AN ABSTRACT OF THE DISSERTATION OF

Benjamin Poris Hershberg for the degree of Doctor of Philosophy in Electrical and Computer Engineering presented on June 4, 2012.

Title: Ring Amplification for Switched Capacitor Circuits.

Abstract approved: _____

Un-Ku Moon

A comprehensive and scalable solution for high-performance switched capacitor amplification is presented. Central to this discussion is the concept of ring amplification. A ring amplifier is a small modular amplifier derived from a ring oscillator that naturally embodies all the essential elements of scalability. It can amplify with accurate rail-to-rail output swing, drive large capacitive loads with extreme efficiency using slew-based charging, naturally scale in performance according to process trends, and is simple enough to be quickly constructed from only a handful of inverters, capacitors, and switches. In addition, the gain-enhancement technique of Split-CLS is introduced, and used to extend the efficacy of ring amplifiers in specific and other amplifiers in general. Four different pipelined ADC designs are presented which explore the practical implementation options and design considerations relevant to ring amplification and Split-CLS, and are used to establish ring amplification as a new paradigm for scalable amplification.

Ring Amplification for Switched Capacitor Circuits

by

Benjamin Poris Hershberg

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 4, 2012
Commencement June 2012

Doctor of Philosophy dissertation of Benjamin Poris Hershberg presented on June 4, 2012

APPROVED:

Major Professor, representing Electrical and Computer Engineering

---

Director of the School of Electrical Engineering and Computer Science

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Benjamin Poris Hershberg, Author

ACKNOWLEDGMENTS

To begin, I would like to express my sincere gratitude and thanks to my advising professor, Dr. Moon. He has always known how to provide the right guidance at the right time - when to apply pressure and when to provide slack, when to let me explore tangents and when to suggest focus; always helpful but never overbearing. My finished work always stood as the ultimate measure of accountability, and I was given the freedom to explore my ideas and take ownership of them. That, in many ways, has made all the difference in my happiness these last six years, and I consider myself truly fortunate.

I wish to thank Kazuki Sobue, Seiji Takeuchi, and Koichi Hamashita at Asahi Kasei Microdevices for providing me with the tools, resources, support, and fabrication access that has been so crucial to the success of my projects. I appreciate the meaningful discussion, insight, and friendship provided by my group-mates, office-mates, colleagues, and friends over the years, especially Peter Kurahashi, Nima Maghari, Tawfiq Musah, Rob Gregoire, and Hariprasath Venkatram. My university experience wouldn't have been nearly as productive or enjoyable if I hadn't shared so much of it with my good friend and colleague Skyler Weaver. In-depth analysis of everything from cascode current sources to Salacious B. Crumb are something that will always be a quintessential part of my college memories.

Lastly, and foremost, I wish to thank my parents for their total commitment to my education and the work ethic that they instilled in me. In a way, this dissertation belongs to them. I was always pushed to do my best, and given the encouragement to believe that I'm capable of my ambitions, for which I am eternally grateful.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

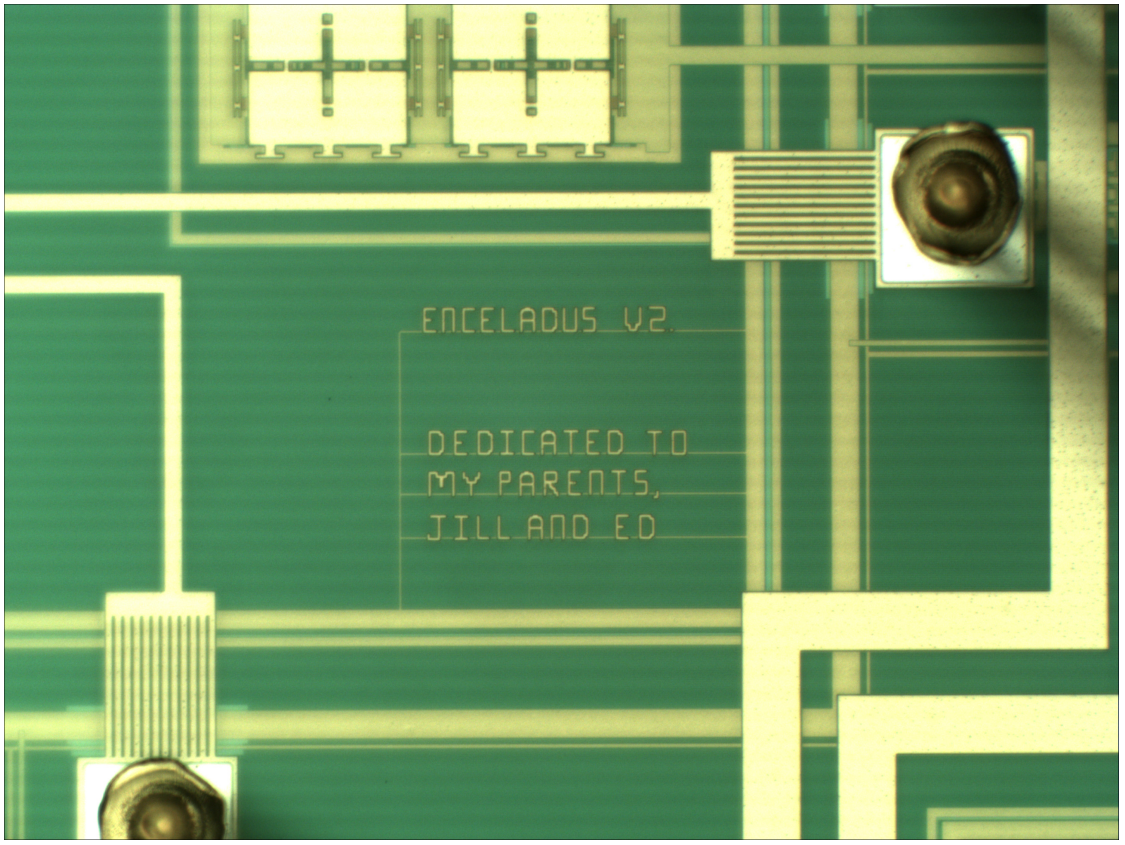
Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF TABLES



Chip micrograph, bottom-right corner of 'Enceladus' test chip (Chapter 6).

For my parents, Jill and Ed.

Ring Amplification for Switched Capacitor Circuits

CHAPTER 1. INTRODUCTION

"Every problem, Mr. Higgins, is an opportunity in disguise."

- Inara Serra, *Firefly*

1.1 Before we begin

Much of the work presented in this dissertation assumes a graduate-level familiarity with analog circuit concepts as well as a general background in electrical engineering. In particular, one should be well versed in the topics of opamp design, switched capacitor circuits, and analog-to-digital converters (particularly pipelined A/D's). Furthermore, much of the discussion in this dissertation assumes a pre-existing understanding of MOSFET device behavior, operational parameters, design tradeoffs, and a little bit of device physics. The classic texts *Analysis and Design of Analog Integrated Circuits* [3] and *Analog Integrated Circuit Design* [4] are good general-purpose references which contain most

of the background knowledge required, and the device physics primer *Semiconductor Device Fundamentals* [5] may also be helpful. With regards to pipelined analog-to-digital converters, one should be familiar with the 1.5b/stage and multi-bit architectures explained in [6] and [7]. In addition, there are many other smaller techniques and concepts which will be necessary to know, but these references will be provided to you throughout the text as necessary. Over time this dissertation may be updated; to download the most current version, please go to http://benjamin.hershberg.com/dissertation.

1.2 An end to scaling?

In 1965, Gordon Moore famously predicted that transistor density would increase exponentially with advancing process technology, in what thereafter became known as "Moore's Law" [8]. By the mid-70's, with an additional 10 years of data available, Moore and his colleagues at Intel altered their prediction slightly to state that transistor density, or device "scaling", would double every two years, and that total computational power would double every 18 months [9][10]. Remarkably, for more than half a century, this scaling phenomenon has held true. And yet, this is not because Moore's Law is an immutable constant of the universe, or even a true law. It is rather something of a self-fulfilled prophecy; one that has more to do with economics than physics [11]. For decades, corporate roadmaps have been planned with the assumption that if one does not stay on pace with Moore's Law, the competition will surge ahead. This has fueled exponential amounts of spending in both manufacturing facilities for next-generation process technologies and in the research and development necessary to enable them [10].

To a great extent, Moore's Law has remained true because we deem it to be true. And yet despite the underlying causes, the faith we have come to place in it is not without reason. The "end of scaling" has been predicted almost as many times and for as many reasons as there have been obstacles encountered, but in the end these

predictions have always proven wrong. The consistent error in their logic has been an underestimation of the power of economic impetus when coupled with human ingenuity (and perhaps the amount of caffeine consumed by device physicists). In the 1990's, at the 1μm node, minimum feature sizes began to approach the wavelength of light being used in the lithography process, and many viewed this as a fundamental physical limit. However, a clever solution was soon found, and the trend marched on. In the 2000's, at the 65nm node, the thickness of gate oxide was on the order of only a few atoms thick, and tunneling effects began to cause unacceptable amounts of gate leakage. After considerable effort, engineers found a way to replace silicon dioxide with a high-κ dielectric material in the 45nm node, and progress ventured onward [12]. In the early 2010's, short-channel effects and off-state leakage had become so bad that a move beyond the traditional planar MOSFET structure was required. Sure enough, SOI (silicon-on-insulator) and tri-gate (FinFET) technologies emerged as new physical architectures which provided much tighter control of the channel and greatly alleviated short-channel effects [13][12].

And yet, as Moore himself once said: "no exponential is forever" [14]. Perhaps in the strictest interpretation of Moore's original thesis, this is true - there are certain fundamental quantum mechanical limits to how small a single transistor can be. But Moore's Law ceased to refer solely to planar transistor dimensional scaling long ago, and is now colloquially used to describe a much more general trend of performance improvements in digital integrated circuits. Sometimes this performance improvement is put in terms of computations-per-second for a given state-of-the-art CPU, sometimes in terms of cost-per-transistor, and sometimes in terms of power dissipation. Such a broad definition of scaling is in fact more useful at this point - innovation in the nanoscale regime has come as much from new materials and physical structures as it has from dimensional scaling (and this broader definition of "scaling" is what we will use throughout the rest of this work). Yet ultimately - even under this broader definition - such exponential growth must at some point slow or end. However, growth has not *yet* shown any sign

of slowing, and is certainly a long ways from the end. The Intel technology roadmap currently forecasts a path down to 4nm, and many supplemental innovations will likely arise in the meantime, such as improved interconnect materials, through-silicon vias, and a shift from 2D chips to 3D cubes, to name a few [12][15]. Given the history of Moore's law to overcome the perceived odds, and what we know about future technologies already in development, it is safe to assume that digital performance scaling will continue on for the foreseeable future.

1.3 Challenges in analog

While all of this digital scaling seems to suggest a very cheerful future for integrated circuit performance, there is a growing problem that isn't described by Moore's Law. This problem is analog. Digital performance scaling is mainly defined by how well the transistors operate as simple on/off switches and how many of those switches can be packed into a given area. Although the digital performance of a transistor is inexorably linked to its analog performance, there is no guarantee that the improvements seen in the digital switching behavior of the transistor will imbue better analog performance as well. To the contrary, in nanoscale CMOS technology the properties of transistors have encountered radical changes, most of which have spawned serious challenges in the design and implementation of analog integrated circuits.

A good place to identify these properties and challenges is in the examination of a conventional opamp - the fundamental analog signal processing block and a good bellwether for analog performance. The classic design tradeoff triumvirate of accuracy, speed, and power manifest in an opamp as the specifications of gain, bandwidth, output swing, and architectural choices. In terms of technology parameters, small-signal output impedance (r_o), small-signal trans-conductance (g_m), threshold voltage (V_T), gate capacitance (C_G), and supply voltage (V_{DD}) are key players, although there are many

more that are also influential. The interplay of all these parameters can be very complex, and trying to isolate and analyze one without considering the others leaves us with an incomplete picture. We can, however, consider the larger picture from the particular point of view of each design requirement (gain, bandwidth, or output swing), and it is in this manner that we shall proceed to summarize the effects of these parameters on opamp performance, and the challenges that they represent for analog design in scaled environments in a broader sense.

1.3.1 Gain

With respect to scaling, the overall gain of the opamp (A_V) is influenced most heavily by three factors: g_m, r_o, and the opamp architecture used. Let's first consider the physical transistor parameters g_m and r_o. For a given transistor in the opamp open-loop gain path, its particular contribution to A_V will be $g_m \cdot r_o$. Although intrinsic g_m has been consistently improving with advancing process technology (as expected, since increasing g_m is a key requirement for digital performance improvement), r_o has been decreasing at a faster rate and overall A_V is going down [16]. In a 0.18μm process it is not uncommon to get a minimum-sized inverter gain of 28dB or higher. By contrast, predictive technology models such as [2] have forecasted minimum-sized inverter gains of less than 9dB by the 22nm node. The channel length can be increased to counteract this effect, but even this will not be enough to reverse the trend since it also depends on static process-defined parameters. A very promising development in recent years is the advent of new physical transistor structures such as silicon-on-insulator (SOI) and tri-gate (FinFET) which enhance g_m and r_o and offer some measure of respite [13][12]. Both techniques alleviate many of the problems associated with nanoscale bulk planar MOSFETs by providing better control of the channel. The tri-gate transistor structure looks particularly well poised for mass adoption, and Intel has already put it into mass

production in its 22nm process and expressed intentions to use it for all subsequent process generations into the foreseeable future. This is good news for analog - although it's too early to tell just how much improvement in $g_m \cdot r_o$ this will buy and for how long, it looks like intrinsic device gain in the 22nm node and beyond are not as big of a concern as previously feared. This will be an important conclusion to keep in mind when reading Chapter 2.

Although these intrinsic gain properties affect the performance of individual transistors, their effects can be compensated for on an architectural level by combining the gains of multiple devices. Additional gain can be generated by adding either additional cascode transistors, cascading multiple gain stages, or some sort of supplemental gain-enhancement technique. With respect to cascoding and cascading, these options have been explored extensively over the years, with some particularly popular structures being telescopic, folded-cascode, and Miller-compensated two-stage opamps [3]. Unfortunately, in nanometer CMOS none of these conventional approaches are particularly feasible for high-accuracy amplification. Modern processes have already dipped below 1V supplies, and this places some very harsh biasing-headroom and output-swing limitations on the architectural choices available in the design of a high-gain opamp. At the crux of the matter is the fact that biasing parameters have not scaled at the same pace as supply voltage, and in some cases they have even scaled in the opposite direction (V_T in low-power digital CMOS, for example). Let's consider for a moment how this limits our design options. First of all, supply voltage and output swing requirements in nanoscale CMOS place a stringent limit on the headroom available for use in cascoding. Two-stage opamps are more compatible with tight output swing requirements, but such *cascading* comes at a power and speed penalty over single-stage opamps, and for nanoscale CMOS will still often be insufficient for generating high gain. Moving to a three stage architecture is an option, but the power/speed penalty paid for stabilizing three dominant poles is often too great to be attractive.

The remaining option for achieving high gain in scaled CMOS is to use supple-

mental gain-enhancement techniques such as gain-boosting [17][18], correlated double sampling (CDS) [19][20], digital calibration [21], or correlated level shifting (CLS) [22]. However, in the presence of low supply voltage even these techniques can still be problematic. Gain-boosting requires the use of cascode transistors (which may not be available), digital calibration becomes quite complex when high-order distortion terms must be canceled, and CDS comes at a price of speed, power, and noise while doing little to alleviate the output swing requirement. Of these techniques, the most promising is arguably CLS, which manages to both boost gain dramatically and reduce the output swing requirement with minimal penalties in speed, power, and noise. As we shall explore in Chapter 3, CLS can be further improved such that the output swing requirement is effectively removed altogether (Split-CLS).

To a certain extent, CLS solves the finite-gain problem, and allows high-accuracy switched capacitor amplification in nanoscale CMOS. But even CLS has its limitations, such as in high-speed applications and when driving active loads. Furthermore, there is an important distinction between techniques that work *in* scaled environments and those that truly scale. These gain-enhancement techniques may allow opamps to *function* in low-voltage environments, but they will not grant them the ability to scale with digital trends. As we shall see in the coming chapters, we can do better than this.

1.3.2 Bandwidth

At first look, opamp bandwidth seems to scale quite well with respect to process. Indeed, the transition frequency (f_T) for a minimum sized transistor (and thus transconductance) improves with respect to process, and we can expect to see opamp bandwidths improve over time. While this seems like good news, closer inspection reveals that a conventional opamp will not scale in speed or power at the same rate as a digital inverter, and therefore we could be doing better. Consider for a moment the key

equations governing digital logic performance:

$$\text{gate delay} = \frac{C_{GATE} \cdot V_{DD}}{I_{DSAT}} \tag{1.1}$$

$$\text{fixed delay} = \frac{C_{WIRE} \cdot V_{DD}}{I_{DSAT}} \tag{1.2}$$

$$\text{energy per transition} = C_{GATE} \cdot V_{DD}^2 \tag{1.3}$$

$$\text{energy delay product} = \frac{C_{GATE}^3 \cdot V_{DD}^3}{I_{DSAT}} \tag{1.4}$$

where I_{DSAT} is the saturation drive current of the MOSFET, C_{GATE} is the MOSFET gate capacitance, C_{WIRE} is the wire and interconnect load, and V_{DD} is the supply voltage. Scaling affects all of these parameters - it increases I_{DSAT}, decreases C_{GATE}, and reduces V_{DD}. As a rule of thumb, this scaling provides a 30-40% improvement per process generation in gate delay, and an even greater improvement in energy-delay product depending on how much the supply voltage is reduced [12].

Just how well do these improvements translate into analog performance scaling? The best news is with respect to I_{DSAT}, which will lead to improvements in opamp bandwidth via improved g_m, and will provide opamps with at least some measure of scalability. The reduction in C_{GATE} and C_{WIRE} is less useful. When the output load capacitance of the opamp is set by thermal noise requirements (as will be the case in a high-accuracy application), the dominant pole(s) and bandwidth will be primarily governed by the size of C_{LOAD} and g_m, and influenced only to a lesser extent by higher-order parasitic poles related to C_{GATE} and C_{WIRE}. The other scaling parameter mentioned (V_{DD}) is of course useful for saving power in digital logic, but creates many complications for analog with respect to gain, bandwidth, and output swing. The net effect is that only some of these parameters will provide improvements in analog performance,

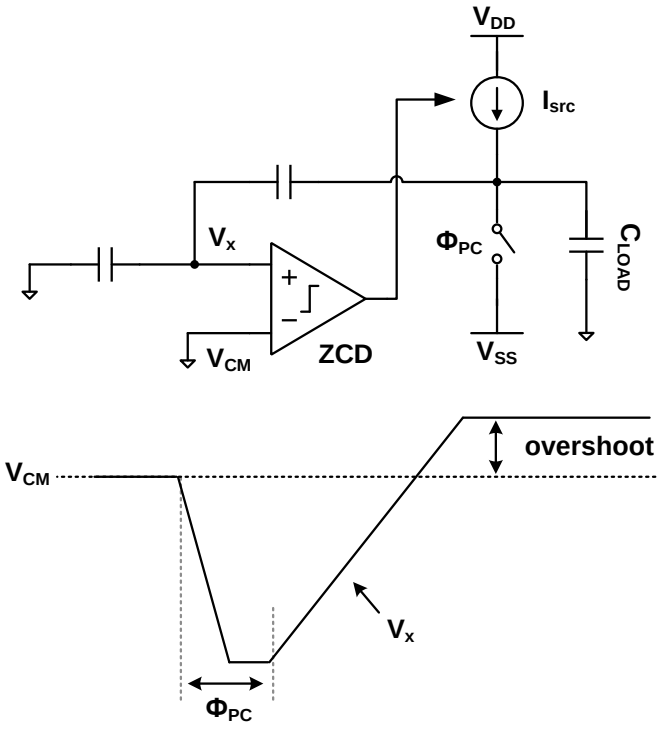


Figure 1.1: Basic theory of operation for a zero-crossing based circuit (ZCBC). Rather than use an opamp to force V_X to V_{CM}, we can use a zero-crossing detector (ZCD) and current source to charge the output and detect when $V_X = V_{CM}$.

whereas all of them will provide improvements to digital, and opamp performance lags behind digital, sometimes by exponential amounts.

One tantalizing solution to this disparity is to use opamp architectures whose dominant poles are unrelated to the output load size. The dominant poles of the system would then be defined by the amplifier's internal parasitics, and this would bring C_{GATE} and C_{WIRE} back into the equation for analog scaling. To do so, however, one would need to decouple the internal circuitry of the opamp from the output load, and this is no trivial task. One approach to this goal was proposed several years ago, an idea known as zero-crossing based circuits (ZCBC) [23]. Shown in Fig. 1.1, a zero-crossing based circuit uses a comparator and current source to charge towards and detect a feedback circuit's desired final output value, rather than explicitly settle to it with an opamp. The current source is digitally switched by inverters that decouple the comparator circuitry from the

output load, and this means that the internal bandwidth and power requirements of the comparator are much more weakly coupled to the size of the output load capacitance than in a regular opamp. This approach has proven its applicability over a range of design specifications in modern processes, and set a new standard for the state-of-the-art in amplifier power efficiency [24][25]. However, the open-loop nature of ZCBC operation makes this technique difficult to design in a way that ensures sufficient linearity, particularly in a setting where significant tolerance to process, voltage, and temperature are important. Furthermore, the ZCBC's dependence on having a linear current source creates significant headroom requirements that are not particularly amenable to voltage scaling, and the ZCBC must rely on additional linearity-enhancement techniques such as CLS (which can also be used to enhance current source linearity in the context of ZCBCs as described in [26]). Despite the ability of ZCBCs to improve some aspects of scalability, its susceptibility to others and a plethora of reliability and complexity challenges bring into doubt whether it will become a broadly applicable solution suitable for mass adoption. ZCBC are, however, a step in the right direction. In Chapter 2 we will explore the concept of ring amplification, which takes the next step.

1.3.3 Output Swing

Of all the design challenges that we have discussed thus far, none have been quite as vexing or deleterious to analog performance in scaled CMOS as that of maintaining high output swing. Supply voltage has scaled faster than threshold voltage and biasing headroom requirements (V_T, V_{DSAT}) and made useable output swing an ever more precious resource. Whereas a five transistor telescopic opamp stack with V_{DSAT} values of 175mV in a 2.5V technology will leave 1.624V of the range free for useable output swing, the same setup in a 1V process can only provide a mere 125mV (92.3% less) for the same task. For high-accuracy amplification where thermal noise limitations are a design

constraint, such a small output swing range will come at a tremendous cost in terms of power.

At the very least, this requires that the opamp utilize a two-stage architecture where the output stage has only one PMOS and one NMOS transistor consuming headroom. Consider for a moment an identical opamp architecture in a 2V process and the same one in a 1V technology. The SNR will be defined by

$$SNR_{dB} = 20\log_{10}\left(\frac{V_{SIGNAL}}{V_{NOISE}}\right) \tag{1.5}$$

and speaking strictly in terms of sampled kT/C noise contributions, this will be

$$SNR_{dB} = 20\log_{10}\left(\frac{\psi V_{SIGNAL}\sqrt{C}}{\sqrt{kT}}\right) \tag{1.6}$$

where ψ is some constant that depends on the particular switched-capacitor feedback structure used. What we see here is that when the output swing halves, the capacitance must be quadrupled in order to preserve the same SNR. For example, if the biasing headroom requirement remains fixed at 400mV, the 2V supply case will have 1.6V output swing and the 1V case will have only 600mV swing. To maintain the same SNR, the opamp in the 1V case would need to use a 7.1x larger capacitor and burn at least 1.8x more power in order to maintain the same bandwidth. This painful tradeoff becomes even worse as we descend further into low-voltage, since the usable output swing accounts for an increasingly smaller amount of the total available voltage range.

A glimmer of hope lies in the actual pace of voltage scaling, which has decreased relative to other scaling trends in recent nanoscale CMOS, and it looks like this slow-down will persist [12]. Although this does not solve the underlying crisis, it does at least buy some sort of reprieve. However, if we wish to truly solve the problem, conventional stand-alone opamp architectures won't do, and new techniques are needed. One interesting option is to use a supplemental technique such as CLS, which can boost the effective output swing to the same size as the supply voltage itself. In this ideal max-swing case,

we find that the migration from a 2V supply down to 1V leads to a more promising result. Although the load capacitance and opamp current must still quadruple, the total power spent will not change (in the ideal case, anyway). This seems like an acceptable option - although voltage scaling may not do much to improve opamp power efficiency, at least it won't hurt it.

1.4 Troubling Trends

In general, things look bad for analog scaling. But the question remains: just how bad? To answer this, we must study the effects, not the causes. In other words, we must look to the large data set of published works, which represent many different realizations of the design tradeoffs we have explored. The only complete data set known to-date that can serve this purpose is that of [1], which contains a comprehensive survey of analog-to-digital converter (ADC) performance in published works throughout the years (and the answer to our question). Although we have limited our discussion to amplifiers thus far, and not all ADCs use amplification, there are certain classes of ADC which do use amplification almost universally. In particular, pipelined ADCs make heavy use of amplification. In fact, their overall performance is so strongly tied to the performance of their internal amplifiers that they serve as a good platform for characterizing amplifier scaling trends.

For easier comparison, the overall power efficiency of an ADC can be assessed by condensing the classic speed-accuracy-power tradeoff into a single scalar figure-of-merit (FoM) value. The most common FoM formula used in the field of ADCs today is:

$$FoM_1 = \frac{P}{2^{ENOB} \cdot f_s} \tag{1.7}$$

where P is the total power of the converter, f_s is the sampling rate (or twice the input bandwidth for oversampling ADCs), and ENOB is the effective-number of bits, which

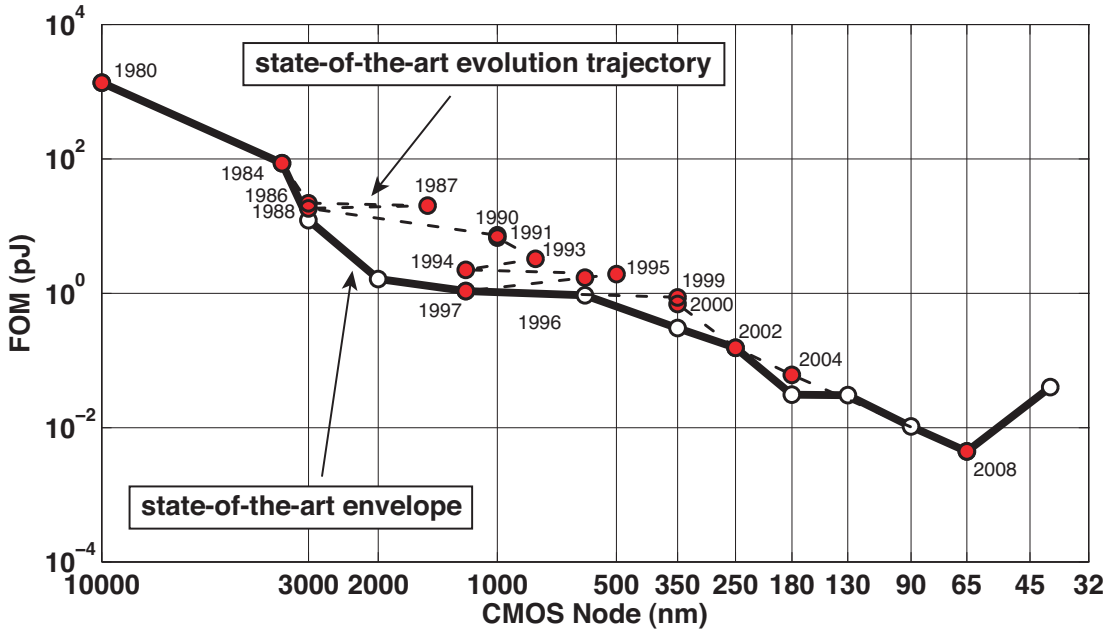


Figure 1.2: State of the art FoM_1 versus CMOS process node. The dashed line is the evolution path with respect to time. ADC performance with respect to FoM_1 appears to be scaling very well. This is because FoM_1 does not account for thermal noise, and therefore strongly favors ADC resolutions and architectures that are in the low and medium resolution range. Amplifier-less, digitally switched SAR ADCs in particular have fueled this trend. (This figure is reprinted with permission from [1].)

for a pure sine-wave input tone is given by:

$$ENOB = \frac{SNDR - 10\log_{10}(3/2)}{20\log_{10}(2)} \tag{1.8}$$

Fig. 1.2 presents a plot of the state-of-the-art FoM_1 for progressing process nodes. What we see is that with respect to FoM_1, ADC performance is scaling very well, even into nanoscale CMOS. This seems to be good news, and indeed it is. But FoM_1 only reveals one part of the story. In the evaluation of the speed-accuracy-power tradeoff, FoM_1 considers the accuracy component of this equation (2^{ENOB}) in terms of non-linearities and distortion, which manifest as voltage errors. In low and medium resolution ADCs, such voltage errors will indeed limit overall accuracy, and the formulation of FoM_1 is logical. However, in high-resolution ADCs, where thermal noise is the fundamental limitation on overall accuracy, it will be noise *power* not error *voltage* that will limit

accuracy. In the case of noise, to reduce the thermal noise voltage by a factor of two, the noise power must be reduced by a factor of four. In other words, for each additional bit of ENOB resolution in a high-resolution noise-limited ADC, 4x more power must be spent. This puts high-resolution ADCs at a disadvantage with respect to FoM_1, and for this reason, the curve of Fig. 1.2 is dominated by low and medium resolution ADCs. In the nanoscale era, medium resolution ADC peak performances have been dominated by SAR ADCs, which are highly digital and inherently scalable and have, to a great extent, displaced amplification-based architectures from the low and medium resolution realm across a wide range of speeds and accuracies. Therefore, with respect to analog circuitry and amplifier scalability, we can learn little from Fig. 1.2 other than to observe that there are some highly-viable alternatives to amplifier-based architectures in the realm of low and medium resolution ADCs.

To better capture the effect of high-resolution noise-limited systems, we can use an alternative formula, FoM_2:

$$FoM_2 = \frac{P}{2^{2 \cdot ENOB} \cdot f_s} \tag{1.9}$$

State-of-the-art performance with respect to this FoM_2 formula is given in Fig. 1.3. In this plot, we find a radically different story than the one told by Fig. 1.2 - rather than scaling well, power efficiency is actually getting progressively *worse* in nanoscale CMOS technologies. Much of what is seen in this trajectory consists of pipelined ADCs and Delta-Sigma ADCs, both of which use amplification, so by proxy, this plot is also showing us amplifier scalability with respect to process. Where is this degradation coming from? We should expect to see some amount of improvement in speed (for the reasons we discussed earlier), so f_s of Eq. 1.9 is unlikely to be the culprit. This leaves us with supply voltage scaling and the degradation of r_o, and the additional strain that they place on meeting gain and output swing requirements, as possible suspects. If they are indeed the culprits, then we should expect to see this noise-linearity-power penalty at

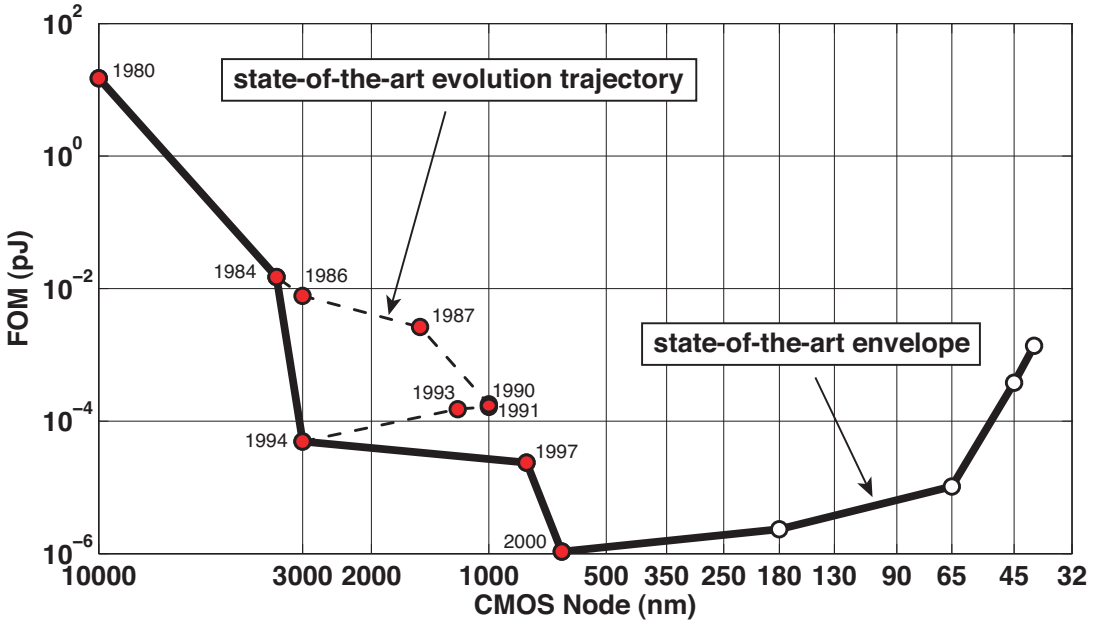


Figure 1.3: State of the art FoM_2 versus CMOS process node. FoM_2 takes into account thermal noise, and favors high-resolution ADCs. Much of what is seen in this trajectory consists of pipelined ADCs and Delta-Sigma ADCs, both of which use amplification. As we can see, amplification-based solutions ceased scaling well by the 0.18μm node, and continue to *degrade* in power efficiency in nanoscale technologies. (This figure is reprinted with permission from [1].)

the opamp level in the form of an SNDR-power degradation at the global level. Sure enough, an increase in the relative noise floor in nanoscale CMOS as depicted in Fig. 1.4 reveals this to be the case. Ultimately, for amplifiers and high-resolution ADCs, any speed or power benefits emanating from technology scaling are being overpowered by the increase in relative noise floor caused by gain and output swing challenges.

This disquieting trend for amplification-based analog design has by no means gone unnoticed in industry. Even as technology marches on, monolithic high-resolution ADC products on the market today tend to linger in the 90nm - 180nm realm, and a survey of ADCs published in the nanoscale era reveals a notable absence of competitive high-resolution amplifier-based ADC architectures implemented in nanoscale technologies [27]. While the reasons for this are quite apparent by this point in our discussion, such stalling isn't a viable long-term option. There is a huge amount of industry momentum to realize

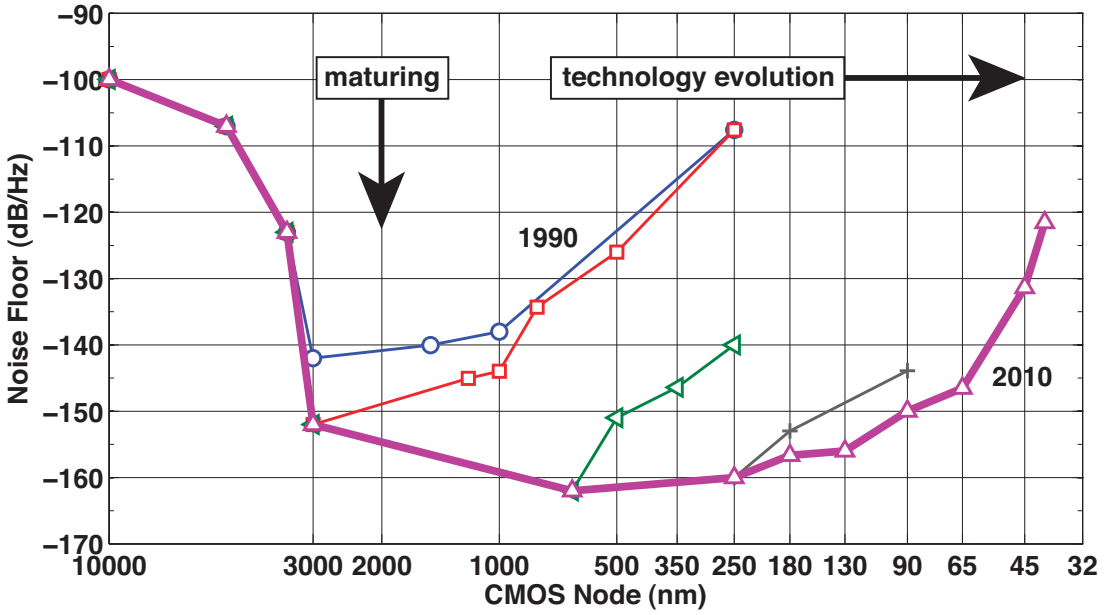


Figure 1.4: State-of-the-art relative noise floor versus technology node. This explains the degradation of FoM_2 seen in Fig. 1.3 - noise floor and SNR is degrading faster than any speed or power savings can compensate. (This figure is reprinted with permission from [1].)

fully integrated system-on-chip (SoC) solutions in nanoscale CMOS, driven by the many tantalizing performance, miniaturization, and economic incentives of doing so. At this point there is little doubt that SoC will be a big part of the future, and the ease with which analog circuitry can be implemented in nanoscale CMOS will play a major role in the types of devices and level of mobility that SoC ultimately delivers.

1.5 A New Amplification Paradigm

Although SoC may indeed be the future, if amplification is "broken" in nanoscale technologies, what choice do we really have other than to burn more power in the analog blocks? The first step that we can take in improving analog SoC integration is to minimize the analog circuitry itself (and amplifiers in particular) as much as possible. Indeed, a great deal of analog research seems to be headed in this direction, with digital-

transceivers and digital-PLLs garnering their fair share of attention in recent years. Even for true-analog circuits, the cost of augmenting them with complex digital assist and calibration techniques is becoming more and more affordable. The choices made in the construction of future communication standards can also make a big difference in this effort, by establishing industry-wide requirements that "play well" with the solutions currently available. For example, if a certain standard called for a 100Msps ADC with better than 8b ENOB, this would not be a particularly hard demand to satisfy, since SAR ADCs, which scale very well, can now meet this requirement. By contrast, if the standard required a 100Msps ADC with better than 12.5b ENOB, it would be much harder to find an amplifier-less solution to meet this need.

The bottom line, however, is that we can't completely rid ourselves of amplifiers. Although we can try to favor scalable structures such as SAR ADCs whenever possible, analog and mixed signal design is still a very big world, and there will always be a need for amplification in many parts of it. We live in an analog world and our devices are given meaning only by their ability to interface with that world. While we can digitize the internal processing tasks, we can never fully digitize the interface between the two domains. Amplifiers must be a part of this interface, because real-world quantities consist of capacitances and resistances, frequencies and phases, currents and voltages, and we require amplifiers (in some form or another) to interface with these quantities.

On a more philosophical level, we should also ask why we are so eager to give up on amplification in the first place. The quest to adopt other approaches is not without its costs, some of which may have very profound and far-reaching implications. Without a doubt, whenever good alternative solutions are found, we should make full use of them. But the existence of these alternative techniques does not necessarily preclude the possibility of scalable amplifiers. Why not have our cake, and eat it too? "Analog" and "amplification" have only inherited a negative connotation because of the conventional solutions that they represent, and assigning the label "digital" to analog design is somewhat meaningless. A "digital-like" amplifier topology may incorporate

digital components or traits, but unless it can scale with process, such "digital" features are really of little use. Ultimately, there is only one term that matters in analog design evolution: "scalable". The advent of scalable amplification can only lead to a greater diversity of solutions available to the future designer.

For the most part, the failure-to-scale seems to have happened because the available scaling techniques tend to focus on treating the symptoms of the problem, rather than on curing the underlying disease. The underlying structure - opamps - are fundamentally ill-suited to scaling. Applying additional techniques can enable opamps to function in nanoscale environments, but it won't grant the opamp the ability to scale at the same pace as digital performance improvements. As long as the underlying structure remains flawed, we will never achieve our objectives. The only option moving forward is to explore entirely new paradigms that depart from the conventional techniques.

One such paradigm (ring amplification) will be the primary focus of this dissertation. Chapter 2 introduces the fundamental theory of ring amplification and explores the mechanisms and benefits that make is a truly scalable solution. In Chapter 3 we will explore the general gain-enhancement technique of Split-CLS, which can both be used generally in all opamp design as well as specifically in the context of ring amplification. In the remaining chapters of this dissertation we will explore four different pipelined ADC implementations that incorporate one or both of the two key ideas put forth in this thesis (ring amplification and Split-CLS). The highly promising results measured from these designs serve to demonstrate the exciting potential of ring amplification as a scalable paradigm for the future.

CHAPTER 2. RING AMPLIFICATION

> "Things just happen in the right way, at the right time. At least when you let them, when you work with circumstances instead of saying, 'This isn't supposed to be happening this way,' and trying harder to make it happen some other way."
>
> - Benjamin Hoff, *The Tao of Pooh*

> "Numquam ponenda est pluralitas sine necessitate."
> (Plurality must never be posited without necessity.)
>
> - William of Ockham

2.1 Introduction

In Chapter 1 we established the motivation and need for scalable CMOS amplification solutions. But the more difficult question of how to construct such an amplifier remains unanswered. To begin with, we can logically deduce that it must operate in a way that implicitly uses the characteristics of scaled CMOS to its advantage, transforming potential weaknesses into inherent strengths. Since process scaling is inherently skewed to favor the time-domain world of high-speed digital, scalable analog techniques will most likely be found in the same realm and are likely to incorporate things like

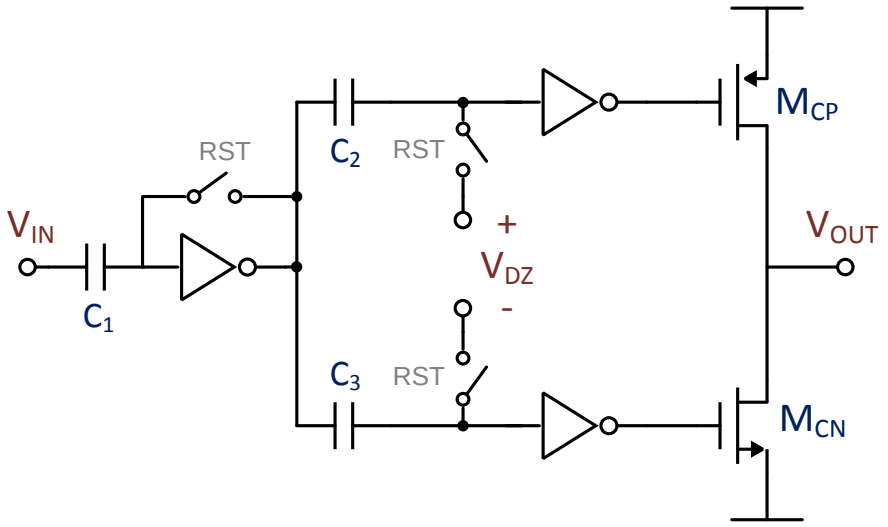


Figure 2.1: Fundamental structure of a ring amplifier. A ring amplifier is a ring oscillator that has been split into two signal paths, with a different offset embedded in each path, creating an input-referred "dead-zone" for which neither output transistor will conduct.

cascades of logic cells, rapid sequential operations, digital signaling, and DACs with tiny unit element sizes. The best of these structures will exploit the complex interplay of steady-state, small-signal, and transient operating characteristics in new ways. By considering all three operational domains simultaneously (AC, DC, and transient) we can unlock an enormous number of hitherto unexplored possibilities.

In this chapter, we introduce such a technique: ring amplification. A ring amplifier (Fig. 2.1) is a small modular amplifier derived from a ring oscillator which naturally embodies all the essential elements of scalability. It can amplify with rail-to-rail output swing, drive large capacitive loads with extreme efficiency using slew-based charging, naturally scale in performance according to process trends, and is simple enough to be quickly constructed from only a handful of inverters, capacitors, and switches.

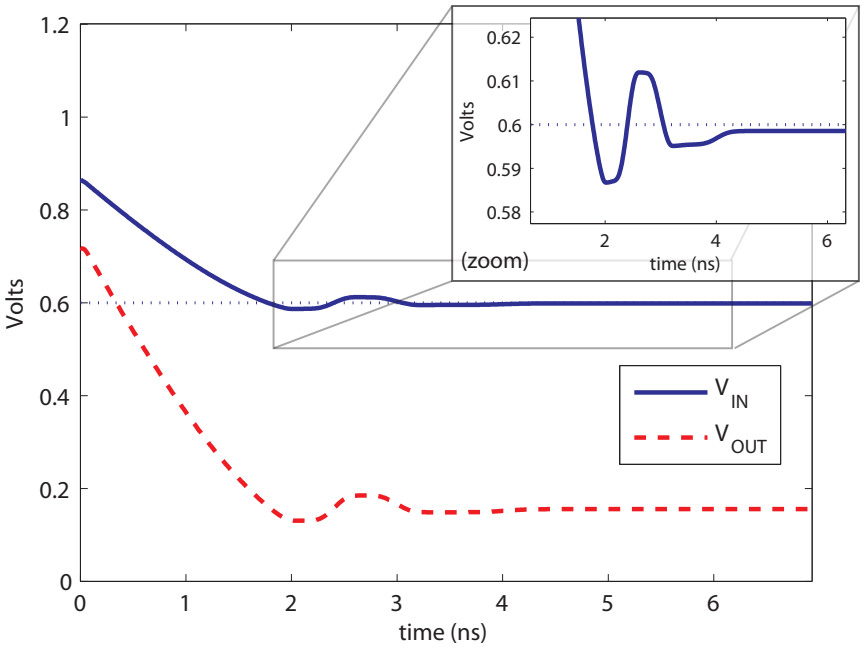


Figure 2.2: Input and output charging waveforms of Fig. 2.1 for a typical case. The ringamp efficiently slews toward the ideal settled value, then rapidly stabilizes and locks into the dead-zone.

2.2 Basic Structure

Ring amplification is, at its core, a set of concepts - concepts which can be realized through a variety of structural implementations and design choices. One such implementation is depicted in Fig. 2.1. This simple structure embodies all of the key features and concepts of ring amplification, and in many ways can be thought of as the quintessential "base case". However, a much wider variety of ringamp implementations and techniques are possible, and can be used to meet a broad range of speed, accuracy, and loading requirements (and some of these options will be explored in later chapters).

Fundamentally, a ring amplifier is a ring oscillator that has been split into two (or more) separate signal paths. A different offset is embedded into each signal path in order to create a range of input values for which neither output transistor M_{CN} nor M_{CP} of Fig. 2.1 will conduct. If this non-conduction "dead-zone" is sufficiently large, the

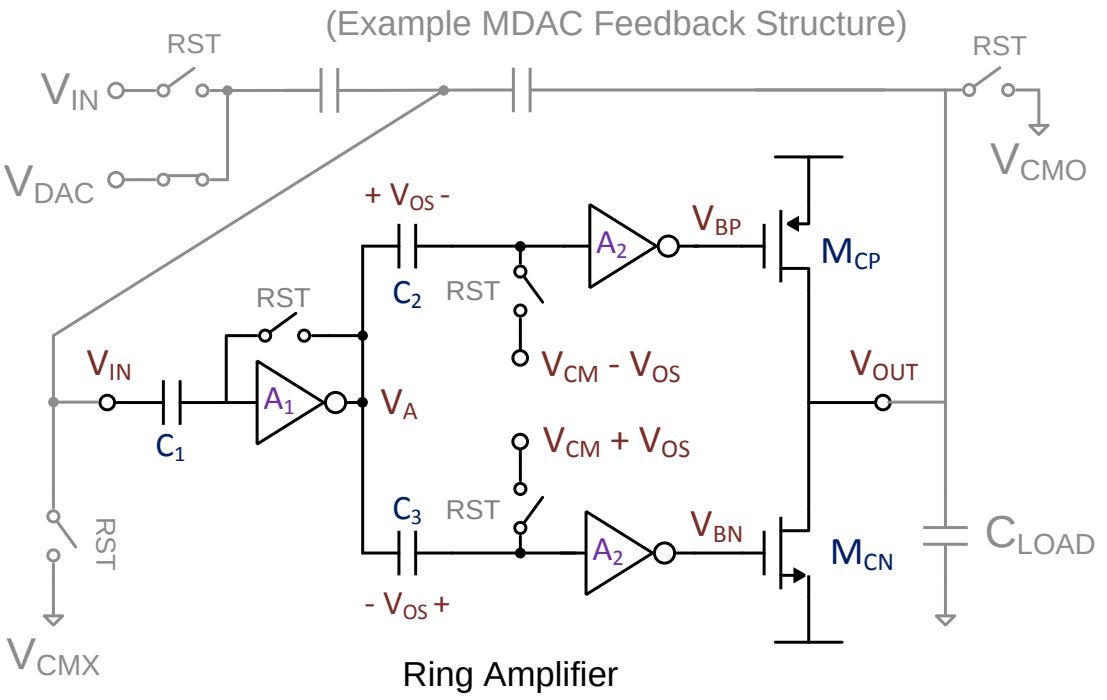


Figure 2.3: The ringamp and basic switched-capacitor feedback network that we will primarily consider in the theoretical discussions of this chapter. Devices and parameters that are referenced throughout the text are labeled.

ring amplifier will operate by slewing-to, stabilizing, and then locking into the dead-zone region. When placed in the example switched capacitor MDAC feedback structure of Fig. 2.3, this charging and settling behavior results in the waveforms of Fig. 2.2.

Before we examine how and why this occurs (in Section 2.3), it is useful to first understand some of the basic characteristics of the structure itself. To begin with, consider the capacitor C_1 of Fig. 2.3. C_1 is used to cancel the difference between the MDAC virtual-node sampling reference (V_{CMX}) and the trip-point of the first stage inverter. This ensures that the ideal settled value for V_{IN} will always be V_{CMX}, independent of the actual inverter threshold. Any sources of offset that are generated after the first stage inverter will not be removed by C_1, but the input-referred value of such offsets will typically be negligibly small.

The dead-zone of the ringamp in Fig. 2.3 is embedded prior to the second stage inverters by storing a voltage offset across capacitors C_2 and C_3. Any value for V_{IN}

within the dead-zone region is a viable steady-state solution for the ring amplifier, and so in almost all practical cases, the input-referred value of the dead-zone will define the overall accuracy of the amplifier. In other words, the maximum error of V_{IN} when the ringamp has stabilized and locked will be

$$\epsilon_{V_{IN}} \leq \frac{V_{DZ}}{A_1} \tag{2.1}$$

where $V_{DZ} = 2 \cdot V_{OS}$ and A_1 is the DC small-signal gain of the first stage inverter.

It is worth briefly noting that there are many additional options for both where and how to embed the dead-zone offset into the ring amplifier, and for different target accuracies and design applications it may be useful to consider additional possibilities and their respective advantages and disadvantages. In this chapter, however, we will focus solely on the embedding scheme of Fig. 2.3, which possesses several key benefits. First of all, embedding it with capacitors allows us to accurately and linearly set the dead-zone offset value, and it can be done with a high-impedance, low-power reference. Second, as we shall soon see in Section 2.3, there are important stability benefits gained by embedding the offset prior to the second stage inverters, rather than the first or third stage. Finally, due to the accuracy limitations imposed by Eq. 2.1, we typically wish to create an input-referred dead-zone value of a few millivolts or less, and for medium accuracy ring amplifiers, embedding the dead-zone offsets immediately after the first gain stage will create input-referred dead-zone sizes small enough to achieve desired accuracies while still keeping the embedded offset large enough to easily tune with a simple DAC or voltage reference.

2.3 Stabilization Theory

Although the fundamental structure of a ring amplifier is quite simple, a full understanding of the operational theory behind ring amplification is considerably more complex. The steady-state, small signal, and transient characteristics of a ring amplifier

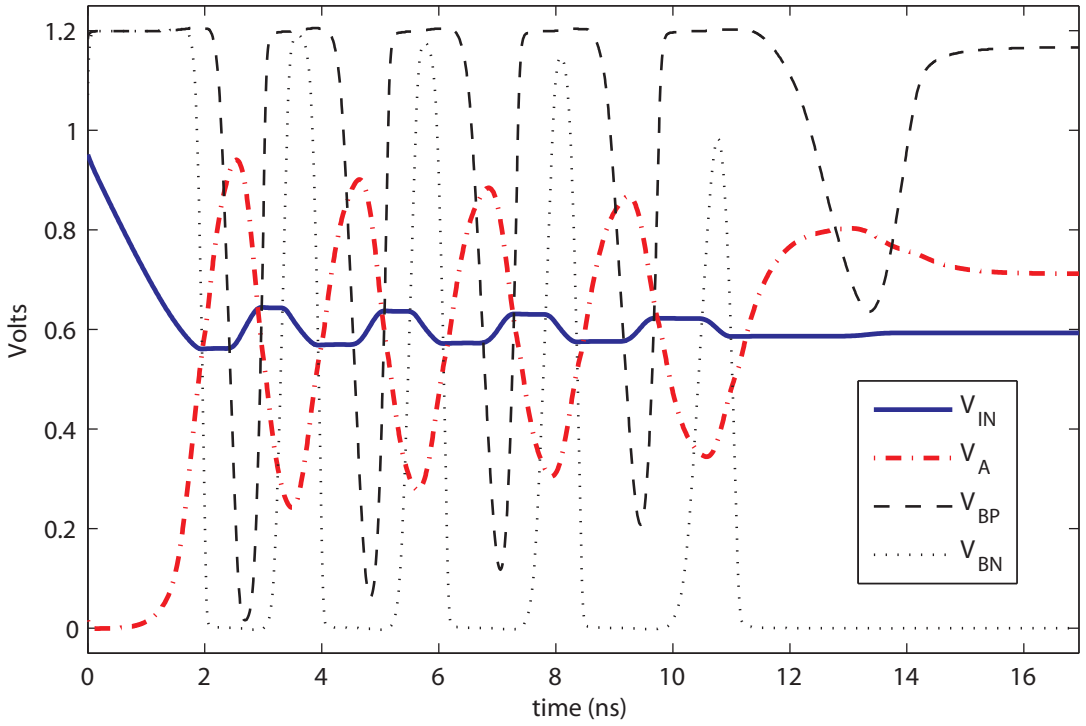


Figure 2.4: Example ring amplifier operation for an exaggerated design biased at the edge of stability, showing the three key phases of operation: 1) initial ramping, 2) stabilization, and 3) steady-state.

are highly co-dependent, and as such, its behavior cannot be accurately explained by considering each operational domain (DC, AC, transient) separately, as is often done in opamp design. However, it is still possible to reduce the theory down to several simple sub-concepts. First, the ringamp operation can be subdivided with respect to different phases of operation in time: slewing, stabilization, and steady-state. Second, the theory of operation within each phase can be reduced to a chain of cause-and-effect mechanisms. It is in this manner that we will introduce the fundamental theory of operation for ring amplifiers.

To illustrate key concepts, we will use the exaggerated charging waveform of Fig. 2.4 (taken from the ring amplifier of Fig. 2.3) that has been designed with relatively low bandwidth, excessive drive current, and a dead-zone size that biases the ringamp right at the edge of stability. Although one would never wish to make a real

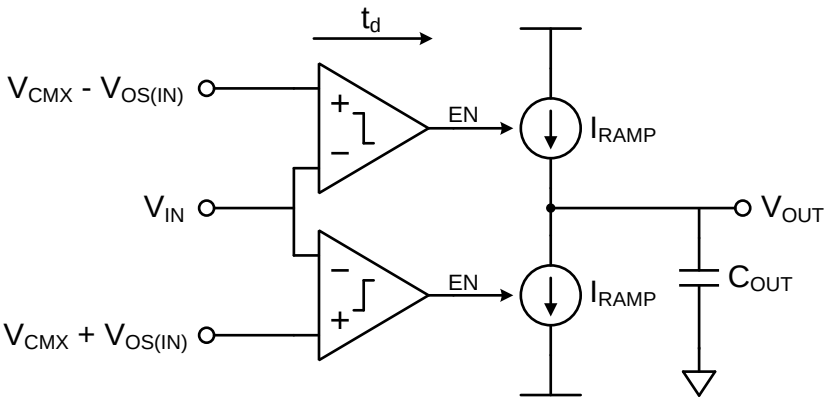


Figure 2.5: Conceptual model of a ringamp during the initial slew-charging phase of operation. This model only applies to the initial charging phase and does not include the key ringamp stabilization mechanisms.

design in this way, as a teaching example it is quite useful. V_{CMX} is set to 0.6V, and thus the settled value of V_{IN} will also be 0.6V. For the sake of simplicity and generality V_{OUT} is not shown, because it is simply a scaled and shifted replica of V_{IN}; whereas V_{IN} will always settle to V_{CMX}, V_{OUT} will settle to a signal-dependent value. Unless otherwise stated, any mention of the amplitude of the fed-back signal will refer to the amplitude seen at V_{IN}.

In Fig. 2.4 we can clearly see three main phases of operation. Initially, from 0ns to 2ns, the ringamp rapidly charges toward the dead-zone. Then, from 2ns to about 14ns it oscillates around the dead-zone region as it attempts to stabilize. By 15ns, with the output transistors M_{CP} and M_{CN} both completely cutoff, the ring amplifier reaches a steady-state solution within the dead-zone, and remains locked.

2.3.1 Initial Ramping

In the initial slew-charging phase of operation, the ring amplifier is functionally equivalent to the circuit of Fig. 2.5. The output transistors M_{CN} and M_{CP} behave like complementary, maximally-biased current sources, and only one of these current

sources will be active during the initial charging ramp. The first two stages of the ring amplifier act like a pair of bi-directional continuous-time comparators that correctly select which current source to use depending on the value of the input signal. In this phase of operation, the ringamp behaves much like a zero-crossing based circuit [23][25]. However, as we will see later in Section 2.4, a ringamp receives all the key benefits of zero-crossing based circuits while suffering from almost none of the drawbacks.

The ramping phase ends when the input signal crosses the threshold of the comparator and the current source turns off. Due to the finite time delay of the comparator, there will be some amount of overshoot beyond the comparator threshold, which will be given by:

$$\Delta V_{overshoot} = \frac{t_d \cdot I_{RAMP}}{C_{OUT}} \tag{2.2}$$

where t_d is the time delay of the comparator decision, I_{RAMP} is the current supplied by the active current source, and C_{OUT} is the total loading capacitance seen at the output. This overshoot is with respect to the trip point, which will be on the boundary of the dead-zone. It will be more useful later on if we consider Eq. 2.1 as well, and express the overshoot with respect to the ideal settled value (the center of the dead-zone):

$$\Delta V_{init} = \frac{t_d \cdot I_{RAMP}}{C_{OUT}} - \frac{V_{DZ}}{2 \cdot \hat{A}_1} \tag{2.3}$$

where $\hat{A}_1$ is the effective gain of the first stage inverter at the end of the ramping operation (explained later).

2.3.2 Stabilization

After the initial charging ramp, the ring amplifier will begin to oscillate around the target settled value with amplitude ΔV_{init}. With no dead-zone, the structure is functionally identical to a three inverter ring oscillator, and will continue to oscillate indefinitely. However, as the size of the dead-zone is increased, the ringamp will eventually reach an operating condition where it is able to self-stabilize, such as depicted in

Fig. 2.4. If the dead-zone size is increased further still, the time required to stabilize decreases substantially, and for most practical designs, a ringamp will stabilize in only one or two periods of oscillation.

The most fundamental mechanism in the process of stabilization is the progressive reduction in the peak overdrive voltage applied to the output transistors M_{CN} and M_{CP} on each successive period of oscillation (visible in Fig. 2.4). When the following relation is true, the amplitudes of the signals V_{BP} and V_{BN} will be limited by the finite-gain of the first two stages, and begin to decrease:

$$\hat{A}_2[(\hat{A}_1 \cdot \tilde{V}_{IN}) - V_{DZ}] \leq V_{DD} - V_{SS} \tag{2.4}$$

(where $\tilde{V}_{IN}$ is the peak-to-peak amplitude, and $\hat{A}_1, \hat{A}_2$ are the effective instantaneous inverter gains). The resulting reduction in V_{OV} applied to the output transistors M_{CN}/M_{CP} will reduce the output current I_{RAMP} quadratically, and thus reduce the amplitude of $\tilde{V}_{IN}$, due to Eq. 2.3, by a quadratic amount. In turn, this reduces the left side of Eq. 2.4 further, and V_{OV} will pinch off even more. This feedback effect will continue to build until the input signal amplitude becomes smaller than the input-referred value of the dead-zone, at which point the ring amplifier will stabilize and lock into the dead-zone.

If we rearrange Eq. 2.4, we see that in order to trigger this pinch-off effect, the input signal amplitude must satisfy the following relation:

$$\tilde{V}_{IN} \leq \frac{1}{\hat{A}_1}(\frac{V_{DD} - V_{SS}}{\hat{A}_2} + V_{DZ}) \tag{2.5}$$

Furthermore, at the beginning of the stabilization phase:

$$\tilde{V}_{IN} = 2 \cdot V_{init} \tag{2.6}$$

Finally, using Eqs. 2.2, 2.3, 2.5, and 2.6, we can express the stability criterion in terms of the dead-zone (i.e. settled accuracy) and the initial slew rate (i.e. speed):

$$\frac{t_d \cdot I_{RAMP}}{C_{OUT}} \leq \frac{1}{2 \cdot \hat{A}_1}(\frac{V_{DD} - V_{SS}}{\hat{A}_2} + 2 \cdot V_{DZ}) \tag{2.7}$$

From this relation we see that there is a clear design tradeoff between accuracy, speed, and power. Let's assume for a moment that only t_d, I_{RAMP}, and V_{DZ} can be adjusted. To increase speed, one can either increase the initial ramp rate or decrease the time required to stabilize. Both options require sacrificing either accuracy (by increasing V_{DZ}) or power (by decreasing t_d). Likewise, to increase accuracy (decrease V_{DZ}), one can either decrease I_{RAMP} or increase t_d. While these simple tradeoffs serve as a good starting point, as we will soon discover, every parameter in Eq. 2.7 is adjustable to some extent.

Let's take a moment to consider this pinch-off effect and resulting stability criterion in the form of a practical example. Consider a ringamp where $A_1 = A_2 = 25\frac{V}{V}$, $V_{DZ} = 100mV$, $V_{DD} = 1.2V$, and $V_{SS} = 0V$. By Eq. 2.1, the input-referred size of the dead-zone will be about 4mV, which for a 2V pk-pk input signal would ideally be accurate enough to achieve an input-referred SNDR of 54dB. By Eq. 2.5, the maximum allowable peak-to-peak amplitude of $\tilde{V}_{IN}$ is approximately 6mV, and by Eq. 2.3, the maximum allowable overshoot at the end of the initial ramping phase must be less than 5mV.

This isn't a very encouraging result, since such a small overshoot will place a tight constraint on the parameters in Eq. 2.2. However, if one were to simulate this same scenario, it will turn out that the amplitude of oscillation can be significantly larger than the predicted 3mV and still achieve stability. A closer look at Fig. 2.4 reveals the answer to this disparity between theory and practice. Although the AC small-signal gain of the first stage inverter, A_1, may be $25\frac{V}{V}$, the effective instantaneous value

$$\hat{A}_1(t) = \frac{V_A(t)}{V_{IN}(t)} \tag{2.8}$$

in the actual transient waveform will be several times smaller at the beginning of stabilization. Thus, although the overall accuracy of the ringamp is determined by the final, settled, small-signal value of A_1, the stability criterion is determined by the initial, transient, large signal effective value of A_1. This reduction in A_1 occurs because the first stage inverter inherently operates around its trip point, where it will be slew limited.

The maximum slewing current that the inverter can provide will be

$$I_{slew} = I_P - I_N \tag{2.9}$$

and for a square law MOSFET model, this will become:

$$I_{slew} = 2k' \frac{W}{L} (\frac{V_{DD} - V_{SS}}{2}) V_{IN} \tag{2.10}$$

Notice here that the slew current is linearly related (not quadratically) to the input voltage, and for small values of V_{IN}, the slew rate will also be quite small. Thus, for the first stage inverter, slew rate limiting has an important impact on determining the effective value of $\hat{A}_1$ during stabilization (and to a lesser extent, the value of $\hat{A}_2$). This dynamic adjustment of the effective inverter gain is a very attractive characteristic, and improves the design tradeoff between speed, accuracy, and power by a significant factor.

The V_{OV} pinch-off effect can be conceptualized as a dynamic adjustment of the ringamp's output pole corner frequency. The decrease in current due to V_{OV} pinch-off increases the output impedance (R_o) of the ringamp, and pushes the output pole (formed by R_o and C_{LOAD}) to lower frequency. As the pinch-off effect gains momentum on each successive oscillation half-period, the output pole progressively pushes to lower and lower frequency. By the time the ringamp is locked into the dead-zone and the output transistors are in cutoff, R_o is infinite and the output pole is at DC. While dynamic current control is not a new idea in itself, ring amplifiers are unique in the elegance and efficiency in which they do it, and in the profound benefits that their particular approach brings to scaled environments, as we shall see in Section 2.4.

2.3.3 Steady State

Thus far, we have defined the steady-state condition for a ring amplifier as the complete cutoff of both output transistors, with the input signal lying solidly within the dead-zone, such as is the case in Fig. 2.4. However, considering the discussion

about pole adjustment in the previous paragraph, it's clear that the ringamp can in fact be stable for a range of low frequency output pole locations down to DC. Such a situation will in practice occur often, even for a large dead-zone, since there is always a finite probability that the ring amplifier will happen to stabilize right at the edge of the dead-zone. If that happens, one of the output transistors will still conduct a small amount of current to the output, and may never fully shut off before the amplification period ends. This behavior isn't an inherent problem for ring amplification, since any low-bandwidth settling will only serve to further improve accuracy, but there are often higher-level structural considerations that make it advantageous to ensure that both output transistors are non-conducting. The design presented in Chapter 6 is one such case, and it is there that we will explore this issue in more detail.

2.4 Key Advantages

Ring amplifiers are both structurally and functionally quite different from conventional opamps in many ways, and it is in these differences that the ringamp finds a unique advantage in the context of modern low-voltage CMOS process technologies. In this section, we will examine several of these important benefits.

2.4.1 Output Compression Immunity

In low-voltage scaled environments, kT/C noise, SNR, and power constraints will typically be dictated by the usable signal range available [1], and therefore any practical amplification solution for scaled CMOS must utilize as much of the available voltage range as possible. As it turns out, ring amplifiers are almost entirely immune to output swing compression, and this enables them to amplify with rail-to-rail output swing.

To understand the basis of this output compression immunity, we must consider

two scenarios. First, imagine a ringamp whose dead-zone is large enough that when the ringamp is locked into the center of the dead-zone, both M_{CN} and M_{CP} will be in cutoff. In other words, when:

$$V_{DZ} \geq \frac{V_{DD} - V_{SS} - 2V_T}{A_2} \tag{2.11}$$

As a rule of thumb, this relation will usually hold for low and medium accuracy ringamps up to about 60dB. Under this scenario, the gain of M_{CN} and M_{CP} is irrelevant. The internal condition of the ringamp depends only on the signal at the input, and it will continue to steer towards the dead-zone until M_{CN}/M_{CP} completely cut-off, regardless of whether M_{CN}/M_{CP} are saturated or in triode. It will ultimately be the size of the input referred dead-zone that will determine accuracy (Eq. 2.1), independent of the characteristics at the output.

Now let's consider the condition where Eq. 2.11 does not hold. This will occur when the dead-zone is very small, and accuracies in the 60dB to 90dB range are desired. Although other practical issues in the ringamp structure of Fig. 2.3 may make such design targets less likely, a theoretical discussion is still quite useful in understanding the issues relevant to high accuracy ringamp topologies in general. In this scenario, the dead-zone is no longer a complete dead-zone, but more of a "weak-zone", where M_{CN} and M_{CP} are pinched off enough to drive the output pole to a stabilizing frequency, but still slightly on. The ringamp's steady state condition will essentially be that of a three stage opamp, and the open loop gain will be the product of the three stage gains. If the loop gain of the amplifier drops below the accuracy limitation imposed by the dead-zone, finite gain will become the fundamental limitation on accuracy. Clearly, this seems to be a possibility, since the gain of the third stage will be affected considerably by the value of the output voltage. Consider the case where all three stages have a gain of 25dB. In the best case, the open loop gain will be 75dB, and in the worst case it will be 50dB (if we assume that A_1 and A_2 are signal independent). Even in the best case, this seems to

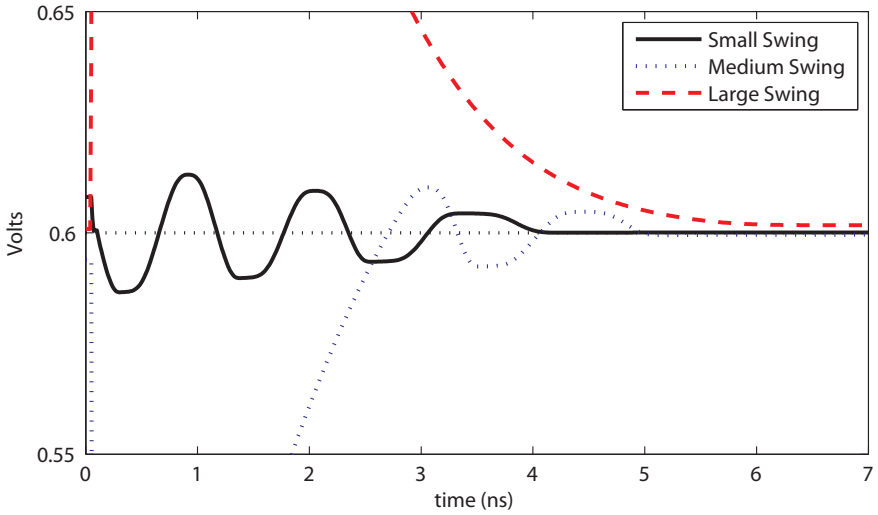


Figure 2.6: Zoomed stabilization waveform of V_{IN} for three output swing cases: small (output near mid-rail), medium, and large (output near the supply).

suggest that to build an 80dB accurate ringamp, one will need more than three stages.

Luckily, there is another effect at play here. In the ideal square-law MOSFET model M_{CN} and M_{CP} will be in saturation when V_{OV} is less than V_{DS}. Furthermore, the small signal output impedance, r_o, is inversely proportional to the drain current, I_D. In the context of the V_{OV} pinch-off that occurs in a ringamp, both V_{OV} and I_D will in fact trend towards zero. This implies that during steady-state, M_{CN} and M_{CP} will remain in saturation even for very small values of V_{DS}, and moreover, that their gain will be enhanced by a dynamic boost in r_o. Thus, even for a nominal open loop gain of 75dB, with a wisely chosen topology it is possible to have an enhanced gain of 90dB at steady-state, even when swinging rail-to-rail.

Although output swing has little effect on ringamp accuracy, it will indeed affect speed, both with respect to slewing and settling. In the initial ramping phase, the selected current source transistor will be biased with the maximum possible V_{OV}, and this guarantees that throughout much of the output range it will be operating in the triode region. As seen in Fig. 2.6, for settled output values near mid-rail, I_{RAMP} will be the highest and the initial ramping will be faster, but more time will be required to

stabilize for the reasons discussed in Section 2.3.2. Likewise, for values close to the rails, I_{RAMP} will be smaller, so the initial ramping will be slower but the stabilization time will be shorter. For the most part, this works out quite nicely, since the total time required in each case to reach steady state turns out to be approximately the same. However, for extreme cases very close to the rails, the RC time constant of the output transistor in triode operation will take a very long time to reach its target value. Ultimately, it is this RC settling that will usually dictate the maximum output swing possible for a given speed of operation.

2.4.2 Slew-Based Charging

Whereas a conventional opamp charges its output load with some form of RC-based settling, the output transistors M_{CN} and M_{CP} in the ring amplifier behave like digitally switched current sources, and charge the output with slew-based settling. This is a much more efficient way to charge, since only one of the current sources in Fig. 2.5 will be active at a time, and the only power dissipated will be dynamic. Furthermore, during the initial ramping operation, M_{CN} or M_{CP} (whichever is selected) will be biased with the maximum V_{OV} possible for the given supply voltage. This is a major benefit, because it means that even for large capacitive loads, small transistor sizes can still provide high slew rates. With small output transistors, the second stage inverters will be negligibly loaded by M_{CN}/M_{CP}, and this effectively decouples the ringamp's internal power requirements from that of the output load size. For typical load capacitances in the femto and pico-farad range, the internal power requirements are more-or-less independent of output loading. Even for very large load capacitances where the size of M_{CN}/M_{CP} does have an appreciable effect on the internal power requirements, the ratio of static-to-dynamic power will scale very favorably. In many regards, one of the "killer apps" for ring amplification lies in designs which require driving analog signals

onto large capacitive loads.

Although zero-crossing based circuits (ZCBC) ramp-charge in a similar way [23], the two techniques share little in common beyond that point. A ringamp is fundamentally defined by its stabilization mechanisms, and this is something that a ZCBC, including even the circuit in Fig. 2.5, does not possess. Whereas ringamps operate in continuous feedback, ZCBCs are inherently open loop, and must deal with many additional challenges including signal-dependent comparator decision delay, integrated current source and switch non-linearity, output swing compression, integrated noise, static offsets, and PVT variability (process, voltage, and temperature). While many effective solutions have been developed to address these concerns [25][26][28][29][30], implementing them ultimately comes at some cost in performance, complexity, or versatility. Furthermore, even in the shared task of ramping, ring amplifiers hold an advantage: whereas ramp linearity is irrelevant to ring amplifier accuracy and the current sources can be driven with a maximum V_{OV}, the overdrive applied to a ZCBC current source must be small enough to keep the current source transistor in saturation. This means that the ZCBC current source transistor must be comparatively much larger, and in order to preserve good decoupling between internal power consumption and output load size, additional buffering must be added (and thus additional comparator decision delay) in the case of the ZCBC.

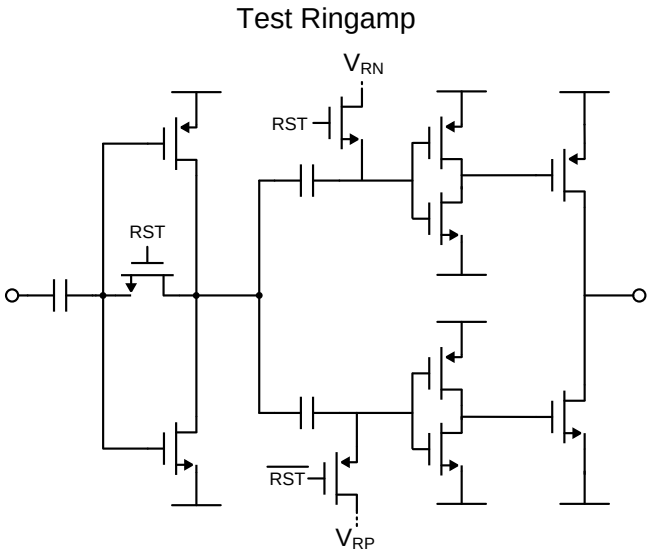
2.4.3 Performance Scaling with Process

In order to designate something as a truly "scalable" technique, it must meet two criteria. First, the given technique must operate efficiently within a scaled environment. This criterion has been the primary focus of this chapter thus far. Second, the technique must inherently improve in performance simply by migrating into a newer process technology. It is this second criteria that we will discuss now.

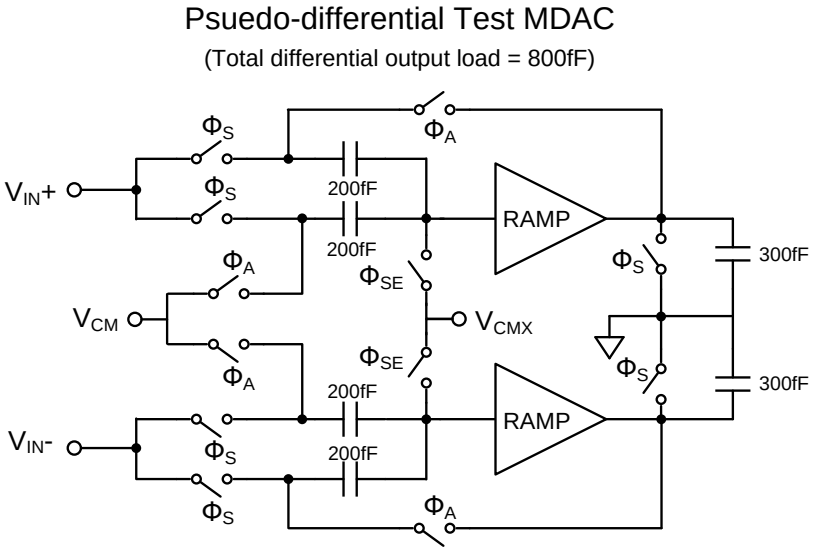
Intuitively, the ring amplifier seems like a prime candidate to benefit from process scaling, simply by its structural similarity to a ring oscillator. After all, the performance of ring oscillators track so closely with process technology that they are often used by foundries as a means of characterizing a given technology. And indeed, the stability criterion of Eq. 2.7 suggests this to be true. As we explored in Section 2.4.2, the internal power consumption in the ringamp is governed much more by the inverter power-delay product and internal parasitics than the size of the output load (in contrast to conventional opamps). Since the power-delay product of an inverter decreases approximately linearly in accordance with decreasing feature size [31], the ringamp's inverter chain propagation delay, t_d, can be expected to scale according to digital process performance as well. With the relationships in Eq. 2.7, this reduction in t_d can be directly traded for an improvement in any of the three main design specifications: speed, accuracy, and power.

The intrinsic device gains that determine the properties of A_1, A_2, $\hat{A}_1$, $\hat{A}_2$, and M_{CN}/M_{CP} will also change with scaling. However, these changes can be compensated with other design variables. Furthermore, the techniques being implemented in the newest processes, such as FinFET, are able to control short-channel effects much better than previous process nodes, and maintain surprisingly good intrinsic device gains [13].

To demonstrate this scalability property of ringamps, the results of a simple experiment designed to predict scaling trends is presented in Fig. 2.8. The setup for the test is very simple: using the ring amplifier structure and psuedo-differential switched-capacitor MDAC of Fig. 2.7, the feedback structure must be tuned to meet the required specs of Table 2.1 for several different CMOS technology nodes while minimizing total power consumption. Only transistor width and length resizing is allowed. The speed and accuracy specs are set at the upper-end of the given structure's practical limits, and the load capacitance is sized to be sufficient for an 11b pipelined ADC with 10b ENOB. Although the ringamp structure could be more efficient (for example, disabling itself during ϕ_s), we are only interested in the *relative* power efficiencies across process



(a)



(b)

Figure 2.7: Test structures used to generate the predictive scaling trends of Fig. 2.8. Two of the bare-bones single-ended ringamps of (a) are used in the pseudo-differential MDAC configuration of (b). The switches used in (b) are ideal.

Table 2.1: Design Requirements for Scaling Test

V_{DD}	process defined
L_{min}	process defined
SNDR	$>$ 66dB (input-referred)
Output Range	$0.8 \cdot V_{DD}$ (per side)
Sampling Speed	$13.5/L_{min}$ MHz
Total Load	800fF(differential)
Power	minimize

technologies, and the bare-bones structure used here illustrates the core concepts clearly.

The predictive technology models provided by [2] were used to implement the design in the 130nm, 90nm, 65nm, 45nm, and 32nm nodes. What we see from the plots in Fig. 2.8 is that the core ring amplifier structure does indeed scale exceptionally well according to process (note the logarithmic scale of the y-axis). In the upper curve, which depicts the total energy-per-cycle with respect to f_s (and thus L_{min}), the slope decreases after 45nm. This is because the ringamp's efficiency below 45nm is good enough that the dynamic switching energy begins to dominate the total energy. The fundamental lower bound for dynamic power given a certain output load can be directly calculated for this design, since the output load is reset every cycle, and is:

$$P_{CV^2f} = C_{LOAD} \cdot \left(\frac{V_{SWING}}{\sqrt{2}}\right)^2 \cdot f_s \tag{2.12}$$

which for this particular design will be:

$$P_{CV^2f} = 800\text{fF} \cdot \left(\frac{0.8V_{DD}}{\sqrt{2}}\right)^2 \cdot f_s \tag{2.13}$$

The energy-per-cycle with respect to the total power minus this CV^2f power is shown in the lower trend line of Fig. 2.8. What we see there is that the ringamp's internal

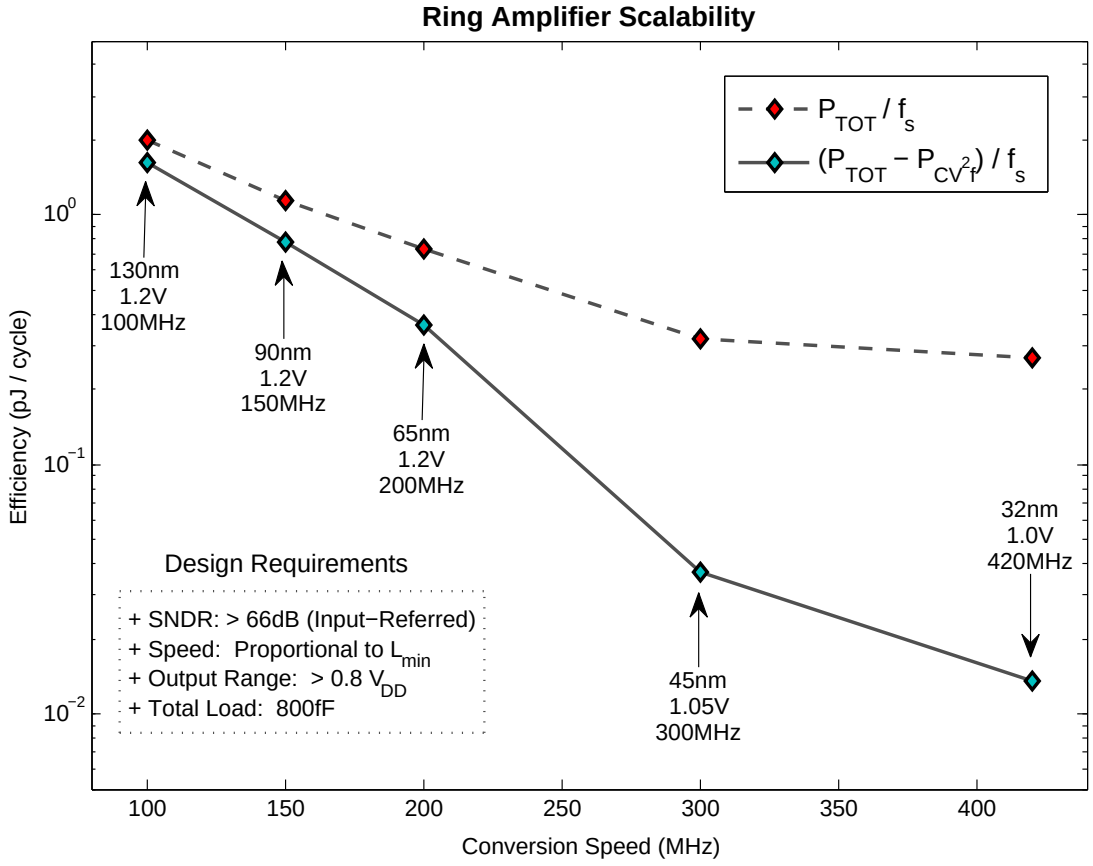


Figure 2.8: Ring amplifier scaling trends, characterized using the simple test structures of Fig. 2.7 and predictive technology models for nanoscale CMOS [2]. Device sizes are adjusted for each process as necessary to meet the fixed design requirements for each technology node.

power continues to scale at pace into deep nanoscale nodes, and somewhat remarkably, scales so well that the primary power contributor eventually becomes the ideal dynamic charging power itself. While this is a very encouraging result, there are some scaling effects that are not represented here that may cause the real trend to differ somewhat. In particular, the interconnect R, L, and C parasitics of the circuit are not modeled, which in upcoming nanoscale processes will become an increasingly dominant effect [12]. However, since digital circuits are also influenced by interconnect parasitics, whatever effect this has on ringamp performance will likely also affect digital circuits, and relative scaling trends will remain the same.

2.4.4 Noise

While accurate noise modeling of conventional opamps can often be a challenging task, performing noise analysis for a ring amplifier is considerably more complex still. Unlike an opamp (which for the sake of simplified analysis can often be approximated as a linear time-invariant system), the AC and DC characteristics of a ringamp's transistors are in a constant process of change. The power and spectrum of the noise generated at its output will be a non-linear function and subject to memory effects. Although we will not embark on the arduous journey of formulating a ring amplifier noise model in this dissertation, some comments about general noise performance trends and advantages are in order.

In the context of switched capacitor amplification, a ring amplifier's noise performance will be much better than an opamp designed for the same purpose. Although it may have worse noise at the beginning of its amplification phase, this initial time in the amplification period will have little influence on the final sampled noise. The more important time will be close to the end of the amplification period, when the ringamp is locked (or locking) into the dead-zone. Whether the output transistors (M_{CN}/M_{CP}) are completely cutoff or in just in weak inversion, the net trend will be the same: the noise generated by the ringamp when it is in the dead-zone will be comparatively quite small. Consider for a moment the thermal noise contribution of the transistor M_{CN} to the output C_{LOAD} when it is in saturation for an ideal square law model of MOSFET noise behavior [3]:

$$\bar{i}_d^2 = \frac{kT\frac{2}{3}g_m}{r_o C_{LOAD}} \tag{2.14}$$

Also, recall that

$$g_m = \sqrt{2\mu_n C_{ox}\frac{W}{L}I_D} \tag{2.15}$$

and

$$r_o = \frac{2L\sqrt{V_{DS} - \Delta + \phi_o}}{I_D k_{rds}} \tag{2.16}$$

When we combine Eqs. 2.14, 2.15, and 2.16 we see that

$$\overline{i_d^2} \propto (I_D)^{\frac{3}{2}} \tag{2.17}$$

In other words, when I_D is small, the noise at the output will also be small. A properly designed ringamp will always have a very small I_D at the end of its amplification phase, even for an implementation that doesn't entirely cut-off but instead locks into a weak-inversion mode (i.e. Chapter 7). By extension, all other noise sources in the ringamp must go through M_{CN} or M_{CP} to reach the output, and thus they will be suppressed by this effect as well. Once again, the ringamp pinch-off effect provides a valuable advantage, this time in terms of noise performance. This enables even very small and simple designs to possess very good noise suppression, and improves power efficiency as a whole.

2.5 Reflection

Looking back at the individual benefits of a ring amplifier we find many parallels with other pre-existing techniques: zero-crossing based circuits also have the ability to charge efficiently and decouple internal power from external load requirements [23][25], dynamic current adjustment and placing different offsets in multiple signal paths is a core attribute of class-AB amplifiers [3], using simple inverter-based amplifiers has long been a topic of interest [20][32][33][34], and CLS can be used to achieve rail-to-rail output swing [22]. Ultimately, it is both logical and encouraging that we should find such similarities. Each one of these parallels represent a specific trait desired in a scalable amplifier topology, and the fact that a ring amplifier encompasses all of them indicates that it is a convergence point for many lines of research in scalable amplification. In

other words, the unique aspect of a ring amplifier lies not in the traits that it possesses, but the particular solution by which it is able to embody all of them.

CHAPTER 3. SPLIT-CLS: A GENERALIZED ENHANCEMENT TECHNIQUE

3.1 Introduction

For the reasons that we have already discussed in Chapter 1, generating large open-loop gains with conventional opamps has become increasingly difficult in scaled CMOS technology, and finite gain error has become a major design challenge. There have been many proposed solutions to this challenge, ranging from digital calibration to analog techniques, including the ones mentioned in Section 1.3.1. One of the most promising solutions for this (in the realm of switched capacitor amplification circuits) is that of correlated level shifting (CLS), first proposed in [22]. CLS makes for a particularly attractive technique in scaled environments because it not only dramatically reduces finite opamp gain error from a factor of $\frac{1}{A}$ to $\frac{1}{A^2}$, but also significantly enhances the useable output swing (for example, the output swing reported in [22] was beyond even rail-to-rail). In Chapter 1 we saw that the two biggest challenges in nanoscale opamp design are arguably gain and output swing, and CLS addresses both issues. Compared to other techniques, these two enticing benefits come at a very reasonable price in terms of noise, speed, and complexity.

In this chapter, we will explore the idea of Split-CLS, a generalized form of CLS that greatly expands the range of applications which can benefit from the accuracy and output swing enhancements provided by CLS. This broader spectrum of possibilities also comes with the potential for higher accuracy, speed, and power efficiency than would be possible with standard CLS. In Chapters 4 and 6, we will study two possible configurations within this spectrum, and see how Split-CLS allows emerging technologies

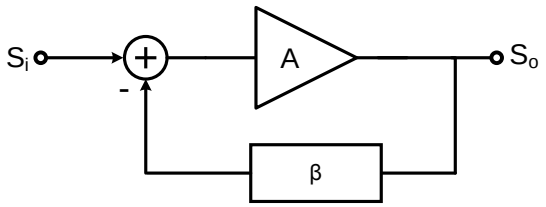


Figure 3.1: Basic signal flow diagram of an opamp in feedback.

such as ring amplification to be used alongside a conventional opamp. In a very practical sense, Split-CLS's ability to combine old and new provides an important level of flexibility to designers looking to make market-ready products. It allows their projects to benefit from newer techniques, while helping to mitigate the risks associated with unproven technologies by allowing conventional opamps to remain the final judge of accuracy and noise in the system. Split-CLS can also be applied to structures consisting entirely of conventional opamps. There is a large body of knowledge that has been accumulated in the last 40 years of opamp design, and rather than throw it all away in nanoscale CMOS, Split-CLS allows us to continue to benefit from the past as we look towards the future.

In the first half of this chapter we will look at the basic theory of operation behind CLS and Split-CLS. In the second half, we will perform noise analysis on a Split-CLS network and see how it affects the overall noise requirements in a switched capacitor feedback network. In Chapters 4 and 6 we will look at two different designs which use Split-CLS, and the specific amplifier choices and implementation details of each Split-CLS design will be covered in their respective chapters.

3.2 Finite Opamp Gain Error

Switched capacitor amplification structures are a nearly ubiquitous building block in the design of modern CMOS analog circuits. These methods allow designers to perform high accuracy, discrete time mathematical functions in the analog domain. The accuracy of this operation is often limited by finite loop gain, manifesting as an effect commonly

referred to as finite opamp gain error. It arises from the fact that, with a finite opamp gain, there must be *some* amplitude of signal at the input of the opamp in order to create the correct amplitude at the output. The ideal feedback system, however, requires zero signal amplitude at the input of the opamp, and thus an error results. In mathematical terms, the input-to-output transfer function of the ideal signal flow diagram given in Fig. 3.1 is:

$$\frac{S_o}{S_i} = \frac{1}{\frac{1}{A} + \beta} \tag{3.1}$$

where A is the gain of the opamp and β is the feedback factor. We see that under ideal circumstances, when $A \to \infty$, the transfer function is simply $\frac{1}{\beta}$. However, with finite gain, there will also be a $\frac{1}{A}$ term, and this term will cause a signal-dependent error of the same amount in the final amplified value. Due to its signal dependence, finite opamp gain error must be either suppressed or canceled in order to minimize non-linearity and distortion in the system. Although there have been some successful digital schemes which cancel the error, there are several benefits to suppressing the error in the analog domain before it is even created. The CLS and Split-CLS techniques that we will discuss here follow this analog approach.

3.3 Basic CLS Structure

The basic CLS MDAC structure introduced in [22] and shown in Fig. 3.2 can be used to reduce finite opamp gain error and increase the opamp's useful output swing. After sampling the input (ϕ_S), the MDAC switches into an amplification configuration (ϕ_A). CLS occurs during this amplification phase in two steps: estimation (ϕ_{EST}) and level shifting (ϕ_{CLS}). In ϕ_{EST} the opamp is connected directly to the output, which allows the CLS capacitor to sample an estimate of the correct output voltage with respect to the opamp output common mode (or any other analog reference). In ϕ_{CLS}, C_{CLS}

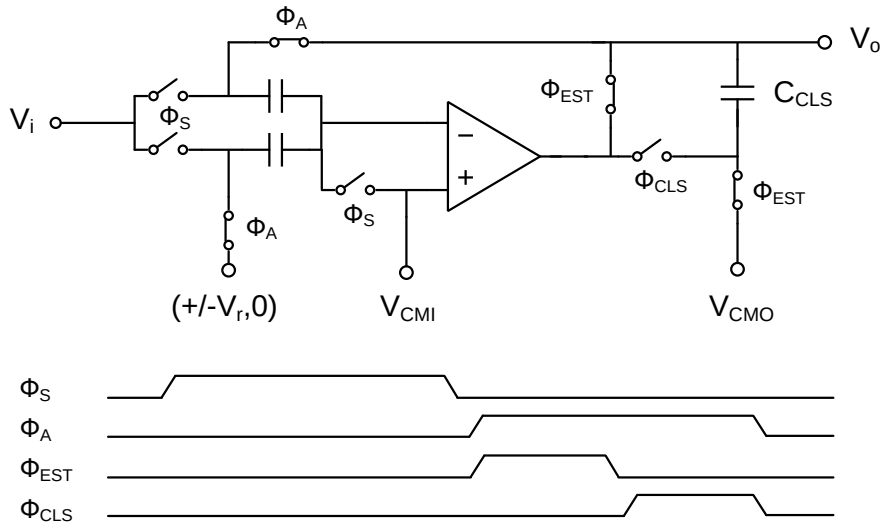


Figure 3.2: Correlated level-shifting (CLS) structure.

is connected between the MDAC output and the opamp output, which level-shifts the opamp output back to V_{CMO}. The opamp now only processes the error of the initial estimate (the signal has been canceled in the feedback path), which will reduce the finite opamp gain error and requires only enough opamp output swing as is necessary to correct this error. Using the same set of assumptions and simplifications made in the derivation of equation 22 of [22] and adding an additional assumption that the opamp is single stage (which we will later see is true for our use), we see that the final effective gain (A_{EFF}) of the structure is related to the opamp gain during ϕ_{EST} and ϕ_{CLS} as

$$A_{EFF} \propto A_{EST} \cdot A_{CLS} \tag{3.2}$$

and thus the finite opamp gain error will be dramatically reduced from a factor of $\frac{1}{A}$ to something roughly equivalent to $\frac{1}{A^2}$.

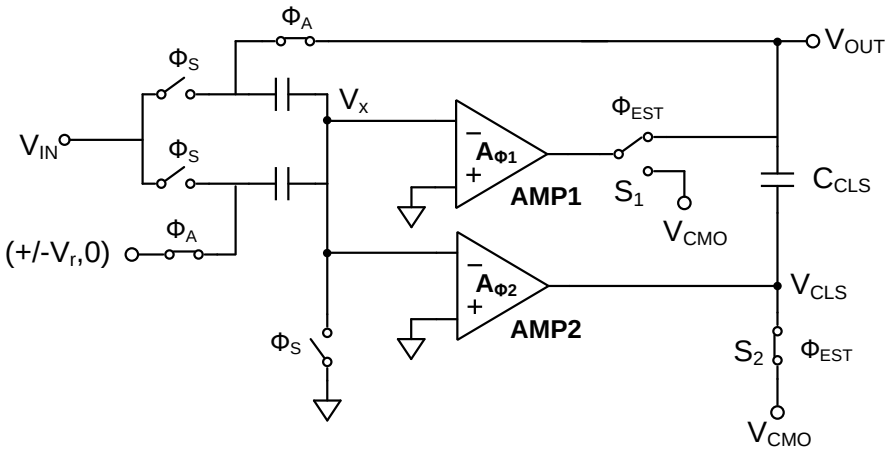


Figure 3.3: The basic Split-CLS structure. An estimate of the signal is generated during ϕ_{EST} and sampled onto C_{CLS}, which is then placed in the feedback path in order to cancel finite opamp gain error during ϕ_{CLS}. The effective gain at the end of ϕ_{CLS} is proportional to $A_{\phi 1} \cdot A_{\phi 2}$.

3.4 Split-CLS

A useful observation that we can make about CLS is that the gains A_{EST} and A_{CLS} do not necessarily need to be the same or even come from the same amplifier. Moreover, the amplifier requirements for ϕ_{EST} and ϕ_{CLS} differ. By splitting the amplifier in Fig. 3.2 into two separate amplifiers and then designing each amplifier with its specific requirements in mind, we can improve the overall performance of CLS in terms of power, speed, and accuracy. These observations lead us to the technique of Split-CLS. Fig. 3.2 shows a generic single ended Split-CLS structure. The amplification phase, ϕ_A, is sub-divided into an estimation phase and a fine settling phase. In the first sub-phase, the switches S_1 and S_2 are asserted (i.e. the configuration shown in Fig. 3.2), and AMP1 charges an estimate of the final settled value onto the output. Meanwhile, the capacitor C_{CLS} samples this estimation value and the output of AMP2 is shorted to mid-rail and held in standby. In the second sub-phase, the switches S_1 and S_2 are de-asserted and AMP1 is disconnected from the output (and can be disabled). Meanwhile, AMP2 is coupled into the output via C_{CLS} and begins to fine-settle the output towards its ideal

value.

In this new split configuration of CLS, let's revisit the design requirements for each amplifier. AMP1 processes the full signal during ϕ_{EST} and is directly connected to the output load. Therefore, for optimum performance it should have as large of an output swing as possible and high slewing capabilities. By contrast, AMP2 must only process the small error term remaining after ϕ_{EST} and charges the output indirectly through C_{CLS}, and if C_{CLS} is sized appropriately, the output swing and slew requirements for AMP2 are much smaller than for AMP1. It is important to notice here that only AMP2 needs to be an opamp in order to benefit from Split-CLS; AMP1 simply charges an estimate of the final output value, and can be any sort of charging device, not necessarily an opamp. With all this in mind, we can see that a ringamp is a near optimal choice for the wide-swing, efficient-charging AMP1, and that AMP2 should be a small-swing fine settling device such as a telescopic opamp. We will look further into implementation options in Chapters 4 and 6.

With regard to a comparison between CLS and Split-CLS, a few more observations are in order. First, the opamp outputs are connected to different nodes in Split-CLS, which comes with some advantages over CLS. To begin with, we can move the switch at the output of AMP2 outside of the signal path. Switch S_2 is used during ϕ_{EST} to both sample C_{CLS} and short the opamp output. Without this shorting switch, AMP2's output would swing heavily during ϕ_{EST} due to AMP1's settling of V_X. This swinging would in turn affect the voltage at V_X via parasitic kickback through AMP2's input transistors, creating a second feedback loop and potential settling issues. Likewise, in the generalized form of Split-CLS, switch S_1 must be used at the output of AMP1 during ϕ_{CLS}, but as we will see later, if the device used for AMP1 can turn its output off internally, S_1 is unnecessary.

Another difference between Split-CLS and CLS is the transient effects that occur during the transition from ϕ_{EST} to ϕ_{CLS}. In traditional CLS, at the beginning of ϕ_{CLS} the amplifier output will quickly transition from the full-swing estimation voltage back to

V_{CMO}, causing a signal dependent glitch at node V_{CLS} that can degrade settling speed. The size of this glitch depends on the swing of the opamp during ϕ_{EST} and the size of the opamp's internal capacitances. While it was mentioned in [22] that this glitch can be tolerated with careful sizing of the opamp transistors, this limits the opamp design space, and it is not easy to predict the actual parasitic effects of such subtle parameters. Split-CLS doesn't have this glitch problem because AMP2 is held during ϕ_{EST} at the exact bias condition which we desire it to be at when beginning ϕ_{CLS}; upon S_2 opening, AMP2 immediately begins settling the MDAC output where AMP1 left off.

3.5 Noise Analysis

The size of C_{CLS} affects many overall aspects of performance, sometimes in complex ways, since it influences both the feedback factor and total charging (load) capacitance of the structure. For example, decreasing C_{CLS} increases the loop bandwidth of AMP2 and reduces the total capacitance which AMP1 must charge, but also decreases the loop gain (and thus the overall accuracy) and increases the output swing requirement of AMP2. Beyond the effect of C_{CLS} on efficiency and accuracy, an important question to answer is the effect of C_{CLS} on noise requirements. The noise from AMP1 during ϕ_{EST} is suppressed by the gain of AMP2 [22], which means that AMP2 is by far the more dominant source of noise at the end of ϕ_a when the final output voltage is sampled. In this section, we will examine the effect of C_{CLS} on the noise due to AMP2 in order to gain a better understanding of the key noise considerations unique to the design of CLS amplification.

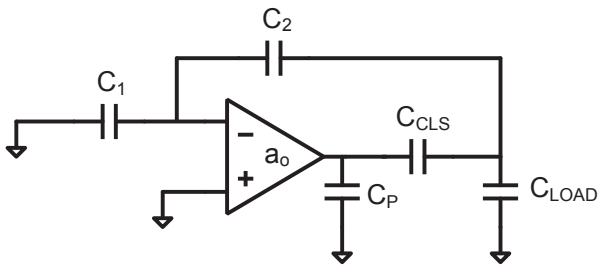


Figure 3.4: CLS configuration during ϕ_{CLS} used for this analysis. Unless stated otherwise, all numerical results in this section are found for C_1=C_2=400fF, C_P=45fF, C_{LOAD}=640fF, and a_o=58.4dB. The values for C_P and a_o are extracted from the transistor level implementation of the opamp of Fig. 3.6.

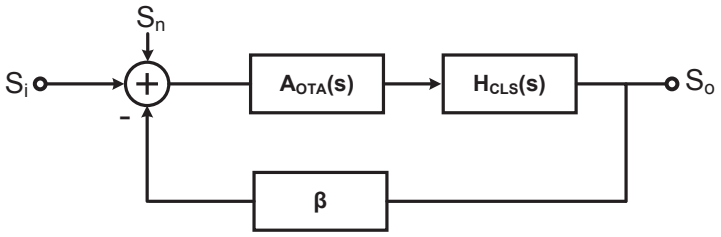


Figure 3.5: Equivalent signal flow diagram for the structure given in Fig. 3.4

3.5.1 Split-CLS Noise Analysis

The CLS configuration during ϕ_{CLS} is shown in Fig. 3.4. An equivalent signal flow diagram of Fig. 3.4 is given in Fig. 3.5. The capacitor C_P represents the lumped internal capacitances of the opamp as seen from its output. C_{LOAD} is the capacitor which samples the final output voltage. In this analysis we will consider a pipelined ADC, where C_{LOAD} is the sampling capacitance of the next stage MDAC and typically given by

$$C_{LOAD} = \alpha(C_1 + C_2) \tag{3.3}$$

where α is a stage-scaling factor. In this analysis, α is chosen to be 0.8. As with a standard MDAC, the feedback factor, β, is:

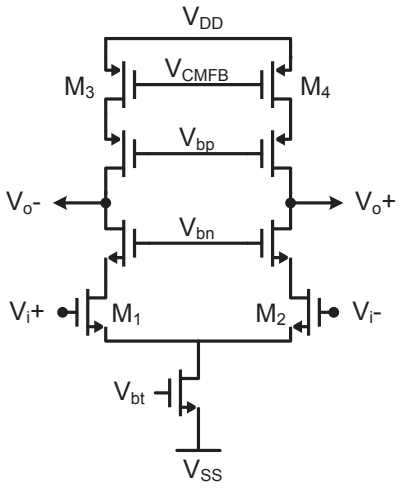


Figure 3.6: Cascoded telescopic opamp used for analysis and transistor level simulation. The dominant noise sources are transistors M1-M4, which can be modeled as a single noise source at the opamp input that is filtered to the output by (3.8).

$$\beta = \frac{C_2}{C_1 + C_2}. \tag{3.4}$$

Furthermore, the transfer function of the CLS network is found to be

$$H_{CLS}(s) = \frac{C_{CLS}}{C_{CLS} + C_{LD}} \tag{3.5}$$

where C_{LD} is the total capacitance at the output of $H_{CLS}(s)$:

$$C_{LD} = C_{LOAD} + \frac{C_1 \cdot C_2}{C_1 + C_2}. \tag{3.6}$$

The behavior of $A_{OTA}(s)$ depends on the type of opamp used. A cascoded telescopic opamp is an ideal architecture to use for AMP2 in Split-CLS[22], and the one shown in Fig. 3.6 is used for this analysis. This opamp has a single dominant pole defined by its output impedance, R_O, and its total output capacitance, C_{OTA}:

$$C_{OTA} = C_P + \frac{C_{CLS} \cdot C_{LD}}{C_{CLS} + C_{LD}} \tag{3.7}$$

$$A_{OTA}(s) = \frac{a_o}{1+\frac{s}{p_1}}\Bigg|_{p_1=\frac{1}{R_O C_{OTA}}} \tag{3.8}$$

where a_o is the dc open-loop gain of the opamp and R_O is the opamp output impedance. The gain of the forward path for this system is

$$A_{fp}(s) = A_{OTA}(s) \cdot H_{CLS}(s) \tag{3.9}$$

and the noise transfer function (which is also the overall closed-loop gain of the system) is

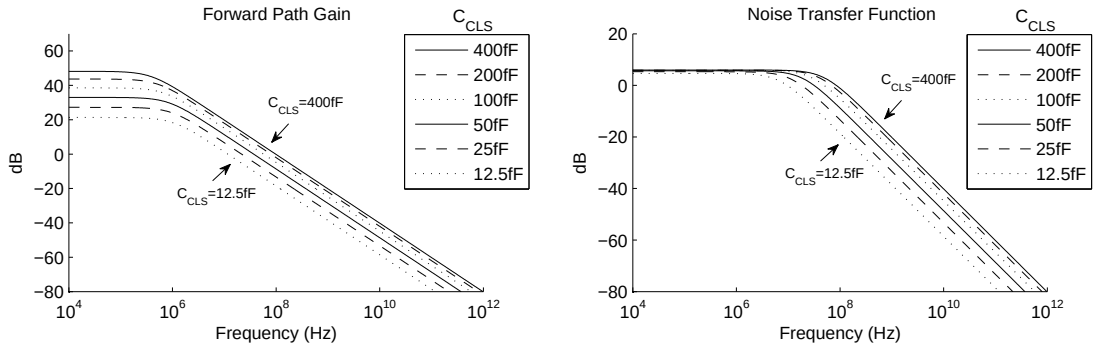
$$H_n(s) = \frac{A_{OTA}(s) \cdot H_{CLS}(s)}{1 + \beta \cdot A_{OTA}(s) \cdot H_{CLS}(s)}. \tag{3.10}$$

The forward path and overall gains for several values of C_{CLS} are shown in Fig. 3.7. In a typical feedback system, the gain-bandwidth product is a constant[3]. As Fig. 3.7(a) shows, this is not the case when C_{CLS} is varied. If C_P were equal to zero, the gain-bandwidth product would be constant for all values of C_{CLS} because the factor

$$\frac{C_{CLS}}{C_{CLS} + C_{LD}} \tag{3.11}$$

is found in both (3.5) (which affects the forward path gain) and (3.7) (which affects the pole frequency) and will cancel each other out in the gain-bandwidth product. However, because C_P only affects (3.7) and (3.8) and not (3.5), the gain-bandwidth product varies with respect to C_{CLS}. The effect that this dependency on C_{CLS} has on the overall noise transfer function is seen in Fig. 3.7(b).

In the process of designing a CLS structure, it is ultimately the degree to which C_{CLS} affects the noise sampled onto C_{LOAD} that we are interested in knowing. This integrated noise power at the output is found by passing the spectrum of the noise source through the noise transfer function and integrating the noise power:



(a) (b)

Figure 3.7: Magnitude response curves for C_{CLS} ranging from 12.5fF to 400fF for C_P=145fF for a) the forward path gain described by (3.9) and b) the noise transfer function $H_n(s)$ described by (3.10). As illustrated by the bandwidth spreading seen in these figures, the gain-bandwidth product is not constant with respect to C_{CLS}.

$$\tilde{v}_{no}^2 = \int_0^\infty S_n(f) \cdot |H_n(2\pi f)|^2 \, df. \tag{3.12}$$

where $S_n(f)$ is the power spectrum of the noise source.

The shape of $S_n(f)$ is not white for an opamp, including the one of Fig. 3.6. Furthermore, there are many noise contributors within the opamp, with each one generating noise. Different noise sources take different paths to the output of the opamp, and so the filtering of each noise source as it appears at the output must be examined on a case-by-case basis. Conveniently, for the opamp of Fig. 3.6, transistors M1-M4 are by far the dominant noise sources, and our analysis can be simplified to the noise contribution of only these transistors. M1-M4 are all filtered on their way to the output by (3.8) (or something very similar in the case of M3 and M4). For this reason, the total opamp noise can be idealized as a single noise generator at the input of the opamp, such as indicated in Fig. 3.5.

For most designs, the spectrum of the noise consists of two main components - white noise effects (i.e. thermal and shot noise) and low-frequency $1/f$ noise (i.e. flicker noise). For practical values used in a typical pipelined ADC, such as the values chosen in this discussion, the -3dB frequency of the noise transfer function is much larger than

the frequency where the $1/f$ noise effects cease to be a significant contributor to noise power. In this case, the shape of the low-frequency noise is unaltered by $H_n(s)$ since the magnitude response of $H_n(s)$ is flat in this frequency range; the non-white noise components manifest themselves in the final integrated output noise as a constant offset. By making these reasonable assumptions, the white and $1/f$ noise components can be separated as such:

$$\tilde{v}_{no}^2 = v_{n(1/f)}^2 + \int_0^\infty S_{n(white)} \cdot |H_n(2\pi f)|^2 \, df \tag{3.13}$$

where $v_{n(1/f)}^2$ is the total integrated noise power of the low-frequency $1/f$ noise component and $S_{n(white)}$ is the noise power per hertz of the white noise component. This equation for integrated output noise is particularly convenient to use because the input noise source is expressed as two frequency-independent constants. For designs where the shape of low-frequency noise is altered by $H_n(s)$, the complete noise spectrum calculation of (3.12) should be used instead.

3.5.2 Numerical Results

The total integrated noise sampled by C_{LOAD} described by (3.13) is plotted in Fig. 3.8 for many values of C_{CLS} and C_P. When the condition

$$C_P >> \frac{C_{CLS} \cdot C_{LD}}{C_{CLS} + C_{LD}} \tag{3.14}$$

is true, the forward path gain pole frequency is mostly defined by the fixed capacitance C_P, so the total integrated noise and C_{CLS} become correlated due to the effects considered in Section 3.5.1. Likewise, when

$$C_P << \frac{C_{CLS} \cdot C_{LD}}{C_{CLS} + C_{LD}} \tag{3.15}$$

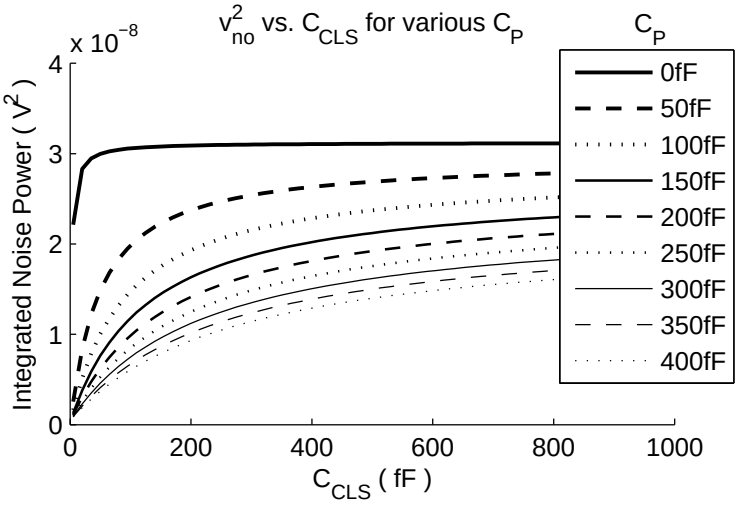


Figure 3.8: The relation between opamp noise at the output and C_{CLS} is shown for many values of C_P. For most designs the condition in (3.15) will be true and noise will not be strongly affected by C_{CLS}.

the pole frequency is most influenced by the factor in (3.11) and leads to little variation in the gain bandwidth product and noise. To ensure a low-swing requirement for AMP2 in Split-CLS and achieve a high level of gain-enhancement, the value of C_{CLS} should usually be on the same order of magnitude as the values of C_1 and C_2 or larger (i.e. 400fF). Moreover, because opamp efficiency and speed is maximized by minimizing it's capacitive load, for optimum performance C_P should be minimized and consist only of the unavoidable parasitic capacitances of the opamp itself. Therefore, the region described by (3.15) is the most commonly encountered scenario in actual designs. The resulting implication is that for most practical designs, there will not be a strong correlation between C_{CLS} and noise sampled at the output, which simplifies the considerations a designer must make when building a CLS system.

Although the derivation of this section was made for a Split-CLS architecture, the theoretical model will not be drastically altered for most other forms of CLS, and the same key conclusions found here are also applicable in a more general sense.

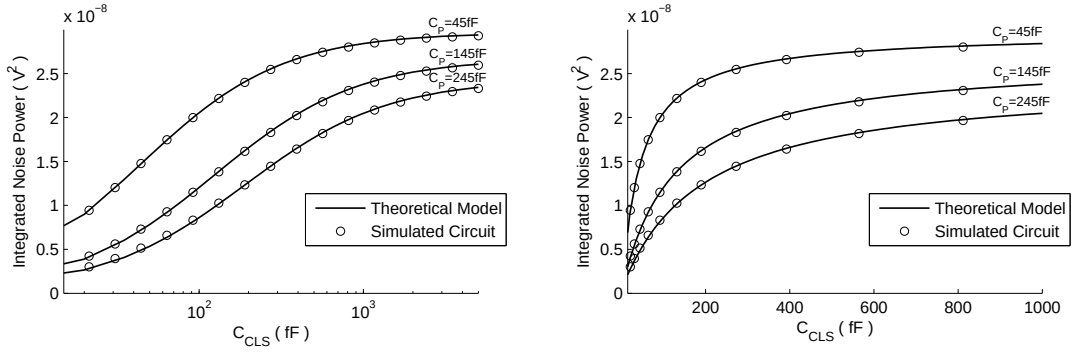


Figure 3.9: Comparison of the theoretical model derived in this section with transistor level simulation in a 0.18μm CMOS process using the opamp of Fig. 3.6 for C_P=$\{$45fF, 145fF, 245fF$\}$. The two are compared on (a) a log scale for C_{CLS} ranging from 5fF to 5000fF, and (b) a linear scale for C_{CLS} from 5fF to 1000fF.

3.5.3 Comparison with Simulated Results

To verify the theoretical model derived in this section, a transistor level simulation of the circuit in Fig. 3.4 was compared with theory [35]. Shown in Fig. 3.9, simulated and theoretical curves are superimposed together and found to match up very closely. The three curves correspond to three values of C_P (45fF, 145fF, and 245fF), with the minimum value C_P=45fF being defined by the extracted parasitic output capacitance of the transistor level opamp in the simulation environment. The extracted transistor-level opamp DC gain of 58.4dB was used for a_o, C_1=C_2=400fF, and C_{LOAD}=640fF in both setups. For the theoretical model, the input-referred noise values $v^2_{n(1/f)} = 6.31 \times 10^{-10}\ V^2$ and $S_{n(white)} = 6.17 \times 10^{-17}\ \frac{V^2}{Hz}$ were used for all three curves. These flicker and white noise values were extracted from the transistor level opamp measured in open-loop. The noise contribution of all opamp transistors were accounted for and all other noise sources such as switches and resistors were excluded (in both theoretical and simulated results).

On the whole, it is found that for most practical designs, there is only a weak correlation between total integrated output noise and the size of C_{CLS}. This conclusion

simplifies the interaction between design variables, and allows the size of C_{CLS} to be chosen based on design parameters other than opamp noise, simplifying the task of the designer wishing to implement such techniques.

CHAPTER 4. A 11.5B SPLIT-CLS ZCBC/OPAMP ADC

4.1 Introduction

In this chapter, we explore one possible implementation of the Split-CLS technique introduced in Chapter 3 in the form of a pipelined ADC. Although from a technical standpoint the Split-CLS design presented later on in Chapter 6 is by all accounts superior to this one (and for the reader on a tight schedule, Chapter 4 can be skipped), this chapter may help the reader to understand the underlying motivation behind the invention of ring amplification. Chronologically, of the ideas and designs explored this dissertation, this is the earliest, appearing in [36] and [28]. The challenges and drawbacks of designing the zero-crossing based circuit in this design, and the lessons learned in retrospect, were a key inspiration behind the invention of ring amplifiers. Indeed, the ringamp was first conceived of after this design had been completed, while searching for a more ideal coarse-charging device to use in Split-CLS.

4.2 Opamp/ZCBC Split-CLS

There are many possible amplifier topologies that could be promising choices for AMP1 and AMP2 in Split-CLS (Fig. 4.1). In this section we will explore the design of one possible configuration, the Opamp/ZCBC Split-CLS structure.

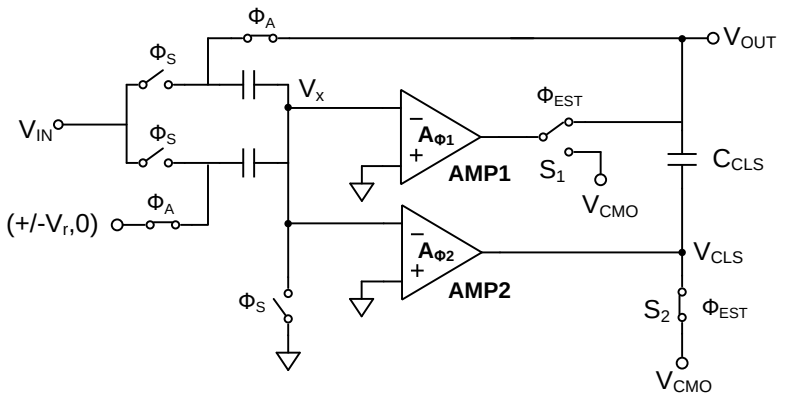


Figure 4.1: Split-CLS structure.

4.2.1 AMP1

As mentioned in Section 3.2, designing a high gain amplifier to have large output swing in deep sub-micron technologies using traditional opamp topologies is difficult. Ultimately, it will probably require a multi-stage approach so that the output stage can be a single PMOS/NMOS pair. The compensation required to make such an amplifier stable will place an upper bound on the best case power-to-speed ratio. On the other hand, if we consider some alternative amplification techniques, we may get better results. A ZCBC is a particularly attractive AMP1 choice. Unlike traditional opamps, the power requirement of the sense circuitry (the ZCD) is ideally unrelated to the size of the output load, and in practice only weakly related. For this reason, ZCBCs are capable of excellent slewing efficiencies. Another key benefit of using ZCBC amplification is that the current source and ZCD can be shut off after ϕ_{EST}, which will save considerable power compared to using an opamp for AMP1. Since the current sources shut off internally, S_1 of Fig. 4.1 is no longer needed.

A major challenge in ZCBC designs is achieving high-accuracy amplification at high speed. Any signal-dependent effects in the ZCD decision time-delay, output current source, or switches will directly impact the overall accuracy. Slower operation reduces the influence of all of these problems, but ideally we would like AMP1 to charge the output

very quickly. Deterministic errors are not the only challenge, however. The open-loop nature of ZCBC makes it susceptible to internal and external transient variations and integrated noise. Whereas an opamp in feedback can suppress these memory effects, a ZCBC is not always able to.

Ultimately, the strength of pure ZCBC appears to lie more in its high power efficiency at medium to high speeds with medium resolution than in the realm of high accuracy. For a conventional single-shot differential ZCBC implementation such as [25], at low speeds, high accuracy is less attractive because of the higher ratio of ZCD static current to current source dynamic current. At higher speeds, high accuracy is hard to achieve due to the linearity challenges mentioned previously. We can relax many of the accuracy challenges of ZCBC by treating it as a coarse, fast, high efficiency charging device, and leave the high accuracy settling to an opamp (AMP2) which is better suited to such a task. From the traditional amplification perspective, this approach can be seen as a way to improve the slewing efficiency of an opamp. From the perspective of ZCBC, this can be seen as a method for extending the use of ZCBC into higher resolution applications by alleviating many of its critical issues through the application of linear feedback.

This idea of using a coarse charging device paired with an opamp was previously explored in [37] by connecting both coarse and fine charging devices directly to the output. Although this could potentially improve power efficiency, errors from low opamp gain and low output swing are not reduced like they are with CLS. In addition, the opamp must be designed so that it will not charge the output while the coarse charging device is operating. By contrast, AMP2 in Split-CLS is held in a fixed state during ϕ_{EST} by simply shorting its output.

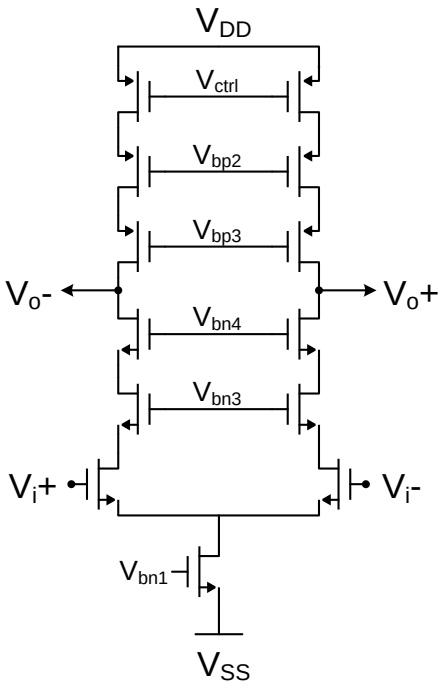


Figure 4.2: Double cascode opamp used for AMP2.

4.2.2 AMP2

The device chosen for AMP2 in the implemented Split-CLS structure is a double-cascode telescopic opamp, shown in Fig. 4.2. This structure's merits lie in its high speed and high gain relative to power consumption. It is much easier to guarantee stability for a single stage opamp than for a two-stage opamp. By trading output swing for gain with additional cascode transistors, high gain is obtained without the use of gain-boosting amplifiers or other techniques that would require additional power. Typically, the drawback of a telescopic opamp is that it has a small output swing due to the cascoding transistors and a low slew rate due to the lack of a high current output stage. However, as discussed earlier, AMP2 requires neither high output swing nor high slew rate. The opamp employs the switched capacitor common-mode feedback of [4].

The amount of error remaining at the end of ϕ_{EST} and the size of C_{CLS} determine AMP2's output swing requirement. C_{CLS} also affects the effective gain of the overall

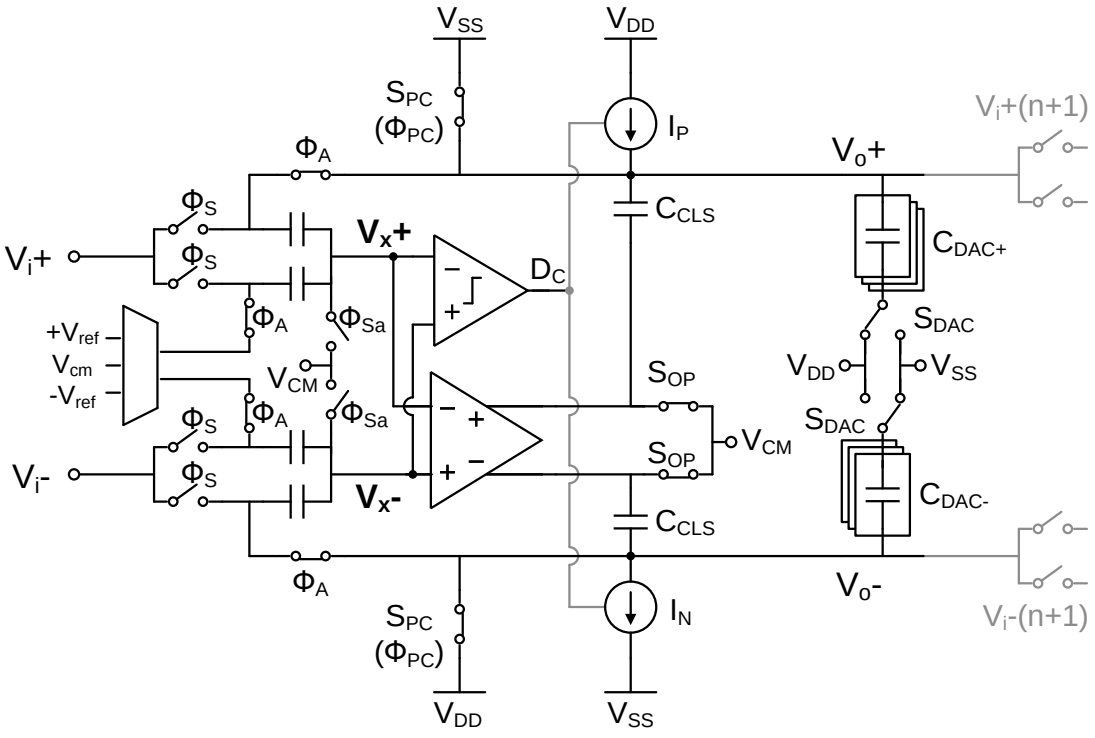


Figure 4.3: Opamp/ZCBC Split-CLS structure.

structure, and for a single stage amplifier, increasing C_{CLS} will increase effective gain (at the cost of higher output load) [22]. Therefore, after sizing C_{CLS} to meet the output swing requirement, only if more effective gain is needed in the structure should the value of C_{CLS} be increased further.

4.2.3 Full Structure

The Opamp/ZCBC Split-CLS MDAC amplifier structure is presented in Fig. 4.3 with corresponding timing and waveform diagrams in Fig. 4.4. In addition to the opamp and ZCBC circuitry, the structure contains a set of capacitive DACs which are used to cancel ZCBC overshoot prior to the start of opamp settling (explained in Section 4.3).

During ϕ_S the input capacitors sample the input signal while the ZCD and current sources are shut off and the opamp idles. At the beginning of ϕ_A the amplification

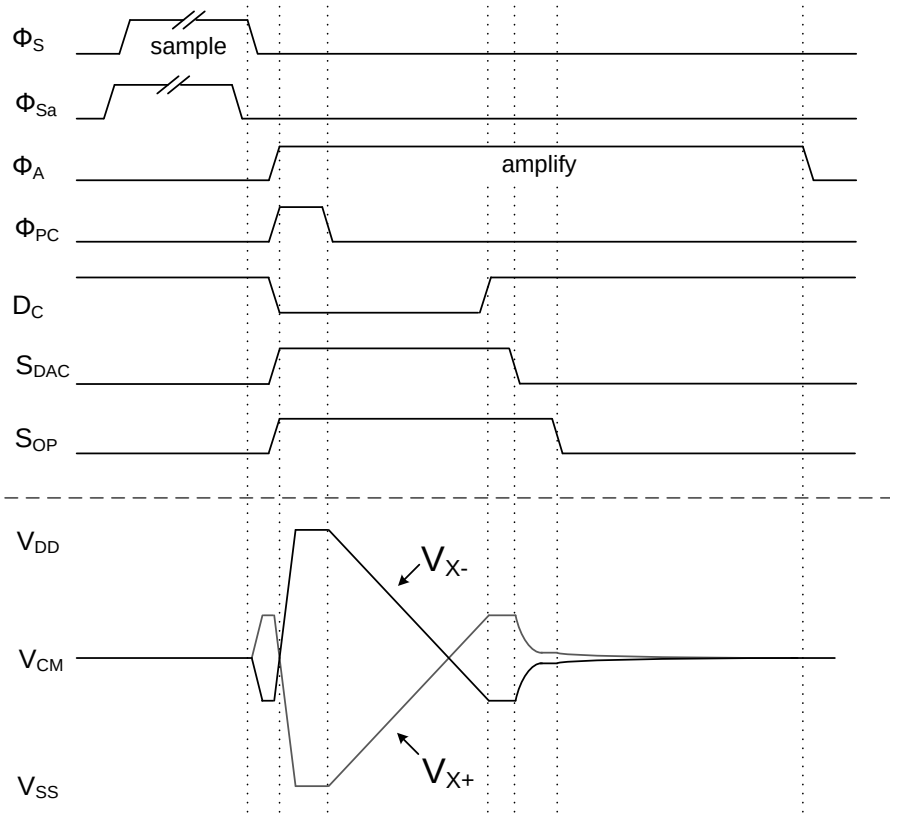


Figure 4.4: Timing and waveforms of the Opamp/ZCBC Split-CLS structure.

operation begins with a short pre-charge phase ϕ_{PC}, during which several things occur simultaneously: the pre-charge switches (S_{PC}) charge the output nodes to V_{DD} and V_{SS}, the current sources (I_N, I_P) are turned on, the ZCD internally switches into a startup configuration (discussed in Section 4.4), the opamp output and bottom plate of C_{CLS} is connected to a low-impedance reference through S_{OP}, and the overshoot cancellation DACs (C_{DAC}) are connected to the output. When the pre-charge switches open at the end of ϕ_{PC} the current sources freely charge the output load with a linear ramp until the zero crossing of V_X+ and V_X- is detected by the ZCD and the current sources are switched off. When the digital output of the ZCD flips, an asynchronous timing block is activated. This block facilitates the cancellation of the ZCBC overshoot by flipping the S_{DAC} switches to the alternate supply voltage. After the overshoot cancellation

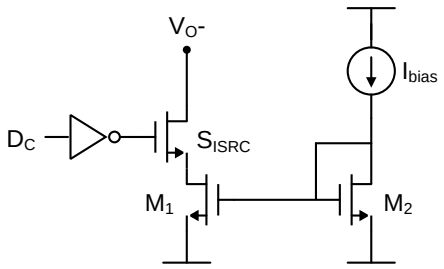


Figure 4.5: ZCBC current source I_N.

is settled, the timing block disconnects C_{DAC} from the output (in order to minimize the capacitive load that the opamp must drive) and opens the shorting switches S_{OP}, allowing the opamp to begin freely settling to the final output voltage until the end of ϕ_A.

4.3 Implementation Details

4.3.1 Current sources

One of the current sources used in the ZCBC implementation (I_N) is shown in Fig. 4.5. The design of I_P is an analogous PMOS current source. Transistor S_{ISRC} is designed to operate more like a switch than a cascode device in order to maximize the ZCBC output swing. As a fast, coarse charging device, a single transistor current source provides sufficient linearity. By switching the current source at its drain, there will be some initial non-linearity as the inversion layer of S_{ISRC} is created and the drain of M1 rises from V_{SS} to a voltage slightly below V_O-. It is for this reason that digital signal D_C (generated by the ZCD) transitions low at the beginning of ϕ_{PC} while the output is being shorted to a supply, allowing the initial switching transients to resolve and ensuring a linear ramp on the output right after ϕ_{PC} goes low.

4.3.2 Switches

Unlike opamp settling, where the amount of current (and any resulting IR drop) flowing through the switches in the MDAC will converge to zero as the output settles, in ZCBC the current through the switches remains high. Even as a coarse charging device, switch linearity during the ZCBC operation is something that must be addressed.

The bootstrapping switch of [38] was used to create a small signal-independent on-resistance for switches in the feedback path. All of the switches that are closed during ϕ_S are bootstrapped. Even though the bottom plate sampling switches controlled by ϕ_{Sa} have a fixed gate-to-source voltage, that voltage is only V_{DD} - V_{CM}. By bootstrapping the switch, V_{GS} is made twice as large, yielding an overall reduction in switching power compared to using a single transistor switch. Of the switches that are asserted in ϕ_A, only the ones in the MDAC feedback loop are bootstrapped. The MDAC reference voltage switches used in ϕ_A already have constant and sufficiently large V_{GS}. S_{OP} is also outside of the feedback path and with fixed V_{GS}, so it is implemented as a single-transistor switch.

4.3.3 Overshoot cancellation

The primary challenge of pairing a ZCBC with an opamp is the differential output offset (or "overshoot") introduced by the finite time delay between when the inputs of the ZCD cross to when the current sources turn off. In a pure ZCBC pipelined ADC this overshoot is passed down the pipeline and the signal-independent portion of it will become a DC offset in the final digitized signal. This static offset can be canceled on a global level, and one such approach to this is shown in [25]. Depending on the ZCBC accuracy requirements, a global correction like this is not necessarily possible in the case of Split-CLS; the overshoot must be canceled completely before being passed on to the

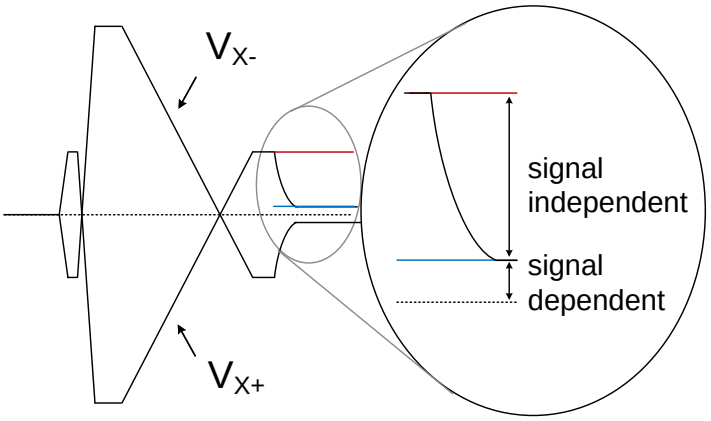


Figure 4.6: All components of ZCBC overshoot will be processed by the opamp as error. To minimize this error, we must cancel the signal-independent portion of the overshoot as best as possible prior to level-shifting.

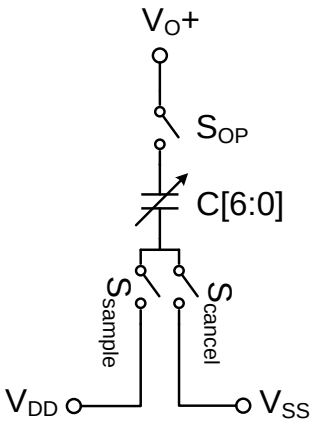


Figure 4.7: Programmable overshoot cancelation DAC.

next stage. This is because the opamp will attempt to settle such that $V_X- = V_X+$. Any differential DC offset at these nodes will be processed by the opamp as if it were error. In practice, this overshoot can be several hundred millivolts, and if the opamp is left to cancel it directly, we must sacrifice an inordinate amount of either opamp output swing or increased size of C_{CLS} in order to keep the opamp transistors in saturation - negating our previous assumptions and benefits related to the low-swing and low-slew requirements for the opamp.

As shown in Fig. 4.6, the overshoot consists of a signal-independent portion and a

smaller signal-dependent portion. The signal-dependent portion is not easily predictable or cancelable, but the signal-independent portion is. Fig. 4.6 illustrates the cancelation method implemented in this design - the explicit cancelation of the overshoot voltage. Cancelation with the digitally programmable capacitor DAC of Fig. 4.7 has the benefit of flexibility in a test environment, and was chosen over a single capacitor for this reason. Explicit cancelation also makes the Split-CLS structure compatible with integrators, where the accumulation of ZCBC overshoot can quickly saturate the integrator output. The DAC illustrated in Fig. 4.7 is C_{DAC+} of Fig. 4.3. C_{DAC-} is similar, but with V_{DD} and V_{SS} swapped on the lower reference nodes.

Rather than canceling the overshoot explicitly, it is also possible to prevent it from occurring in the first place by designing the ZCD with an input referred offset so that it will trip earlier. Likewise, we could let the overshoot occur, but design the opamp with an input referred offset to match the magnitude of the overshoot at its inputs. Yet another option, suggested in [39], is to charge both V_O+ and V_O- from the same direction during ZCBC operation. The resulting overshoots would ideally have the same magnitude and direction, with zero differential overshoot.

4.3.4 Asynchronous timing block

There are several signals that must be generated between the time that the ZCD transitions and the opamp begins settling the output. These signals are generated by an asynchronous timing block, shown in Fig. 4.8. The switch S_{DAC} in Fig. 4.3 is actually two switches: S_{sample} and S_{cancel} of Fig. 4.7. These switches are either NMOS or PMOS depending on the supply that they're connected to (which are opposite for C_{DAC+} and C_{DAC-}), so all polarities must be generated. The asynchronous timing block allows the level shifting phase to be started automatically by the ZCD's output (DC), maximizing the time that the opamp has to settle and greatly simplifying the clocking scheme

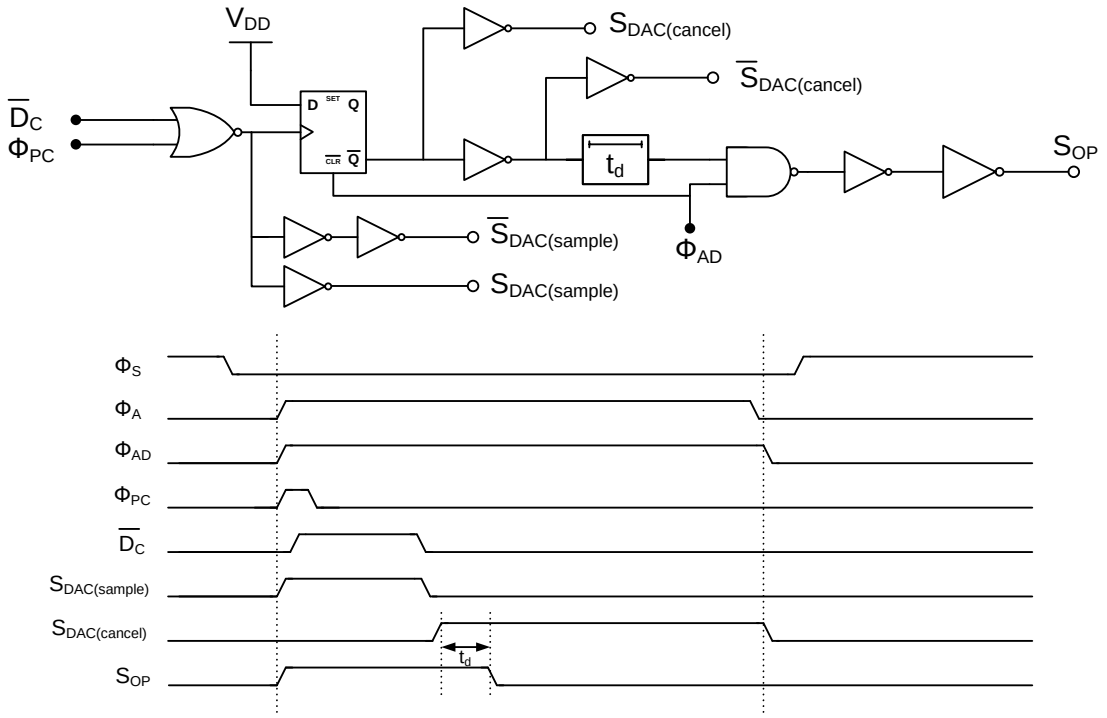
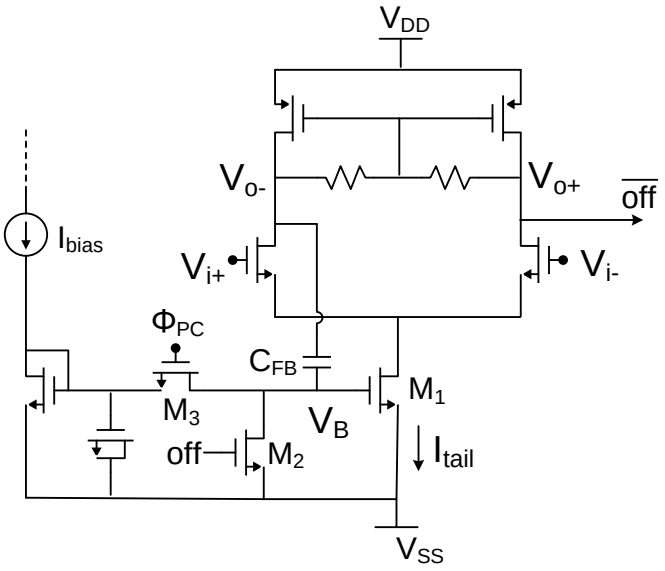


Figure 4.8: Asynchronous timing block used to facilitate the transition from ZCBC to opamp operation.

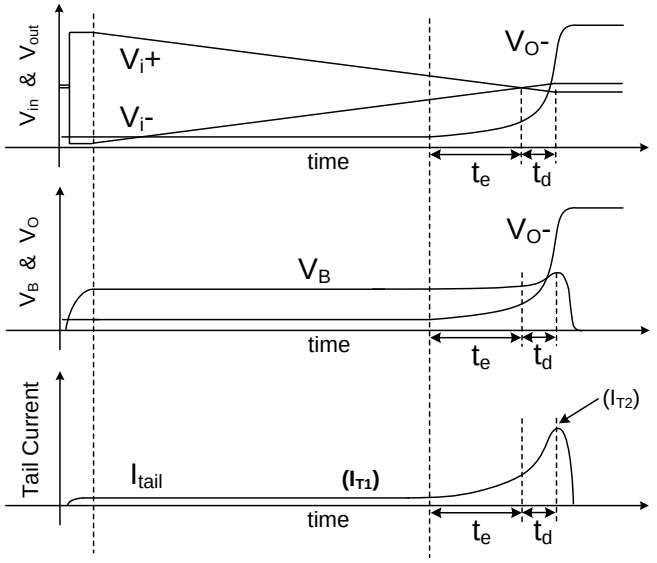
compared to [22].

4.4 Dynamic Zero Crossing Detector

The ZCD in a differential ZCBC system is typically statically biased while the current sources are ramping the output [23][25]. A useful observation that we can make in this regard is that the ZCD's bandwidth requirements depend on the relative position of its inputs. When the inputs of the ZCD are close to crossing, we know that the ZCD must trip soon, so we would like to have a high bandwidth in order to minimize both the ZCD's absolute time-delay and variation in time-delay. By contrast, when the inputs are still far apart we can infer that the ZCD is not yet close to tripping, so there is very little benefit to having a high bandwidth in the detection circuit. Therefore, if we redistribute the power consumption of the ZCD to concentrate the use of current around



(a)



(b)

Figure 4.9: Differential dynamic zero-crossing detector. (a) Simplified schematic of proposed ZCD. (b) Example ZCD waveforms demonstrating dynamic rationing of power.

the detection instant itself, we can optimize power efficiency and minimize static power dissipation.

4.4.1 Basic Structure and Operation

A simple way to dynamically scale the power of a ZCD is to adjust the tail current (M1 in the ZCD of Fig. 4.9). A more challenging question is how to detect when this current should be scaled. One possibility lies in the output nodes themselves; because the ZCD's gain (AZCD) is finite, the outputs will begin to transition before the actual zero-crossing. The output itself provides us with a transitioning signal that always precedes the zero-crossing by a fixed amount. Fig. 4.9 presents the basic structure and theory of operation of the proposed dynamic ZCD. Initially, during ϕ_{PC}, tail transistor M1's gate node (V_B) is charged via switch M3 to a static bias. On the falling edge of ϕ_{PC}, V_B is disconnected from this reference and left at a fixed but floating voltage. The inputs (V_{i+}, V_{i-}) are still relatively far apart, and if AZCD is sufficiently large, the outputs V_{O+} and V_{O-} are saturated at the maximum value of the ZCD's output swing. Node V_{O-} will remain close to V_{SS} in this saturated state until

$$A_{ZCD}(V_{i+} - V_{i-}) \approx (V_{O+} - V_{O-}) \tag{4.1}$$

At this point the voltage at V_{O-} will begin to rise and, via the feedback capacitor C_{FB}, the floating node V_B will also rise by some proportional amount, increasing the tail current (I_{tail}). Around the detection instant, the gm of the ZCD is at a maximum. Once the ZCD decision is registered by the dynamic latch at the output of V_{O+}, switch M2 connects V_B to V_{SS} and shuts off the ZCD.

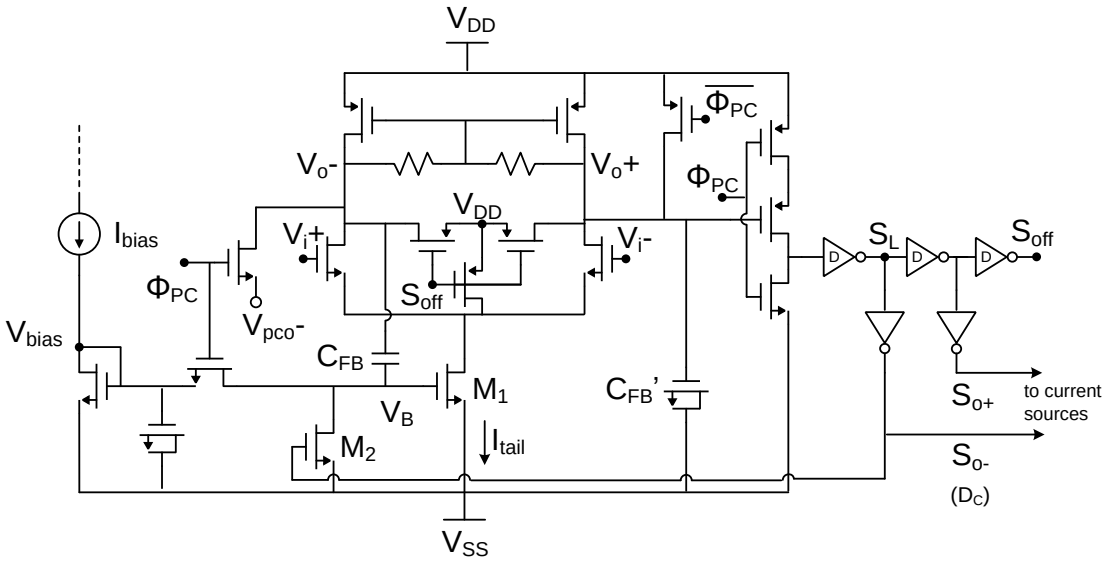


Figure 4.10: Detailed view of the dynamic ZCD, including pre-charge switches, output load matching, dynamic output latches, and shutoff switches.

4.4.2 Implementation Details

The complete schematic of the dynamic ZCD is given in Fig. 4.10. Specific implementation considerations are discussed in the following sub-section.

Startup Behavior

For the ZCD to be properly biased at the end of ϕ_{PC}, both V_B and V_{O-} must be at a constant, settled voltage. At the beginning of ϕ_{PC} the ZCD is brought out of shutoff, and $V_{O-} = V_{DD}$. V_{O-} must then settle during the short ϕ_{PC} phase to its initial steady-state voltage, $V_{O-(sat)}$ (which will be a few hundred millivolts above V_{SS}). Incomplete settling of V_{O-} is problematic: V_{O-} is coupled to V_B through C_{FB} and if V_{O-} continues to fall after the end of ϕ_{PC}, V_B and I_{tail} will drop as well. This scenario is very likely, because the initial bandwidth of the ZCD during ϕ_{PC} is intentionally low. Due to these concerns, V_{O-} is pre-charged to V_{pco-}. Ideally, V_{pco-} would be equal to $V_{O-(sat)}$, but for simplicity we can set it to any convenient voltage smaller than $V_{O-(sat)}$, such as V_{SS}. This ensures fast and signal-independent settling and guarantees that I_{tail}

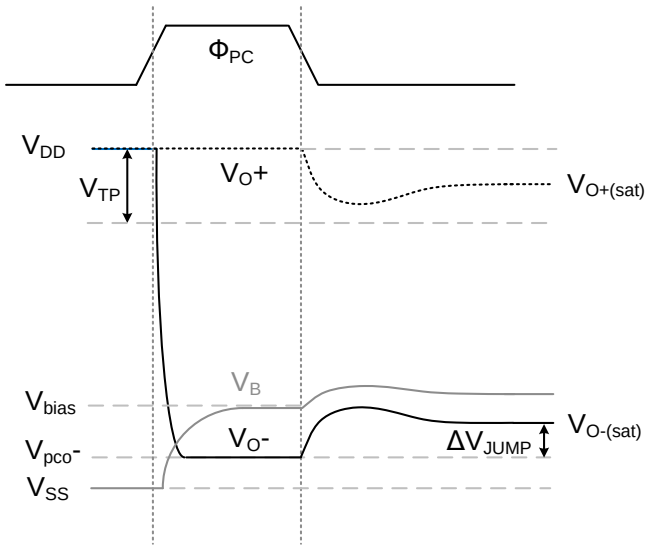


Figure 4.11: Transient behavior of V_{O+}, V_{O-}, and V_B during and immediately after pre-charge. To ensure proper operation, V_{O+} must not dip below V_{TP} and V_B must stay above V_{bias} after ϕ_{PC} goes low.

will be greater than zero. The resulting transient ZCD waveforms are shown in Fig. 4.11.

The initial voltage of the other output node, V_{O+}, is also critical. During ϕ_{PC} V_{O+} is set to V_{DD} for two reasons. First, V_{O+} must remain above the PMOS threshold voltage (V_{TP}) of the dynamic output latch. Second, with V_{O-} simultaneously held at V_{pco-}, this will provide a good approximation of the common mode feedback (CMFB) bias condition that will exist right after ϕ_{PCends}. It is important to match this CMFB condition as best as possible in order to minimize any settling ripple which could occur. Any transient ripple larger than V_{TP} will also cause the dynamic latch to trip in error. Another such scenario where this could occur is in a design where the CMFB bandwidth is not high enough to track the dynamic increase in tail current.

Detection and Shutoff

If not for the presence of the shutoff switches controlled by S_{off}, the ZCD's internal voltages would continue to change even after the ZCD trips because the CMFB will begin to pull V_{O+} and V_{O-} toward V_{DD} after M1 is turned off. To ensure that the ZCD doesn't

affect the charge at nodes V_{X+}/V_{X-} (in Fig. 4.3) except during its own operation, the internal nodes of the input amplifier are brought to V_{DD} immediately after the zero-crossing is detected, and held there until the next pre-charge phase.

Dynamic Latches

After the zero-crossing, the ZCD decision must be latched before V_{O+} is pulled back to V_{DD}. Dynamic latches were used to accomplish this. The dynamic latch connected to V_{O+} is shown explicitly, and the rest are denoted with a "D". Alternating pull-high and pull-low latch styles are used in the output chain. During ϕ_{PC} the latches are reset.

Symmetric Loading

V_{O+} and V_{O-} as shown in Fig. 4.9 have unbalanced capacitive loads, which affect the transient behavior of the CMFB and time-delay of the ZCD. We can better balance the outputs by adding a small capacitance $C_{FB'}$ to V_{O+}. Even more importantly, the dip in V_{O+} directly following pre-charge due to settling can be reduced by the additional capacitance of $C_{FB'}$, ensuring that V_{O+} will remain above V_{TP}.

4.4.3 Design Considerations

The ramp rate of the inputs, the values of A_{ZCD}, C_{FB} and V_{bias}, and the pre-charge value of V_{O-} are inter-dependent design variables which all factor into how much power savings can ultimately be achieved with the dynamic ZCD versus a more typical ZCD. We will begin by looking at the relation between ramp rate and A_{ZCD}. As shown in Fig. 4.9b, the time when V_{O+} begins rising to when the zero-crossing occurs is denoted t_e, and the time delay between the zero-crossing moment and the output of the ZCD latching is denoted t_d. Considering the condition defined in Eq. 4.1 when the outputs begin transitioning, the value of t_e can be defined as:

$$t_e = \frac{V_{O+(sat)} - V_{O-(sat)}}{A_{ZCD} \cdot \frac{d(V_{i+}-V_{i-})}{dt}} \tag{4.2}$$

where $V_{O(sat)}$ is the saturated full-swing value of V_O. If t_e is larger than some optimal value, we are increasing the bandwidth earlier than necessary, because the value of t_d depends mainly on the bandwidth of the ZCD at the zero-crossing instant. If t_e is smaller than this optimal value, we will not have peak bandwidth at the zero-crossing instant due to practical speed limitations of the ZCD itself. In other words, the lower bound of t_e depends mostly on internal design constraints rather than external factors such as ramp rate. Because of this, we can approximate the optimal value of t_e for a given design to be constant with respect to different ramp rates. From Eq. 4.2, the relation between A_{ZCD} and the ramp rate now becomes

$$\frac{1}{A_{ZCD}} = c_1 \cdot \frac{dV_i}{dt} \tag{4.3}$$

where c_1 is a constant. The conclusion that we can draw from Eq. 4.3 is that for an optimum design, A_{ZCD} and ramp rate are tightly coupled variables, and the slower the ramp rate, the larger A_{ZCD} should be made. If A_{ZCD} is not made large enough to satisfy Eqs. 4.2 and 4.3, then power is not being optimally utilized.

Not only does efficiency depend on how optimally we distribute the power consumption in time, it also depends on the ratio between the initial ZCD current shortly after pre-charge (I_{T1} of Fig. 4.9b) to the final ZCD current when it latches (I_{T2} of Fig. 4.9b). Using the simplified square-law model for a MOSFET in saturation, this ratio is:

$$\frac{I_T1}{I_T2} = \left(\frac{(C_{FB} + C_P)(V_{bias} - V_{TN}) + C_{FB} \cdot \Delta V_{jump}}{(C_{FB} + C_P)(V_{bias} - V_{TN}) + C_{FB}(V_{trip} - V_{pco-})} \right)^2 \tag{4.4}$$

C_{FB} is the sum of all explicit and parasitic capacitances between V_B and V_{O-}, C_P is a capacitor from V_B to V_{SS} that represents all other parasitic loading on V_B, V_{pco-} is the pre-charge value of V_{O-}, $\Delta V_{jump} = V_{O-(sat)} - V_{pco-}$, and V_{trip} is the value of V_{O-} when

the ZCD latches. To maximize efficiency, we want to minimize I_{T1}/I_{T2}. From Eq. 4.4 we see that this can be done foremost by minimizing $V_{bias} - V_{TN}$ and ΔV_{jump} so that they are much smaller than $V_{trip} - V_{pco-}$ and then by increasing the size of C_{FB} relative to C_P.

In reality, there are many other considerations which play into the overall efficiency of the dynamic ZCD. For example, the assumption that A_{ZCD} is constant is an approximation, because current is being dynamically changed. While this discussion has explored general design concepts, the inter-dependency of the design variables and time-varying nature of important circuit parameters means that no simple analytical model is available and transient simulations of the dynamic ZCD are a vital part of the design process.

4.4.4 Dynamic vs. Static Comparison

Much of the dynamic ZCD structure exists to deal with issues related to the dynamic biasing. If we wish to compare the dynamically biased ZCD to an equivalent statically biased ZCD, simply removing C_{FB} will not provide a functional or fair comparison. As we modify the structure to operate with static biasing, we will eventually arrive at a design nearly identical to the ZCD of [25]. In simulation, these two ZCDs were compared within the context of this work: for an input ramp duration of 7ns, total period of 50ns, and $t_d = 410$ps, the average power dissipation is 51μW for the dynamic ZCD and 175μW for the static ZCD. The tail current of the dynamic ZCD scales from $I_{T1} = 102\mu$A when the inputs are far apart to a peak current of $I_{T2} = 425\mu$A at the detection instant.

This comparison is by no means a generalized conclusion, because it was done within the context of a design that does not require high accuracy, high supply rejection, or low noise. In designs where these factors are more critical, the relative efficiencies

Table 4.1: Split-CLS ZCBC/Opamp ADC Summary of Performance

Technology	0.18μm CMOS			
Supply Voltage	1.8V			
Input Voltage Range	1.4V			
Sampling Frequency	f_S = 10 MHz		f_S = 20 MHz	
ENOB	11.3b		11.1b	
SNR	69.6dB		68.3dB	
SNDR	69.5dB		68.3dB	
SFDR	78.8dB		76.3dB	
Power (analog/digital)	7.2mW	1.2mW	15.0mW	2.2mW
FoM	343.5 fJ/step		405.5 fJ/step	

may be different. In particular, the floating gate of the tail source transistor M1 makes the Dynamic ZCD sensitive to external transient variations, and may be unsuitable for applications which demand high power supply rejection.

4.5 Experimental Results

A pipelined ADC incorporating the Split-CLS structure (with dynamic ZCD) was fabricated in a 0.18μm CMOS technology. Designed for testability and proof-of-concept, the prototype ADC is composed of 10 identical 1.5b/stage pipeline stages followed by a 1.5b flash backend. The telescopic opamp was designed to maintain better than 70dB open-loop gain for a 300mV differential output swing. Including the contribution of the ZCBC amplification, the total effective open-loop gain of the Split-CLS structure was designed to be more than 110dB. Measurement results, previously given in [36], are presented in Table 4.1 and Fig. 4.13. The SFDR is limited by even harmonics that were not present on a previous version of the test board, which suggests that the source of

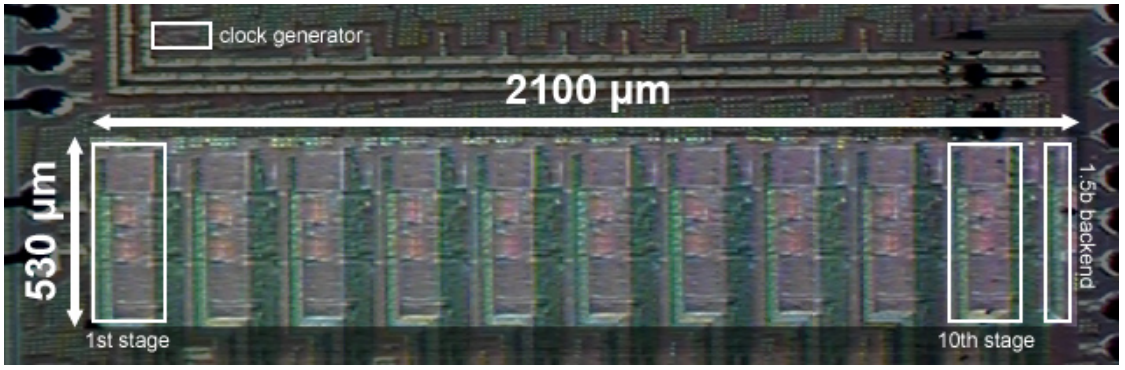
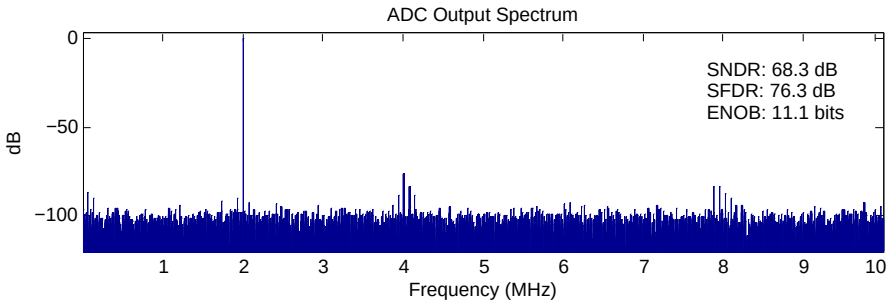


Figure 4.12: Die micrograph of Split-CLS ZCBC/Opamp test chip.

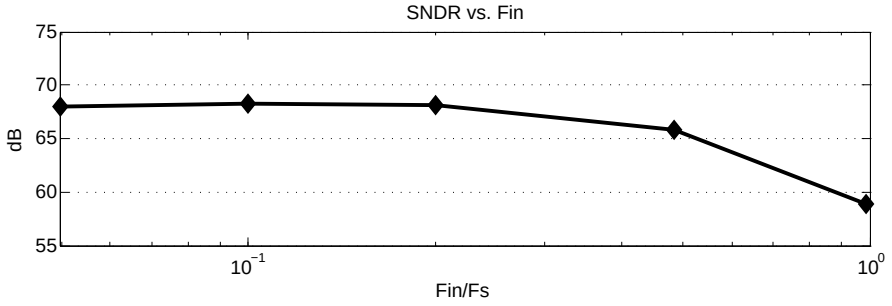
error may originate from outside of the chip. The ADC maintains better than 66dB SNDR for fin up to $f_{Nyquist}$, and the roll-off above this frequency seen in Fig. 4.13b is limited primarily by sampling jitter, which can be improved by more careful design of the input clock buffer, clock generation circuitry, and sampling network.

The overshoot cancellation DACs were implemented with independent digital controls for each stage as well as independent control of the two DACs within each stage. The high bandwidth of the dynamic ZCD helped to create a very consistent amount of overshoot across all stages of the pipeline. All measured results presented in this section were taken with a single static DAC code applied to all DACs in the pipeline. This consistency across all stages indicates that the overshoot cancellation can be implemented as a single global control in the future.

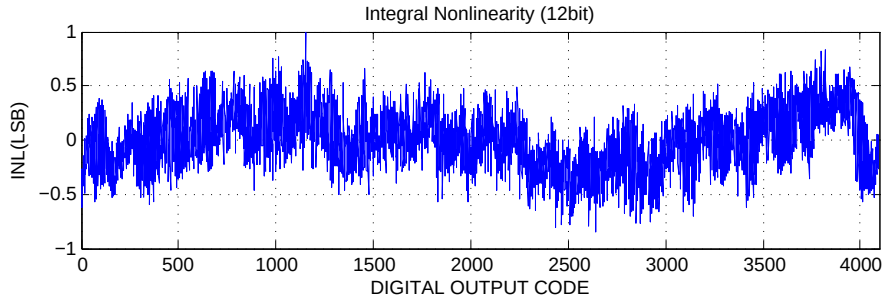
The test chip did not support disabling of the opamps, so a characterization of the dynamic ZCD without the influence of the opamp was not possible. However, the high tolerance to certain internal and external variation in t_d provided by the high bandwidth at the critical zero-crossing instant was measurable. In test, the ZCD bias current I_{bias} was varied between 5μA and 50μA with no observable change in performance besides power, with all other controls held constant. The FoM is 344fJ/conversion-step at f_s=10MHz and 406fJ/conversion-step at 20MHz. Application of common design techniques such as stage scaling, multi-bit quantization, and opamp sharing, as well as



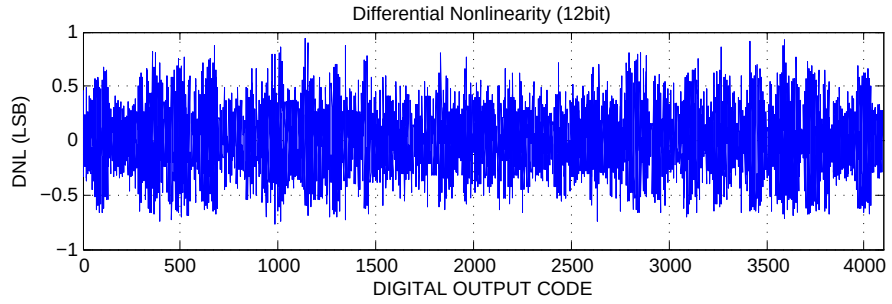
(a)



(b)



(c)



(d)

Figure 4.13: Measured results. (a) ADC output spectrum for $f_s = 20$ MHz (b) SNDR versus f_{in} up to $2f_{Nyquist}$ (c) INL (d) DNL.

the removal of the cancellation DACs in favor of a global ZCD input offset control could be applied to significantly extend power efficiency.

CHAPTER 5. A 10.5B RINGAMP CHARACTERIZATION ADC

5.1 Introduction

In this chapter we consider a pipelined ADC design that uses nothing but ringamps to perform amplification. The intentionally simple design allows the behavior of ringamps to be characterized independent of any other supplemental techniques (such as Split-CLS, as is the case in Chapter 6). The results are very encouraging, and show that with only a handful of tiny inverters, capacitors and switches, it is possible to achieve 10-bit accuracy. The ability to get such good SNR with such small inverters is a feat in itself, and hints at the subtle but important noise advantages that ringamps possess. This design was initially published in [40] and most of the text in this chapter is based on [41].

5.2 Design Details

To demonstrate the properties discussed in Sections 2.3 and 2.4, a 10.5-bit pipelined ADC was implemented and tested [40]. The ADC consists of nine identical 1.5-bit switched capacitor MDAC stages followed by a 1.5-bit backend flash. The transistor-level ringamp used is shown in Fig. 5.1. As we can see from the device sizes listed in this figure, this ringamp is quite small, and even the largest transistor in the design is only 2x the minimum W/L allowed by the process. In each stage, two of these single-ended ringamps are placed in the pseudo-differential configuration shown in Fig. 5.2. Due to the lack of common-mode feedback (CMFB) in such a configuration, the MDAC employs the 1.5b flip-around pseudo-differential float sampling scheme of [20]. This scheme sets the differential-mode gain to 2 and the common-mode gain to 1. Thus, any common-mode

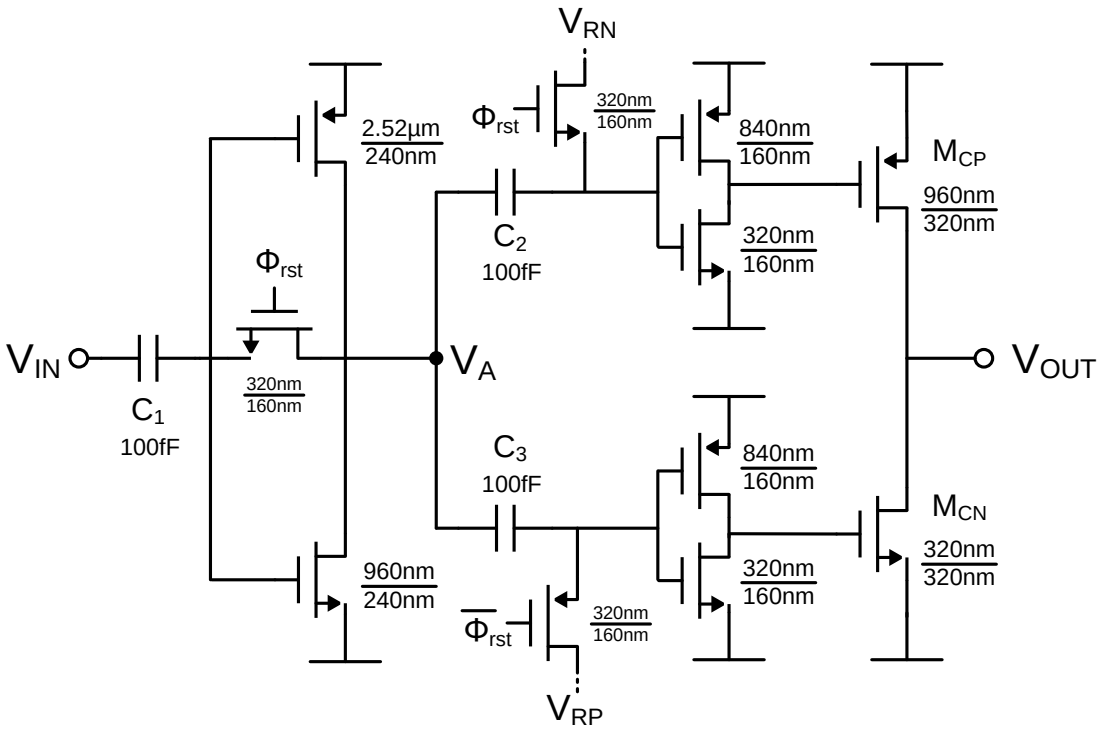


Figure 5.1: Complete transistor-level ring amplifier structure of the 10.5b characterization ADC, with the actual component values used in the fabricated design.

errors along the pipeline will simply add, rather than multiply and potentially saturate the common-mode level.

5.3 Experimental Results

The pipelined ADC was fabricated in a 1P4M 0.18μm CMOS technology. The output spectrum is shown in Fig. 5.3, and the performance with respect to input frequency is given in Fig. 5.4a. At 30MHz sampling rate, the ADC achieves 61.5dB SNDR, 61.9dB SNR, and 74.2dB SFDR. Total power consumption is 2.6mW, with approximately 90μW consumed per ring amplifier. The measured ERBW is greater than 15MHz, and the Figure-of-Merit (FoM) is 90 fJ/conversion-step. Neither speed nor power were prioritized in this design (only accuracy), and with a more aggressive design and use of the

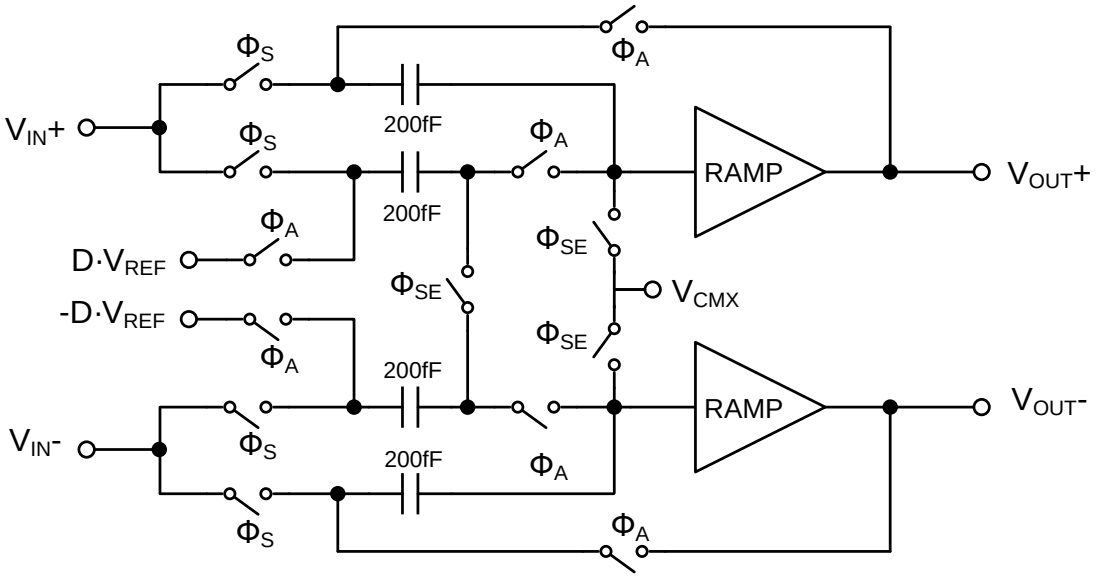


Figure 5.2: Pseudo-differential float-sampled 1.5b flip-around MDAC. In this structure, the differential gain is 2 and the common-mode gain is 1, eliminating the need for additional common-mode feedback.

power saving techniques discussed in Chapter 6, both can be improved significantly.

Besides serving as a proof-of-concept, a key motivation behind this ADC is to characterize noise and PVT sensitivity, which can be particularly difficult to accurately characterize from simulation alone. As seen in the performance summary given in Table 5.1, noise is the fundamental limitation on accuracy and the majority of this comes from quantization noise. Despite such small device sizes, the actual noise contribution from the ringamp is quite small. This highlights the key noise advantages of ring amplifiers that were discussed in Section 2.4.4.

To demonstrate the key scalability benefit of output-swing compression immunity, the analog supply was reduced to 1.2V, the MDAC references were set to the supply voltages (0V and 1.2V), and the sampling frequency was reduced to 4MHz, with all other settings left unchanged. The results of an input amplitude sweep under this test setup is presented in Fig. 5.4b. Because the transfer function of a 1.5b MDAC spans the entire supply range, a rail-to-rail input signal will cause the ringamps to swing rail-to-

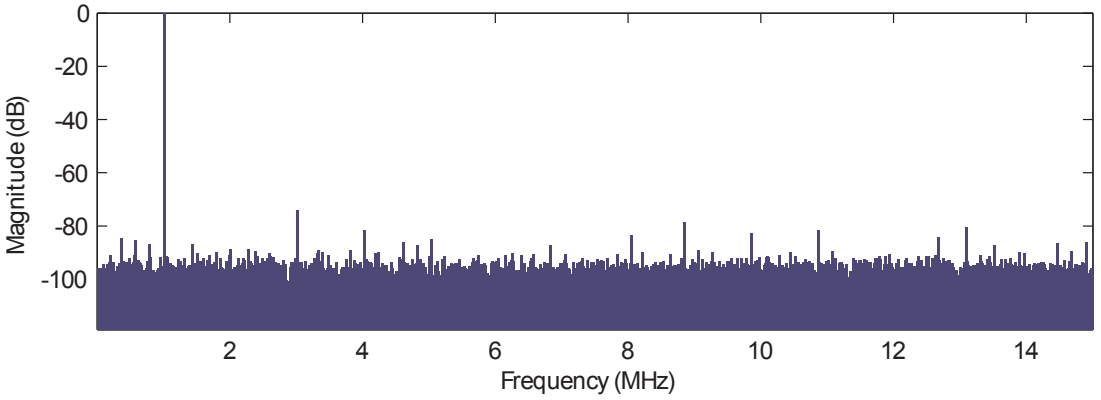


Figure 5.3: Measured output spectrum of the characterization ADC for a 1MHz input tone sampled at 30Msps.

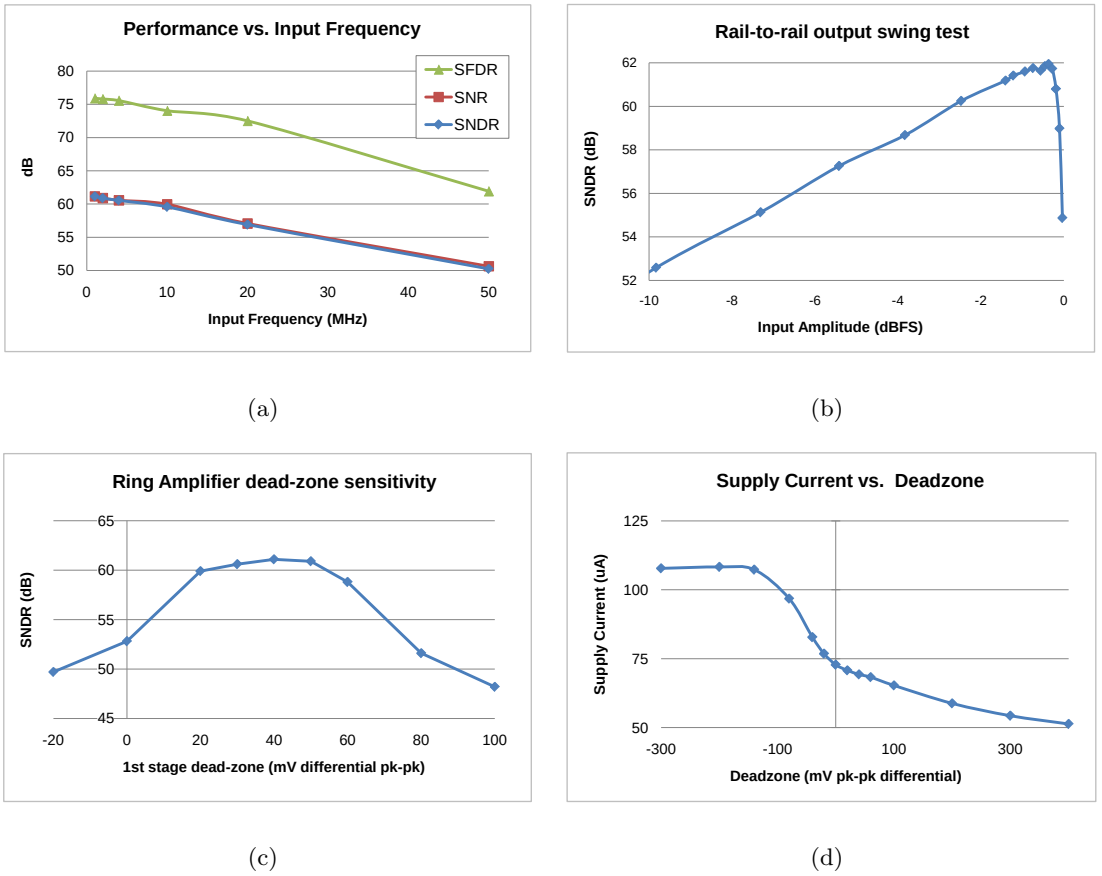


Figure 5.4: Measured performance and characterization data for the 10.5b ADC.

Table 5.1: Characterization ADC Summary of Performance

Resolution	10.5 bits
Analog Supply	1.3V
Sampling Rate	30 Msps
ERBW	15 MHz
Input Range	2.2 V pk-pk diff.
SNDR	61.5 dB
SNR	61.9 dB
SFDR	74.2 dB
ENOB	9.9 bits
Total Power	2.6 mW
FoM	90 fJ/c-step
Technology	0.18μm 1P4M CMOS
Active Area	0.50mm^2 (2.00mm x 0.25mm)

rail at their outputs as well [7]. The ringamp maintains linearity even in true rail-to-rail operation, and only begins to degrade within ±15mV of V_{DD}/V_{SS} due to insufficient RC settling time.

Fig. 5.4c shows the effect of dead-zone variation on SNDR. As can be seen, there is a wide, stable range of dead-zone values for which performance is largely unchanged. The roll-off on the right is due to Eq. 2.1, and the roll-off on the left is due to the ringamp becoming unstable. To a large extent, the plateau in the middle is a reflection of quantization noise limiting SNR, and not the accuracy limit of the ringamp. The plot of ringamp power consumption with respect to dead-zone size in Fig. 5.4d indicates that the faster a ringamp stabilizes, the less power it will consume (as one would expect). On the left half of this plot, where the dead-zone is less than 0mV, the structure becomes a

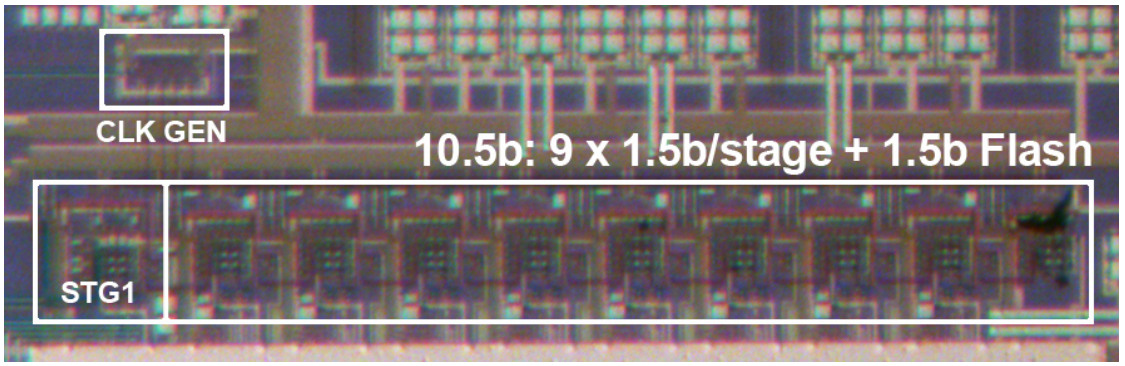


Figure 5.5: Die micrograph of the 10.5b characterization ADC ring oscillator.

CHAPTER 6. A 15B SPLIT-CLS RINGAMP/OPAMP ADC

6.1 Introduction

In Chapter 3 we observed that the ideal amplifier for the Split-CLS estimation phase ought to have maximal output swing and efficient load driving capabilities, and that it does not necessarily need to take the form of a conventional opamp. The choice in Chapter 4 to use a zero-crossing based circuit only partially meets these requirements; it is efficient at charging large output loads and can asynchronously transition into the fine-settling phase, but it has substantially limited output swing in low-voltage. Furthermore, the need to cancel the ZCBC overshoot before fine-settling adds quite a bit of complexity and susceptibility to PVT variation, and places a practical limit on ZCBC accuracy in the context of Split-CLS. For all these reasons, ZCBC makes for only an "acceptable" option. The persisting lack of an "ideal" solution for the estimation phase amplifier was what initially motivated the invention of ring amplification, and the ringamp is indeed a nearly ideal structure for this use. It boasts wide output swing, high efficiency charging, low complexity, and uses feedback (and thus does not require overshoot cancelation like a ZCBC).

The high-resolution pipelined ADC design that we will discuss in the following chapter is in many ways at the center of this dissertation. It brings the two ideas of ring amplification and Split-CLS together, and does so in a way that makes an important statement: although there many ways to make a medium-resolution ADC, there are far fewer options in the realm of high-resolution, where most approaches are frequently confounded by either linearity or noise. If a technique can be proven to operate in high-resolution applications, it can also work across the entire spectrum of accuracies. In this

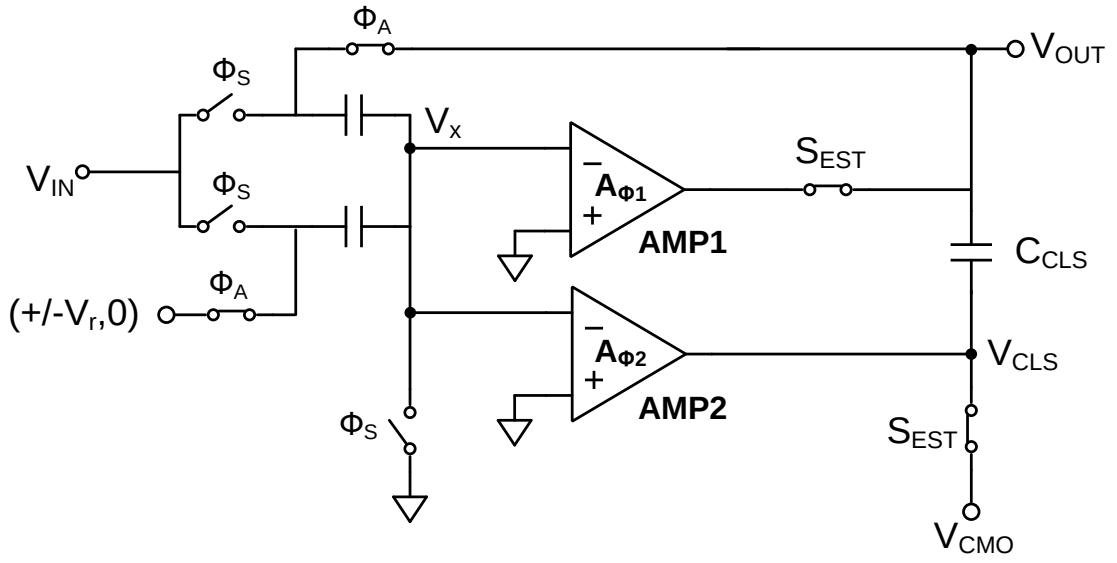


Figure 6.1: A basic implementation of the Split-CLS technique. The amplification phase, ϕ_A, is sub-divided into an estimation phase and a fine settling phase. In the first sub-phase, the switches S_{EST} are asserted, and AMP1 charges an estimate of the final settled value onto the output. Meanwhile, the capacitor C_{CLS} samples this estimation value and the output of AMP2 is shorted to mid-rail and held in standby. In the second sub-phase, the switches S_{EST} are de-asserted and AMP1 is disconnected from the output (and can be disabled). Meanwhile, AMP2 is coupled into the output via C_{CLS} and begins to fine-settle the output towards its ideal value. The effective gain at the end of ϕ_A will be $A_{AMP1} \cdot A_{AMP2}$.

sense, this design provides a way to incorporate ring amplification (and its associated benefits) into any hypothetical accuracy requirement, from low to very high. Not only does it make accurate amplification in nanoscale CMOS *possible*, it also makes it practical and attractive compared to other alternatives. In fact, the design that we will explore here achieves the highest power efficiency of any high-resolution ADC ever reported (at its time of publication) [27][42][41].

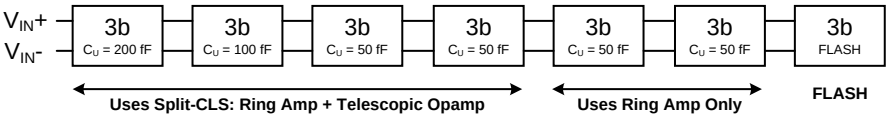
6.2 Structure Overview

Recall from Chapter 3 and Fig. 6.1 that the key benefit derived from Split-CLS will be an accuracy "stacking" effect, where the accuracy of AMP1 and gain of AMP2 will multiply to form a large final effective gain:

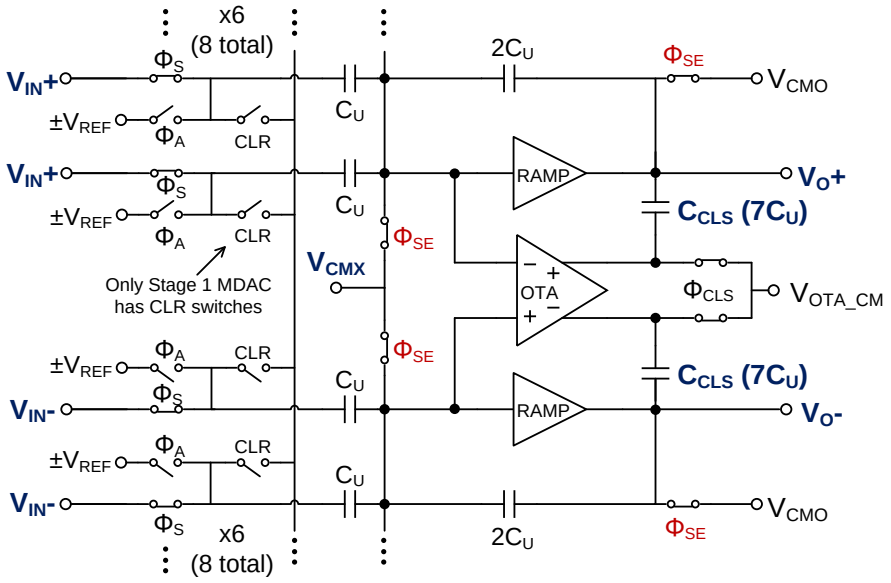
$$A_{effective} \approx A_{AMP1} \cdot A_{AMP2} \tag{6.1}$$

Thus, for a 55dB accurate AMP1, and a 65dB gain AMP2, the total effective gain at the end of ϕ_A will be approximately 120dB. Recall for a moment the requirements of AMP1 and AMP2 during their respective phases of operation. AMP1 charges the output load directly, and should have a high slew rate and wide output swing. By contrast, AMP2 only processes the small error term left over from AMP1, and requires a much smaller output swing and slew rate. Ring amplifiers are clearly an attractive choice for AMP1, particularly in high-resolution designs where kT/C noise constraints require a relatively large loading capacitance. A single stage telescopic opamp is a good candidate for AMP2, due to the small output swing requirement, and will work well even in low supply voltages. For the cautious designer, this structure is particularly appealing - one can both benefit from the advantages of ring amplifiers and still have the final settled accuracy be determined by a conventional opamp.

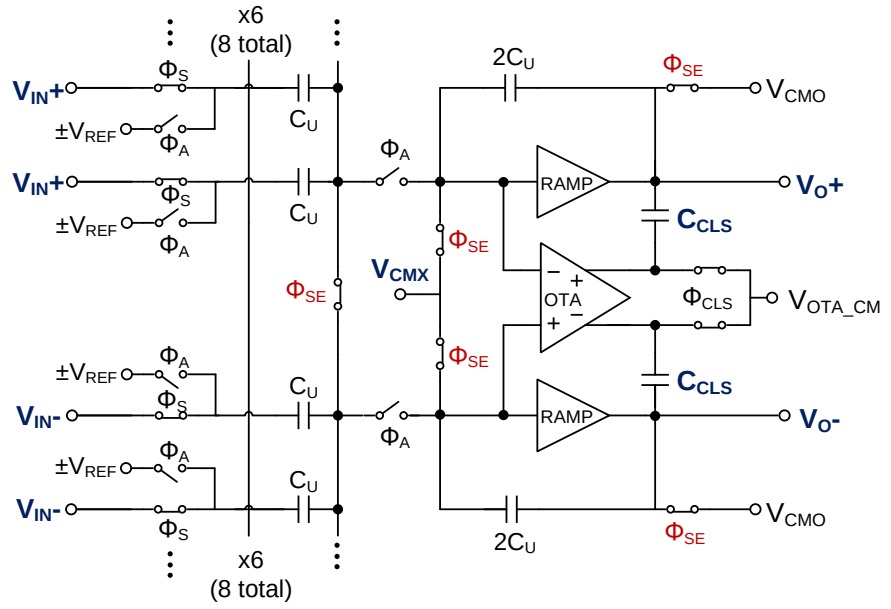
To demonstrate the effectiveness of this pairing, the high resolution pipelined ADC of Fig. 6.2a is presented. The pipeline resolves 15 bits with 6 MDAC stages and a 3b back-end flash. The first four stages employ Split-CLS (and stages 5 and 6 use ring amplifiers only). The fully-differential 3-bit/9-level first stage MDAC, shown in Fig. 6.2b, uses two pseudo-differentially configured ringamps of Fig. 6.3 as the AMP1 coarse charging device and the opamp of Fig. 6.4 as the AMP2 fine settling device. A digitally programmable delay line controls the time allotted to the ringamp's coarse charge operation (ϕ_{CLS}), and the remainder of ϕ_A is used by the opamp. Due to the fact that the opamp common-mode feedback (CMFB) is applied at the opamp output (and not the stage output), the pseudo-differential ring amplifiers must control the stage output's common-mode voltage. The simple capacitive CMFB network depicted in Fig. 6.3 provides an effective solution. The CMFB gain must be several times smaller than the gain of the primary feedback paths, and can be configured by selecting the appropriate capacitor ratio between C_{SIG} and C_{CM}. The reference voltage $V_{CMSHIFT}$ allows a static common-mode



(a)



(b)



(c)

Figure 6.2: Structural details of the 15b ADC. (a) Top level architecture, (b) the fully-differential Split-CLS 3b/9-level MDAC used for stage 1, and (c) the float-sampled, pseudo-differential Split-CLS MDAC used in stages 2-4. The stage 5-6 MDACs are the same as Fig. 6.2c except with the opamp and C_{CLS} removed.

Power Save Controls	
ENABLE	Asserted only during the first portion of Φ_A, when ring amp is used
REFRESH	Only asserted on a refresh cycle during CLR (for MDAC 1) or during Φ_{SE} (for MDACs 2-6)
FRONT ENABLE	Always asserted from the end of Φ_S to the end of ENABLE, and also whenever REFRESH is asserted

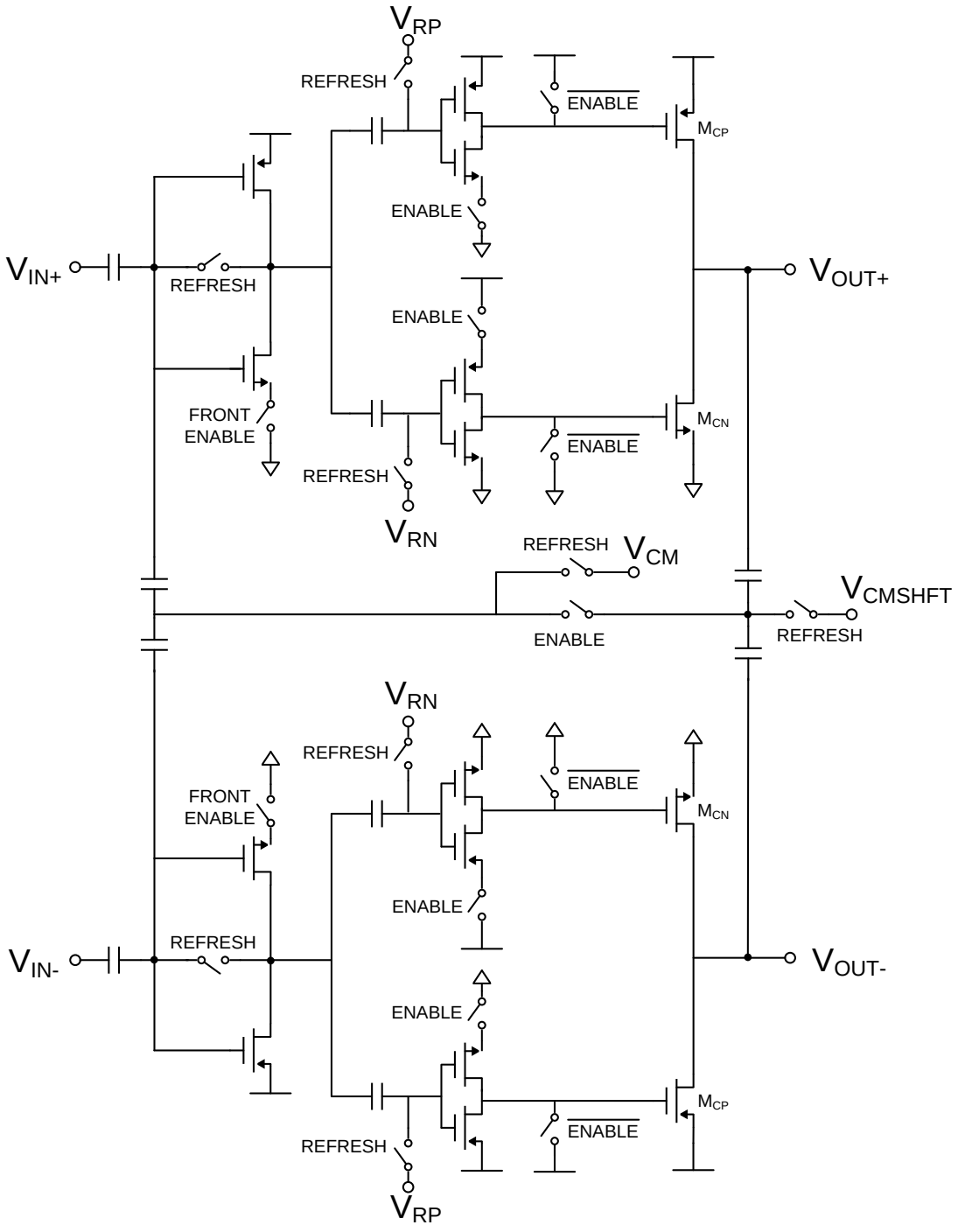


Figure 6.3: Complete pseudo-differential ringamp used with power-save features, showing the two coarse single-ended ringamps of Fig. 6.2(b) and the CMFB linking the two together.

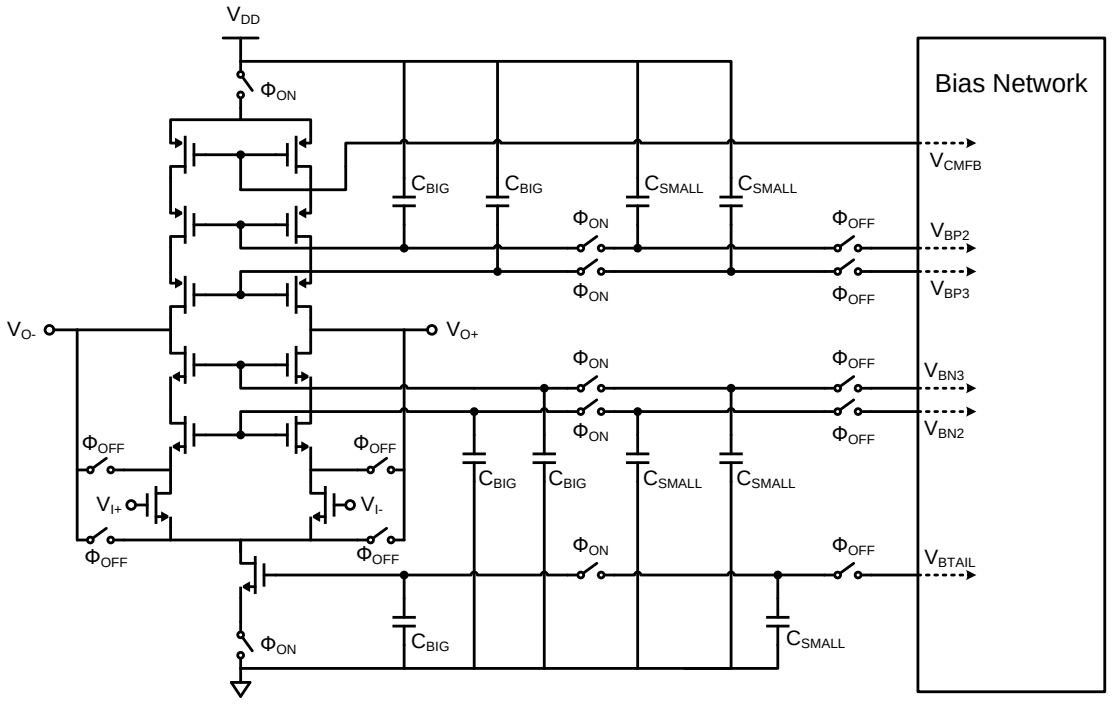


Figure 6.4: The double-cascoded fully differential opamp used in the first four MDACs. The charge-biased opamp switching scheme shields the bias network from power-up and power-down voltage kickback.

offset to be programmed into the network, and can be used to cancel any systematic common-mode offset and thereby increase the common-mode settling speed and useable range. While this simple CMFB is adequate for most uses, it has inherently low common-mode accuracy, and under extreme circumstances may be insufficient. To further relax the CMFB requirement, the stage 2-6 MDACs employ a 3-bit/9-level pseudo-differential MDAC similar to the 1.5b flip-around MDAC of Fig. 5.2, but with a fixed feedback capacitor. Shown in Fig. 6.2c, this scheme will allow the differential-mode gain to be 4 and the common-mode gain to be 1. Split-CLS is not needed in stages 5 and 6, where the accuracy requirement is low, and to save power the opamp and level-shifting capacitors (C_{CLS}) are removed from these stages.

6.3 Ringamp Power-Save Features

The ringamp shown in Fig. 6.3 is implemented almost identically in all stages, with only M_{CN} and M_{CP} sized differently in order to maintain the same slew rate across the differing output load sizes. For the vast majority of each period, the ringamp is unused, and can be disabled. This power-saving feature is implemented with the additional 'ENABLE' and 'FRONT ENABLE' switches depicted in Fig. 6.3. It is also important that the ring amp be completely disconnected from the output while the opamp is settling, and this feature guarantees that as well.

The input-offset and dead-zone voltages stored across capacitors C_{SIG}, C_{DZ1}, and C_{DZ2} must be refreshed periodically, and in order to do so, the first stage inverter must be active (although the second and third stages can remain off). Therefore, during a refresh, the signals 'REFRESH' and 'FRONT ENABLE' will be asserted. These refresh operations must be done at a time when the ringamp is not in use. Although there are several options, a good choice for when to do the refresh in stage 1 is during the short window when 'CLR' is asserted and the input sampling capacitors are shorted together and cleared. At the end of this clearing pulse, the virtual node will be exactly equal to V_{CMX}, allowing an accurate value to be sampled on C_{SIG}. Doing the refresh operation while the input capacitors are sampling is also possible, but somewhat less accurate, because there will be a signal-dependent voltage drop across the virtual node sampling switch (ϕ_{SE}) that will also be sampled onto C_{SIG}. For the later stages this problem does not exist, since the sampled signal is completely settled at the end of ϕ_S, and thus ϕ_S is used as the refresh signal in these stages. The charge stored on C_{SIG}, C_{DZ1}, and C_{DZ2} will only be corrupted by small parasitic leakage currents, and it is sufficient to perform a refresh only once every N cycles (where N is controlled by an on-chip digital counter). Such an approach is used in this design, and reduces the contribution of additional static power during refresh periods down to negligible levels. For the majority of conversion cycles, the ring amp is only briefly on at the beginning of ϕ_A, and completely off during

the rest of the period.

6.4 Charge-biased Switched Opamp

To save additional power in the opamps, the switched opamp scheme of Fig. 6.4 is used. The opamp is only enabled during ϕ_A, and uses the time at the beginning of ϕ_A (when the ringamps are amplifying) to power up. Shorting switches are placed on the source and drain nodes of the input transistors to ensure that any kickback onto the MDAC virtual node during power-on will be signal independent.

In the conventional switched opamp approach, the bias lines are directly connected to the opamp, and both the power-up and power-down operations will kick charge from the parasitic capacitances of the main opamp transistors onto the bias network. The bias network must then try to absorb this kickback and re-settle to the correct bias voltages, which will increase the required power-on time of the opamp [43]. The proposed solution to this problem is to bias the opamp in charge-domain rather than voltage-domain, as shown in Fig. 6.4. Imagine for a moment that ϕ_{ON} is asserted, the opamp is amplifying, and the C_{BIG} capacitors hold the correct charge to properly bias the opamp. When the amplification period ends and ϕ_{OFF} asserts, the opamp will switch off and a large amount of voltage will kickback onto C_{BIG}. However, since the top-plate of C_{BIG} has no DC path to ground during ϕ_{OFF}, the charge stored on it is trapped. Meanwhile, a set of small capacitors (C_{SMALL}) sample the bias network. Then, when the opamp is powered on again, C_{BIG} will be shorted to C_{SMALL} and some amount of charge transfer will occur. Since this charge transfer occurs during ϕ_{ON}, there is no unwanted voltage kickback present, and C_{SMALL} will update C_{BIG} with an incremental piece of charge corresponding to the correct bias voltage. This bucket-brigade that C_{SMALL} provides between the bias network and C_{BIG} isolates the bias network from opamp kickback and enables rapid power-up.

Table 6.1: Split-CLS Ringamp/Opamp ADC Summary of Performance

Resolution	15 bits
Analog Supply	1.3V
Sampling Rate	20 Msps
ERBW	10 MHz
Input Range	2.5 V pk-pk diff.
SNDR	76.8 dB
SNR	77.2 dB
SFDR	95.4 dB
ENOB	12.5 bits
Total Power	5.1 mW
FoM	45 fJ/c-step
Technology	0.18μm 1P4M CMOS
Active Area	1.98mm^2 (3.05mm x 0.65mm)

6.5 Experimental Results

The 15-bit pipelined ADC was fabricated in a 1P4M 0.18μm CMOS process. At 20MHz sampling rate it achieves 76.8dB SNDR (12.5 ENOB), 77.2dB SNR and 95.4dB SFDR, consuming 5.1mW. The ERBW is found to be above 10MHz, which results in a Figure-of-Merit of 45fJ/conv-step. The MDAC references are set at 25mV and 1275mV, allowing the input signal to utilize 96% of the available supply range. Capacitor matching was good enough that no digital calibration was needed. Total accuracy is fundamentally limited by noise, and the SNR is almost exactly equal to the kT/C noise limit predicted for this design. Plots of various performance metrics are given in Figs. 6.5, 6.6, and 6.7.

All analog portions of the circuit operate at 1.3V, including the ringamps and

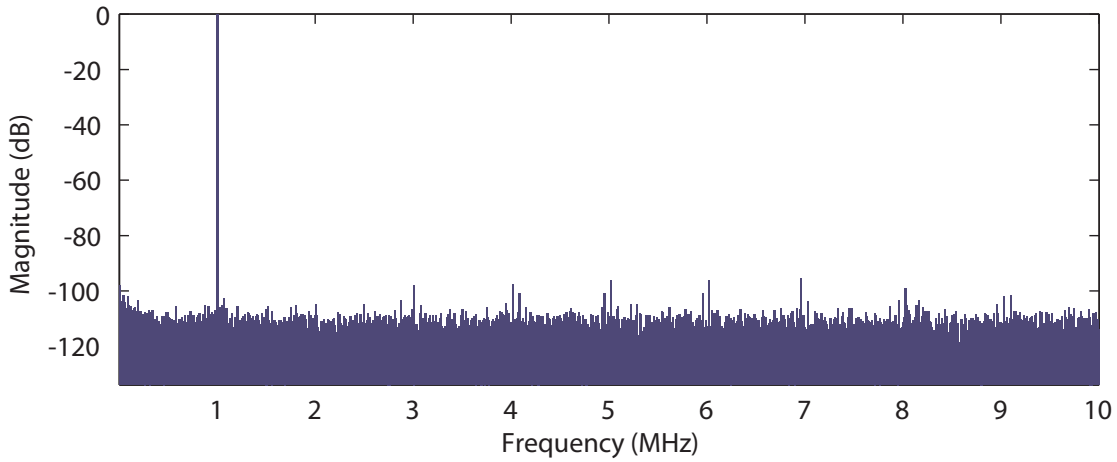
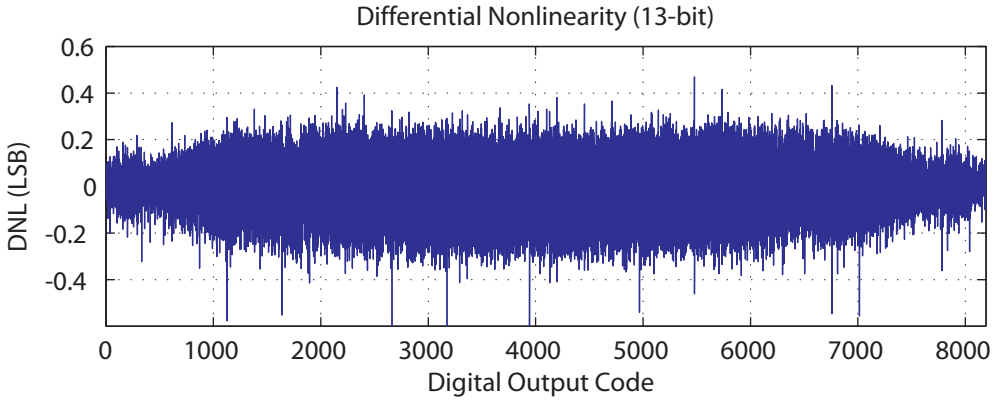


Figure 6.5: Measured output spectrum of the high-resolution Split-CLS ADC for a 1MHz input tone sampled at 20Msps.

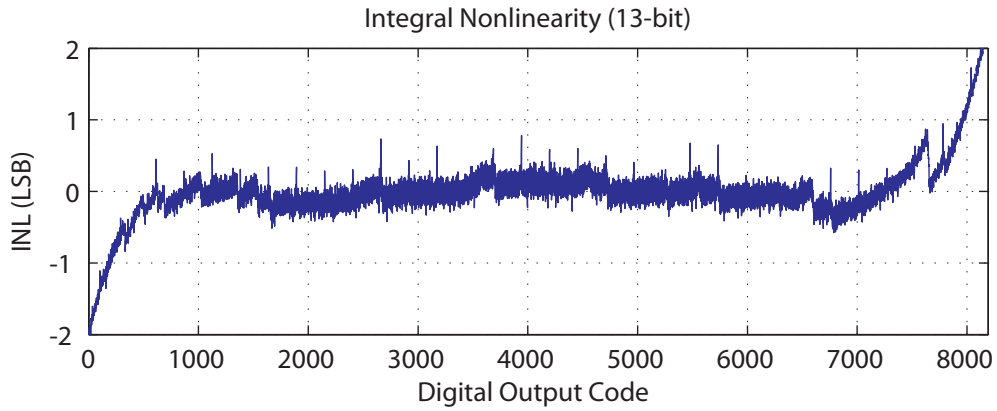
opamps. Opamp switching is measured to reduce the total opamp power by 35%, and improve SNDR by 0.6dB. An unfortunate issue with the physical layout caused a node in the bootstrapped input sampling switches to be shorted into a deep-nwell substrate, and in order to compensate the digital and switch supply had to be operated above 1.3V. Based on follow-up simulations, this issue is estimated to have increased total power consumption by more than 20%, and is included in the 5.1mW reported.

When the sampling frequency is increased to the point that the opamps never have a chance to amplify, the contribution of the ringamps to overall accuracy can be measured, and is found to be 55dB SNDR at f_s = 80MHz. This conclusion is also confirmed by the result of Fig. 6.8d. In this plot, the digitally controlled timing generator that sets the ringamp amplification window is swept with the opamps disabled. What we find is that ring amp performance decreases logarithmically with respect to incomplete settling time, and that the ring amp settles to about 55dB SNDR in roughly 6ns.

The ringamps exhibit a high tolerance to variation, as shown in Figs. 6.8a and 6.8b. In Fig. 6.8a, the dead-zone of the ring amps in the 1st stage MDAC is swept with everything else held constant. The roll-off above +50mV is due to the dead-zone becoming large and reducing accuracy and the roll-off below -50mV is due to the amplifier go-



(a)

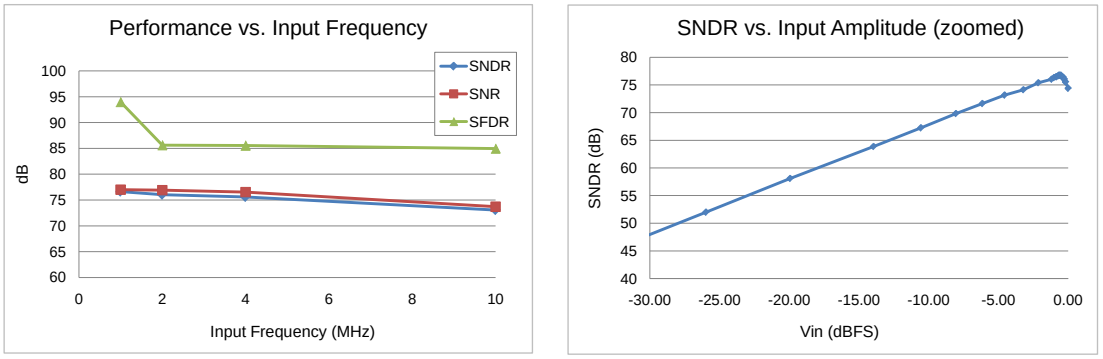


(b)

Figure 6.6: Measured DNL (top) and INL (bottom) normalized to LSB=2^{-13}.

ing unstable. The curve's peak has a systematic offset of roughly -100mV from what might otherwise be expected, and is due to an additional embedded offset caused by the asymmetry of the 'ENABLE' switches that control the second stage inverters of Fig. 6.3. The opamp, via Split-CLS, helps to absorb small errors and further flattens the peak of Fig. 6.8a into a wide stable plateau. When the ring amplifier supply voltage is swept with all other voltages and biases held constant (Fig. 6.8b), SNDR is virtually unchanged. At 1275mV the supply voltage equals the positive MDAC reference, and explains the roll-off seen below that voltage.

The benefit of the ringamp power-save feature is demonstrated in Fig. 6.8c. Total



(a) (b)

Figure 6.7: Measured input signal performance data for the high-resolution Split-CLS ADC.

ringamp power consumption is reduced by about 4.5x when the ringamps are refreshed only once every 64 cycles. For the ADC presented in Chapter 5, switching the ringamps off during ϕ_S could have reduced ringamp power in that design by almost 50%.

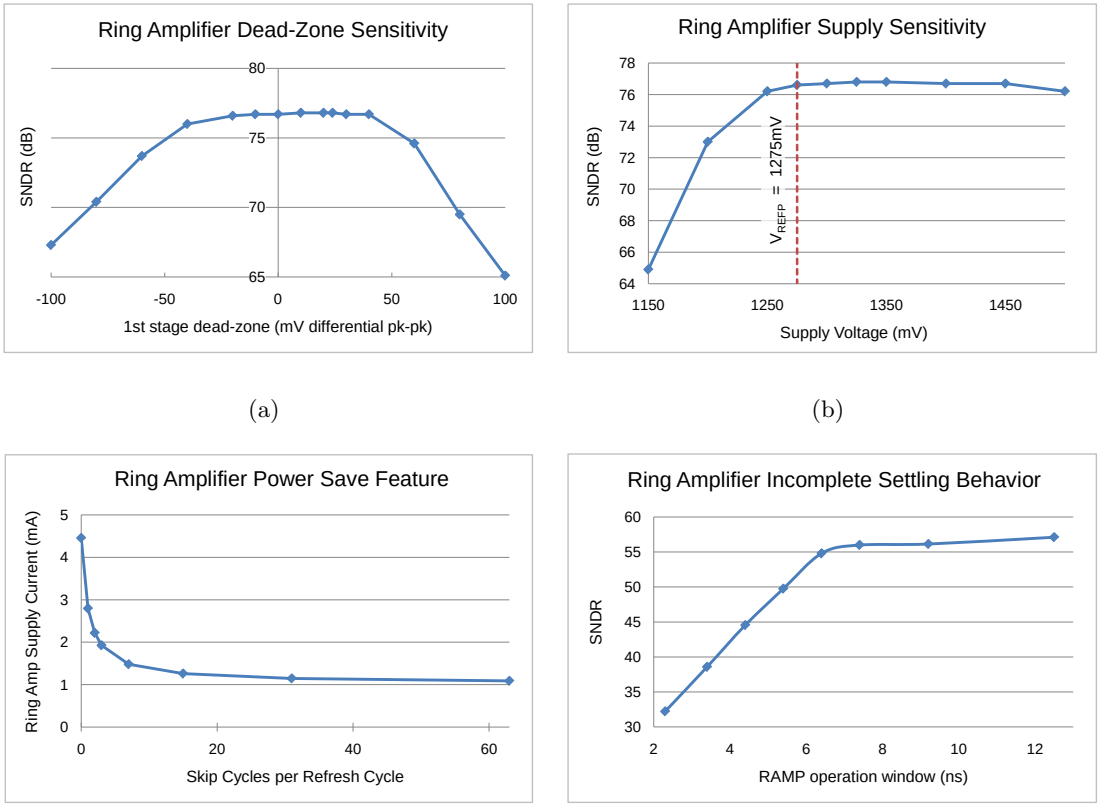


Figure 6.8: Measured ringamp sensitivity and characterization data for the high-resolution Split-CLS ADC.

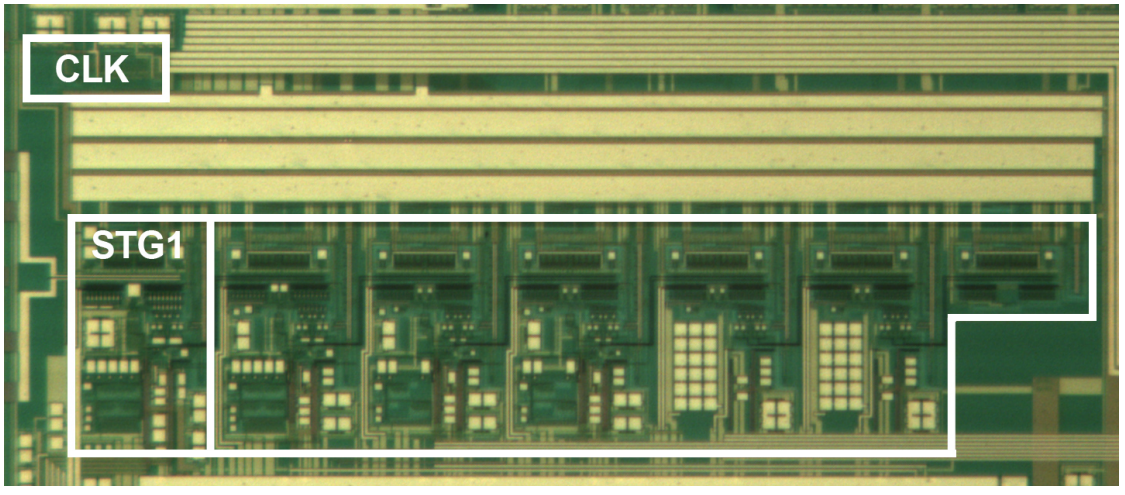


Figure 6.9: Die micrograph of the active area of the high-resolution Split-CLS ADC.

CHAPTER 7. A 15B PARALLELIZED RINGAMP ADC

7.1 Introduction

The Split-CLS ringamp/opamp ADC that we just explored in Chapter 6 enables high-accuracy amplification in nanoscale CMOS, and provides a comprehensive solution to many of the scaling concerns raised in Chapter 1. However, as we also observed in Chapter 1, there is an important distinction between techniques that work *in* scaled environments and those that truly scale. If we are going to completely reverse the trend depicted in Fig. 1.3, power efficiency must scale according to process at the same pace as digital performance scaling. With regard to the Split-CLS ringamp/opamp structure of the last chapter, the telescopic opamp used for fine-settling is only partially compatible with this scalability goal. Although the opamp does in fact possess a good measure of scalability (the output swing requirement can be made arbitrarily small and intrinsic improvements in g_m will help the power-speed tradeoff), its internal power requirement will still be directly related to the external load size (per the discussion in Section 1.3.2). Although this Split-CLS design goes a long way towards reversing the trend of Fig. 1.3, logic tells us that we can, in theory, do even better.

Ring amplifiers, on the other hand, possess all critical elements of scalability, but so far this potential remains unfulfilled in the realm of high-accuracy amplification. The single-ended and pseudo-differential ringamp structures that we have covered in previous chapters are not suitable for high-accuracy amplification. Although these structures can in *theory* generate very high accuracies on their own, several practical issues involved will likely limit any attempts to exceed about 70dB THD. Ultimately, to ensure high accuracy, a differential ringamp structure is required. In this section, we look at how to

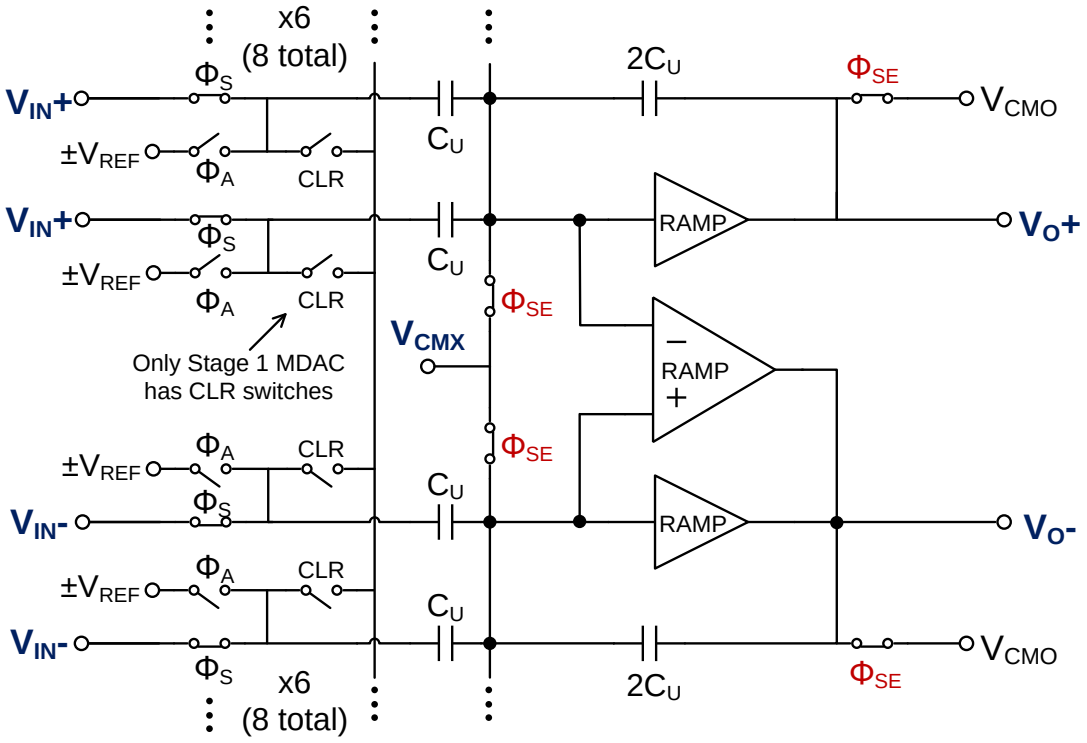


Figure 7.1: Differential MDAC using the parallelized ringamp structure. The common-mode voltage is set by the CMFB of the coarse ringamps, allowing the differential-input fine ringamp to have a single-ended output (and thus avoid CMFB).

build such a ringamp and incorporate it into a high-resolution ringamp-only pipelined ADC structure. The result is a 100% scalable high-accuracy amplification solution for nanoscale CMOS.

7.2 Structure Overview

Illustrated in Fig. 7.1 is a parallelized ringamp structure suited for high-accuracy amplification. The composite amplifier block consists of the parallel combination of two pseudo-differentially configured single-ended "coarse" ringamps, and one differential-input, single-ended output "fine" ringamp. At the beginning of the amplification phase, all three ringamps are active and charge the output. The coarse ringamps are designed

with a much larger dead-zone and output slew current than the fine ringamp, and dominate the behavior of the initial charging operation. After a fast initial ramp and stabilization, the coarse ringamps lock into their dead-zones and automatically disconnect from the output. The fine ringamp (which has a much smaller input-referred dead-zone that lies well within the dead-zone region of the coarse ringamps) then continues to settle until it stabilizes and locks into its own dead-zone.

Designing an efficient fully-differential ringamp with CMFB can quickly become an challenging puzzle. Luckily, due to the unique traits of ring amplification, a fully-differential fine amplifier is not required. This is made possible for two reasons. First, the coarse ringamps incorporate a simple CMFB network (shown in Fig. 6.3), and this sets the correct output common-mode voltage as the coarse ringamps enter their dead-zones. Second, once the coarse ringamps are in their dead-zones, there will be no resistive conduction path to V_O+, and whatever charge (and voltage) was set on this node will remain unaltered. With the common-mode voltage already set to a sufficient accuracy, the fine ringamp must only settle out the remaining differential error, and this can be done by simply detecting differentially but charging single-ended onto V_O-.

When we compare this MDAC to the MDAC of Fig. 6.2(b), we see a striking benefit of the parallelized ringamp approach: no Split-CLS. Although it would be theoretically possible to design a fine ringamp which could benefit from Split-CLS, it turns out that this is not necessary. Thanks to the inherent high gain of the multi-stage ringamp structure and gain-boosting property of the ringamp's pinch-off effect (discussed in Section 2.4.1), it is possible to generate very high gains with the ringamp alone. There is no need to "stack" the gains of the coarse and fine ringamps using Split-CLS, and the coarse ringamps can be used purely as charge-assist devices for increasing settling speed. Furthermore, the wide-swing capabilities also bestowed by the ringamp's pinch-off effect enable the fine ringamp to maintain its high gain across a large output voltage range. With inherently high-gain and wide-swing, the fine ringamp can be connected directly to the output. The removal of the Split-CLS network comes with several benefits. First of

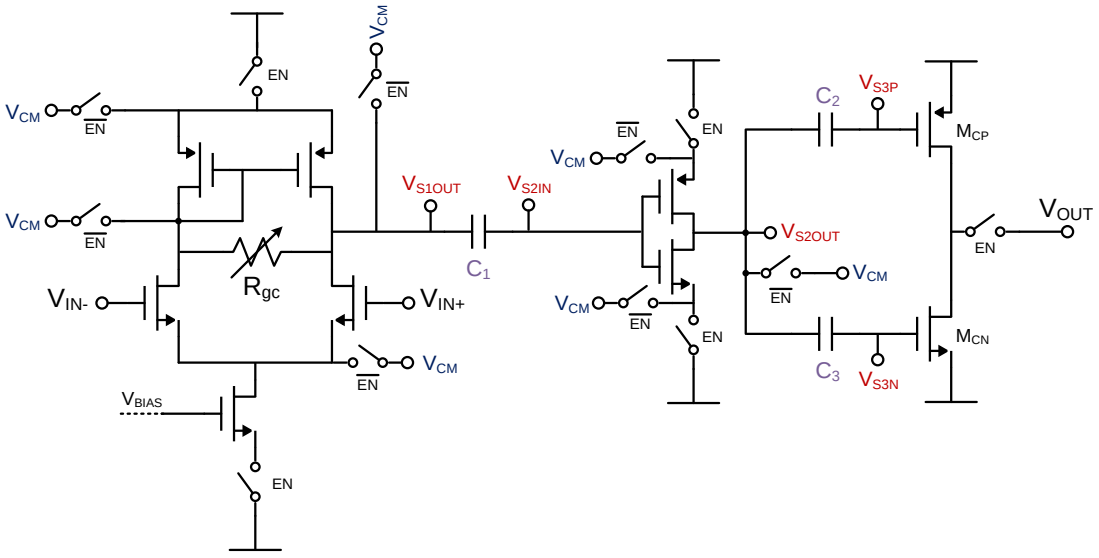


Figure 7.2: Complete schematic of the differential ringamp in Fig. 7.2. The dead-zone is set after the second stage, allowing for tight control of the pinch-off gain boosting effect and a small input-referred dead-zone.

all, C_{CLS} often constitutes a significant percentage of the total output capacitance that the coarse ringamps must charge, and its removal improves both the static and dynamic power consumption of the coarse ringamps. Second, Split-CLS requires two sub-phases of amplification whose ideal transition point is nearly impossible to time perfectly, since it must be guaranteed that the coarse ringamps are completely settled before beginning the fine settling opamp phase. By contrast, the parallelized ringamps operate concurrently in one unified phase, and the differential ringamp begins fine settling even while the coarse ringamps are still locking into the dead-zone. The simplicity of this concurrent, parallelized approach results in an obvious speed advantage over Split-CLS.

7.3 Differential Ringamp

The complete structure of the high-accuracy dual input, single output ringamp that is implemented in this design is depicted in Fig. 7.2. Like the power-save features of the ringamp in Fig. 6.8(c), this ringamp is only enabled during the amplification phase,

and completely powered down and disconnected from the output otherwise. Many of the implementation details of this ringamp deviate from those of the simple single-ended structure that we have been dealing with so far (i.e. Fig. 2.1). To begin with, the ringamp must have accurate low-noise differential detection with good common-mode rejection. A basic source-coupled input pair meets this requirement well, and is easily biased. Staying true to the ringamp paradigm, the input stage requirements are very well decoupled from the output load in this design, and the power requirement for the input pair is small.

Achieving a sufficiently small dead-zone and high gain is also a key concern in a design such as this. As we discussed in Section 2.4.1, it is possible in theory to design a ringamp with both high gain and wide output swing, thanks to the pinch-off effect that biases the output transistors in weak-inversion as they enter the dead-zone. For a three-stage ringamp, achieving gains in excess of 90dB is entirely possible, assuming that the dead-zone can be injected in such a way that the transistors are guaranteed to be in weak-inversion when the ringamp reaches steady-state. In this sense, the steady-state condition of a high-accuracy ringamp is analogous to a very low bandwidth three-stage opamp. Unlike the coarse ringamp, whose accuracy is typically limited by the input-referred dead-zone size (rather than finite gain), the fine ringamp settles into a "weak-zone", and never fully cuts off, leaving gain to be the final judge of settled accuracy.

Although the input-referred weak-zone size will not influence accuracy in the way that a true dead-zone does, it *will* define stability. Therefore, the weak-zone must be injected in such a way that it can both ensure weak-inversion in the output transistors and also set a large enough input-referred weak-zone to guarantee stability. The ringamp structure of Fig. 7.2 addresses these issues by moving the offset storage capacitors (C_2 and C_3) to the output of the second stage. This allows the pinch-off effect to be easily and confidently set. To ensure that the output transistors will be in weak-inversion, the differential offset that should be injected is simply:

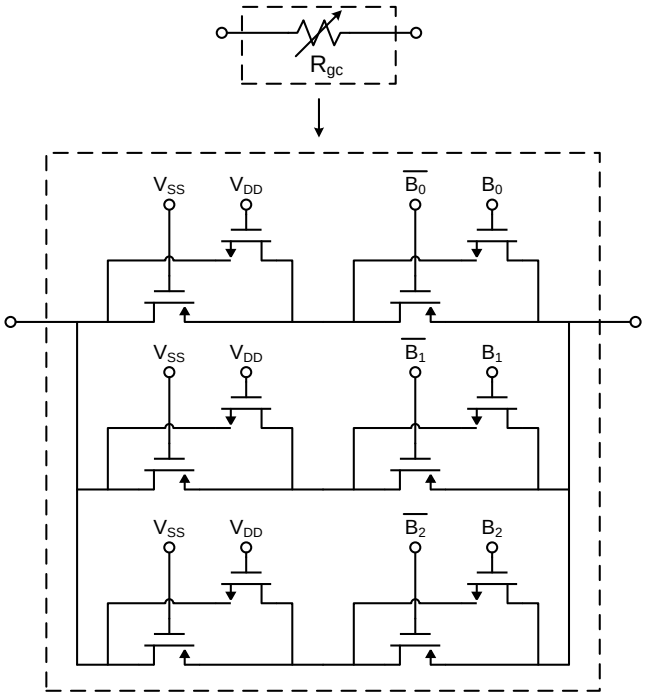


Figure 7.3: Expanded view of the digitally programmed gain-control resistor used in both the main and replica ringamps of Figs. 7.2 and 7.4(b). Neither high linearity nor accuracy is required, which allows small MOSFETs to be used as the resistive DAC elements.

$$V_{DZ} = V_{DD} - V_{TP} - V_{TN} - \alpha \tag{7.1}$$

where α is some small voltage that will determine how far in weak-inversion the output transistors will be in steady state. Although injecting the offset at the input of the third stage is very helpful for ensuring sufficient pinch-off and gain boosting, it comes with a few complications. First of all, the size of the offset will now be determined by Eq. 7.1, and this adds an additional challenge to ensuring stability. For a given internal bandwidth and power consumption of the ringamp, stability will only be guaranteed for an input-referred weak-zone above a certain minimum value, and the input-referred offset dictated by Eq. 7.1 will not necessarily meet this criteria. This co-dependence means that an additional degree of tuning freedom is required if stability and accuracy are to be independently defined. Although the tail source bias (V_{BIAS}) could be used

for this purpose, it would be much better to have a way to control stability that isn't so closely coupled to power consumption. The presence of a tunable gain-control resistor (R_{gc}) across the output of the first stage adds this additional degree of freedom with very little overhead; by adjusting the front stage's gain, the input referred value of V_{DZ} can be adjusted without changing the value of V_{DZ} itself. Setting stability in this manner turns out to be a simple and attractive option. R_{GC} can be implemented as a DAC with small resistive elements, and the digital control allows for a very wide tuning range. The linearity of R_{gc} is not critical, because it is principally the signal-independent steady-state gain that we are interested in tuning. There is also no strict matching requirement for the DAC elements, since ringamp stability can be guaranteed for any input-referred weak-zone value above a certain threshold. Fig. 7.3 shows the particular implementation of R_{gc} used in this design. The resistive elements are built out of tiny CMOS transmission gates. To minimize the parasitic loading on node V_{S1OUT} of Fig. 7.2, each composite DAC element is composed of a resistor component (left column) and a switch component (right column). V_{S1OUT} will only see the parasitics of the three small switches plus the resistive elements that are selected. Even in the worst case, the parasitics that R_{gc} adds to V_{S1OUT} are minimal, and has only a minor impact on the bandwidth and power of the input stage.

Another consideration that must be made with respect to the structure of Fig. 7.2 is the interconnect between the first stage and the second stage. Notice that the trip voltage of the second stage inverter will be approximately mid-rail, whereas the settled output voltage of the first stage will depend on the first stage's common-mode output voltage. Because of this voltage mismatch in the interface between stage 1 and stage 2, if the two stages are directly connected, the ringamp will have a static differential offset. The input-referred value of this offset can easily be as large as 10mV, and this causes several problems. First of all, it will usually lie outside the dead-zone of the coarse ringamp connected to V_{O-} in Fig. 7.1. Left as-is, the coarse ringamp would fight the fine ringamp's efforts to charge to the offset value, and repeatedly force V_{O-} back

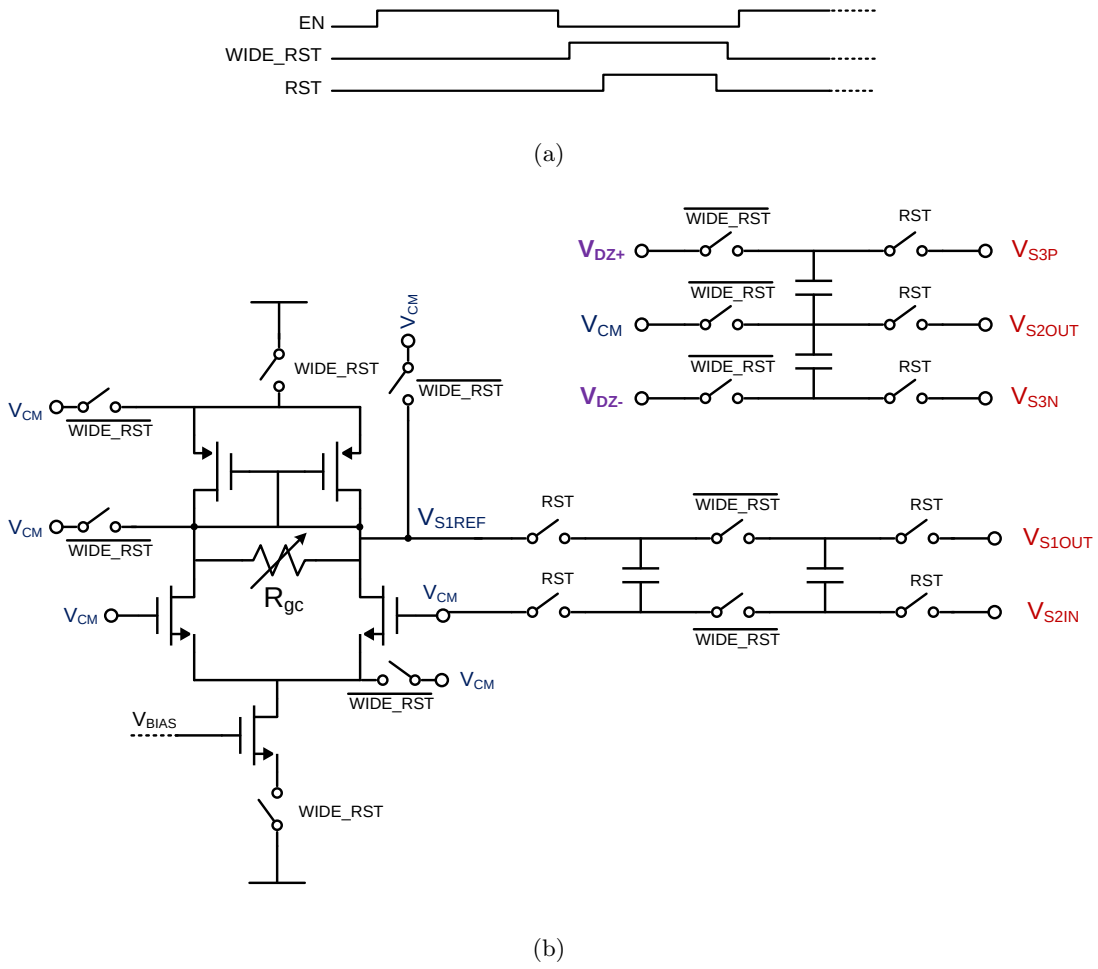


Figure 7.4: Differential ringamp bias circuitry showing a) relation of clock phases EN, RST, and WIDE_RST during a refresh period, b) complete network for generating the charge-transfer signals V_{S3P}, V_{S2OUT}, V_{S3N}, V_{S1OUT}, and V_{S2IN} applied to their respective nodes in Fig. 7.2

into its own dead-zone. While this could be solved by deliberately disabling the coarse ringamp before the fine ringamp begins, we would then lose the benefits of parallelized concurrent operation. Moreover, even if the coarse ringamp is disabled, the fine ringamp will now have to spend much more time slewing to the offset value before it can begin fine settling, and this would come at a severe performance penalty. For these reasons, it is best to just remove the offset altogether and avoid the associated complications. The capacitor C_1 in Fig. 7.2 does exactly that, and stores the difference between the common-mode output voltage of stage 1 and the trip voltage of stage 2 across its plates.

Although the trip voltage of the second stage inverter can be set equal to V_{CM} by design, the common-mode output voltage of stage 1 cannot be easily predicted. It is therefore generated on-chip using the input stage replica circuit shown in Fig. 7.4(b). Rather than charge C_1 directly, a switched-capacitor passive charge-transfer network is used to periodically update the value stored on C_1 in charge domain. This is made possible by the fact that V_{S2IN} of Fig. 7.2 is a floating node. To ensure an accurate and signal-independent charge transfer, the source and drain nodes of the first and second stage transistors are held at V_{CM} when the refresh capacitor is connected. The dead-zone storage capacitors C_2 and C_3 are also handled in a similar way: V_{S3P} and V_{S3N} are floating nodes, and via charge transfer the dead-zone offsets are accurately set across these capacitors using a switched capacitor network shown in Fig. 7.4(b). Compared to the direct voltage charging method that we have used thus far (Fig. 2.1, for example), charge transfer can set very accurate capacitor offsets without requiring those capacitors to be held at an exact, settled, signal-independent voltage every time. This leads to a more accurate and significantly faster refresh cycle, and for high and low accuracy ringamps alike, it is an attractive technique for programming internal ringamp offsets.

Although the charge-transfer biasing scheme assures that the values generated by the bias circuitry will be accurately transferred to the capacitors in Fig. 7.2, these transferred values will still only be as accurate as the references themselves. This is a particular concern for the reference generated by the stage 1 replica circuit (V_{S1REF}), since any noise on V_{BIAS} will translate into error in the common-mode output voltage that is provided. Unfortunately, without sufficient decoupling of V_{BIAS}, this can easily become the case due to the power-save feature's cyclic switching of the main and replica input stages. All of these switching actions kick a certain amount of charge back onto V_{BIAS}, and a component of it will be cycle-dependent. This will ultimately result in transient variations in the common-mode output value generated by the replica stage, which will in turn be transferred to the offset stored across C_1 of Fig. 7.2 (with some amount of attenuation). Although the input-referred value of this error will be further

attenuated by the gain of the first stage, it can still be large enough to limit overall accuracy if V_{BIAS} is not sufficiently decoupled. Although the solution to this issue is simple (add decoupling capacitance to V_{BIAS}), its importance should not be underestimated.

The architectural changes that enable high accuracy do come with some drawbacks, but the price paid turns out to be fairly minimal. In particular, injecting the dead-zone offsets after the second stage inverter means that the maximum V_{OV} that can be provided to the output transistors will be approximately one V_T less than the maximum value possible. In theory, this will lead to worse decoupling between the internal power and external load requirements. However, in practice, this is often not an issue; the fine ringamp will have a much smaller output slew rate compared to a coarse ringamp, and its output transistor W/L's will be many times smaller. Although the decoupling factor is worse, the transistors are already so small that it won't influence internal power requirements. In this 15b ADC design, for example, the W/L of M_{CN} is only 1.7x the minimum size (in the front stage MDAC).

7.4 Experimental Results

The 15b pipelined ADC which incorporates the parallelized ringamp structure was fabricated in a 0.18μm CMOS process . For the sake of both design reuse and direct comparison, the ADC is nearly identical to the design in Chapter 6, with the key difference being the substitution of the fine ringamp in place of the telescopic opamp and Split-CLS network in the stage 1 through 4 MDACs.

At 20MHz sampling rate the ADC achieves 76.0dB SNDR (12.3 ENOB), 76.4dB SNR and 90.6dB SFDR, consuming 4.2mW. The ERBW is found to be above 10MHz, which results in a Figure-of-Merit of 41fJ/conv-step. The MDAC references are set equal to the fine ringamp supply at 0V and 1.2V, allowing the input signal to utilize the entire available supply range. Total accuracy is fundamentally limited by noise, and

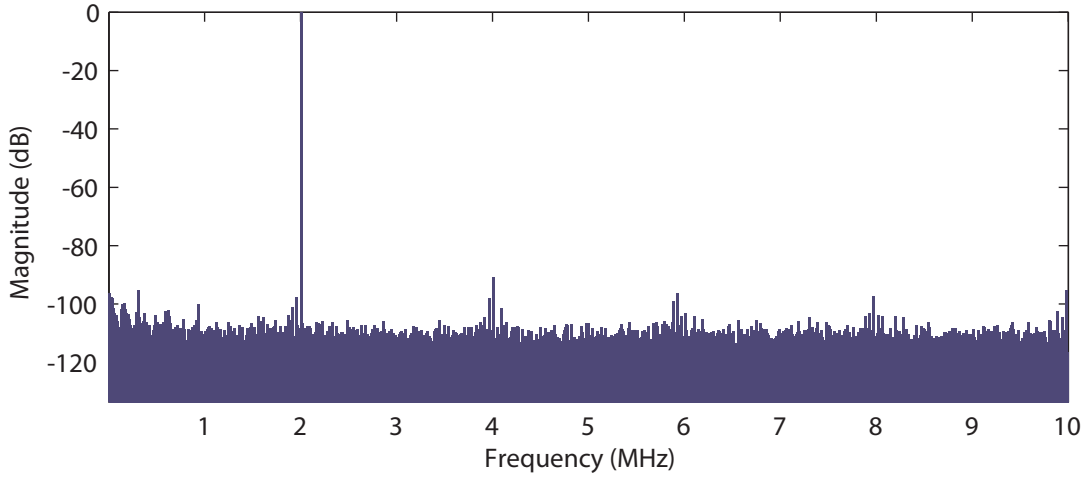


Figure 7.5: Measured output spectrum of the high-resolution parallelized-ringamp ADC for a 2MHz input tone sampled at 20Msps.

the SNR is almost exactly equal to the kT/C noise limit predicted for this design. The input spectrum for a 2MHz input signal sampled at 20Msps is given in Fig. 7.5, and key performance metrics are listed in Table 7.1

The parallelized ringamp design presented here and the Split-CLS design presented in Chapter 6 share an identical ADC structure, and operate at the same conversion frequency. This allows us to make a direct comparison of the two circuit approaches. With regard to power, the telescopic opamp consumes 1500μW while the fine ringamp uses only 790μW to perform the same task (a 47% improvement). This clear power advantage will only become more pronounced in nanoscale CMOS, due to the superior scaling properties of ring amplification. In terms of ultimate accuracy, the experimental results obtained are less conclusive. Although the Split-CLS design achieves the highest *peak* SFDR, this is only for a single data point, at $f_{in} = 1MHz$. By contrast, looking at data for $f_{in} = 2MHz$, we see that the parallelized ringamp has the best SFDR. Most likely, SFDR and THD in both designs are limited by input sampling switch linearity, and the final quantized ADC output is not a reliable measure of amplifier accuracy beyond a certain level. Although the experimental results are inconclusive, from a theoretical viewpoint Split-CLS should be capable of the highest accuracies. The gain "stacking"

Table 7.1: Parallelized Ringamp ADC Summary of Performance

Resolution	15 bits
Analog Supply	1.2/1.3V
Sampling Rate	20 Msps
ERBW	10 MHz
Input Range	2.4 V pk-pk diff.
SNDR	76.0 dB
SNR	76.4 dB
SFDR	90.6 dB
ENOB	12.3 bits
Total Power	4.2 mW
FoM	41 fJ/c-step
Technology	0.18μm 1P4M CMOS
Active Area	1.98mm^2 (3.05mm x 0.65mm)

property of Split-CLS makes achieving 120dB effective gain a realistic endeavor, whereas achieving the same accuracy with an unassisted three-stage fine ringamp is not particularly feasible, even with full use of the pinch-off effect. Ringamp performance can be boosted to even higher accuracies, however, by simply moving from a three-stage to a four-stage architecture.

During the testing of this design, a serious design flaw was discovered. Due to a misunderstanding of how deep-nwell trenches are fabricated in the process technology used, unintentional conductive paths into the substrate were fabricated in all of the bootstrapped switch circuits (30 total throughout the design), and eight decoupling MOSCAPs connected to the fine ringamp's off-chip dead-zone references. This problem also exists in all of the bootstrap circuits in the designs of both Chapter 5 and Chapter 6,

and explains why the switch supply voltage for all of the affected designs needed to be raised well above the expected value in order to achieve necessary accuracies (at the cost of increased power). The fine ringamp's shorted positive dead-zone voltage reference is also a problem. Without the ability to set this reference, the supply voltage of the fine ringamps had to be adjusted in order to compensate (since the pinch-off effect can be tuned by adjusting gate *or* source voltage). Thus, although the coarse ringamps operate at 1.3V, the fine ringamps are forced to operate at 1.2V. This required the main MDAC references to be set smaller than the 1.3V design in Chapter 6, at the cost of SNR. At the time of this chapter's writing, a revision that solves these problems is in the process of being fabricated. The updated design is expected to yield an improvement in both power and accuracy, and push FoM down to approximately 35fJ/conversion-step. When the revision IC has been fully tested, this chapter will be updated with a detailed set of measurement results and made available for download at http://benjamin.hershberg.com/dissertation.

CHAPTER 8. CONCLUSION

8.1 A Comprehensive Solution

In the preceding chapters, we established the groundwork for a comprehensive solution in scalable switched capacitor amplification. It is now possible to build high performance amplifiers in nanoscale CMOS across the entire spectrum of accuracies and speeds. As we saw in Chapter 5, a simple pseudo-differential ringamp structure can cover accuracies up to about 60dB, and in the scaling test of Section 2.4.3 we found that extremely high speeds will be possible in nanoscale CMOS (and even in the fabricated 0.18μm CMOS designs of Chapters 6 and 7 high speed was demonstrated). The parallelized ringamp structure of Chapter 7 is a good candidate for high-resolution, medium-to-high speed applications, and extends ringamps into the realm of high-accuracy in a way that is entirely scalable. At the ultimate highest end of accuracy lies the Split-CLS ringamp/opamp structure of Chapter 6. This is the best option for applications which require precision analog in extreme environments. The tandem use of a ringamp to achieve high efficiency with an opamp to provide the final determination of accuracy strikes an important balance between robustness and performance for critical applications.

8.2 Future Work

Although the designs we have explored throughout this work demonstrate the clear potential of ring amplification as a future technology, it is still, after all, a very new topic. And as with any technique, ring amplification comes with its upsides and downsides. While the benefits far outnumber the drawbacks in this case, there is still a

need for further theoretical analysis and new practical solutions which can enhance its robustness and modularity. In addition, there is a great deal of un-recognized potential in the IC design space for ring amplification which deserves further consideration. In the following section, we will explore some of the key areas for future work and research.

8.2.1 Automatic Tuning

For an actual commercial-grade design, ringamp dead-zones will need to be set on-chip. Luckily, there are many dead-zone injection options that only require simple tunable high-impedance voltage references, and the tuning method does not need to have high linearity or accuracy. For these reasons dead-zone injection is a trivial matter, and can be done digitally with a small, simple, and low power DAC.

Determining the best digital DAC code to use, however, is a more interesting question. The dead-zone should ideally be large enough to ensure stability, but not so large that it limits accuracy. Moreover, in a real product stability must be guaranteed over a wide range of process, voltage, and temperature variation (PVT). One solution to this requirement is to simply set the dead-zone large enough to ensure stability across all PVT corners. Just as extra phase margin will only improve opamp stability, extra dead-zone will only serve to improve ringamp stability. While this conventional solution works well enough, the unique characteristics of ringamps suggest that we could do better. Encoded within the oscillatory behavior of a ringamp is information about its stability: the more periods the ringamp oscillates for before locking into the dead-zone, the less stable it is. Using this information, it is possible to build a simple digital tuning loop which can track PVT and keep the ringamp optimally biased. The basic algorithm for one potential tracking scheme is depicted in Fig. 8.1. This scheme requires very little digital hardware, continuously tracks PVT, and has the option of doing an edge-detection only once every N periods to save power.

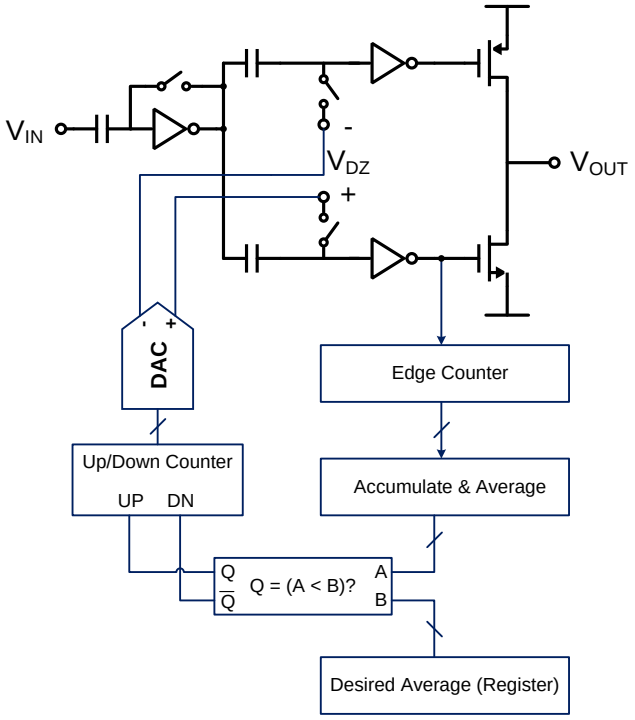


Figure 8.1: A scheme for maintaining optimal ringamp biasing across PVT. The average number of oscillations per period is detected (the signal used for detection here is just one option) and compared against a user-defined average value. Based on the comparison result, the value of the dead-zone DAC reference is adjusted.

8.2.2 Active and Dynamic Loads

The very nature of the pinch-off effect, which lies behind so many of the ringamp's benefits, requires that the charge placed onto the output node will be preserved. For this reason, the designs in this dissertation have had a purely capacitive load. While this isn't a concern for the majority of switched-capacitor circuits, it does limit the ability of ringamps to work in the broader application space where active loads are involved. Luckily, this is probably a structural limitation, not a fundamental one. A possible solution to this problem is a ringamp structure that incorporates a secondary integral output path. The main path would both drive the capacitive component of the output impedance and provide (via some form of integration) the integral path with the correct

bias needed to drive the real component of the output impedance. This approach could, in theory, preserve the beneficial effects of overdrive pinch-off. However, as tends to be true with ringamp design in general, the biggest challenge probably won't be in finding a solution, but in finding an *elegant* solution that retains the speed and efficiency of the core structure.

Another issue of output loading that deserves further consideration is that of dynamic, or unpredictable loads. Even if the output impedance is purely capacitive, the ability of a ringamp to adjust to dynamic variations in capacitance is limited. At the heart of the issue is the fact that the ringamp's maximum slew rate is a function of the output transistor device size, and statically set. A load capacitance above the expected design value will result in decreased speed, and a capacitance below the expected value will result in reduced stability. Although the latter case can be dynamically compensated for by increasing the size of the dead-zone, this will still come at a price of reduced accuracy. For this reason, the ringamp accuracy and stability must be designed with the smallest possible load in mind, and the maximum speed will be limited by the largest possible load. A ringamp with a digitally controlled variable output drive strength seems like a clear solution to this, but the practical implementation becomes less obvious. There are several possible approaches, but so far no clear winner has emerged. And yet, considering how many applications require driving unknown capacitive loads (such as certain types of sensors, displays, and imagers), a simple solution with a minimal impact on core aspects of operation is an important topic for future research.

8.2.3 More Techniques and Analysis

The small handful of designs and techniques that we have explored in this dissertation are really only a scratch on the surface, and there is quite a bit of room for development both in the core ringamp structure itself as well as the higher-level structures (i.e.

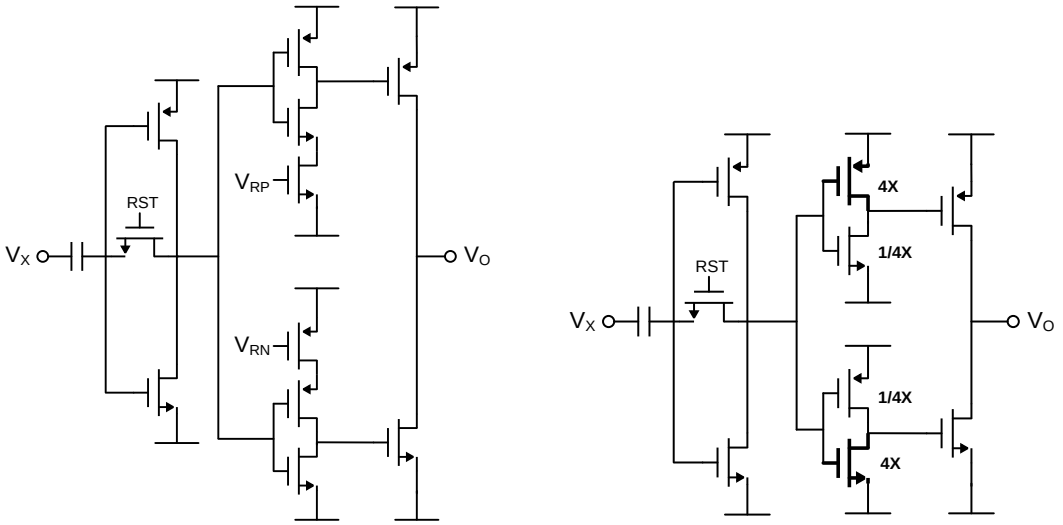


Figure 8.2: Two dead-zone embedding options that directly adjust the threshold of the second stage inverters using series resistance. The approach in (a) comes with the interesting possibility of dynamic adjustment of the dead-zone during the amplification process. Although (b) is not adjustable, it is compact and does not require any external reference.

MDAC, integrators, etc). In the core ringamp circuit, there are several yet-unrealized implementations that would be nice to have, such as a fully differential ringamp with common-mode feedback. Common-mode feedback, while adequately managed in the designs we have discussed, would benefit from more robust and accurate solutions. Another circuit topic worth consideration is the options available for dead-zone injection. Thus far, the dead-zone has always been stored across a set of linear capacitors. However, there are many other ways to induce offsets between the two signal paths, such as the options depicted in Fig. 8.2. Another enticing path for further investigation is that of charge-assist techniques. The parallelized ringamp structure in Chapter 7 is one such charge-assist technique. In that case, the charge-assist occurs on a topological level, but possibilities for charge-assist techniques exist on both a circuit and structural level.

There is also a great deal of core theory and analysis work still to be done. Ringamp noise is a particularly complex subject to accurately model, and although the key points

of interest are identified in Section 3.5.1, further work is needed to formalize the theory. Supply-rejection is another interesting theory question that remains unanswered. On one hand, the effect of supply noise on the stabilization process raises certain concerns. On the other hand, the pinch-off and/or cutoff state of the output transistors at steady-state may cause some unique supply rejection characteristics with respect to both amplitude and frequency. A third question warranting further investigation is the noise-shaping properties of ringamp distortion in over-sampled systems. Unlike an opamp, which is very deterministic, ringamp distortion tends to be more random and broadly distributed.

8.3 Restoring Moore's Law to Analog

In even this initial foray in ring amplification, a reversal in the scaling trends of Fig. 1.3 has begun. At the time of its original publication, the Split-CLS design of Chapter 6 became the highest efficiency high-resolution ADC ever reported ([42]). This record did not stand for long, however, because the parallelized ringamp design in Chapter 7 soon replaced it as the new champion. Considering that this was all achieved in 0.18μm CMOS technology, this is an exciting preview of what ring amplification can accomplish in the future. As we saw in Section 2.4.3, by the 32nm node ringamp based pipelined ADCs will be capable of operating at speeds in excess of 400MHz.

This reversal of trends has pushed amplification-based solutions back into a competitive position in many application spaces that they were beginning to die out in. For example, in the range of medium resolution ADC architectures, SAR ADCs have been the most successful at scaling. Their sequential, digitally switched operation naturally scales with process, and improved element matching has enabled capacitor DAC sizes to be reduced to thermal noise limits. However, SAR ADCs are also limited in speed by the number of sub-operations required in each conversion cycle, and in speed-accuracy product by the costly power/noise tradeoff that occurs in the SAR decision comparator

beyond a certain speed and resolution. Yet, with no good scalable competitors, SAR ADCs have dominated. Until now. Consider, for example, the task of building a 500MHz 10b ENOB ADC in 22nm. From the scaling trends that we have seen in Section 2.4.3, a single pipelined ADC using ringamps could accomplish this task. By contrast, this speed-accuracy product is too high for a SAR ADC to achieve, even in 22nm. To meet the design requirements, multiple SARs would need to be time-interleaved, and with this will come an additional set of challenges and complexities, all at the cost of additional power and area. While SAR ADCs remain an excellent scalable solution, and are by no means doomed for desertion, we can also see that ringamp-based ADCs are now a viable and competitive alternative.

Moving on to high-resolutions, ring amplifiers do, in fact, appear well poised to dominate. In high-resolution Nyquist ADCs, they can take the form of pipelined ADCs such as the already state-of-the-art Split-CLS and parallelized ringamp structures that we have seen here. In oversampling ADCs, ringamps are also capable of changing the application-space landscape considerably. A major speed and power bottleneck in $\Delta\Sigma$ ADCs is the opamps used in the integrator and adder blocks. In a switched-capacitor $\Delta\Sigma$ architecture, ringamps can remove this bottleneck. The result will be an increase in the feasible speed-accuracy product, and we can expect to see high-accuracy $\Delta\Sigma$s achieving input bandwidths that were once the sole domain of Nyquist ADCs.

While ADCs and ring amplifiers will always tend to come up in conversation together, there are many other applications that ring amplifiers show potential in, such as active filters, analog line drivers, equalizers, display technologies, sensor interfaces, and synthesized ADCs [44], to name a few. It will be exciting to see the future of ringamps unfold. Whether they are to become a mass adopted technique or doomed for obscurity, it is still too early to say. But either way, there is a valuable lesson that we can take away from the study of ring amplifiers. In the pursuit of scalable analog techniques, a holistic approach to circuit design is required - one which views the AC, DC, and transient characteristics as simply dimensions of a more complex, highly-interrelated, and

dynamic system. With this in mind, and the techniques developed through the course of this dissertation in hand, the future for both ring amplification and scalable analog design as a whole looks bright indeed.

BIBLIOGRAPHY

[1] B.E. Jonsson, "On cmos scaling and a/d-converter performance," in *NORCHIP, 2010*, nov. 2010, pp. 1 –4.

[2] Yu Cao, "Predictive technology model (ptm)," 2012.

[3] Paul R. Gray, Paul J. Hurst, Stephen H. Lewis, and Robert G. Meyer, *Analysis and Design of Analog Integrated Circuits*, John Wiley And Sons, 2001.

[4] K. Martin D.A. Johns, *Analog Integrated Circuit Design*, John Wiley And Sons, 1997.

[5] Robert F. Pierret, *Semiconductor Device Fundamentals*, John Wiley And Sons, 1996.

[6] S.H. Lewis, H.S. Fetterman, Jr. Gross, G.F., R. Ramachandran, and T.R. Viswanathan, "A 10-b 20-msample/s analog-to-digital converter," *Solid-State Circuits, IEEE Journal of*, vol. 27, no. 3, pp. 351 –358, mar 1992.

[7] I. Mehr and L. Singer, "A 55-mw, 10-bit, 40-msample/s nyquist-rate cmos adc," *Solid-State Circuits, IEEE Journal of*, vol. 35, no. 3, pp. 318 –325, mar 2000.

[8] G.E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82 –85, jan. 1998.

[9] G.E. Moore, "Progress in digital integrated electronics," in *Electron Devices Meeting, 1975 International*, 1975, vol. 21, pp. 11 – 13.

[10] R.R. Schaller, "Moore's law: past, present and future," *Spectrum, IEEE*, vol. 34, no. 6, pp. 52 –59, jun 1997.

[11] C.A. Mack, "Fifty years of moore's law," *Semiconductor Manufacturing, IEEE Transactions on*, vol. 24, no. 2, pp. 202 –207, may 2011.

[12] S.E. Thompson, R.S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M.T. Bohr, "In search of "forever," continued transistor scaling one new material at a time," *Semiconductor Manufacturing, IEEE Transactions on*, vol. 18, no. 1, pp. 26 – 36, feb. 2005.

[13] M. Shrivastava, R. Mehta, S. Gupta, N. Agrawal, M.S. Baghini, D.K. Sharma, T. Schulz, K. Arnim, W. Molzer, H. Gossner, and V.R. Rao, "Toward system on chip (soc) development using finfet technology: Challenges, solutions, process co-development amp; optimization guidelines," *Electron Devices, IEEE Transactions on*, vol. 58, no. 6, pp. 1597 –1607, june 2011.

[14] G.E. Moore, "No exponential is forever: but "forever" can be delayed! [semiconductor industry]," in *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International*, 2003, pp. 20 – 23 vol.1.

[15] W.R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A.M. Sule, M. Steer, and P.D. Franzon, "Demystifying 3d ics: the pros and cons of going vertical," *Design Test of Computers, IEEE*, vol. 22, no. 6, pp. 498 – 510, nov.-dec. 2005.

[16] B. Murmann, P. Nikaeen, D.J. Connelly, and R.W. Dutton, "Impact of scaling on analog performance and associated modeling needs," *Electron Devices, IEEE Transactions on*, vol. 53, no. 9, pp. 2160 –2167, sept. 2006.

[17] K. Bult and G.J.G.M. Geelen, "A fast-settling cmos op amp for sc circuits with 90-db dc gain," *Solid-State Circuits, IEEE Journal of*, vol. 25, no. 6, pp. 1379 –1384, dec 1990.

[18] Yun Chiu, P.R. Gray, and B. Nikolic, "A 14-b 12-ms/s cmos pipeline adc with over 100-db sfdr," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 12, pp. 2139 – 2151, dec. 2004.

[19] C.C. Enz and G.C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584 –1614, nov 1996.

[20] Jipeng Li and U Moon, "A 1.8-v 67-mw 10-bit 100-ms/s pipelined adc using time-shifted cds technique," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 9, pp. 1468 – 1476, sept. 2004.

[21] A.N. Karanicolas, Hae-Seung Lee, and K.L. Barcrania, "A 15-b 1-msample/s digitally self-calibrated pipeline adc," *Solid-State Circuits, IEEE Journal of*, vol. 28, no. 12, pp. 1207 –1215, dec 1993.

[22] B.R. Gregoire and U Moon, "An over-60 db true rail-to-rail performance using correlated level shifting and an opamp with only 30 db loop gain," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 12, pp. 2620 –2630, dec. 2008.

[23] J.K. Fiorenza, T. Sepke, P. Holloway, C.G. Sodini, and Hae-Seung Lee, "Comparator-based switched-capacitor circuits for scaled cmos technologies," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 12, pp. 2658 –2668, dec. 2006.

[24] L. Brooks and Hae-Seung Lee, "A zero-crossing-based 8-bit 200 ms/s pipelined adc," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 12, pp. 2677 –2687, dec. 2007.

[25] L. Brooks and Hae-Seung Lee, "A 12b, 50 ms/s, fully differential zero-crossing based pipelined adc," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 12, pp. 3329 –3343, dec. 2009.

[26] H. Venkatram, B. Hershberg, and U.-K. Moon, "Asynchronous cls for zero crossing based circuits," in *Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on*, dec. 2010, pp. 1025 –1028.

[27] Boris Murmann, "Adc performance survey 1997-2012," 2012.

[28] Benjamin Hershberg, Skyler Weaver, and U Moon, "Design of a split-cls pipelined adc with full signal swing using an accurate but fractional signal swing opamp," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 12, pp. 2620 –2630, dec. 2010.

[29] J. Chu, L. Brooks, and Hae-Seung Lee, "A zero-crossing based 12b 100ms/s pipelined adc with decision boundary gap estimation calibration," in *VLSI Circuits (VLSIC), 2010 IEEE Symposium on*, june 2010, pp. 237 –238.

[30] A. Chow and Hae-Seung Lee, "Offset cancellation for zero crossing based circuits," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 30 2010-june 2 2010, pp. 1719 –1722.

[31] H. Iwai, "Cmos scaling towards its limits," in *Solid-State and Integrated Circuit Technology, 1998. Proceedings. 1998 5th International Conference on*, 1998, pp. 31 –34.

[32] F. Krummenacher, E. Vittoz, and M. Degrauwe, "Class ab cmos amplifier micropower sc filters," *Electronics Letters*, vol. 17, no. 13, pp. 433 –435, 25 1981.

[33] M. Figueiredo, R. Santos-Tavares, E. Santin, J. Ferreira, G. Evans, and J. Goes, "A two-stage fully differential inverter-based self-biased cmos amplifier with high efficiency," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 58, no. 7, pp. 1591 –1603, july 2011.

[34] Youngcheol Chae and Gunhee Han, "Low voltage, low power, inverter-based switched-capacitor delta-sigma modulator," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 2, pp. 458 –472, feb. 2009.

[35] B. Hershberg, Tawfiq Musah, S. Weaver, and U Moon, "The effect of correlated level shifting on noise performance in switched capacitor circuits," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, may 2012.

[36] Benjamin Hershberg, Skyler Weaver, and U Moon, "A 1.4v signal swing hybrid cls-opamp/zcbc pipelined adc using a 300mv output swing opamp," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pp. 302 –303, feb. 2010.

[37] C.T. Peach, U Moon, and D.J. Allstot, "An 11.1 mw 42 ms/s 10 b adc with two-step settling in 0.18 m cmos," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 2, pp. 391 –400, feb. 2010.

[38] M. Dessouky and A. Kaiser, "Very low-voltage digital-audio $\delta\sigma$ modulator with 88-db dynamic range using local switch bootstrapping," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 3, pp. 349 –355, mar 2001.

[39] T. Musah and U Moon, "Pseudo-differential zero-crossing-based circuit with differential error suppression," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 30 2010-june 2 2010, pp. 1731 –1734.

[40] Benjamin Hershberg, Skyler Weaver, Kazuki Sobue, Seiji Takeuchi, Koichi Hamashita, and U Moon, "A 61.5db sndr pipelined adc using simple highly-scalable ring amplifiers," *VLSI Circuits (VLSIC), 2012 Symposium on*, june 2012.

[41] Benjamin Hershberg, Skyler Weaver, Kazuki Sobue, Seiji Takeuchi, Koichi Hamashita, and U Moon, "Ring amplifiers for switched capacitor circuits," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 12, dec. 2012.

[42] Benjamin Hershberg, Skyler Weaver, Kazuki Sobue, Seiji Takeuchi, Koichi Hamashita, and U Moon, "Ring amplifiers for switched-capacitor circuits," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, feb. 2012, pp. 460 –462.

[43] Hwi-Cheol Kim, Deog-Kyoon Jeong, and W. Kim, "A partially switched-opamp technique for high-speed low-power pipelined analog-to-digital converters," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 53, no. 4, pp. 795 – 801, april 2006.

[44] S. Weaver, B. Hershberg, and U Moon, "Digitally synthesized stochastic flash adc using only standard digital cells," in *VLSI Circuits (VLSIC), 2011 Symposium on*, june 2011, pp. 266 –267.