{on} and can be mimicked by appropriately sized diode-connected transistors. As implied by Equation (A.3) and shown in Figure A.1c, a parallel combination of NMOS and PMOS devices is desirable to track variations between the device types.

A diode-connected transistor is an attractive replica load because of its static current and small parasitic capacitance, but it may not match the I-V curves of all VCO topologies well – particularly VCOs with delay interpolating stages (such as [30] and [49]) that draw large amounts of crowbar current (making Equations (A.1) and (A.3) invalid). In these cases, the replica load may need to be augmented by additional structures (e.g. an element which mimics crowbar current) in order to improve its I-V curve tracking.

A.2 Mismatch Analysis

A replica-compensated regulator relies on matching between the replica load and the VCO I-V curves to minimize output voltage offset and to achieve optimal supply rejection; therefore in this section we analyze the effects of mismatch (both random and systematic) on these parameters. The analysis shows that as long as the mismatch is small enough that the small signal characteristics of the real and replica output drivers remain matched, variations in the replica load cause only a voltage offset in the output and a shift in the effective feedback gain k_s. Hence, as long as the regulator is designed with enough margin to withstand small variations in k_s, the effects of mismatch will be relatively benign.

As shown in Figure A.2, mismatch in the replica I-V curve can be split into two components – a static error current (I_{err}) and a change in the small signal output resistance (r_{rep}) of the replica load. We will first describe the voltage offset created by the error current, and then examine the effects of mismatch in r_{rep} on the stability and supply rejection of the regulator.

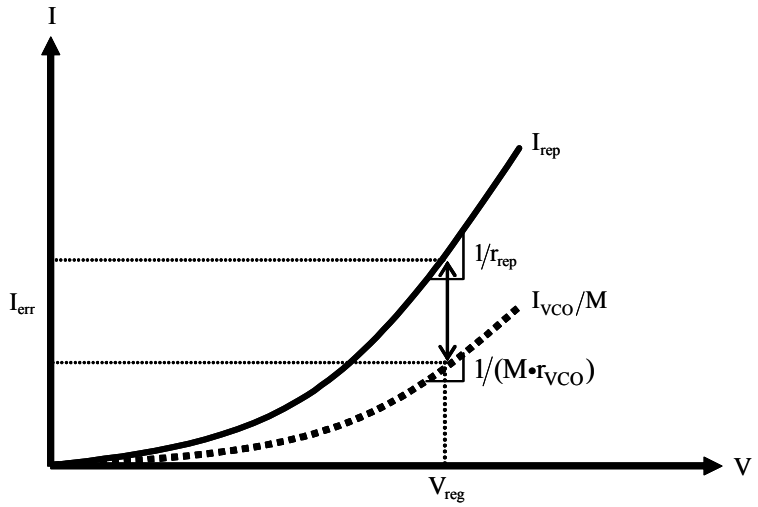


Figure A.2: I-V curves of a VCO (scaled by M) and a replica load showing the offset (I_{err}) and output resistance (r_{rep}) mismatch components.

A.1.1 Effect of Mismatch on Voltage Offset

If I_{err} is small enough that the linearized characteristics of the replica PMOS driver do not significantly deviate from their matched values, we can model the effect of I_{err} by simply converting it into an error voltage (V_{n_rep}) through the small-signal resistance at V_{rep}; the effect of this error voltage on the regulator output has already been shown in Equation (4.12). As k_s is increased a larger percentage of this error voltage is transferred to the regulator output, but for reasonable offsets this is not a large concern because the PLL will adjust the regulator's input voltage to keep the VCO oscillating at the proper frequency.

A.1.2 Effect of Mismatch on Stability

Mismatch in the replica load has two effects on the stability of the overall regulator; the gain of the replica output stage no longer matches that of the main output stage, and the

location of the parasitic pole ω_{rep} is shifted. Since ω_{rep} should be well above the regulator's closed-loop bandwidth, the more dominant effect is the change in the gain of the replica output stage.

If the DC gain of the replica output stage is $A_o+\Delta A_o$, where A_o is the gain of the output stage with perfect matching and ΔA_o is the gain error term, the regulator transfer function becomes

$$\frac{V_{reg}(s)}{V_{in}(s)} = \frac{N_{matched}(s)}{D_{matched}(s) + k_s A_a \Delta A_o (1 + s\tau_o)}, \tag{A.4}$$

where $N_{matched}(s)$ and $D_{matched}(s)$ are the numerator and denominator of the transfer function with perfect matching. The main effect of the additional term is that it adjusts the feedback gain of the replica loop, shifting the entire root locus of Figure 4.12 to the right if $\Delta A_o < 0$ or to the left if $\Delta A_o > 0$. For small k_s values, this is equivalent to changing the gain allocation to k_{eff}, where

$$k_{eff} = \left(1 + \frac{\Delta A_o}{A_o}\right) k_s. \tag{A.5}$$

As long as the nominal k_s is chosen so that small perturbations in its value do not affect the nature of the regulator's closed loop poles (e.g. two real poles become underdamped and leave the real axis), mismatch does not limit the stability of the regulator.

A.1.3 Effect of Mismatch on Supply Rejection

As explained in the previous subsection, mismatch causes the gain of the replica output driver to differ from that of the main output driver. To understand how this affects the supply rejection of the regulator, consider a simple example in which mismatch increases the gain of the replica output driver. The increased gain of the replica loop attenuates the effect of the supply noise on the replica load, decreasing the error signal fed to the amplifier and hence increasing the effect of the noise on the regulator's true output.

Mismatch also causes the supply sensitivity of the replica (S_{Vdd_rep}) to differ from that of the VCO. An increase in the sensitivity of the replica load causes the amplifier to pull the true output in a direction opposite that of the supply noise; with large enough mismatch the sign of the coupling from V_{dd} to V_{reg} can even become negative.

We can formally model these two effects by re-deriving the supply sensitivity transfer function with $A_{o\_rep} = A_o+\Delta A_o$ and $S_{Vdd\_rep} = S_{Vdd}+\Delta S_{Vdd}$,

$$\frac{\Delta V_{reg}(s)}{V_{dd}(s)} = \frac{N_{vdd\_matched}(s) + k_s A_a A_o(\Delta A_o/A_o - \Delta S_{Vdd}/S_{Vdd})}{D_{vdd\_matched}(s) + k_s A_a \Delta A_o(1+s\tau_o)}. \quad (A.6)$$

The effect of mismatch on supply sensitivity can be further simplified if we continue to assume that I_{err} is small enough in magnitude that it does not cause a significant shift in the small signal characteristics of the replica output driver ($g_{m\_rep} = g_m/M$ and $r_{o\_rep} = r_o \cdot M$).[34] In this case, the source of both error terms (change in gain and change in supply sensitivity) is the mismatch in the small signal output resistance of the replica load, r_{rep}.

With this simplifying assumption we can derive the error terms $\Delta S_{Vdd}/S_{Vdd}$ and $\Delta A_o/A_o$ in terms of the output resistance mismatch Δr_{rep} ($r_{rep} = r_{vco} \cdot M+\Delta r_{rep}$) and the small signal parameters of the output driver. Starting from the matched gain $A_o = g_{m\_rep} \cdot (r_{rep} \| r_{o\_rep})$ and supply sensitivity $S_{Vdd} = r_{rep}/(r_{rep}+r_{o\_rep})$, we find that the two error terms are in fact equal to each other.

$$\frac{\Delta A_o}{A_o} = \frac{\Delta S_{Vdd}}{S_{Vdd}} = \left(\frac{\Delta r_{rep}}{r_{rep}}\right) \cdot \left(\frac{r_{o\_rep}}{r_{rep} + \Delta r_{rep} + r_{o\_rep}}\right) \quad (A.7)$$

This result simplifies the analysis of mismatch on supply rejection because the two error terms in the numerator of Equation (A.6) cancel each other and the only remaining perturbation lies in the denominator, which as discussed above can be approximated as a change in the gain allocation k_s.

[34] If the replica and real output devices are well-matched, this assumption could be actively enforced by a slow feedback loop that adjusts I_{rep} until $V_{rep} = V_{reg}$. As long as this feedback loop is slow relative to the regulator's bandwidth, it will not impact the regulator's stability.

Appendix B

Series Regulator Efficiency Optimization

With the increasing integration of sensitive analog and mixed-signal components into power-limited, mostly-digital chips, maximizing the overall efficiency of the series regulators that provide the isolation required for these components is of significant interest. Having described the requirements on amplifier GBW for a given level of rejection in Chapter 4, we build upon those results here to develop a strategy for output device sizing (i.e., overdrive selection) and amplifier power dissipation (i.e., GBW) that maximizes the overall efficiency of a series regulator at a specified rejection. Since they are generally the easiest to integrate (and hence quite common), we will first perform the analysis for a regulator with a common-source output stage. We will then briefly comment on optimizing the efficiency of a design with a source-follower output stage.

In all regulators, the two principal components of power dissipation are the power spent in the output devices, and the power spent in the feedback circuitry. Hence, we can write the regulator's power dissipation as:

$$P_{reg} = V_{od} I_L + (V_{out} + V_{od}) I_{amp}, \tag{B.1}$$

where V_{od} is the overdrive of the output device (which for simplicity we have assumed is exactly equal to the regulator's dropout voltage), I_L is the load current, V_{out} is the desired output (regulated) voltage (which we will assume is fixed, so that the input supply voltage is $V_{out}+V_{od}$), and I_{amp} is the current consumed by the feedback amplifier. For the sake of clarity, in this analysis we will not be including the fact that the input supply itself may vary (forcing the input supply to be margined for the worst-case), although including this effect is a fairly straightforward extension.

In order to optimize this expression, we will next model the dependence of I_{amp} on V_{od}. This involves first calculating the amplifier's required GBW as a function of the output device overdrive, and then establishing a model for the current that must be dissipated by the amplifier to achieve this GBW. The required GBW to achieve a certain minimum rejection LR_{min} was already derived in Chapter 4,[35] and is $2\cdot\omega_o\cdot LR_{min}^2/A_o(V_{od})$ – where we have explicitly shown that the output stage gain A_o depends upon V_{od}. Therefore, we must now establish the model for the dependence of I_{amp} upon this GBW.

For a simple single-stage amplifier (or any amplifier whose response is dominated by a single low-frequency pole), it is straightforward to show that the GBW is set by $g_{m,amp}/C_{load,amp}$ [47]. Assuming a differential-pair based amplifier with a fixed current density in the amplifier's transistors, $g_{m,amp}$ will increase linearly with I_{amp}:

$$g_{m,amp} = (\alpha_a/2) I_{amp}/V_{amp}, \tag{B.2}$$

where V_{amp} is the overdrive of the input pair and α_a is the power-law dependence of the input pair's current on V_{amp}.[36] Assuming that the capacitive loading on the amplifier ($C_{load,amp}$) is dominated by the gate capacitance of the output device (i.e., ignoring any of the amplifier's diffusion capacitance), I_{amp} can therefore be written as

[35] We will only present in detail here the general case where a replica of the load is not available. The use of a replica simply reduces the dependence of GBW on LR_{min} from quadratic to linear, but does not significantly alter the mathematical procedures of the analysis presented here.

[36] Throughout this analysis we will assume an alpha-power law model of the IV characteristics of the transistors [88].

$$I_{amp} = \left(\frac{4}{\alpha_a} \omega_o V_{amp} \right) \frac{C_{go}(V_{od})}{A_o(V_{od})} LR_{min}^{\ 2}, \tag{B.3}$$

where $C_{go}(V_{od})$ is the (overdrive-dependent) gate capacitance of the output device.

To complete the modeling of I_{amp} as a function of V_{od}, we must now find the functional forms of $A_o(V_{od})$ and $C_{go}(V_{od})$. Ignoring the dependence of the power device's output resistance (r_o) on V_{od}, and using the fact that we know that the DC current through the power device is I_L, the output gain is

$$A_o(V_{od}) = \alpha_o I_L (R_{load} \| r_o) / V_{od} = \alpha_o V_L / V_{od}, \tag{B.4}$$

where we have defined V_L as $I_L \cdot (R_{load} \| r_o)$ to simplify the following expressions. We can then write C_{go} as being proportional to the load current:

$$C_{go}(V_{od}) = \frac{I_L}{I_o} C_{g,Io} \frac{V_{od,Io}^{\ \alpha_o}}{V_{od}^{\ \alpha_o}}, \tag{B.5}$$

where I_o is some reference current level, and $C_{g,Io}$ and $V_{od,Io}$ are the gate capacitance and overdrive of a transistor whose drain current is I_o.

We can now combine Equations (B.3), (B.4), and (B.5) to express I_{amp} as a function of V_{od} and a set of technology and design-dependent constants:

$$I_{amp} = \left(\frac{4}{\alpha_o \alpha_a} \frac{V_{amp}}{V_L} \frac{C_{g,Io} \omega_o}{I_o} V_{od,Io}^{\ \alpha_o} \right) \frac{LR_{min}^{\ 2}}{V_{od}^{\ \alpha_o - 1}} I_L = k_{amp} \frac{LR_{min}^{\ 2}}{V_{od}^{\ \alpha_o - 1}} I_L. \tag{B.6}$$

After using Equation (B.6) to re-express P_{reg} and taking the derivative of the resulting expression with respect to V_{od}, it is found that the condition for V_{od} to minimize P_{reg} is:

$$\frac{V_{od}^{\ \alpha_o}}{(1-\alpha_o)V_{out} + (2-\alpha_o)V_{od}} = -k_{amp} LR_{min}^{\ 2}. \tag{B.7}$$

This equation cannot be symbolically solved for any value of α_o, but for the case of a quadratic output device (i.e., $\alpha_o = 2$), this condition simply reduces to:

$$V_{od} = \sqrt{k_{amp} V_{out}} \cdot LR_{min}. \tag{B.8}$$

This linear increase in V_{od} with LR_{min} for a quadratic output device has an intuitive explanation as well; if V_{od} were left constant, the power consumption of the amplifier would have to rise quadratically as LR_{min} is increased. However, by increasing V_{od} linearly at the same time (which reduces A_o linearly, but C_{go} quadratically) the amplifier current need only increase linearly with LR_{min}. As long as V_{od} is significantly smaller than V_{out}, this makes the overall power dissipation of the regulator also scale roughly linearly. In fact, at the optimal V_{od}, P_{reg} scales relative to the load power $P_L = V_{out} \cdot I_L$ as

$$\frac{P_{reg}}{P_L} = \left(2 + \frac{V_{od}}{V_{out}}\right) \cdot \frac{V_{od}}{V_{out}}, \tag{B.9}$$

simplifying to $P_{reg}/P_L \approx 2(V_{od}/V_{out})$ when $V_{out} \gg V_{od}$. Therefore, the regulator's power dissipation can be simply predicted by

$$P_{reg} \approx 2\sqrt{k_{amp}/V_{out}} \cdot LR_{min} \cdot P_L. \tag{B.10}$$

B.1 Source-Follower Output Stage

Since a source-follower intrinsically applies its own g_m against noise at the regulator's output, lowering the source-follower's overdrive not only reduces the dropout of the regulator, but also reduces the required GBW of the amplifier. Because of this, unless the regulator is targeting a very high level of rejection (LR_{min} > ~40), P_{reg} is essentially always minimized by using the smallest V_{od} possible.

Therefore, ignoring any potential area limitations, V_{od} should be set just high enough to guarantee reliable operation (and modeling) of the device (typically, ~100 mV).[37] For LR_{min} less than A_o, no feedback amplifier is required; as the desired rejection is increased, the GBW of the feedback amplifier (and hence I_{amp}) should simply be increased quadratically – i.e., $I_{amp} \alpha (LR_{min} - A_o)^2$.

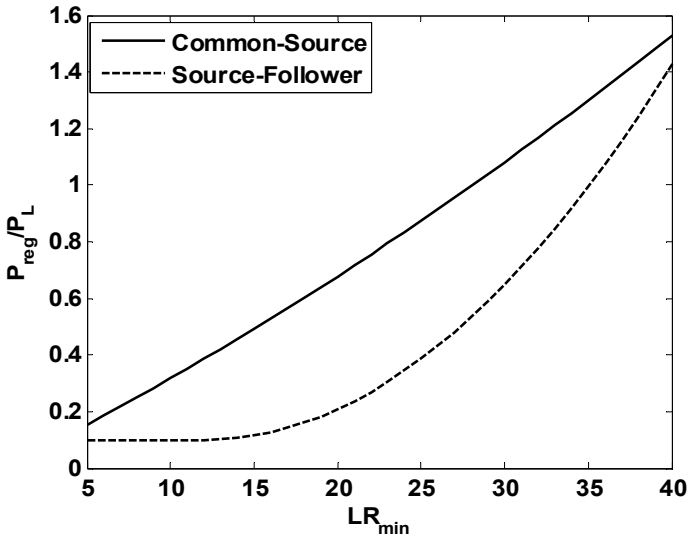


Figure B.1: Comparison of regulator power dissipation for common-source and source-follower output stages.

As a final note, we can use the results of this analysis to compare the power dissipation of regulators based on the two different types of output stages. As shown in Figure B.1, even when the supply of the source-follower regulator's amplifier is statically increased by V_{th} to enable low dropout, the power dissipation of the source-follower regulator is significantly lower than that of the common-source at the same rejection.

[37] Particularly in cutting-edge processes, the subthreshold region is not modeled very accurately. In addition, at overdrives near or in the subthreshold region of operation, the output offset of the regulator will become extremely sensitive to any biasing errors (random or systematic).

Appendix C

Derivative Feedback for Regulators

As described in Chapter 5, additive derivative feedback can be used in a regulator to reduce the phase shift of the feedback path and improve its effective bandwidth. Therefore, in this appendix we derive strategies for designing the derivative filter which minimize the effective impedance of a regulator. As in Chapter 4, we will perform this analysis for both deterministic and random, broad-band current noise. Although the analysis will be performed on regulators with linear feedback amplifiers (where the potential performance improvement of this scheme is limited) to simplify obtaining analytical results, the filter is essentially unchanged for a regulator with equivalent bandwidth comparator-based feedback (the improvement is simply much larger).

C.1 Deterministic Noise

Since the implementations of the derivative filter we are considering do not make use of peaking inductors (and hence cannot truly increase the high-frequency gain), the ratio between the filter's zero and pole frequencies will exactly equal the filter's DC gain reduction. To explicitly take this behavior into account in the design of the filter, we will define a "gain peaking" parameter A_{pk} such that the filter's transfer function $H_d(s)$ is

$$H_d(s) = \frac{1}{A_{pk}} \cdot \frac{1 + s/\omega_d}{1 + s/\left(A_{pk}\omega_d\right)}. \tag{C.1}$$

With this derivative filter at its input, the effective amplifier transfer function $A_{amp}(s)$ becomes:

$$A_{amp}(s) = \frac{A_a}{A_{pk}} \cdot \frac{\left(1 + s/\omega_d\right)}{\left(1 + s/(A_{pk}\omega_d)\right)\left(1 + s/\omega_a\right)}. \tag{C.2}$$

making the regulated impedance $Z_{reg}(s)$:

$$Z_{reg}(s) = \frac{Z_o(s)\left(1+s/\omega_a\right)\left(1 + s/(A_{pk}\omega_d)\right)}{\left(1+s/\omega_a\right)\left(1 + s/(A_{pk}\omega_d)\right) + g_m\left(A_a/A_{pk}\right)\left(1 + s/\omega_d\right)Z_o(s)}. \tag{C.3}$$

Ideally, the analysis would proceed by following the same methodology as that presented in Chapter 4 – i.e., by deriving an expression for the worst-case impedance, and then finding expressions for κ, A_{pk}, and ω_d by minimizing this impedance. Unfortunately, the expressions that would result from attempting to carry this out for Equation (C.3) analytically are extremely difficult to work with.[38] Therefore, we will analyze the derivative filter design with an alternative method that, as we will describe below, builds upon the results from Chapter 4, and is based on insight into the response required of the amplifier to achieve a voltage positioned regulator. Although the following analysis will be approximate, the values derived for κ, A_{pk}, and ω_d are within 5-10% of those found from a numerical optimization with typical regulator and distribution network parameters. More importantly, the worst-case impedance is very shallow near the optimum, and the values derived from the analysis are within this shallow region.

A voltage-positioned response requires there to be minimal frequency domain peaking in the regulated output impedance. In other words, the impedance provided by

[38] For example, even solving for the frequency at which the worst-case impedance occurs involves 8th order polynomials.

an ideal regulator would look purely resistive, with only perhaps a capacitor in parallel to provide further filtering of high frequency noise. From the standpoint of the feedback amplifier (including the derivative filter as part of the amplifier), if the regulator is to behave purely as a resistor in parallel with the rest of the supply network, a step at the amplifier's input should result exactly in a (scaled) step at the amplifier's output. Of course, the feedback bandwidth is limited and hence a true step output can never be achieved – but this indicates that at the same gain, an amplifier whose output closely resembles a step will result in better performance than one whose response does not.

Therefore, in order to find settings for the filter that approximately optimize the regulator's response to deterministic (i.e., step or a worst-case sinusoid) noise current, we will derive the step response of the amplifier as a function of the filter parameters ω_d and A_{pk}, and then choose these parameters by minimizing the mean-squared error of the feedback path's response with respect to an ideal step. This methodology relies on maintaining a constant feedback gain as the filter parameters are varied, and hence the amplifier's allocation between gain and bandwidth will need to vary along with A_{pk}.

Since we know that an A_{pk} of 1 (i.e., no derivative filter) would result in the allocation of Equation (4.3), we will use this allocation here as well, but simply scale the amplifier gain by A_{pk} (making the net gain constant) and the bandwidth by $1/A_{pk}$. While reducing the bandwidth of the amplifier in this manner may at first seem counter-productive, as we will see shortly, the bandwidth boost provided by the derivative filter is actually larger than A_{pk}, making the net result an increased bandwidth at the same gain.

Having fixed the feedback gain to a constant (which can therefore be normalized out), the transfer function we will be optimizing is:

$$H(s) = \frac{\left(1 + s/\omega_d\right)}{\left(1 + s/(A_{pk}\omega_d)\right)\left(1 + s/(\omega_a/A_{pk})\right)}, \tag{C.4}$$

where ω_a here is the amplifier bandwidth with no derivative filter (i.e., ω_a is directly set by Equation (4.3)). The step response of this transfer function is:

$$V_{step}(t) = 1 - \frac{\left(A_{pk} - 1\right)\omega_a}{\left(A_{pk}{}^2\omega_d - \omega_a\right)} e^{-\omega_d A_{pk} t} - \frac{A_{pk}\left(A_{pk}\omega_d - \omega_a\right)\omega_a}{\left(A_{pk}{}^2\omega_d - \omega_a\right)} e^{-\omega_a t / A_{pk}}, \tag{C.5}$$

and the mean-squared error e_{ms} can be found from:

$$e_{ms} = \int_0^\infty \left(1 - V_{step}(t)\right)^2 dt. \tag{C.6}$$

After taking the derivative of e_{ms} with respect to ω_d and A_{pk}, setting both equal to zero and solving the system of two equations, the resulting filter parameters are:

$$\omega_d = \omega_a / A_{pk}{}^2 \qquad A_{pk} = 5/3. \tag{C.7}$$

Before showing the impact of Equation (C.7) on the feedback path's response, it is important to note here that the resulting pole of the filter ($\omega_p = 1/\tau_p = A_{pk}\omega_d$) is set at exactly the amplifier's bandwidth (i.e., $\omega_p = \omega_a/A_{pk}$). Intuitively, this is expected for essentially the same reasons as already described in Chapter 5 – for a fixed ω_d, if ω_p was greater than the amplifier bandwidth the DC gain would be unnecessarily reduced. If ω_p was less than the amplifier bandwidth, the positive phase shift (and hence bandwidth extension) provided by the filter would be reduced. We will make use of the fact that the filter pole should match the amplifier bandwidth to simplify the random noise analysis presented in the next section.

As shown in Figure C.1, despite the initial reduction in amplifier bandwidth necessary to compensate for the reduction in DC gain, the derivative filter design of Equation (C.7) extends the overall bandwidth of the feedback path (typically, by ~15%). However, the DC gain reduction does limit the improvement in worst-case regulated impedance to less than ~5%.

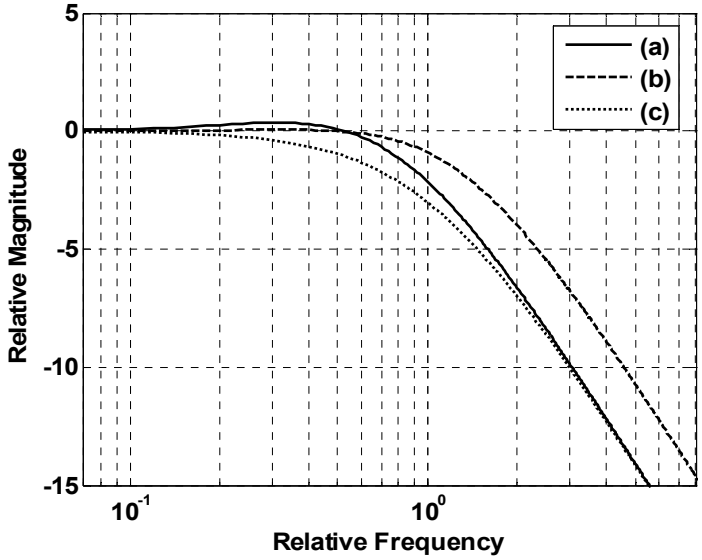


Figure C.1: Magnitude response of the feedback path with a) a derivative filter designed using Equation (C.7), b) an optimized derivative filter without DC gain reduction (A_{pk} = 1.5), and c) the original amplifier (i.e., no derivative filter).

Figure C.1 also shows the response of the feedback path with an optimized derivative filter whose DC gain is not reduced (i.e., the filter actually provides increased gain at high frequency).[39] This response provides an indication of the impact of the derivative filter in a regulator with comparator-based feedback, where the excess gain of the comparator automatically compensates for the reduction in DC gain. In this case, the improvement in feedback bandwidth is ~68%.

C.2 Random Noise

Fortunately, by making use of the result from the previous section that the derivative filter's pole should be identical to that of the amplifier, a direct analysis of the effective impedance of the regulator as a function of A_{pk} and κ can be performed.[40] Specifically, $Z_{reg}(s)$ is now:

[39] The procedure used to optimize this filter is essentially identical to that of the filter with DC gain reduction, but the amplifier gain and bandwidth (A_a and ω_a) no longer need to be altered along with A_{pk}. The resulting filter has an A_{pk} of 1.5.

[40] As in Chapter 4, we will only show the results for the series regulator to facilitate symbolic analysis, but the results for a shunt regulator are essentially identical.

$$Z_{reg}(s) = \frac{Z_o(s)\left(1+s/\omega_a\right)^2}{\left(1+s/\omega_a\right)^2 + g_m\left(A_a/A_{pk}\right)\left(1+s/\left(\omega_a/A_{pk}\right)\right)Z_o(s)}, \tag{C.8}$$

and hence the square of the normalized effective impedance to random, white noise current noise is:

$$\frac{Z_{wn}{}^2}{R_{load}{}^2} = \frac{A_{pk}\left(A_{gbw}{}^2A_o{}^2 + 2A_{pk}\kappa^2(1+\kappa)^2 + A_{gbw}A_o\kappa\left(2+A_{pk}+3\kappa\right)\right)\omega_o}{4\left(A_{gbw}A_o + A_{pk}\kappa\right)\left(2A_{pk}\kappa(1+\kappa)^2 + A_{gbw}A_o\left(A_{pk} - \kappa + 2A_{pk}\kappa\right)\right)}. \tag{C.9}$$

Taking the derivative of Equation (C.9) with respect to A_{pk} and to κ, setting them equal to zero and solving the system of two equations results in:

$$\kappa = \sqrt{\tfrac{1}{3}A_{gbw}A_o} \qquad A_{pk} = 3\cdot\frac{3\kappa+1}{4\kappa+1}. \tag{C.10}$$

Although Equation (C.10) indicates that A_{pk} should be tuned along with the amplifier's GBW, the dependence is very small – A_{pk} asymptotes to 2.25 at infinite GBW, and is less than 2.4 for the typical range of GBW and output gain. It is also interesting to note that Equation (C.10) increases the gain of the amplifier by less than A_{pk} (reducing the DC gain of the feedback path), but the improvement in Z_{reg} near the worst-case frequency reduces the overall white noise impedance by ~1%.

Appendix D

Non-Linear Analyses of the Push-Pull Regulator

In this appendix we will analyze some of the non-linear characteristics of the push-pull regulator in order to glean insight into their impact on the regulator's design. Specifically, we will first examine the limit cycle behavior of the regulator in order to find the dead-band magnitude necessary to eliminate such behavior. Since countering small variations near the average voltage will not significantly impact the worst-case voltage droop, it may be desirable to increase the dead-band beyond the size necessary to eliminate limit cycles in order to improve the efficiency of the regulator. Therefore, we next make use of describing function techniques to provide intuition into how the dead-band should be scaled along with the magnitude of the noise on the supply. Finally, we will briefly present an analysis of the width of the sliding region in a push-pull shunt regulator with derivative feedback.

The basic model of the regulator we will use for these analyses is shown in Figure D.1. The model assumes essentially ideal comparator behavior (but with a fixed delay t_d), with a symmetric dead-band whose width is $2 \cdot \Delta V$ (i.e., the comparator output is zero for $-\Delta V < V_{in} < \Delta V$). $Z(s,v_c)$ is used to represent the fact that with a switched source-follower output stage, the linear impedance at the supply node varies as the comparator turns on and off. In other words, the $1/g_m$ resistance in the model accounts for the source

follower's intrinsic negative feedback (when it is turned on); ΔI accounts for the bias current of the source follower when $V_{reg} = V_{ref}$ (i.e., $M \cdot I_{bias}$ from Figure 5.12a).

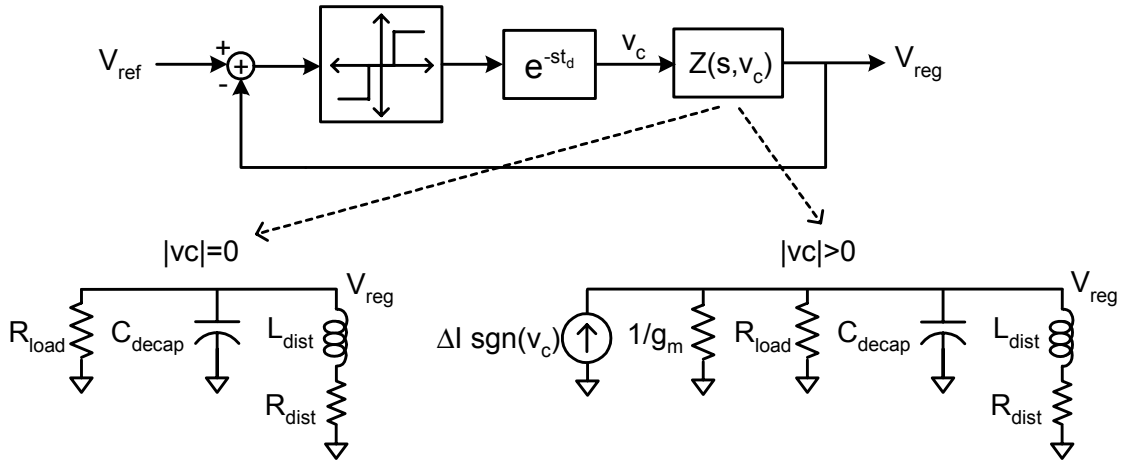


Figure D.1: Model of the push-pull shunt regulator for non-linear analysis.

D.1 Limit Cycle Analysis

While many techniques have been developed for the exact or approximate calculation of limit cycles in comparator-based systems [70,73,89], analyzing the push-pull shunt regulator is somewhat more challenging because of the time-varying nature of the linear portion of the system. Since one of the most common techniques for limit cycle analysis is to make use of describing functions (DFs) to quasi-linearize[41] the non-linearities (and, in this case, the time-varying linear portion of the system), in Section D.1.1 we will show how to formulate a model of the regulator that allows describing functions to be applied. However, this DF analysis is limited to numerical solutions because of the transcendental equations involved.

In our experience, for low-order feedback systems based on comparators with delay (such as the push-pull regulator or band-bang PLLs [49]), it is often more straightforward to derive the limit cycle behavior by directly solving for the time-domain (periodic)

[41] The describing function is defined as the linear system whose output most closely resembles (in the mean-squared error sense) the output from the actual non-linear system for a given form of input. Although the DF is a linear system, its parameters are allowed to depend upon the magnitude of its input – and hence the term "quasi-linearize" [73].

waveforms within the loop. This approach is particularly useful in systems whose linear portion has only a single-pole, or consists of a series of integrations. For the push-pull regulator, the underdamped response of the supply network (caused by the distribution inductance L_{dist}) would once again limit the direct application of such an approach to numerical solutions of transcendental equations. However, in a well-designed regulator (with low feedback delay), the behavior of the loop is mostly set by the impedance of the decoupling capacitance in parallel with the load resistance, and therefore in Section D.1.2 we will perform this type of analysis with the assumption that L_{dist} is infinite.[42]

D.1.1 Describing Function Analysis

In order to calculate a describing function that properly models the intrinsic feedback provided by the source-follower, we will split the output current provided by the regulator into two paths as shown in Figure D.2. With this model, $Z_o(s)$ now simply represents the purely linear impedance at the supply node (i.e. $R_{load}||(1/sC_d)||(sL_{dist}+R_{dist})$).

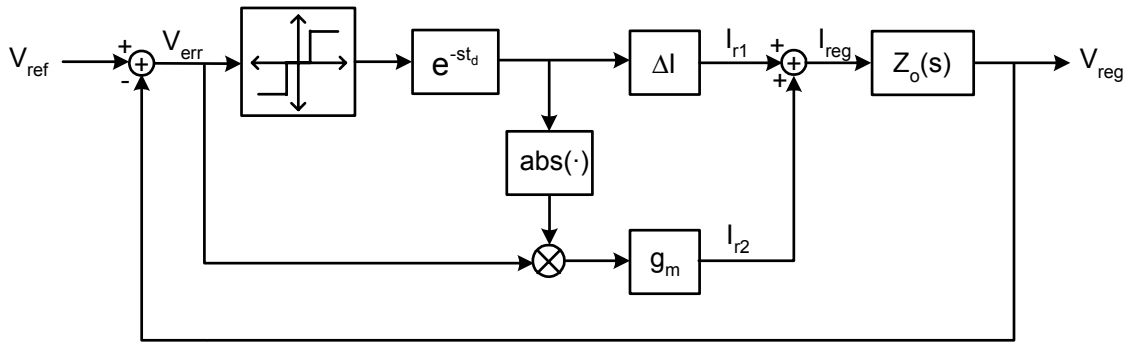


Figure D.2: Model of the push-pull shunt regulator for limit cycle analysis using describing functions.

To find the overall DF from V_{err} to I_{reg}, we can separately calculate the DFs for each of the two current paths and then sum them together to find the overall DF. The path from V_{err} to I_{r1} is through a standard dead-band comparator (with delay), and hence has the well-known DF

[42] To simplify the limit cycle analysis we will not explicitly include the derivative filter. However, the filter can easily be included when calculating limit cycles with the describing function-based technique (since it will already rely on numerical solutions), and the impact of the derivative filter can be approximated in the direct analysis by accounting for the filter's reduction of the comparator delay.

$$N_{I_{rl}}(j\omega) = \frac{4\Delta I}{\pi A_{osc}} \sqrt{1-(\Delta V/A_{osc})^2}\left(\cos(\omega t_d) - j\sin(\omega t_d)\right) \quad (A_{osc} > \Delta V), \tag{D.1}$$

where A_{osc} is the magnitude of the (presumed sinusoidal) limit cycle at the input of the comparator.

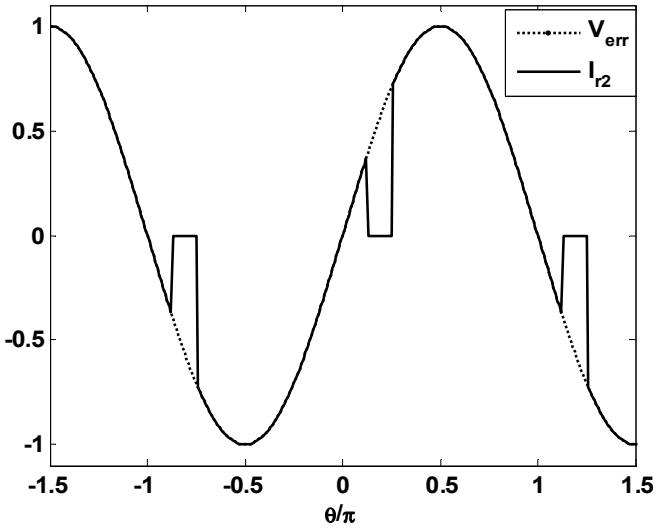


Figure D.3: Example I_{r2} waveform with $V_{err} = \sin\theta$, $g_m = 1$, $\Delta V = 0.2$, and $\omega t_d = \pi/5$.

The non-linearity represented by the path from V_{err} to I_{r2} is not as commonly encountered, and therefore we will present its derivation here. For a general, symmetric non-linearity $f_{nl}(x)$, the sinusoidal-input DF is found from [73]

$$N_{f_{nl}}(j\omega) = \frac{1}{\pi A_{osc}} \int_0^{2\pi} f_{nl}\left(A_{osc}\sin\theta,\omega\right)\cdot\sin\theta \; + j f_{nl}\left(A_{osc}\sin\theta,\omega\right)\cdot\cos\theta \; d\theta. \tag{D.2}$$

To carry out this integral symbolically, it is often instructive to examine the waveforms at the output of the non-linearity for a sinusoidal input. Figure D.3 shows such an example; during most of the period, the output current simply linearly tracks the error voltage. When V_{err} crosses through the origin and its magnitude becomes less than ΔV, the comparator output goes to zero and shuts down the output stage – but because of the comparator delay, this shut down is delayed by ωt_d radians on the waveform.

Since the output current I_{r2} simply tracks V_{err} over the majority of the cycle (such that $f_{nl}(A_{osc}\sin\theta)\cdot\sin\theta = \sin^2\theta$), we can evaluate the integral of Equation (D.2) in a straightforward manner by appropriately selecting the limits of integration:

$$N_{I_{r2}}(j\omega) = \frac{2}{\pi A_{osc}}\left[\int_{-\pi/2}^{\theta_{off}} I_{r,sin}(\theta) + jI_{r,cos}(\theta)\,d\theta + \int_{\theta_{on}}^{\pi/2} I_{r,sin}(\theta) + jI_{r,cos}(\theta)\,d\theta\right] \tag{D.3}$$

$$\theta_{off} = \sin^{-1}(-\Delta V/A_{osc}) + \omega t_d \qquad \theta_{on} = \sin^{-1}(\Delta V/A_{osc}) + \omega t_d$$

$$I_{r,sin}(\theta) = g_m A_{osc}\sin^2\theta \qquad I_{r,cos}(\theta) = g_m A_{osc}\sin\theta\cos\theta$$

Assuming that $A_{osc} > \Delta V$, the result of this integral is:

$$N_{I_{r2}}(j\omega) = g_m\left(N_{I_{r2},I} + jN_{I_{r2},Q}\right) \tag{D.4}$$

$$N_{I_{r2},I} = 1 - \frac{2}{\pi}\sin^{-1}\left(\frac{\Delta V}{A_{osc}}\right) + \frac{2}{\pi}\frac{\Delta V}{A_{osc}}\sqrt{1-\left(\frac{\Delta V}{A_{osc}}\right)^2}\cos(2\omega t_d)$$

$$N_{I_{r2},Q} = -\frac{2}{\pi}\frac{\Delta V}{A_{osc}}\sqrt{1-\left(\frac{\Delta V}{A_{osc}}\right)^2}\sin(2\omega t_d)$$

The overall DF for the feedback path of the regulator N_{Ireg} can then be found by combining Equations (D.1) and (D.4), and the existence (and magnitude/frequency) of limit cycles (for $V_{ref} = 0$) predicted by solving $Z_o(j\omega) = -1/N_{Ireg}(j\omega)$.

D.1.2 Direct Time-Domain Analysis

If we assume that the series inductance of the supply network has a negligible effect on the regulator's limit cycle behavior, it is straightforward to derive time domain equations for the regulated supply voltage. The magnitude of the limit cycle (as a function of the regulator design parameters) can then be calculated by enforcing periodicity, and this will lead us to the magnitude of ΔV necessary to guarantee that the regulator cannot sustain such a limit cycle.

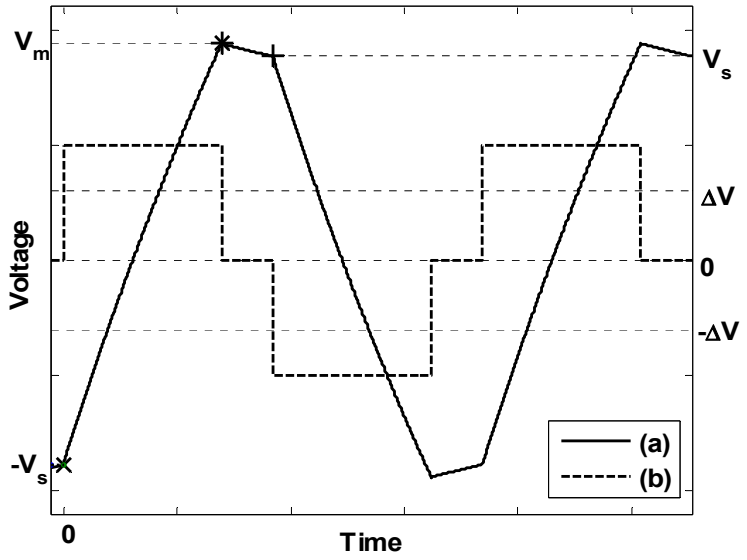


Figure D.4: Example regulator limit cycle waveforms for a) V_{reg} and b) v_c (i.e., the delayed comparator output).

To illustrate the steps we will be taking in the following derivation, Figure D.4 shows example limit cycle waveforms at V_{reg} and v_c. Note that for convenience, we have chosen the time origin to the instant at which the comparator output goes high.

When the comparator output rises, the regulator simultaneously connects the equivalent of a $1/g_m$ resistor to the supply network and pushes ΔI of current into the supply. Since we have assumed infinite L_{dist}, the impedance of the supply network is first-order, and hence the supply voltage has a simple exponential step response that would reach $\Delta V_{IR} = \Delta I \cdot R_{load}/(1+g_m R_{load})$ if the comparator were left on indefinitely.

The magnitude of the limit cycle (V_m) is set by the length of time that the comparator output is high; this time is simply t_d plus the time it takes for the supply voltage to go from $-V_s$ (the voltage when the comparator turns on) to $-\Delta V$. Knowing that the response of the supply voltage is a simple exponential step with a final value of ΔV_{IR}, this time is

$$t_{on} = \ln\left(\frac{\Delta V_{IR} + V_s}{\Delta V_{IR} + \Delta V}\right)\tau_{o,on} + t_d \qquad (D.5)$$

where $\tau_{o,on} = (R_{load}C_{decap})/(1+g_mR_{load})$ is the time constant of the supply network when the source follower is on. Therefore, the limit cycle magnitude is

$$V_m = \left(1-\left(1+\Delta V/\Delta V_{IR}\right)e^{-t_d/\tau_{o,on}}\right)\Delta V_{IR} \tag{D.6}$$

After the supply voltage reaches V_m, the comparator output returns to zero – disconnecting the output stage from the supply node. Therefore, during this time V_{reg} simply exponentially decays from V_m with a time constant of $\tau_{o,off} = (R_{load}C_{decap})$. Knowing this behavior, to complete the half-cycle we need only calculate V_s (the voltage at the instant before the comparator output becomes negative). The amount of time the comparator output stays at zero is set by the time it takes for V_{reg} to go from $-\Delta V$ to ΔV (while the comparator output was high):

$$t_{off} = \ln\left(\frac{\Delta V_{IR}+\Delta V}{\Delta V_{IR}-\Delta V}\right)\tau_{o,on} \tag{D.7}$$

and hence V_s is:

$$V_s = \left(\frac{1-\Delta V/\Delta V_{IR}}{1+\Delta V/\Delta V_{IR}}\right)^{1/(1+A_o)} V_m \tag{D.8}$$

where we have used the fact that $\tau_{o,on}/\tau_{o,off} = 1/(1+g_mR_{load})$ and defined $A_o = g_mR_{load}$.

Having completed characterizing the limit cycle, we can now use these equations to find the ΔV necessary to make these equations inconsistent (i.e., so that no limit cycle can exist). Clearly, a limit cycle could not exist if $V_m < \Delta V$; in other words:

$$\frac{\Delta V_{min}}{\Delta V_{IR}} = \frac{1-e^{-t_d/\tau_{o,on}}}{1+e^{-t_d/\tau_{o,on}}} \tag{D.9}$$

This ΔV_{min} is actually somewhat pessimistic, because it can be shown that a limit cycle cannot exist even if $V_s < \Delta V$:

$$\frac{\Delta V_{min}}{\Delta V_{IR}} = \left(\frac{1 - \Delta V_{min}/\Delta V_{IR}}{1 + \Delta V_{min}/\Delta V_{IR}} \right)^{1/(1+A_o)} \left(1 - \left(1 + \Delta V_{min}/\Delta V_{IR} \right) e^{-t_d/\tau_{o,on}} \right) \tag{D.10}$$

Equation (D.10) must be solved numerically to find ΔV_{min}, but as shown in Figure D.5, the ΔV_{min} found from Equation (D.9) is a very close approximation for most of the range of interest. Furthermore, simulations indicate that these bounds (calculated with infinite L_{dist}) are pessimistic and within 10% of the ΔV_{min} for typical L_{dist} values.

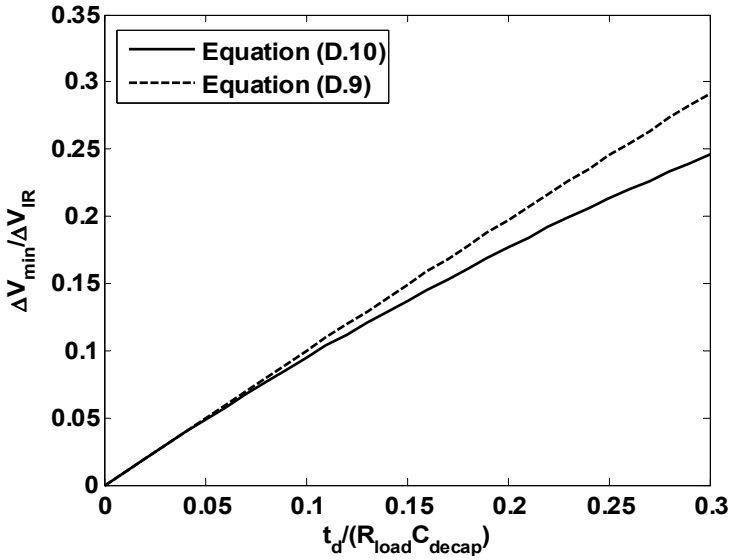


Figure D.5: Example of ΔV_{min} vs. t_d calculated using Equations (D.10) and (D.9). The minimum dead-band and comparator delay are normalized by ΔV_{IR} and $\tau_{o,off} = R_{load}C_{decap}$ respectively. For this example, $g_m = 1/R_{load}$.

D.2 Dead-Band Scaling with Supply Voltage Noise

As mentioned in Chapter 5, eliminating limit cycles is not the only consideration in setting the size of the dead-band. Specifically, since performance in a synchronous digital system is set by the worst-case supply voltage, a push-pull regulator with a dead-band which is too small may spend power unnecessarily to remove small variations which do not significantly impact the worst-case voltage.

The exact relationship between the magnitude of the dead-band and the supply voltage noise magnitude strongly depends upon the distribution of the current noise, and may not be straightforward to calculate analytically. Therefore, our goal in this section will not be to derive the precise relationship between ΔV and the magnitude of the noise. Rather, we will present an approximate analysis to provide further intuition behind the general rule that ΔV should be scaled to remain roughly constant relative to the standard deviation of the noise.

In Section D.1.1, we derived the sinusoidal-input describing function of the regulator's feedback path, and we saw that the terms of the DF that included ΔV were always scaled by A_{osc} (i.e., there were no terms where ΔV appeared alone – only $\Delta V/A_{osc}$). This fact alone already points to keeping ΔV fixed relative to the output noise, but the current noise profiles of interest are likely to be more random in nature (rather than a pure sinusoid). Therefore, we will next briefly derive an approximate DF for the regulator's feedback path in response to random, Gaussian noise on V_{reg}.

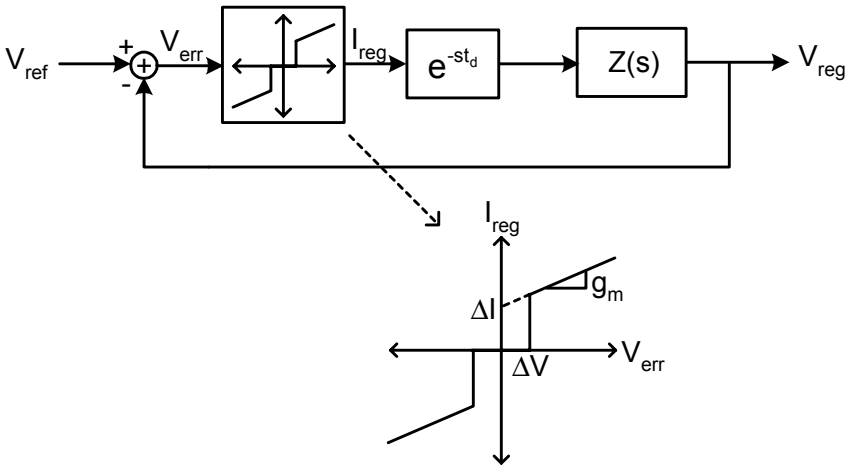


Figure D.6: Simplified model of the push-pull regulator enabled by artificially delaying the intrinsic feedback of the source-follower by t_d.

Since the goal here is to provide intuition (and not an exact analysis), to avoid calculations that would involve the autocorrelation of the noise on the supply voltage, we will calculate the feedback path's DF with the intrinsic feedback of the SF (artificially) delayed by the same amount as the comparator output (i.e., I_{r2} from Figure D.2 is delayed by t_d). This simplification allows us to model the regulator as shown in Figure D.6,

where the non-linearity (which combines the linear response from the source-follower's g_m and the bias current ΔI) is now memoryless (and hence will have no phase shift associated with it).

For a general, symmetric, memoryless non-linearity $f_{nl}(x)$, the random-input DF is a scalar (i.e., simply a gain term), and is found from [73]

$$N_{f_{nl}} = E\left[f_{nl}(x)\cdot x\right]\Big/\sigma_x{}^2 \tag{D.11}$$

Therefore, for the push-pull regulator with zero-mean Gaussian noise whose standard deviation is σ_{vr} on V_{reg}, the DF is found from:

$$N_{I_{lreg}} = \frac{2}{\sigma_{vr}{}^2}\int_{\Delta V}^{\infty}\left(g_m x + \Delta I\right)\cdot x\cdot\frac{1}{\sigma_{vr}\sqrt{2\pi}}e^{-x^2/\left(2\sigma_{vr}{}^2\right)}dx \tag{D.12}$$

Defining $k_{db} = \Delta V/\sigma_{vr}$, Equation (D.12) evaluates to

$$N_{I_{lreg}} = \sqrt{\tfrac{2}{\pi}}e^{-k_{db}{}^2/2}\left(\Delta I/\sigma_{vr}\right) + \left(1 + \sqrt{\tfrac{2}{\pi}}k_{db}e^{-k_{db}{}^2/2} - \mathrm{erf}\left(\tfrac{1}{\sqrt{2}}k_{db}\right)\right)g_m \tag{D.13}$$

Hence, if k_{db} is fixed (i.e., ΔV scales directly with σ_{vr}) and ΔI scales along with σ_{vr} (which is a reasonable rule to follow since an increase in σ_{vr} would correspond to an increase in the noise current), the gain of the regulator's feedback path will remain constant.

Finally, to provide some insight into the choice of the relative magnitude of the dead-band (and how each of the gain terms is affected by this choice), Figure D.7 shows how the two components of N_{Ireg} (i.e., the gain from ΔI, and the gain from g_m) scale with k_{db}. The ΔI-contributed gain remains above 90% of its maximum value for k_{db} less than ~0.46, while the g_m-contributed gain remains above 90% of its maximum value for k_{db} less than ~0.76.

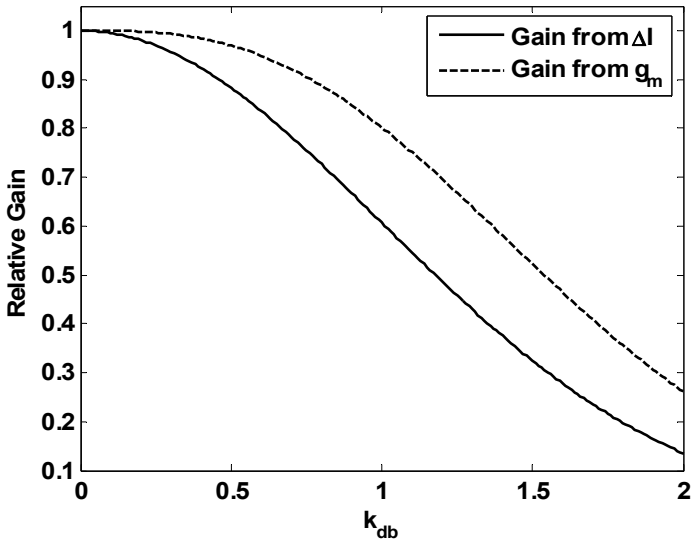


Figure D.7: Scaling of the two components of the feedback gain (N_{Ireg}) with the relative magnitude of the dead-band (k_{db}).

D.3 Sliding Mode Analysis

Since it does not directly affect the boundaries of sliding mode operation, in the following analysis we will ignore the comparator's dead-band and delay (i.e., we will assume $\Delta V=0$). Therefore, either the push or pull output stage will always be connected to the supply node, allowing us to simply treat the source-follower's intrinsic feedback as a resistance of $1/g_m$ in parallel with R_{load}. As in Chapter 5, we define the zero of the derivative filter as ω_d, and the pole of the filter as ω_p.

To guarantee sliding mode operation (i.e., that the output trajectory "slides" along the surface defined by the feedback filter), the following condition must be satisfied [70]:

$$\lim_{S\to 0} S\cdot(dS/dt) < 0 \qquad \text{(D.14)}$$

where S is the sliding surface defined by the feedback filter (e.g., for a feedback filter of the form $(1+s/\omega_d)$, $S = (1/\omega_d)\cdot dV_{out}/dt + V_{out}$). To translate Equation (D.14) into boundaries on the state variables of the system, the system's dynamics are used to define S and dS/dt in terms of these variables.

Choosing V_{reg} (the output voltage[43]) and I_{Ldist} (the current through L_{dist}) as our state variables, the dynamics of the push-pull regulator can be written as:

$$\begin{aligned} dV_{reg}/dt &= \left(-\Delta V_{IR}\,\mathrm{sgn}(S) - V_{reg} - I_{Ldist}R_{load,eff}\right)/\tau_{o,on} \\ dI_{Ldist}/dt &= \left(R_{dist}/L_{dist}\right)\left(V_{reg}/R_{dist} - I_{Ldist}\right), \end{aligned} \tag{D.15}$$

where ΔV_{IR} and $\tau_{o,on}$ are as defined in Section D.1.2, and $R_{load,eff} = R_{load}/(1+ g_m R_{load})$. With a single-pole, single-zero feedback filter, the dynamics of the sliding surface are:

$$S + \left(1/\omega_p\right) dS/dt = V_{reg} + \left(1/\omega_d\right) dV_{reg}/dt. \tag{D.16}$$

Hence, we can write the condition of Equation (D.14) as:

$$\lim_{S\to 0} S \cdot \omega_p \left(V_{reg} + \left(1/\omega_d\right) dV_{reg}/dt - S\right) < 0 \tag{D.17}$$

Using Equation (D.15) to expand dV_{reg}/dt, and making use of the fact that the limit must hold as S approaches zero, Equation (D.17) leads to the following two boundaries on the regulator's sliding region:

$$\begin{aligned} I_{Ldist} &< \Delta I + \left(\omega_d \tau_{o,on} - 1\right)\frac{V_{reg}}{R_{load,eff}} \qquad (S < 0) \\ I_{Ldist} &> -\Delta I + \left(\omega_d \tau_{o,on} - 1\right)\frac{V_{reg}}{R_{load,eff}} \qquad (S > 0) \end{aligned} \tag{D.18}$$

The key point to notice from Equation (D.18) is that for a given derivative filter, the size of the sliding region is set by ΔI. However, since load current noise also directly drives the supply network, this current noise is effectively added to ΔI. Hence, unless ΔI is significantly larger than the magnitude of the expected noise current (which is typically

[43] It is important to note that V_{reg} is assumed to have a mean of zero in this analysis; the actual value of the supply voltage is unimportant in this analysis as long as V_{ref} is correctly generated.

not the case in a regulator optimized to maximize the efficiency of the overall chip), the regulator will not remain within the sliding region for any significant period of time.

Bibliography

[1] R. H. Dennard *et al.*, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, Oct. 1974, pp. 256-268.

[2] M. Horowitz, E. Alon, D. Patil, S. Naffziger, R. Kumar, and K. Bernstein, "Scaling, Power, and the Future of CMOS," *IEEE International Electron Devices Meeting*, Dec. 2005.

[3] W. S. Song and L. A. Glasser, "Power Distribution Techniques for VLSI Circuits," *IEEE Journal of Solid-State Circuits*, vol. 21, no. 1, Feb. 1986, pp. 150-156.

[4] B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley, 1990.

[5] L. A. Arledge, Jr. and W. T. Lynch, "Scaling and Performance Implications for Power Supply and Other Signal Routing Constraints Imposed by I/O Limitations," *IEEE Symposium on IC/Package Design Integration*, Feb. 1998.

[6] P. Larsson, "Power supply noise in future IC's: a crystal ball reading," *IEEE Custom Integrated Circuits* Conference, May 1999.

[7] S. R. Nassif and O. Fakhouri, "Technology Trends in Power-Grid-Induced Noise," *ACM International Workshop on System Level Interconnect Prediction*, Apr. 2002.

[8] A. V. Mezhiba and E. G. Friedman, "Scaling Trends of On-Chip Power Distribution Noise," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 12, no. 4, Apr. 2004, pp. 386-394.

[9] R. Schmitt and C. Yuan, "Power Distribution Design Considerations and Methodology for Multi-Gigabit I/Os," *IEEE International Symposium on Electromagnetic Compatibility*, Aug. 2005.

[10] J. Darnauer *et al*., "Electrical Evaluation of Flip-Chip Package Alternatives for Next-Generation Microprocessors," *IEEE Transactions on Advanced Packaging*, vol. 22, no. 5, Aug. 1999, pp. 407-415.

[11] P. E. Gronowski, W. J. Bowhill, R. P. Preston, M. K. Gowan, and R. L. Allmon, "High-Performance Microprocessor Design," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 5, May 1998, pp. 676-686.

[12] Q. K. Zhu and D. Ayers, "Power Grid Planning for Microprocessors and SoCs," *IEEE International Symposium on Quality of Electronic Design*, Mar. 2005.

[13] A. Krstić and K.-T. Cheng, "Vector Generation for Maximum Instantaneous Current Through Supply Lines for CMOS Circuits," *IEEE Design Automation Conference*, Jun. 1997.

[14] H. H. Chen and J. S. Neely, "Interconnect and Circuit Modeling Techniques for Full-Chip Power Supply Noise Analysis," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B*, vol. 21, no. 3, Aug. 1998, pp. 209-215.

[15] S.R. Nassif and J.N. Kozhaya, "Fast power grid simulation," *IEEE Design Automation Conference*, Jun. 2000.

[16] G. Steele, D. Overhauser, S. Rochel, and S. Z. Hussain, "Full-Chip Verification Methods for DSM Power Distribution Systems," *IEEE Design Automation Conference*, Jun. 1998.

[17] S. Lin *et al.*, "Full-Chip Vectorless Dynamic Power Supply Integrity Analysis and Verification Against 100μV/100ps-Resolution Measurement," *IEEE Custom Integrated Circuits Conference*, Oct. 2004.

[18] M. Takamiya, M. Mizuno, and K. Nakamura, "An On-chip 100GHz-Sampling Rate 8-channel Sampling Oscilloscope with Embedded Sampling Clock Generator," *IEEE International Solid State Circuits Conference*, Feb. 2002.

[19] A. Muhtaroglu, G. Taylor, and T. Rahal-Arabi, "On-Die Droop Detector for Analog Sensing of Power Supply Noise," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 651-660, Apr. 2004.

[20] K. Shimazaki, M. Nagata, T. Okumoto, S. Hirano, and H. Tsujikawa, "Dynamic Power-Supply and Well Noise Measurement and Analysis for High Frequency Body-Biased Circuits," *IEEE Symposium on VLSI Circuits*, Jun. 2004.

[21] T. Okumoto, M. Nagata, and K. Taki, "A Built-In Technique for Probing Power-Supply Noise Distribution Within Large-Scale Digital Integrated Circuits," *IEEE Symposium on VLSI Circuits*, Jun. 2004.

[22] K. Poulton *et al.*, "A 20GS/s 8b ADC with a 1MB Memory in 0.18μm CMOS," *IEEE International Solid-State Circuits Conference*, Feb. 2003.

[23] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley Pub. Co.; 2nd edition, July 1993.

[24] B.J. Skinner, F.M. Ingels, and J.P. Donohoe, "Stationary and cyclostationary random process models," in *Proceedings of the IEEE Southeastcon '94*, pp. 450-454, Apr. 1994.

[25] E. Alon, V. Stojanović, and M. Horowitz, "Circuits and Techniques for High-Resolution Measurement of On-Chip Power Supply Noise," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 4, Apr. 2005, pp. 820-828.

[26] R. Ho *et al.*, "Applications of on-chip samplers for test and measurement of integrated circuits," *IEEE Symposium on VLSI Circuits*, Jun. 1998.

[27] V. von Kaenel *et al*, "A 320MHz, 1.5 mW @ 1.35 V CMOS PLL for Microprocessor Clock Generation," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 11, Nov. 1996, pp. 1715-1722.

[28] R. Naiknaware and T. Fiez, "Time-Referenced Single-Path Multi-Bit ADC using a VCO-Based Quantizer," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 2, May 1999, pp. 33–36.

[29] A. Iwata, N. Sakimura, M. Nagata, and T. Morie, "The Architecture of Delta Sigma Analog-to-Digital Converters Using a Voltage-Controlled Oscillator as a Multibit Quantizer," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 46, no. 7, Jul. 1999, pp. 941–945.

[30] K.-Y.K. Chang *et al.*, "A 0.4-4 Gb/s CMOS Quad Transceiver Cell using On-Chip Regulated Dual-Loop PLLs," *IEEE Journal of Solid-State Circuits*, vol. 38, May 2003, pp. 747-754.

[31] V. Abramzon, E. Alon, B. Nezamfar, and M. Horowitz, "Scalable Circuits for Supply Noise Measurement," *IEEE European Solid-State Circuits Conference*, Sept. 2005.

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

[33] V. Stojanović *et al.*, "Adaptive Equalization and Data Recovery in Dual-Mode (PAM2/4) Serial Link Transceiver," *IEEE Symposium on VLSI Circuits*, June 2004.

[34] W. Gardner, "Stationarizable Random Processes," *IEEE Transactions on Information Theory*, vol. 24, no. 1, Jan. 1978, pp. 8-22.

[35] T. Ghani *et al.*, "Scaling Challenges and Device Design Requirements for High-Performance Sub-50 nm Gate Length Planar CMOS Transistors," *IEEE Symposium on VLSI Technology*, Jun. 2000.

[36] T. Fischer *et al.*, "A 90nm Variable Frequency Clock System for a Power-Managed Itanium® Architecture Processor," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 1, Jan. 2006, pp. 218-228.

[37] R. McGowen *et al.*, "Power and Temperature Control on a 90-nm Itanium Family Processor," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 1, Jan. 2006, pp. 229-237.

[38] Semiconductor Industry Association (SIA), *International Technology Roadmap for Semiconductors 2003*, Austin, TX, 2003.

[39] A. V. Oppenheim and A. S. Willsky, with S.H. Nawab, *Signals and Systems, Second Edition,* Prentice-Hall Inc., 1996.

[40] A. M. Abo and P. R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS Pipeline Analog-to-Digital Converter," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, May 1999, pp. 599-606.

[41] R. Suarez, P. Gray, and D. Hodges, "All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques – Part II," *IEEE Journal of Solid-State Circuits*, vol. 10, no. 6, Dec. 1975, pp. 379-385.

[42] R. M. Gray and T. G. Stockham Jr., "Dithered quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, May 1993, pp. 805-812.

[43] M. F. Wagdy and M Goff, "Linearizing Average Transfer Characteristics of Ideal ADC's via Analog and Digital Dither," *IEEE Transactions on Instrumentation and Measurement*, vol. 43, no. 2, Apr. 1994, pp. 146-150.

[44] K. Chang *et al.*, "Clocking and Circuit Design for a Parallel I/O on a First Generation CELL Processor," *IEEE International Solid-State Circuits Conference*, Feb. 2005.

[45] K. Nose and T. Sakurai, "Optimization of V_{DD} and V_{TH} for Low-Power and High-Speed Applications," *IEEE Design Automation Conference*, Jan. 2000.

[46] J. Lee, G. Hatcher, L. Vandenbergh, and C.-K. K. Yang, "Evaluation of Fully-Integrated Switching Regulators for CMOS Process Technologies," *IEEE International Symposium on System-on-Chip*, Nov. 2003.

[47] B. Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw Hill, 2000.

[48] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[49] J. Kim, "Design of CMOS Adaptive-Supply Serial Links," *PhD. Thesis*, Stanford University, Dec. 2002.

[50] M.-J. E. Lee *et. al*, "Low-Power Area-Efficient High-Speed I/O Circuit Techniques," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 11, Nov. 2000, pp. 1591-1599.

[51] R. Redl, B. P. Erisman, and Z. Zansky, "Optimizing the Load Transient Response of the Buck Converter," *IEEE Applied Power Electronics Conf. Expo. (APEC)*, Feb. 1998.

[52] A. Waizman and C.-Y. Chung, "Extended Adaptive Voltage Positioning (EAVP)," *IEEE Conference on Electrical Performance of Electronic Packaging*, Oct. 2000.

[53] P. Hazucha, T. Karnik, B. A. Bloechel, C. Parsons, D. Finan, and S. Borkar, "Area-Efficient Linear Regulator with Ultra-Fast Load Regulation," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 4, Apr. 2005, pp. 933-940.

[54] H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms*, vol. 25 of *MIT Radiation Laboratory Series*. McGraw-Hill, New York, 1947.

[55] G. W. den Besten and B. Nauta, "Embedded 5V-to-3.3V Voltage Regulator for Supplying Digital IC's in 3.3V Technology," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, July 1998, pp. 956-962.

[56] H. Werker *et al.*, "A 10 Gb/s SONET-Compliant CMOS Transceiver with Low Crosstalk and Intrinsic Jitter," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 12, Dec. 2004, pp. 2349-2358.

[57] S. Sidiropoulos *et al,* "Adaptive Bandwidth DLLs and PLLs using Regulated Supply CMOS Buffers," *IEEE Symposium on VLSI Circuits*, June 2000.

[58] J. G. Maneatis, "Low-Jitter Process-Independent DLL and PLL Based on Self-Biased Techniques," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 11, Nov. 1996, pp. 1723-1732.

[59] J. G. Maneatis, J. Kim, I. McClatchie, J. Maxie, and M. Shankaradas, "Self-Biased High-Bandwidth Low-Jitter 1-to-4096 Multiplier Clock Generator PLL," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 11, Nov. 2003, pp. 1795-1803.

[60] R. Farjad-Rad *et al.*, "A Low-Power Multiplying DLL for Low-Jitter Multigigahertz Clock Generation in Highly Integrated Digital Chips," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 12, Dec. 2002, pp. 1804-1812.

[61] Y.-S. Sohn *et al.*, "A 512Mbit, 3.2Gbps/pin Packet-Based DRAM with Cost-Efficient Clock Generation and Distribution Scheme," *IEEE Symposium on VLSI Circuits*, Jun. 2004.

[62] E. Alon, J. Kim, S. Pamarti, K. Chang, and M. Horowitz, "Replica Compensated Linear Regulators for Supply-Regulated Phase-Locked Loops," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 2, Feb. 2006, pp. 413-424.

[63] M. Mansuri and C.-K. K. Yang, "Jitter Optimization Based on Phase-Locked Loop Design Parameters," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, Nov. 2002, pp. 1375-1382.

[64] M. Ang, R. Salem, and A. Taylor, "An On-Chip Voltage Regulator Using Switched-Decoupling Capacitors," *IEEE International Solid-State Circuits Conference*, Feb. 2000.

[65] J. Gu, R. Harjani, and C. Kim, "Distributed Active Decoupling Capacitors for On-Chip Supply Noise Cancellation in Digital VLSI Circuits," *IEEE Symposium on VLSI Circuits*, Jun. 2006.

[66] A. M. Wu and S. R. Sanders, "An Active Clamp Circuit for Voltage Regulation Module (VRM) Applications," *IEEE Transactions on Power Electronics*, vol. 16, no. 5, Sept. 2001, pp. 623-634.

[67] F. N. K. Poon, C. K. Tse, and J. C. P. Liu, "Very Fast Transient Voltage Regulators Based on Load Correction," *IEEE Power Electronics Specialists Conference*, Jun. 1999.

[68] Intersil Corporation, "Data Sheet for HIP6200 and HIP6201," Tech. Rep., Feb. 1998.

[69] "Advanced Configuration and Power Interface Specification," Hewlett-Packard Corp., Intel Corp., Microsoft Corp., Phoenix Technologies Ltd., and Toshiba Corp., http://www.acpi.info/DOWNLOADS/ACPIspec30.pdf, Sept. 2004.

[70] S. S. Sastry. *Nonlinear Systems: Analysis, Stability, and Control*, Springer-Verlag, 1999.

[71] P. Hazucha *et al.*, "A Linear Regulator with Fast Digital Control for Biasing Integrated DC-DC Converters," *IEEE International Solid-State Circuits Conference*, Feb. 2006.

[72] S. R. Norsworthy, R. Schreir, and G. C. Themes, *Delta-Sigma Data Converters: Theory, Design, and Simulation*, Wiley-IEEE Press, 1996.

[73] A. Gelb and W. E. Vander Velde, *Multiple-Input Describing Functions and Non-Linear System Design*, McGraw-Hill, 1968.

[74] B. Casper *et al.*, "An 8-Gb/s Simultaneous Bidirectional Link with On-Die Waveform Capture," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 12, Dec. 2003, pp. 2111-2120.

[75] R. Farjad-Rad *et al.*, "0.622-8.0 Gbps 150mW Serial IO Macrocell with Fully Flexible Preemphasis and Equalization," *IEEE Symposium on VLSI Circuits*, Jun. 2003.

[76] Y. Tomita *et al.*, "A 10Gb/s Receiver with Equalizer and On-Chip ISI Monitor in 0.11μm CMOS," *IEEE Symposium on VLSI Circuits*, Jun. 2004.

[77] R. Venkataramanan, A. Sabanovic, and S. Ćuk, "Sliding Mode Control of DC-to-DC converters," *IEEE Conf. Industrial Electronics, Control Instrumentations (IECON)*, 1985.

[78] J. Kim and M. Horowitz, "An Efficient Digital Sliding Controller for Adaptive Power-Supply Regulation," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 5, May 2002, pp. 639-647.

[79] S.-C. Tan, Y. M. Lai, M. K. H. Cheung, and C. K. Tse, "On the Practical Design of a Sliding Mode Voltage Controller Buck Converter," *IEEE Transactions on Power Electronics*, vol. 20, no. 2, Mar. 2005, pp. 425-437.

[80] Y. Zheng and K. L. Shepard, "On-Chip Oscilloscopes for Noninvasive Time-Domain Measurement of Waveforms in Digital Integrated Circuits," *IEEE Transcations on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 3, Jun. 2003, pp. 336-344.

[81] M. Takamiya and M. Mizuno, "A Sampling Oscilloscope Macro Toward Feedback Physical Design Methodology," *IEEE Symposium on VLSI Circuits*, Jun. 2004.

[82] V. Petrescu, M. Pelgrom, H. Veendrick, P. Pavithran, and J. Wieling, "A Signal-Integrity Self-Test Concept for Debugging Nanometer CMOS ICs," *IEEE International Solid-State Circuits Conference*, Feb. 2006.

[83] T.-C. Chen, "Where CMOS is Going: Trendy Hype vs. Real Technology," *IEEE International Solid-State Circuits Conference*, Feb. 2006.

[84] S. Naffziger, "High-Performance Processors in a Power-Limited World," *IEEE Symposium on VLSI Circuits*, Jun. 2006.

[85] J. W. Tschanz *et al.*, "Dynamic Sleep Transistor and Body Bias for Active Leakage Power Control of Microprocessors", *IEEE Journal of Solid-State Circuits*, vol. 38, no. 11, Nov. 2003, pp. 1838-1845.

[86] C. R. Sullivan and S. R. Sanders, "Design of Microfabricated Transformers and Inductors for High-Frequency Power Conversion," *IEEE Transactions on Power Electronics*, vol. 11, no. 2, Mar. 1996, pp. 228-238.

[87] A. Hajimiri and T. H. Lee, "A General Theory of Phase Noise in Electrical Oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, Feb. 1998, pp. 179-194.

[88] T. Sakurai and A. R. Newton, "Alpha-Power Law MOSFET Model and Its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, Apr. 1990, pp. 584-594.

[89] Ya. Tsypkin, "On the Determination of Steady-State Oscillations of On-Off Feedback Systems," *IRE Trans. Circuit Theory*, vol. CT-9, no. 3, Sept. 1962.

[90] P. Hazucha *et al.*, "A 233-MHz 80-80% Efficienct Four-Phase DC-DC Converter Utilizing Air-Core Inductors on Package," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 4, Apr. 2005, pp. 838-845.