IMPACTS OF CMOS SCALING ON THE ANALOG DESIGN

By

XUNYU ZHU

Bachelor of Engineering
University of Science and Technology Beijing
Beijing, China
1995

Master of Engineering
University of Science and Technology Beijing
Beijing, China
1998

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2005

IMPACTS OF CMOS SCALING ON THE ANALOG

DESIGN

Thesis Approved:

Dr. Chris Hutchens
Thesis Advisor
Dr. Louis G. Johnson

Dr. Yumin Zhang

Dr. Jack Cartinhour

Dr. A. Gordon Emslie
Dean of the Graduate College

ACKNOWLEDGMENTS

The completion of this dissertation marks the end of my formal education. Throughout my life, I have been unconditionally and lovingly supported by numerous people. At this moment, I will take this opportunity to express my sincere gratitude to each and every one who have supported and loved me all along.

First and foremost, I sincerely would like thank my parents for their unconditional love and support throughout my life. Their concern and caring always extends to the deepest of my heart. Six years in the U.S., I would not able to achieve this milestone of my life without expressing my sincere gratitude to my Dad and my Mom.

Especially, I would like to express my deep appreciation to my wife, Xuehan Zhang, who has mentally encouraged me in all aspects of my life all these years. Her encouragement and support were the driving force for the write-up of this dissertation.

Living in Mixed-Signal VLSI Lab has definitely been a lot of fun, and an unforgettable part of my experience. I probably leaned as much in the MSVLSI as I did in the classroom. I would like to thank all the members of the MSVLSI past and present, especially Dr. Liu. He taught me everything from the basic schematic drawing to using test instruments. Narendra, did you still remember the days you slept on the floor of the

lab? I'll never forget you guys, Sameer, Jianning, Venket, Ansuhman, Vinay, Vijay, Bharath, Srinivasan, Ira, and Prasanna.

I would also like to thank the faculty of the ECEN Department at OSU, especially Dr. Louis G. Johnson, Dr. Yumin Zhang, and Dr. Jack Cartinghour of Electrical Engineering Technology for being on my Graduate Committee. I am also indebted to Dr. Weili Zhang who has helped me numerous times in the dicing and bonding room. In addition, I am grateful to Space and Warfare (SPAWAR) Systems Command (formerly NRaD), San Diego, CA for its support of this project.

Finally, I would like to thank my advisor, Dr. Chris Hutchens. I appreciate all of the opportunities and assistance he has provided me over the past six years from the English grammar to the design knowledge. His numerous guidance and advice have deeply enhanced my chances for success at OSU and my subsequent career and life. Thank you!

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

NOMENCLATURE

ADC	Analog to digital converter
A_c	Close loop DC gain
A_o	Open loop DC gain
B	Quantizer number of bit
B	Resolution bits
BTBT	Band-to-Band Tunneling
BW	Bandwidth
C_F	Feedback/Integration capacitance
C_{gs}	Gate to source capacitance
C_{gd}	Gate to drain capacitance
C_L	Load capacitance
C_{Leff}	Effective load capacitance
C_m	Miller capacitance
CMFB	Common mode feedback
C_{ox}	Gate oxide capacitance
C_S	Sampling capacitance
D	Duty cycle
DIBL	Drain-Induced Barrier Lowering
DNL	differential non-linearity
DR	Dynamic range
ΔL	Mismatch of the channel length
ΔW	Mismatch of the channel width
ΔV	Drain to source saturation voltage needed at the operation point
E_g	Barrier of height assuming a triangular barrier
ENOB	Effective number of bits

E_p	Electric field at the junction
ε	Component mismatch
f	Feedback factor
f_B	Base-band frequency
FFT	Fast Fourier Transform
f_{max}	Power gain frequency
f_s	Sampling frequency
f_T	Unit current gain frequency
Φ_s	Potential at the surface
GBP	Gain bandwidth product
GIDL	Gate-induced drain leakage
g_{ds}	Drain to source conductance
g_m	Mutual transconductance
$\hbar$	Modified Planck's constant
INL	Integral non-linearity
ID	Darin current
J_t	Tunneling current density
k	Boltzmann's constant ($1.38 \times 10^{-23}\ ^{J}/_{K^\circ}$)
K_p	transconductance parameter
Ksps	Kilo samples per second
L	Channel length
L_{eff}	Effective channel length
LSB	Least significant bit
MOSFET	Metal oxide semiconductor field effect transistor
μ	Self-gain of the transistor
N_A	Substrate doping level
NCMC	Nested cascode miller compensated
NGCC	Nested Gm-C compensated
OTA	Operational transconductance amplifier
Π	Power settling product efficiency
R_{in}	Input resistance

R_{on}	Transistor on resistance
rms	Root mean square
S	Sub-threshold slope
SC	Switched capacitor
SCE	Short channel effect
SFDR	Spurious free dynamic range
SH	Sample and Hold
S_i	Noise current spectrum
SINAD	Signal to noise and distortion ratio
SNR	Signal to noise ratio
SOI	Silicon on insulator
SOS	Silicon on sapphire
SR	Slew Rate
q	Electron charge ($1.6\times10^{-19}C$)
T	Kelvin temperature degree
t_f	Fall time
t_{ox}	Oxide layer thickness
t_r	Rise time
t_s	Sampling period
U_T	Thermal voltage
V_{cm}	Common mode voltage
VCO	Voltage controlled oscillator
V_{DS}	Drain to source voltage
V_{DSsat}	Drain to source saturation voltage
V_{GS}	Gate to source voltage
V_{GS}	Gate to source voltage
V_{in}	Input voltage
V_n	Noise voltage
V_{ref}	Accurate reference voltage
V_{res}	Residue voltage
V_{os}	Offset voltage

V_{out}	Output voltage
v_{sat}	Saturation velocity of an electron or hole
V_T	Threshold voltage
W	Channel width
W_{dm}	Maximum depletion width
x_s	Low-doped width

Chapter 1

Introduction

1.1 Background

In recent years, advances in micro electronics makes the wide application of wireless technology possible, changing the way people work, live, and entertainment [1]. For instance, cellular phone, Global Positioning System (GPS), and RF ID all have become a part of our daily routine. These products require wider bandwidth, higher resolution, smaller size, and demand less power consumption [1-3].

Figure 1.1 is the block diagram of a typical wireless application. Briefly speaking, circuits from the antenna to mixer are referred to as the RF analog circuitry; while circuits from the modulator to the ADC/DAC are referred to as the baseband analog circuit. Conventionally, CMOS technology dominates in the digital circuits, and bipolar and/or SiGe technology dominates the RF circuit for their extremely high f_{max}. Thanks to the continuous device dimension scaling, CMOS technology can also achieve f_T up to 95 GHz [4] and f_{max} up to 150 GHz [5].In order to minimize the power and manufacture

cost, and increase the circuit reliability, it highly desirable to integrate the entire system into a single chip (System on Chip) resulting in lower power demands, lower manufacturing cost and less weight. From this point of view, it seems that CMOS technology is the most favorite choice.

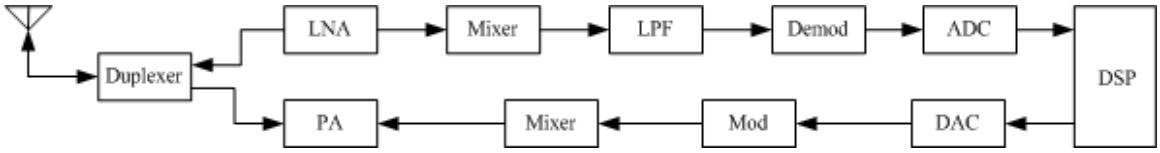


Figure 1.1 The block diagram of a typical wireless application

CMOS SOI technology first emerged for the radiation harden and later for high temperature environment applications [6]. In the recent years, SOI is becoming a mainstream technology having almost the same fabrication process and costs as CMOS. The total dielectric isolation leads to the elimination of the latch-up, and to lower parasitic capacitance and lower leakage current. In addition high Q inductors are now available on the SOI chip as a direct result of the insulation layer allowing for low loss (high resistitvity substrates) from the substrate [7]. This enables or even favors SOI technology for RF applications. A final result of full dielectric isolation is the extension of CMOS for use at ultra high temperatures in excess of 240 ^{0}C.

On the other hand, continuous scaling decreases the per unit manufacturing cost while increasing circuit speed [8]. This places an ever increasing burden on the analog RF and baseband designers as more restrict conditions are faced – lower supply voltage, less self gain, and less consistent device parameters. At the same time, device scaling itself encounters more and more problems, such as short-channel effect, and larger

leakage current [9]. SOI technology gives a promising solution to these emergent problems especially the FinFET or dual gate process [8].

1.2 Research goals

What is the impact of the CMOS scaling to the analog design and analog circuit topology? This research is attempting to answer these question by investigating the analog design limits as a result of the scaling impact on the device performance and on the circuit performance. A more general goal of this research is to develop guidance with regard to the power efficiency analog circuit design under the scaling scenario. In the context of these goals, some key research results are summarized below:

- Demonstrate the minimum threshold voltage V_T for analog design is larger than 300mV; the minimum V_{DSsatQ} is approximate 100mV; and the minimum supply voltage is around 1.1V.
- Figure of merit ─ Power settling product efficiency (Dynamic range / (Power · Settling time) is proposed as a guide to choose an OTA in the approach to analog design in the presents of device scaling.
- For the first time, all the classical OTA topologies are analyzed in the close loop form in a single document and compared by using the power settling product efficiency as the criteria as the supply voltage scales. Nested gain boost topology is demonstrated as an optimum topology especially for operation in the weak inversion.

- Designed a low-voltage, low-power pipelined ADC. Spice and Simulink simulation demonstrating both its feasibility and the power settling product figure of merit approach.

1.3 Organization

In Chapter 2, the impact of CMOS scaling to the device performance is discussed. In Chapter 3, the analog design limits are investigate as the CMOS scales and power settling product efficiency figure of merit is proposed. In Chapter 4, different topologies are compared by their power settling product efficiency as the supply voltage and overdrive voltage scale. Optimum topology is given under different specific supply voltage and overdrive voltage. In Chapter 5, a low voltage pipelined ADC is designed as an example. A Matlab Simulink library for the pipelined ADC is built, and the behavior simulation is completed. Components of the pipelined ADC simulation are done in the spice environment.

Chapter 2

Challenges to the Analog Design with CMOS Scaling

2.0 Introduction to the CMOS scaling

The silicon CMOS architecture has successfully become the dominant technology in the integrated circuits [10], as a result of the closely matched device mobility (1.5 to 3), quality native oxide, and its scalability providing decreased power consumption at enhanced performance levels. CMOS gate lengths have reduced from 10um in the 1970's to the present day geometry of less than 0.09um under the guideline set by the Dennard et al [11] , which is based on the constant electric field in the device, shown in the Table 2.1.

Table 2.1 Scaling rules for CMOS

Device Parameters	Scaling Factor
Device dimension t_{ox}, L, W,	$1/\kappa$
Doping concentration N_A, N_D	κ
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance C_{gs}	$1/\kappa$
Delay time VC/I	$1/\kappa$
Power consumption VI	$1/\kappa^2$
Electric field of in the device E	1

Table 2.2 ITRS roadmap 2002

YEAR OF PRODUCTION	*2003*	*2004*	*2005*	*2006*	*2007*
DRAM ½ Pitch (nm)	100	90	80	70	65
MPU ½ Pitch (nm)	107	90	80	70	65
MPU Printed Gate Length (nm)	65	53	45	40	35
MPU Physical Gate Length (nm)	45	37	32	28	25
V_{DD} *(V)*	1.0	1.0	0.9	0.9	0.7
T_{ox} *Equivalent oxide thickness (electrical) (nm)*	2	2	1.9	1.9	1.4

The international Technology Roadmap for Semiconductors (ITRS) [12] points out that the CMOS gate length will continue scaling to below 50nm (Table 2.2 Figure 2.1) in 2005. However, ITRS focuses heavily on the digital applications performance ignoring analog applications. As to analog or mixed-signal design, device scaling imposes more challenges to the analog engineers, when compared to their digital colleagues, in order to maintain the circuit performance.

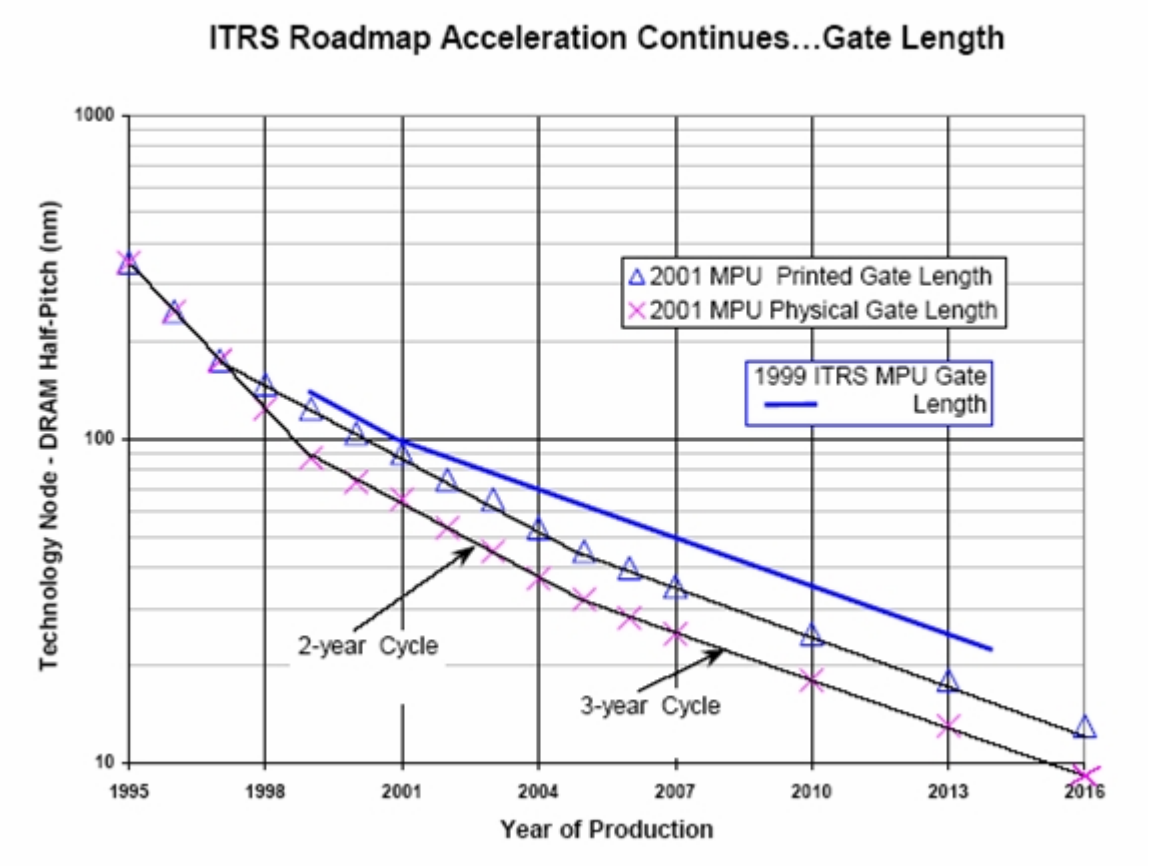


Figure 2.1 ITRS roadmap acceleration continues — gate length trends.

This chapter first presents an overview of the problems induced by the CMOS scaling, focusing on the challenges in analog circuit design. Then, the FinFET device is introduced, as it is believed to be the most promising device to replace the traditional planar CMOS devices in the next decade, for it alleviates many of the problems faced in process scaling.

2.1 Challenges to the analog design in the scaling process

2.1.1 Threshold voltage

The MOSFET threshold voltage should scale at the same rate as the other device parameters according to the scaling theory. However, it scales at a reduced rate, which is clearly understood when taking standby power consumption into consideration for digital applications [13]. The off-state leakage current increases 10 times for every 0.1V decrease in the threshold voltage. Additionally, threshold voltage rolls off as the gate length shrinks due to the loss of gate voltage control over the depletion charge [9], which is referred to as short-channel effects (SCE). SCE is more apparent when the device scales down below sub-100nm, which is one of the main sources of the leakage current increase in deep submicron (DSM) devices [14].

The minimum threshold voltage is determined by considering the acceptable sub-threshold leakage current in the digital circuits under the maximum operating temperature and supply voltage conditions simultaneously with maximum I_{Don}. The sub-threshold leakage is the dominant contributor to the stand-by power in the 0.1um process [10]. Stand-by power occupies 0.01% of the active power in a 1um process while occupying 10% in a 0.1um process. In order to control the stand-by power, the sub-threshold slope should be as large as possible along with an adequate threshold voltage. Presently, the sub-threshold swing of the FinFET is 64-67mV/decade (Figure 2.12). Assuming sub-threshold occupies 4-5 decade (resulting in an on to off ration approaching 10^5), the minimum threshold voltage should be greater than 0.25 ~ 0.3V. With a threshold

temperature coefficient of -1mV/°C and a maximum operation temperature of 90°C, a threshold voltage of 300mV would ensure an acceptable on/off ratio for digital logic.

The variation of the threshold voltage is another concern for both the digital and analog design. The fluctuation of the threshold voltage results in the fluctuation of the propagation delay, leading to intolerable clock skews and low circuit yields [15]. At the same time, it results in a problematic offset voltage in analog processing, such as comparators of ADCs. In the [16], Wong revealed that the fluctuation is not only related to the fluctuation of the number of the dopants, but also to the distribution of the dopants. For the uniformly doping, the standard deviation of the threshold voltage is

$$\sigma_{V_T} = \frac{q}{C_{ox}} \sqrt{\frac{N_A W_{dm}}{3LW}} \tag{2.1}$$

While for the retrograde doping,

$$\sigma_{V_T} = \frac{q}{C_{ox}} \sqrt{\frac{N_A W_{dm}}{3LW}} (1 - \frac{x_s}{W_{dm}})^{3/2} \tag{2.2}$$

where N_A is the substrate doping level, C_{ox} is the gate oxide capacitance, W_{dm} is the maximum depletion width, x_s is the low-doped width.

However, for the ultra-thin films used in fully depleted (FD) SOI or dual gate (DG) FinFET devices, the conventional approach via highly doped film to control the threshold voltage is not viable because the threshold voltage is severely sensitive to the gate oxide and the film thickness variations [17]. Therefore, to improve on threshold performance un-doped film is proposed to achieve good control of the threshold voltage

while improving channel mobility and potentially improving noise performance as well. Trivedi [17] found for the sub-30nm ultra-thin, lightly doped devices, the variation of the threshold voltage is affected by the quantum mechanical (QM) as

$$\sigma_{V_T}^2 \cong \frac{S}{(kT/q)\ln 10} \cdot \frac{0.3763}{(m_x/m_0)t_{si}^2} \tag{2.3}$$

where S is the sub-threshold slope, m_x/m_0 is the ratio of the carrier effective mass in the direction of the confinement to the free electron mass. For the mismatch current caused by the mismatch of the threshold voltage, (2.4.a) and (2.4.b) are for the device operation in the strong inversion and the weak inversion respectively.

$$\frac{\sigma^2(I_{ds})}{I_{ds}^2} = \frac{4\sigma^2(V_T)}{(V_{GS}-V_T)^2} \tag{2.4.a}$$

$$\frac{\sigma^2(I_{ds})}{I_{ds}^2} = \frac{\sigma^2(V_T)}{(nU_T)^2} \tag{2.4.b}$$

where, $(nU_T)^{-1}$ is the slope of the I_{DS}-V_{GS} curve in weak inversion plotted in a log scale.

From (2.2-2.4), we can find that under the same electric field scaling rule, the mismatch of the current is inverse proportional to the scaling factor for the retrograde doping or the square of the scaling factor for the un-doped film. This means in order to get the same offset result from the fluctuation of the threshold voltage, κ^2 time's area and power consumption are paid as price for the latter case.

2.1.2 Leakage current

As to the logic and memory applications, the most important performance gauges are On/Off ratio (leakage current), delay, power and reliability [18]. Leakage current will not only disrupt the dynamic nature of the pass transistor logic but increase the stand-by power consumption as well. Leakage current includes two parts: one is the gate leakage current; the other is the Drain to the Source leakage current (off-state leakage).

2.1.2.1 Gate leakage current

Gate leakage results from the gate silicon dioxide tunneling phenomenon. As the devices scale, gate oxide thickness scales as well to keep the transistor short-channel performance. When gate oxide is sufficiently thin, there is substantial current flowing from the gate to the channel even under the low voltage operation conditions shown in Figure 2.2. I_{gs} and I_{gd} is the tunneling current between the source-drain extension (SDE) and the gate overlap, which is called edge direct tunneling (EDT). I_{gc} is the tunneling current to the channel. Gate leakage current increases exponentially with the decrease of the gate oxide thickness [13, 19, 20]. I_{gc} is the dominant leakage current when $V_g > 0$, while EDT is dominant when $V_{fb} < V_g < 0$ [21]. For the digital circuit, the dissipation power associated with the gate leakage is acceptable until gate oxide thickness reduces less than 2nm Figure 2.3 [13, 22]. This is under the assumption that the maximum tolerable gate leakage current is less than $1A/cm^2$ or $10nA/um^2$. When high κ materials are used as the gate insulator in later generation devices, gate leakage currents are projected to become greatly reduced.

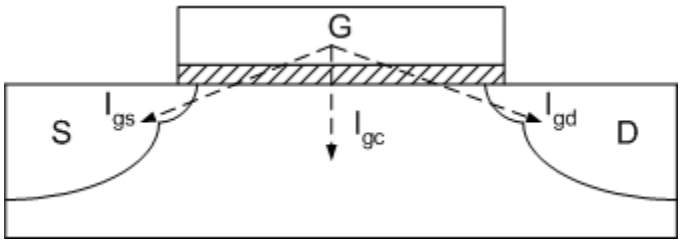


Figure 2.2 Component of the gate leakage current.

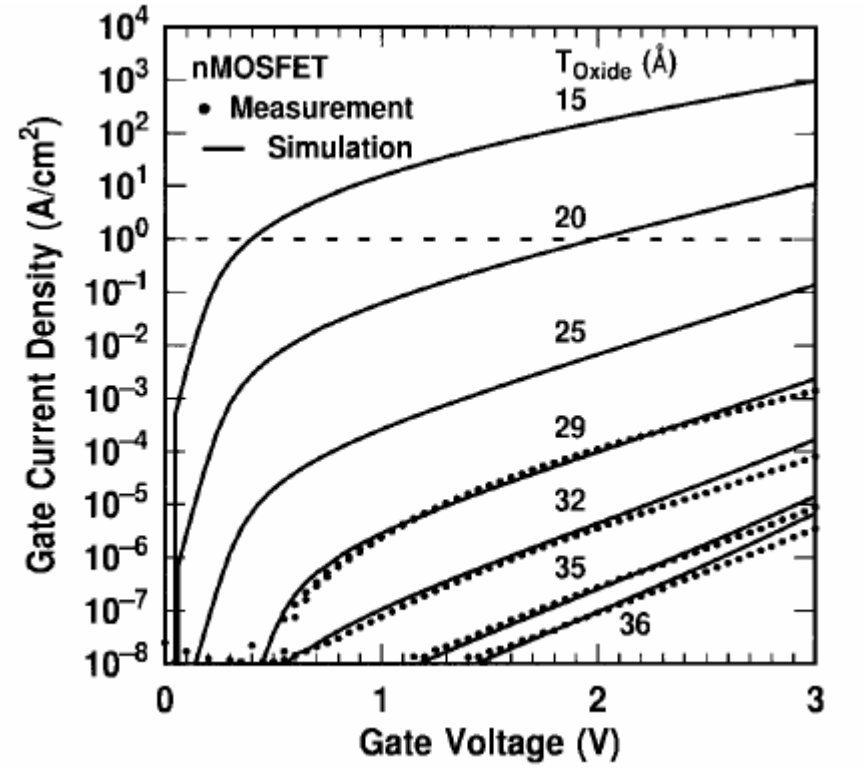


Figure 2.3 Measured and simulated I_G–V_G characteristics under inversion conditions of the nMOSFET's. The dotted line indicates the 1 A/cm^2 limit for leakage current [13, 22].

2.1.2.2 Off-state leakage current

The off-state current is composed of 1) the surface Band-to-Band Tunneling (BTBT) or gate-induced drain leakage (GIDL); 2) the bulk BTBT [23]; 3) the sub-threshold leakage current; and 4) leakage current due to short-channel effects owing to the Drain-Induced Barrier Lowering (DIBL) .

a) GIDL leakage current

GIDL leakage is the leakage current occurring in the deep depletion layer in the gate-drain overlap region due to the band-to-band tunneling. It can be expressed as [24]

$$I_{GIDL} = AE_s \exp(-B / E_s)$$
$$E_s \approx \frac{V_{DG} - 1.2}{3T_{ox}} \tag{2.5}$$

where E_s is the vertical electric field at the silicon surface, A is a preexponential constant and B=21.3MV/cm [24]. A band bending of 1.2eV is the minimum potential for band-to-band to occur [24]. *For the sub-100um device, the GIDL effect can be ignored for supply voltages less than 1.2V according to the ITRS roadmap*.

b) Substrate BTBT leakage current

The substrate BTBT is the junction leakage current occurring in the Drain/Source depletion area with the substrate. The tunneling current density can be expressed as [25]

$$J_t = \frac{\sqrt{2m^*}q^3 E_p V_a}{4\pi^3 \hbar^2 \sqrt{E_g}} \exp\left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\hbar E_p} \right) \tag{2.6}$$

where m^* is the electron effective mass and V_a is the reverse bias voltage across the p-n junction, E_g is the barrier of height assuming a triangular barrier, E_p is the electric field at the junction, $\hbar$ is the modified Planck's constant.

The substrate BTBT current is critical dependant on the substrate doping concentration [23, 26]. Following the rule described in Table 2.1, substrate BTBT current will increase dramatically as the device scales to sub-100nm. We will find in the later,

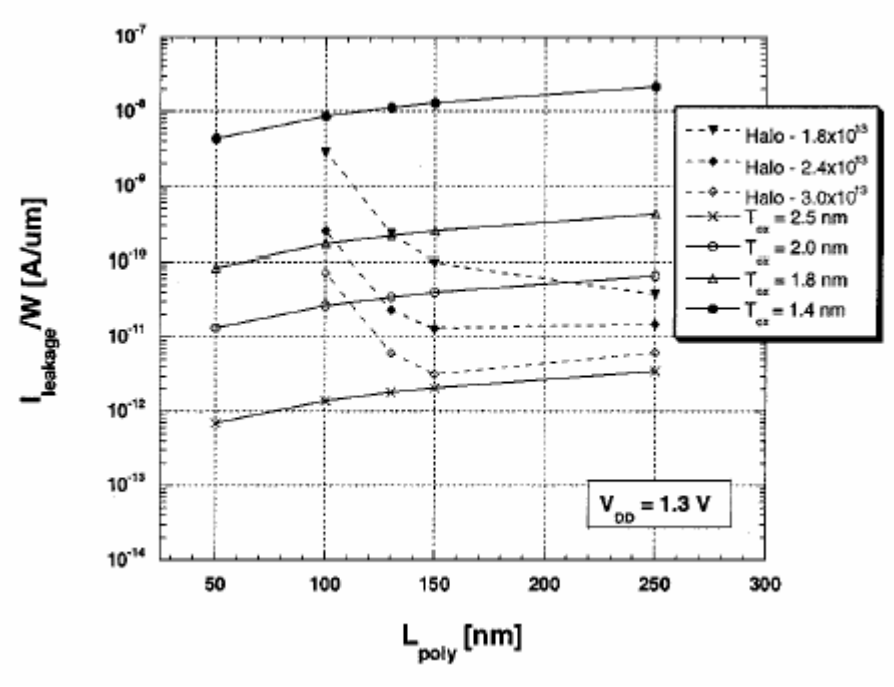
this can be solved by new device — the FinFET, which uses nearly intrinsic Si as the substrate resulting in substantially less substrate BTBT current.

c) DIBL leakage current

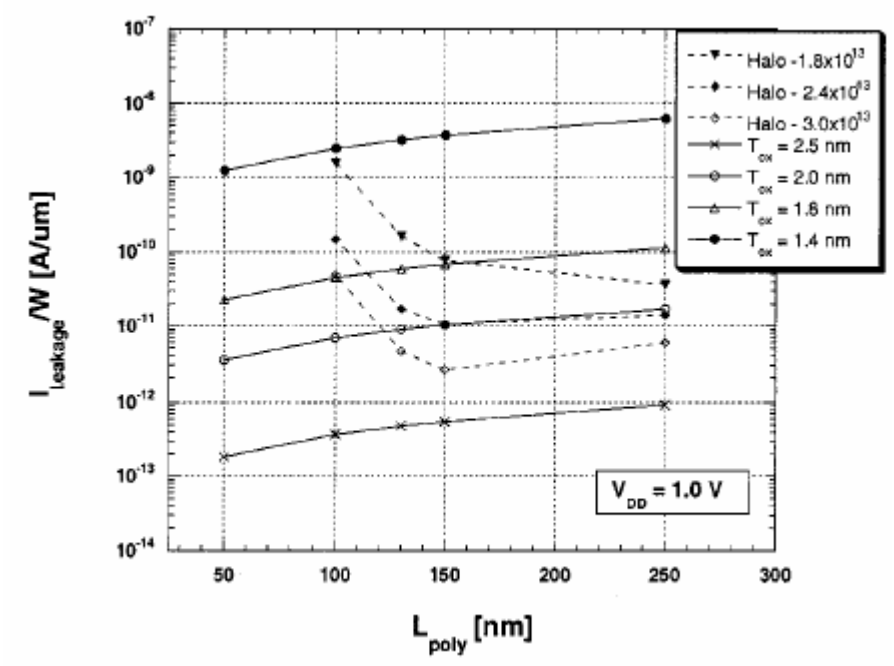
DIBL leakage is a surface punch-through phenomenon occurring in the short channel device due to the barrier lowering at the source junction by the drain field and the band-bending under the effect of the gate-induced surface space charges [27].

$$I_{off} = W_{tot} I_o \exp(\frac{qm|V_{ds}|}{kT}) / L_{eff}$$
$$I_o \propto \exp(\frac{-E_g + q\phi_s}{kT}) \quad (2.7)$$

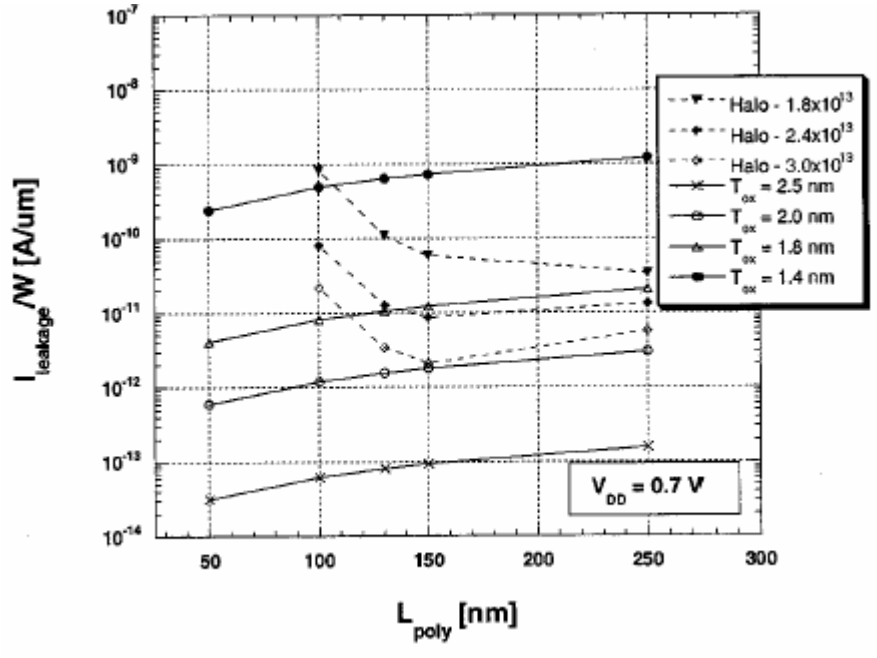
where m is a dimensionless factor, W_{tot} is the total off-state device width, and L_{eff} is the effective channel length, E_g is the bandgap, and ϕ_s is the potential at the surface [27]. The DIBL leakage with the relationship of channel length and supply voltage is shown in Figure 2.4 [28].


$I_{leakage}/W$ [A/um]
L_{poly} [nm]
Halo - 1.8x10^13
Halo - 2.4x10^13
Halo - 3.0x10^13
T_{ox} = 2.5 nm
T_{ox} = 2.0 nm
T_{ox} = 1.8 nm
T_{ox} = 1.4 nm
V_{DD} = 1.3 V

(a)


$I_{Leakage}/W$ [A/um]
L_{poly} [nm]
Halo -1.8x10^13
Halo - 2.4x10^13
Halo - 3.0x10^13
T_{ox} = 2.5 nm
T_{ox} = 2.0 nm
T_{ox} = 1.8 nm
T_{ox} = 1.4 nm
V_{DD} = 1.0 V

(b)



(c)

Figure 2.4 The dashed lines represent the DIBL leakage and the solid one represent the gate leakage.

Although Figure 2.4 cannot be used to represent the actual general leakage current values, it does demonstrate the trends. When channel lengths are less than 150nm, the DIBL leakage increases dramatically.

d) Sub-threshold leakage current

The channel current will not abruptly become zero when the gate voltage is below the threshold voltage. In the sub-threshold region, the channel current is dominated by the diffusion current. It can be expressed as [29]

$$I_{ds\_sub} = \mu C_{ox} \frac{W}{L}(m-1)\left(\frac{kT}{q}\right)^2 \exp\left(\frac{V_g - V_T}{mkT/q}\right)\left(1-\exp\left(-\frac{V_{ds}}{kT/q}\right)\right) \qquad (2.8)$$

Sub-threshold leakage current versus channel length and threshold voltage is shown in Figure 2.5.

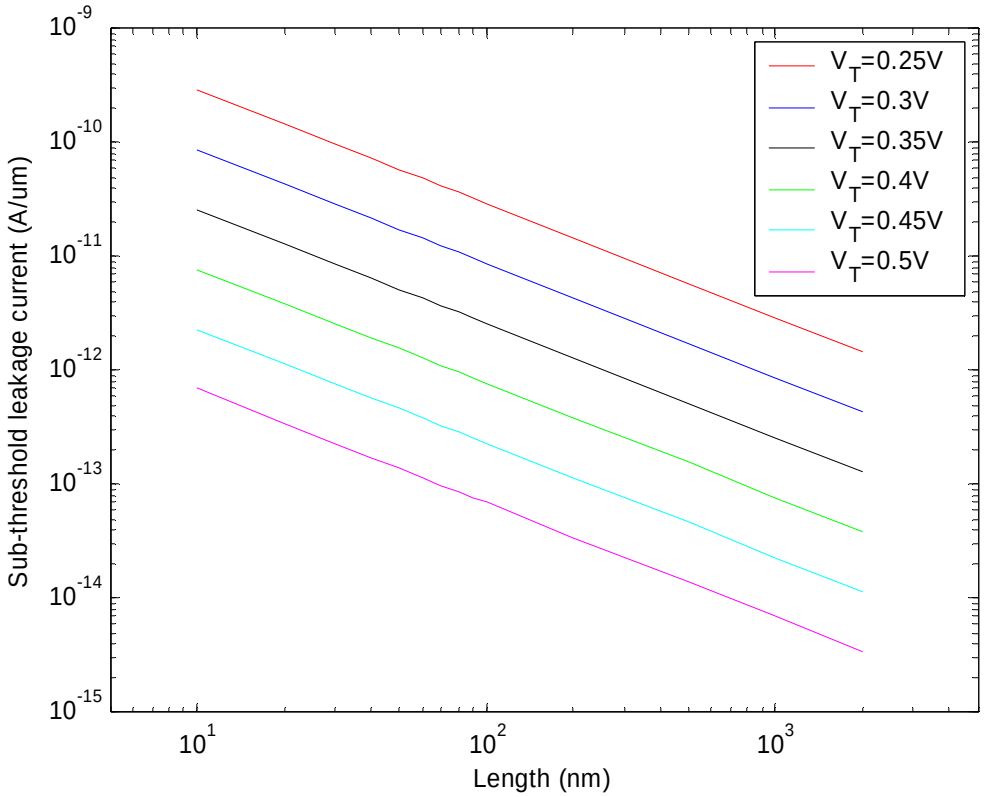


Figure 2.5 The sub-threshold leakage current, where V_{DS}=1V, m=1.1, μ=300 $cm^2/V \cdot s$, T=300K.

From Figure 2.5, we can find the sub-threshold leakage current increases with the reduction of the channel length and threshold voltage.

As mentioned above, during the device scaling, GIDL and BTBT current can be ignored by lowering the supply voltage or adopting the FinFET device. However, DIBL and subthreshold voltage will increase dramatically. For the digital applications, GIDL and BTBT current will increase the off-state power or stand-by power with the device scaling. From the analog point of view, it will affect the sampled voltage accuracy on the

capacitor in the SC (switched-capacitor) circuit during the switch hold period. It will be revealed in the next section.

2.1.2.3 The leakage current impact on the SC circuits

SC circuits are widely used in the data acquisition system. Leakage current has a heavy effect on the resolution. In Figure 2.6 it shows half of a differential sample and hold circuit in the hold period, in which, I_1 is the gate leakage current, and the I_2 is the sum of off-state leakage current. Leakage will result in a differential voltage error V_{error} because of the differential nature of analog voltages used in analog signal processing. The total voltage error result from the leakage is found as

$$\begin{aligned} V_{error} &= \frac{I_{Lk}\Delta t}{C_s} \\ I_{LK} &= I_1 WL + I_2 W \end{aligned} \tag{2.9}$$

Where C_s is the sampling capacitor, Δt is is the hold period of the switch, I_{Lk} is the total leakage current.

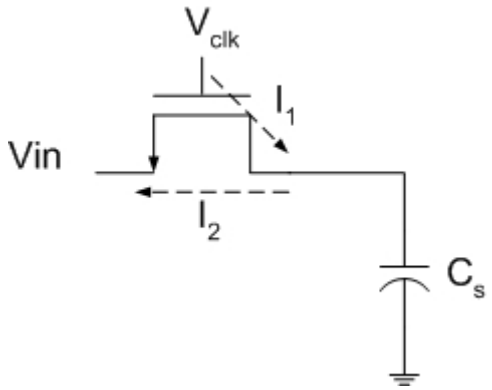


Figure 2.6 Leakage current in the sample and hold.

For data acquisition applications, C_s is set by the desired dynamic range or noise floor

$$V_{n(rms)} = \sqrt{\frac{kT}{C_s}} < \frac{V_{signal(rms)}}{2^{B+1}} \tag{2.10}$$

where $V_{n(rms)}$ is the thermal noise voltage. Low-frequency noise is ignored here for simplicity. B is the resolution bits of the application. In order to prevent leakage current degrading the resolution,

$$V_{error} < V_{n(rms)} \tag{2.11}$$

or

$$I_{Lk} < \frac{\sqrt{kTC_s}}{\Delta t} < \frac{kT2^{B+1}}{V_{signal(rms)}D} f_s \tag{2.12}$$

where D is the duty cycle, and f_s is the sampling frequency.

The maximum off-state leakage current is plotted verses sampling frequency for the desired resolution bits B equal 9 to 17 and is shown in Figure 2.7

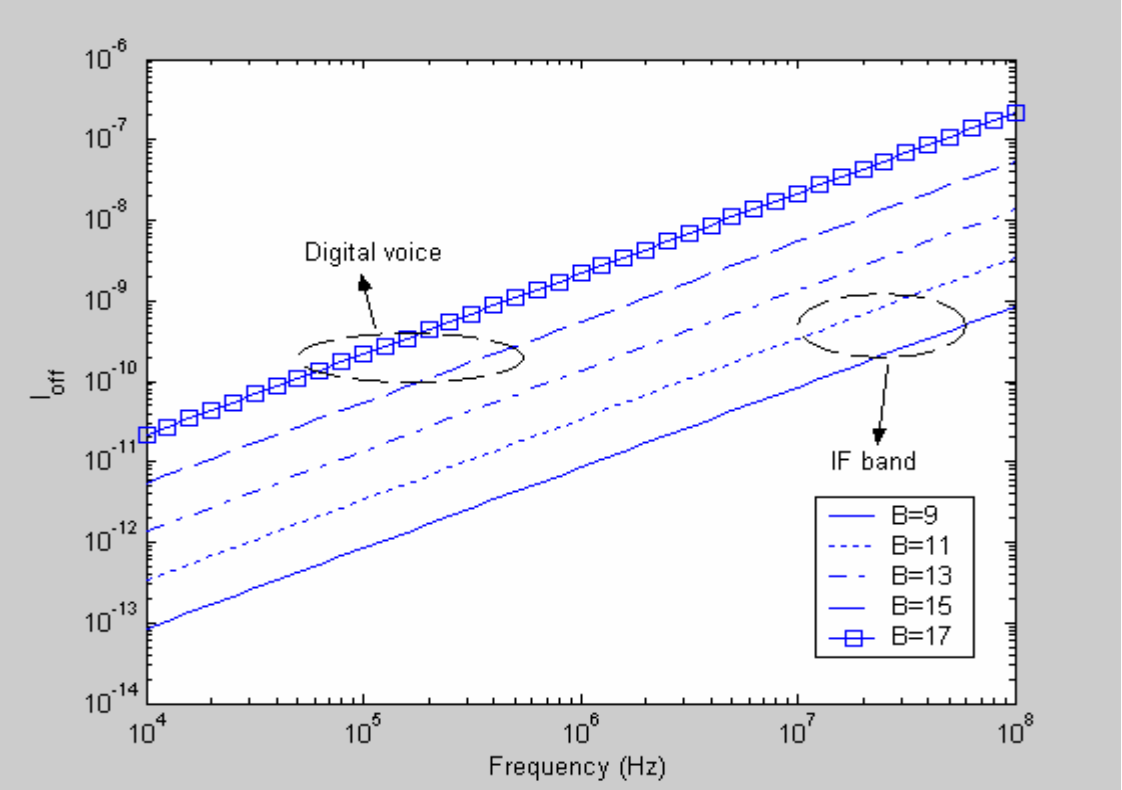


Figure 2.7 Maximum off-state leakage current versus sampling frequency, assuming $V_{signal(rms)}$=1V, sampling frequency duty =0.5, T=300K.

Here, we observe that a lower resolution requires a smaller leakage current. This seems to contradict with our normal concept. This is a direct result of greater resolutions requiring larger capacitors to maintain the dynamic range or lower noise floor.

Because the physical characteristic is very different for each process, it is very difficult to estimate the leakage current value for a specific size transistor. Here we use data of Figure 2.3 to estimate how much the gate leakage current affects the circuit accuracy. Assuming the width of the device is 0.5um and length is 0.1um, a 2nm gate oxide will result in 5nA gate leakage current. This will result in a SH error larger than 15-bit resolution allows when sampling at 1MHz or 9-11bit resolution sampling at 100MHz.

This means that the leakage current sets the lowest sampling frequency for the specific accuracy objective in the SC application. As to the low frequency applications, over-sampling has to be adopted resulting in more power consumption than they may have historically. One of those applications is the audio ADCs with the 10-12 bits resolution.

2.1.3 Noise of the MOS transistor

2.1.3.1 Noise sources of the MOS transistor

The noise sources of the MOS transistor include: 1) Channel thermal noise; 2) Gate noise; 3) 1/f noise; 4) shot noise. The noise induced by parasitic resistances is not considered here. From the frequency domain perspective, the noise sources can be classified into two types: white noise (independent of the frequency), which includes channel thermal noise and shot noise, and 'colored' noise (dependent of the frequency), which includes gate and 1/f noise. Because of the small amplitude of shot noise, it has little effect on the dynamic range, and is not considered here [30]. Gate noise results from the charge in the channel fluctuating due to the channel thermal noise and as a result gate noise is correlated to the channel thermal noise. In this section, the effect of scaling on device noise behavior only considers the channel thermal noise and 1/f noise.

2.1.3.2 Channel thermal noise (Simplified to thermal noise)

As noted above, thermal noise is white in nature with an extremely wide bandwidth. It is the dominant noise source when considering many baseband and RF applications. For long-channel devices, the noise current spectrum of the thermal noise is [31]

$$S_i = 4kT\frac{\mu_{eff}}{L^2}Q_{inv} \tag{2.13}$$

where Q_{inv} is the total inversion layer charge. In particular, (2.13) can be simplified to (2.14) based on the device region of operation, triode or saturation.

$$S_i = 4kTg_{ds}, \qquad V_{DS} = 0 \tag{2.14a}$$

$$S_i = \frac{8}{3}kTg_m, \qquad V_{DS} > (V_{gs} - V_{th}) \tag{2.14b}$$

Recently, for the deep sub-micro devices, researchers have observed that thermal noise increased [32-36] as a result of the hot carriers [36] or velocity saturation [32, 35] or both [33]. The measurement data shows that the thermal noise power increases by the factor of two for the 0.25um channel length devices [33, 34]. To maintain an equivalent SNR, gm and in turn device power must be increased by a factor of 2.

2.1.3.3 1/f noise

1/f noise dominants the low frequency bandwidth as it rolls off with the increase in frequency. In baseband application, 1/f noise can be reduced to the near thermal levels by techniques, such as chopper stabilization, and correlated double sampling [37]. However, for the RF application, it affects the phase noise of the nonlinear circuits, such as mixers and VCOs, resulting from the up-conversion to the high frequency giving rise to a $1/f^3$ sideband around the carrier frequency[38].

1/f noise origins are from the carrier number fluctuation and the mobility fluctuation [39]. With the device scaling, CMOS process are projected to transfer to the SOI processes especially in the RF application, not only as a result of the lower sub-

threshold leakage currents, but also due to the availability of the high Q on-chip inductor, high resistivity substrates and reduced substrate cross talk. The increase or decrease of 1/f noise spectral density with the device scaling is still a contradicted issue. Several groups report an increase [38, 40, 41], while the others report a decrease for the reduced the gate oxide [42]. The reasons for this may be that 1/f noise is heavily dependent on the Si-SiO2 interface quality and as a result immature processes make comparisons difficult. However, all agree on one thing — the minimum size device-to-device noise fluctuation becomes worse and worse along with scaling. This is a direct result of the smaller gate area and finite number of traps which leads to greater dispersion.

2.1.4 Supply voltage scaling

Supply voltage scales with the gate oxide scaling to prevent too large of an electrical field across the oxide. On the other hand, besides the architectural innovation, the most effective way to save the active power is to reduce the power supply voltage [19]. However, threshold voltage does not scale proportionally to the supply voltage scaling as shown in Figure 2.8 due to the leakage current consideration. The dramatic decrease of the $V_{CC} - V_T$ voltage degraded the device performance greatly.

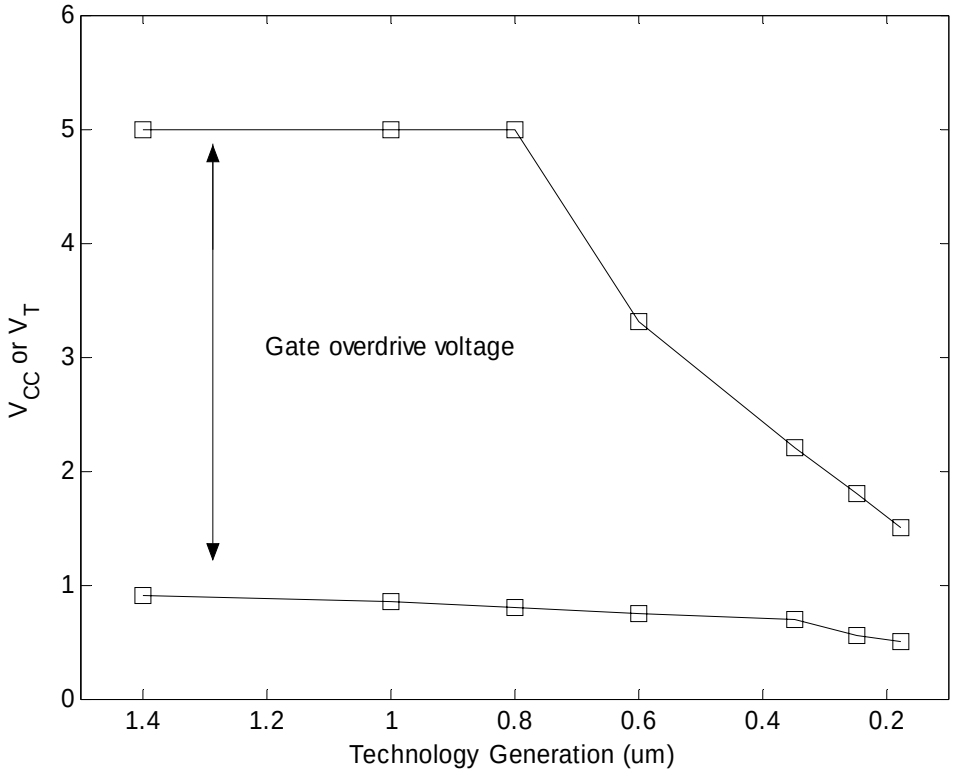


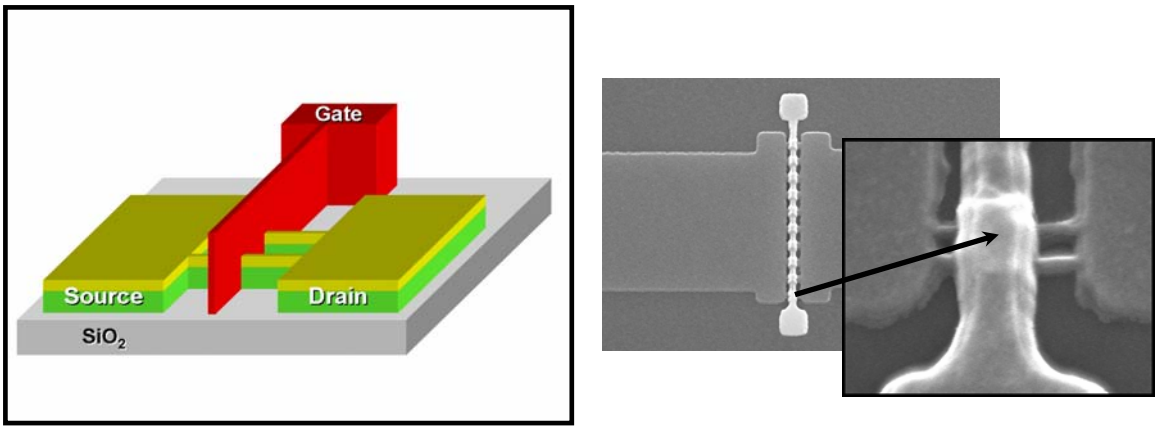
Figure 2.8 Supply voltage and threshold voltage scale with the channel length [10].

Another direct result of supply voltage scaling is the reduction of the signal swing in the analog circuit, which means more power needed to achieve the same signal-to-noise ratio at an identical bandwidth and. This will be discussed in detail in chapter 3.

2.2 Promising new device — FinFET

2.2.1 Introduction of FinFET

The structure of the FinFET is shown in Figure 2.9.



(a) (b)

Figure 2.9 The 3-D structure of the FinFET.

In Figure 2.9, the red represents the gate, blue is drain and source. The device sits on the silicon dioxide, which is white in the plot. The current is flowing parallel to the surface of the SO_2. The Figure 2.10 is the electron micrograph of the FinFET.

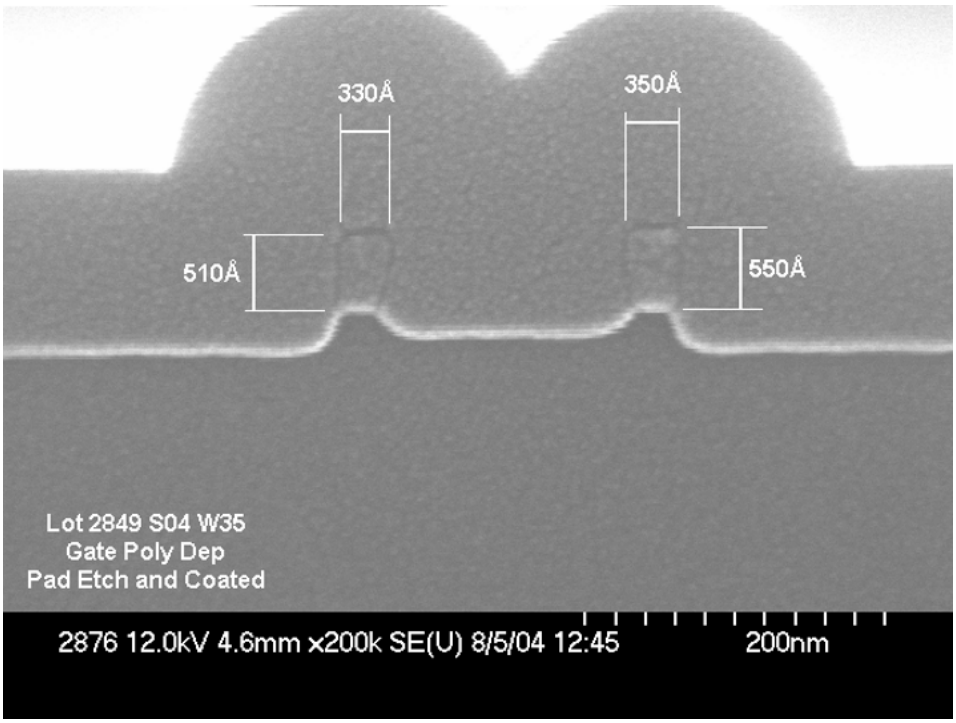


Figure 2.10 The electron micrograph of the FinFET, whose width is 50nm, length is 35nm [SPAWAR, San Diego].

The FinFET provides excellent control of the short-channel effects. Furthermore, the adoption of lightly or even un-doped silicon film achieves higher mobility [14].

2.2.2 Performances of the FinFET

2.2.2.1Suppression SCE

Because a double gate is used to control the channel and ultra-thin film used, SCE is greatly suppressed. Excellent control of the SCE of FinFET is shown in Figure 2.11, where threshold voltage rolling off with the decreases of the channel length is compared between bulk, TFSOI, and FinFET devices. This is also shown in Figure 2.12, in which the solid lines are measured ID-VG curve at low VD (VD=0.025V) and high VD (VD=0.5V) biasing, and the dashed ones are projected curves. A low subthreshold swing of 63-67mV/dec was achieved, which indicated that SCE is well controlled by the FinFET.

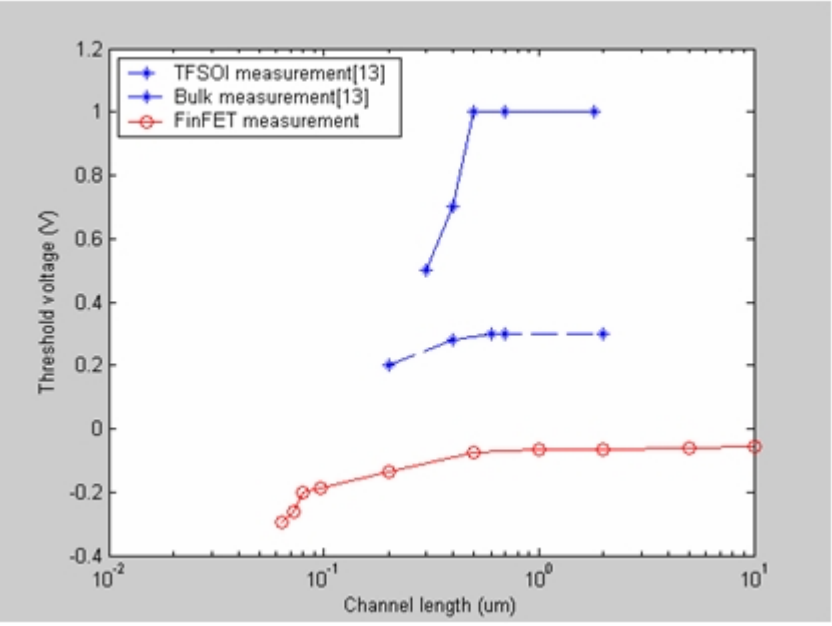


Figure 2.11 Short-channel effect in bulk, thin-film SOI (TFSOI) and FinFET n-channel devices.

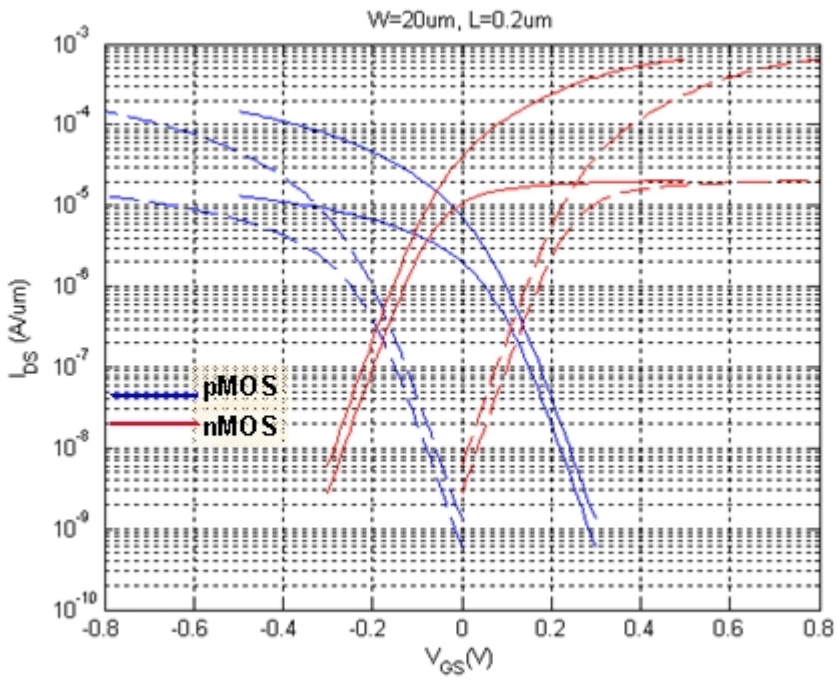


Figure 2.12 The I_{DS} vs. V_{GS} curve biasing in the low V_{DS}=25mVand high V_{DS}=0.5V. The dash lines are projected I_{DS} when using Mo doped gate to raise the threshold voltage.

2.2.2.2 Application in the subthreshold range

FinFET inherits the benefits of the devices scaling — increase of the f_T (unit current gain bandwidth). Figure 2.13 is the measured f_T of 80nm FinFET, which goes up to 20GHz at the 50mV overdrive voltage. As a rule of thumb, the operation bandwidth is 1/10 f_T, which means the device may be considered for designs in the 2-3GHz application for the transistors working in the moderate inversion. The FinFET also demonstrated less low-frequency noise in weak inversion than in the strong inversion Figure 2.14 and 2.15 [43].

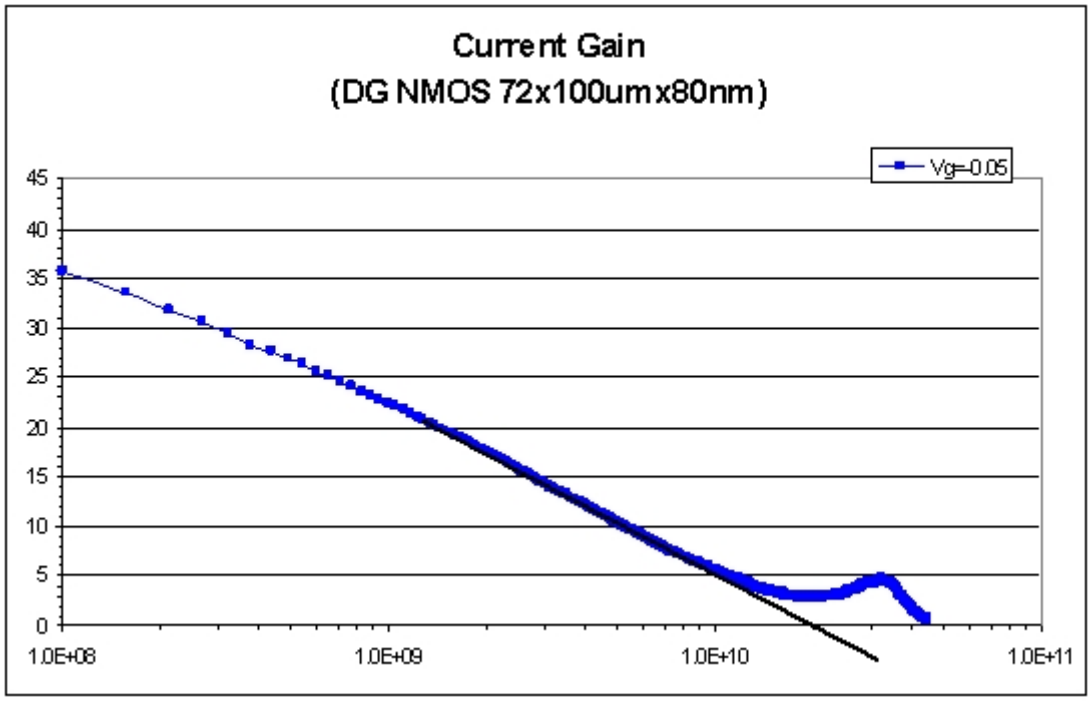


Figure 2.13 Unit current gain of the 80nm length FinFET.

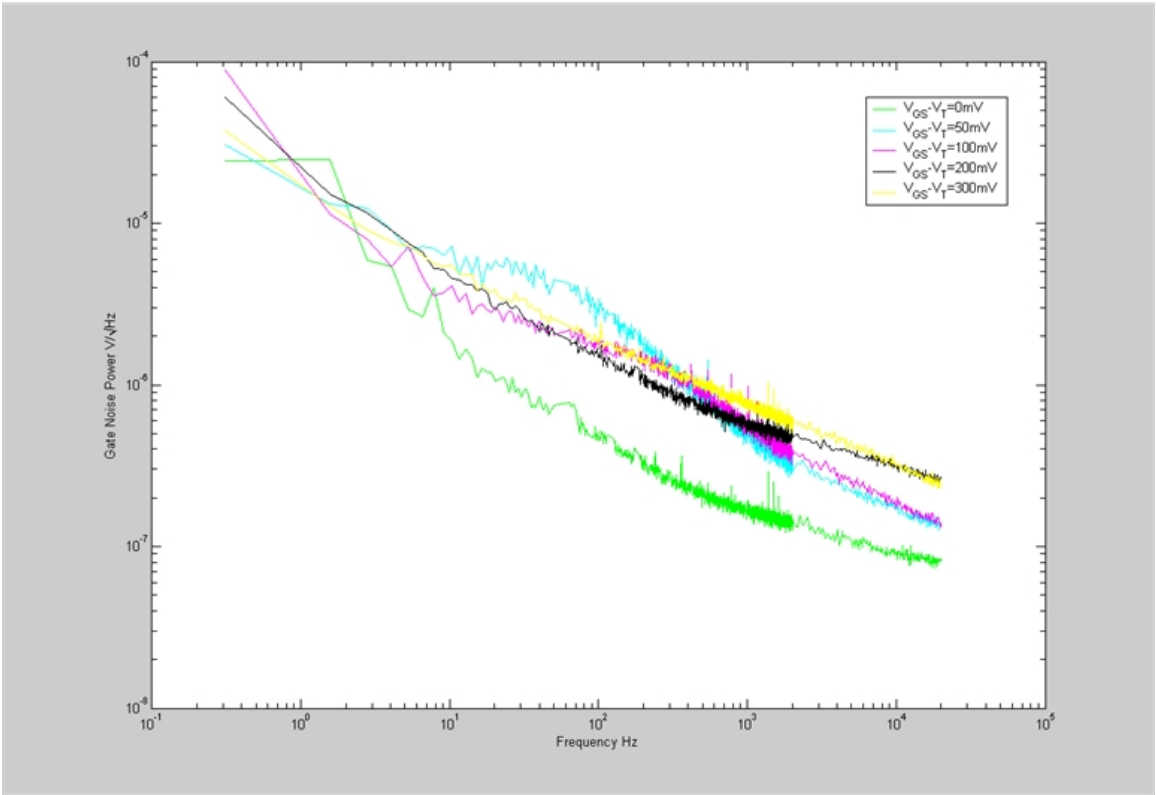


Figure 2.14 The low frequency noise of the FinFET for L=120nm, W=4.2um.

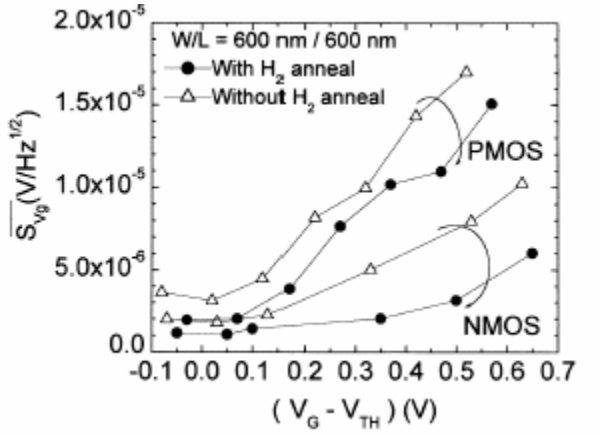


Figure 2.15 Average equivalent gate voltage noise as a function of gate voltage [43]

The FinFET demonstrates good control of the SCE result in less off-state leakage current as a result of the lessoning the TOX requirements compared to planar device of equivalent length. With the high κ material used in the gate, gate leakage problem can also be alleviated further. It also can operate in the weak-inversion range for most state-of-art applications due to its high f_T. Operation in the weak inversion reduces the overdrive voltage, which alleviate the signal swing problem. Less 1/f noise in the weak-inversion enable it good choice in the RF application. All above advantages enable FinFET an excellent candidate for the sub-100nm devices.

Chapter 3

CMOS Scaling Affects to the Analog Design

3.0 Introduction

In the previous chapter, the impact of scaling on single CMOS device performance was discussed. In this chapter the impact of CMOS scaling on analog design will be discussed in detail. Finally concept of a power settling product efficiency is introduced as criteria to choose the optimum process or topologies in design of sampled analog circuits under the circumstance of CMOS scaling.

3.1 CMOS scaling impacts to the analog design

3.1.1 Saturation voltage, V_{DS}

The voltage saturation of V_{DS} is an important parameter in analog design for it determines output stage overhead, and headroom directly affecting output swing and circuit efficiency. For a long-channel device, as the drain voltage V_{DS} increases, the drain current increases as well until channel pinch-off appears at the drain side. This V_{DS} voltage is referred to as the saturation voltage.

$$V_{DSsat} = \frac{V_{gs} - V_{th}}{m} \tag{3.1}$$

where m is the body-effect coefficient.

However, when the device is scaled down to the deep sub-micro range, carrier velocity saturation effects must be considered. For operation in the strong inversion under the constraint of velocity saturation, the saturation voltage should be modified to [1]

$$V_{DSsat} = \frac{2\Delta V / m}{1 + \sqrt{1 + 2\mu_{eff} \cdot \Delta V / (m v_{sat} L)}} \tag{3.2}$$

where v_{sat} is the saturation velocity of an electron or hole. V_{DSsat} versus the channel length with and without considering velocity saturation are plotted in Figure 3.1 to demonstrate the effect of velocity saturation on the saturation voltage.

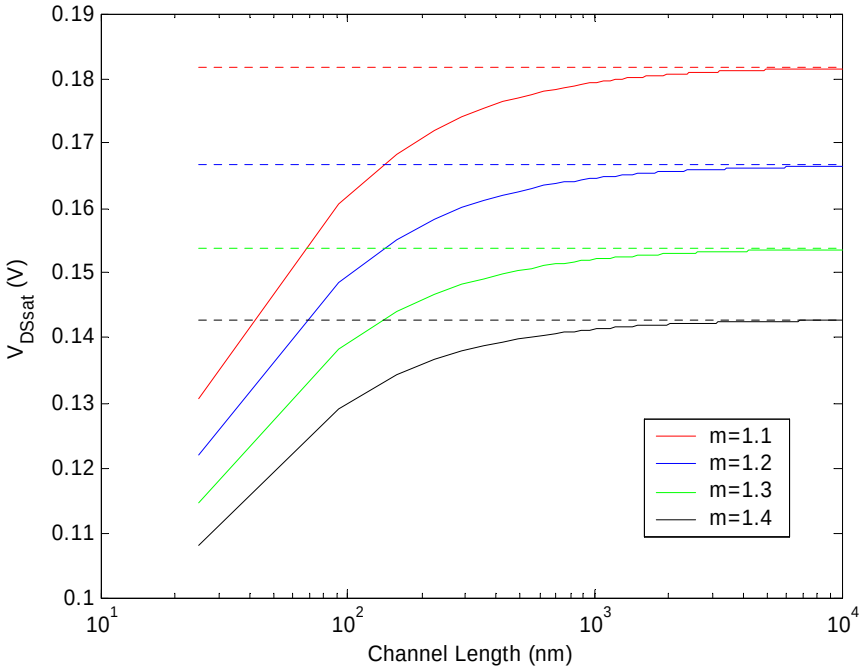


Figure 3.1 V_{DSsat} vs. the device channel length for over-drive voltage of 0.2V and μ_{eff} of 150 $cm^2/V \cdot s$. The solid line is V_{DSsat} voltage considering velocity saturation; while the dashed line is the ideal V_{DSsat} voltage for a long channel device.

In the Figure 3.1, the saturation voltage falls away the ideal saturation voltage as channel length approaches the sub-micro range. As the channel length enters the sub-100nm range, the difference in saturation voltage can be as high as 40mV.

Effect of the velocity saturation is only discussed for strong-inversion operation in [29]. Next, we will discuss the effect of the velocity saturation for the weak-inversion operation. For operation in the weak-inversion, drain current equation for long device is expressed as [29]

$$I_{DS} = \frac{\mu_{eff} W}{L} \sqrt{\frac{\varepsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right)^2 \left(\frac{n_i}{N_a}\right)^2 e^{q\psi_s / kT} \left(1 - e^{-qV_{DS}/kT}\right) \tag{3.3}$$

where ψ_s is the surface potential, Na is the acceptor impurity density, n_i is the intrinsic carrier density, and ε_{si} is the silicon permittivity.

In (3.3), kT/q is approximate to 26mV at the 27ºC. As V_{DS} is larger than 100mV, the value in the right bracelet is almost equal to 1, that is, I_{DS} is almost constant. From this point of view, saturation voltage of long channel device for operation in weak-inversion is referred to 100mV. However, when considering velocity saturation, (3.3) can be similarly as (3.1) modified to

$$I_{DS} = \frac{\mu_{eff}(W/L)}{1 + \frac{\mu_{eff} V_{DS}}{v_{sat} L}} \sqrt{\frac{\varepsilon_{si} q N_a}{2\psi_s}} \left(\frac{kT}{q}\right)^2 \left(\frac{n_i}{N_a}\right)^2 e^{q\psi_s / kT} \left(1 - e^{-qV_{DS}/kT}\right) \tag{3.4}$$

where v_{sat} is the carrier saturation velocity.

In order to determine the saturation voltage, the derivative of I_{DS} w.r.t. V_{DS} is set equal to zero and solved for V_{DS},

$$\frac{dI_{DS}}{dV_{DS}} = K' \frac{d\left(\frac{1 - e^{-qV_{DS}/kT}}{1 + \frac{\mu_{eff} V_{DS}}{v_{sat} L}} \right)}{dV_{DS}} = 0 \Rightarrow$$

$$\frac{1}{U_T} e^{-V_{DS}/U_T} \left(v_{sat} L + \mu_{eff} V_{DS} \right) - \left(1 - e^{-V_{DS}/U_T} \right) \mu_{eff} = 0 \tag{3.5}$$

where U_T is equal to kT/q.

Unfortunately, (3.5) cannot be solved symbolic. It can only be solved numerically. The simulation or numerical results are compared to the measured results for the FinFET device saturation voltages in Figure 3.2. In the Figure 3.2, the saturation voltage is found to be slightly lower than the theoretical value when compared to the long channel device.

Here we define

$$V_{DSsatQ} = V_{DSsat} + V_{sm} = \Delta V \tag{3.6}$$

where V_{sm} is the safe margin voltage. The concept of V_{DSsatQ} is shown in Figure 3.3. From above, The minimum saturation voltage of V_{DS} may or should still be 100mV or greater than $4U_T$ when considering the safety margin.

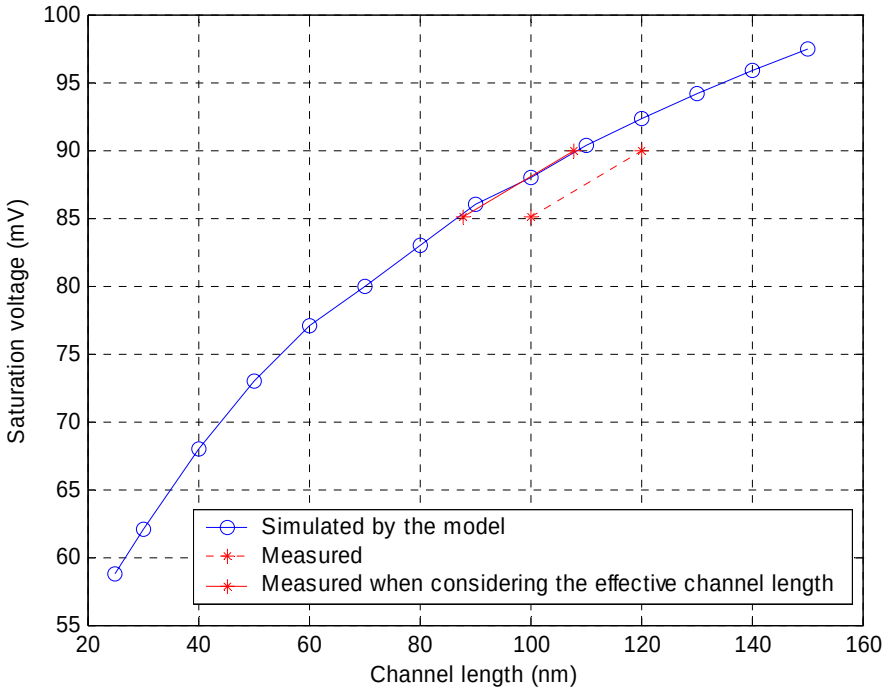


Figure 3.2 Saturation voltage for the weak inversion. μ_{eff} is assumed 300 $cm^2/V \cdot s$, which is measured from the FinFET and v_{sat} is $2*10^7$ cm/s.

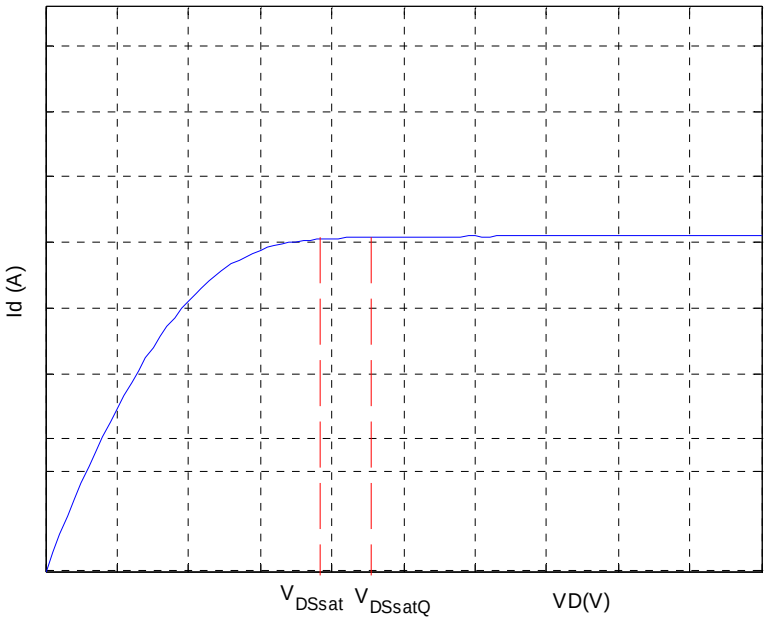


Figure 3.3 The concept of V_{DSsat} and V_{DSsatQ}.

3.1.2 Minimum threshold voltage for the analog design

Threshold voltage is another important parameter in the design of analog circuits for it directly determined the minimum supply voltage. In chapter 2, minimum threshold voltage is discussed based on digital application and from I_{ON}/I_{OFF}. Next, the minimum threshold voltage is now discussed from the analog design perspective. The current mirror circuit is a basic cell for the analog or mixed signal application. A simple current mirror circuit is shown in Figure 3.4.

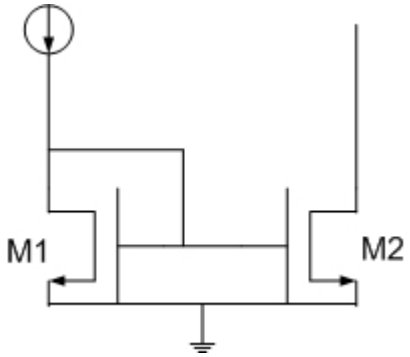


Figure 3.4 The schematic of a simple current mirror.

Because saturation voltage of V_{DS} in strong-inversion is larger than that in the weak-inversion, in order to determine the minimum threshold voltage, M1 in Figure 3.4 is assumed to operate in weak-inversion for this analysis. As a result

$$V_{gs1} - \delta V_T(T) \cdot \Delta T = V_{DSsatQ1} > 100mV \tag{3.7}$$

Where $\delta V_T(T)$ is the Temperature co-efficiency, which is 0.8-1.2mV/ºC; ΔT is the temperature fluctuation, and V_{sm} is the safe margin voltage required to ensure V_{DSsat} and in the presents of effective threshold variability. For $V_{gs1}=V_T$-50mV (operation in the weak-inversion), $\delta V_T(T)$=1 mV/ºC, ΔT=150ºC, (3.7) simplifies to

$$V_T > 300mV \tag{3.8}$$

For the wide swing current mirror of Figure 3.5, the analysis results in (3.9).

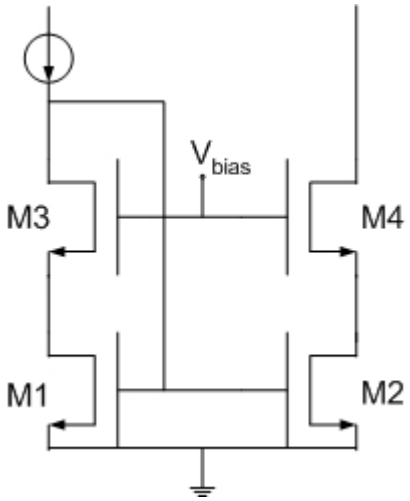


Figure 3.5 The schematic of a wide swing current mirror.

$$\begin{aligned} &V_{gs1} = V_T - 50 - \delta V_T(T)\cdot\Delta T > V_{DSsatQ\,1} + V_{DSsatQ\,3} = 2\cdot 100 \\ &\Rightarrow V_T > 250 + \delta V_T(T)\cdot\Delta T = 400mV \end{aligned} \tag{3.9}$$

From (3.8) and (3.9), the minimum threshold voltage required by analog applications is larger than 300mV.

3.1.3 Analog circuit operation space

From the discussion in Chapter 2, we know that as the CMOS is scaled, 1) the thermal noise floor is increasing; 2) the increasing leakage current requires the sampling frequency f_s to be increased; 3) f_T, current unit-gain bandwidth, is increasing with shrinking channel length; 4) Signal swing reduces as the supply voltage reduces. All these effects can be graphically represented as Figure 3.6. The sampling frequency must increase faster than f_T as a result of the exponential nature of gate leakage with oxide scaling. The author believes that analog supply voltage scaling will likely stall at 1V.

As shown in Figure 3.6, the shaded area is the analog circuit operation space. With the device scaling, the shaded area representing the analog design space becomes smaller and is pushed to the higher frequencies. In the following section, it will be shown that to achieve the same dynamic range and bandwidth, more power is needed as a result of scaling.

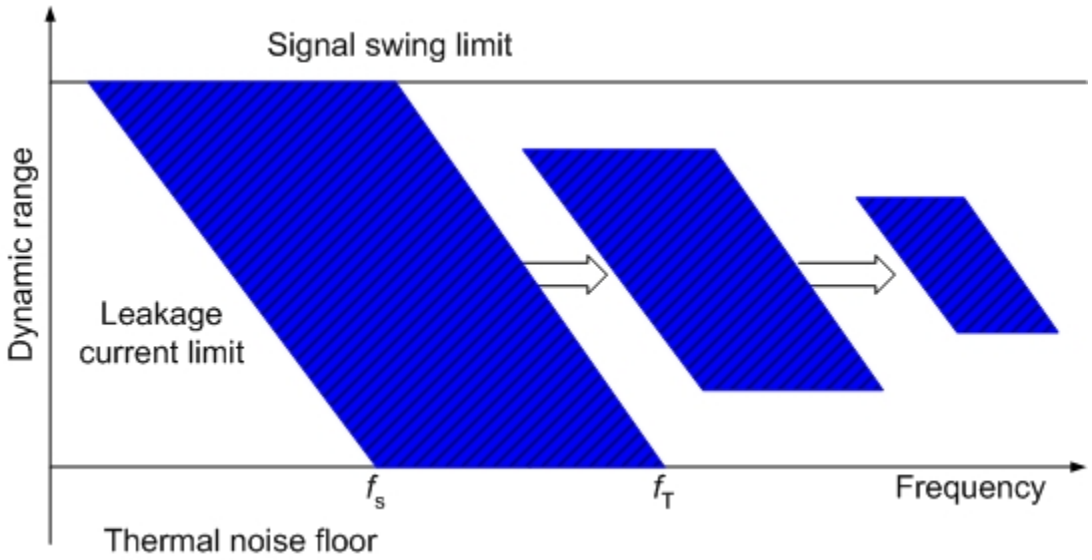


Figure 3.6 Analog circuit operation space.

3.2 Power settling product efficiency

As shown above, with CMOS scaling, the analog designer will face a more and more restrictive design space. Substantial amount of power is required to maintain the same dynamic range and bandwidth. In the case of the battery technology, further scaling provides no great improvement. Power is expensive in this portable era. Power settling product efficiency is proposed as a means to give the analog designer guidance as to which OTA structure or parameter is the better choice for the deep sub-micro device applications.

3.2.1 Introduction of the power settling product efficiency

For the analog designer, scaling of the supply voltage means the output swing decreases for the same over-drive voltage shown in Figure 3.7.

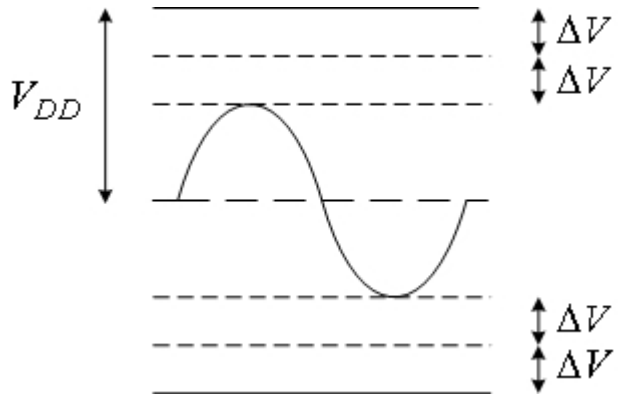


Figure 3.7 The output swing. $V_{out\text{-}swing}=2*V_{DD} - 4\cdot\Delta V$.

The output swing is

$$v_{out-swing} = 2V_{DD} - N \cdot \Delta V \tag{3.10}$$

where the N is the circuit topology factor, by example N=2, 4 and 5 represent a common source, cascode stage and telescopic stage respectively. Where a symmetrical power supply equal to $\pm V_{DD}$ volts has been assumed.

From (3.10), one direct result of voltage scaling is the potential reduction of the Signal-to-Noise Ratio (SNR) for the same noise floor. In order to evaluate the performance limitations due to the device scaling, the power efficiency K is defined in [44]

$$K = \frac{kT \cdot \Delta f \cdot SNR}{P} \tag{3.11}$$

where k is the Boltzmann's constant, T is the temperature in Kelvin degrees, Δf is the signal bandwidth, and P is the consumption power.

However, in the SC circuit or discrete time application, settling time is more important than the bandwidth for wider bandwidth may not mean less settling time. Similar as power efficiency K, power settling product efficiency is defined as

$$\Pi = \frac{kT \cdot SNR}{P \cdot t_{settle}} \tag{3.12}$$

This is used to estimate how much SNR at the unit power settling time product.

Most analog circuits operated in the Class A mode whose power consumption power can be expressed as

$$P = mI_{bias}(2V_{DD}) \tag{3.13}$$

where m is a dimensionless factor, determined by the circuit topology; I_{bias} is the input transistor biasing current.

Function (3.13) can be rewritten as

$$P = m \cdot g_m \cdot \Delta V \cdot V_{DD} \tag{3.14}$$

$$P = 2mg_m \cdot n \cdot U_T V_{DD} \tag{3.15}$$

operation in the strong-inversion or weak-inversion, where $(nU_T)^{-1}$ is the subthreshold slope.

The main noise sources are 1/*f* noise and broadband thermal noise. 1/*f* noise can be reduced to near negligible levels by circuit techniques such as chopper stabilization [45]. Only thermal noise considered here. SNR is defined

$$SNR = \frac{V_{out-swing}^2}{V_{n(rms)}^2(f) \cdot f_n} \tag{3.16}$$

where f_n is the noise bandwidth, $V^2_{n(rms)}(f)$ is the noise spectral density.

For the baseband applications, most circuits can be readily represented as a single or dominant-pole system. As to the multi-poles system, it will be discussed in detail in the Chapter 4. For a single pole system, the settling time is proportional to the bandwidth, which is

$$t_{settle} = \frac{k}{\omega_{3dB}} = \frac{k \cdot C_L}{A \cdot g_m} \tag{3.17}$$

where C_L is the effective load capacitor, g_m is the system transconductance, A is the system gain and k is constant which is related to the settling error.

In order to give a clear and analysis expression of the power settling product efficiency, two examples are given in the later.

3.2.2 Power settling product efficiency of the OTA

As shown in Figure 3.8, the noise of the OTA is determined by the load capacitor. The output noise voltage is

$$V_{n(rms)}^2 = \gamma kT / C_L \tag{3.18}$$

where γ is a constant determined by the circuit process and topology.

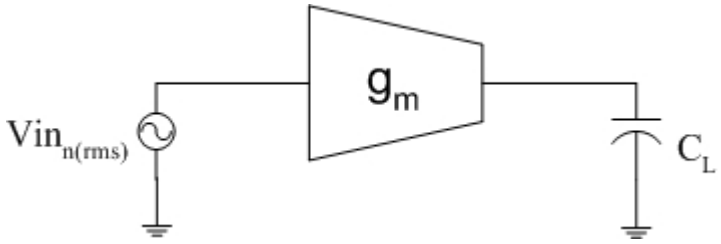


Figure 3.8 OTA block diagram with input referred noise

Combining (3.14), (3.16), (3.17), the power efficiency factor for the square law operation is,

$$\Pi = \frac{A(2V_{DD} - N \cdot \Delta V)^2}{km\gamma \cdot \Delta V \cdot V_{DD}} \tag{3.19}$$

and for the weak-inversion operation is

$$\Pi = \frac{A(2V_{DD} - N \cdot \Delta V)^2}{km\gamma \cdot nU_{T_t} \cdot V_{DD}} \tag{3.20}$$

The impact of the scaling of the supply voltage to the normalized power settling product efficiency versus the topology at the constant overdrive voltage is revealed in Figure 3.9. Power settling product efficiency of N=2 (i.e. Common Source structure) topology is affected least in the three topologies. In Figure 3.9, the overdrive voltage was set to 0.2V to demonstrate that the larger overdrive voltage will cause the power consumption to dramatically increase at the low supply voltages. For the plots even it is idealistictly assumed that Vdssat is equal ΔV to make the point. N equal 4 and 5 type structures can be represented by folded and telescopic cascode topologies. Figure 3.10

predicts that future analog circuit supply voltages based on DMS scaled devices will typically be greater than 1 V and be restricted to N type 2 to 4 devices.

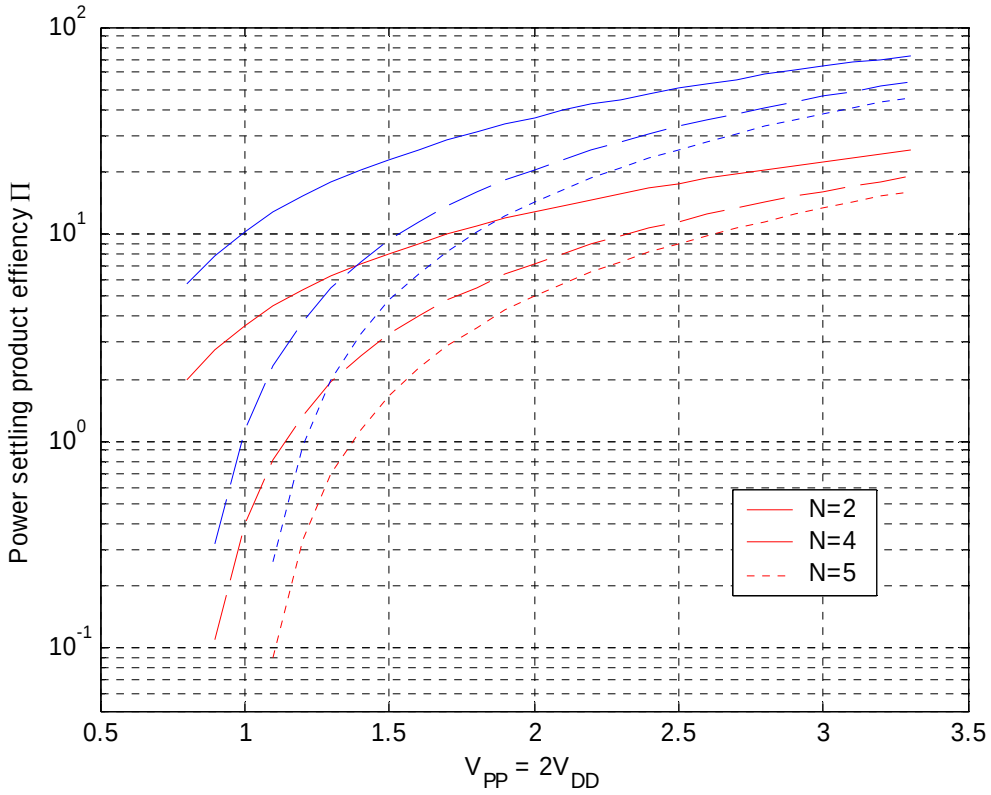


Figure 3.9 Normalized the efficiency factor versus the power supply, ΔV=0.2V. Red lines are for the square law range and blue lines for the weak-inversion range.

Figure 3.10 is the plot of (3.20) with the normalized power settling product efficiency versus the power supply voltage in the cascode topology, resulting in efficiency plots independent to the process. From Figure 3.10, we observe that when V_{PP} less than 1.3V, the power increases dramatically for 0.3V overdrive voltage when N=4, which means that high overdrive voltages are not suitable for use in the cascode topology when considering low supply voltage application. It is apparent from Sections 2.2 that future DSM devices will operate in velocity saturation or subthreshold to achieve both low V_{dssat}, 75 to 150mV, and low 1/f noise. As a result the relevant plots are for ΔV are in

the 0.1 to 0.2V range. From Figures 3.9 and 3.10 one concludes that power efficient analog circuits (N=4) must operate with 1.0 to 1.2V supplies for 100 mV and 200mV V_{dssats} respectively. However, the price of 1.5V supplies is potential increase in DIBL and the return of *GIDL. Given assuming a 1.1 Volt supply and* V_{dssat}*, equals 100mV, N = 2 through 4 circuits will be feasible, while if only a 200mV* V_{dssat} *can be achieved the bulk of analog circuits will be N = 2 for non-weak inversion operation.*

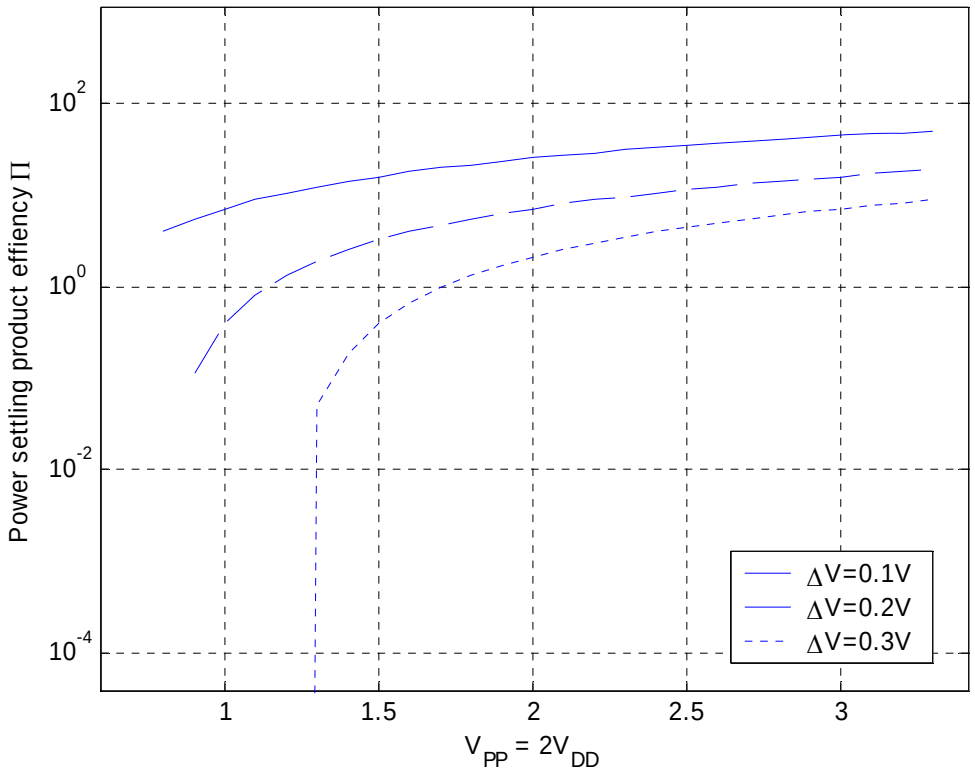


Figure 3.10 Normalized the efficiency versus the power supply, N=4.

3.2.3 Power efficiency factor of the switched-capacitor gain stage

A representative switched-capacitor gain stage is shown as Figure 3.11. Its noise power is composed of two parts: thermal noise induced by the switches and OTA itself. The thermal noise induced by the switches is

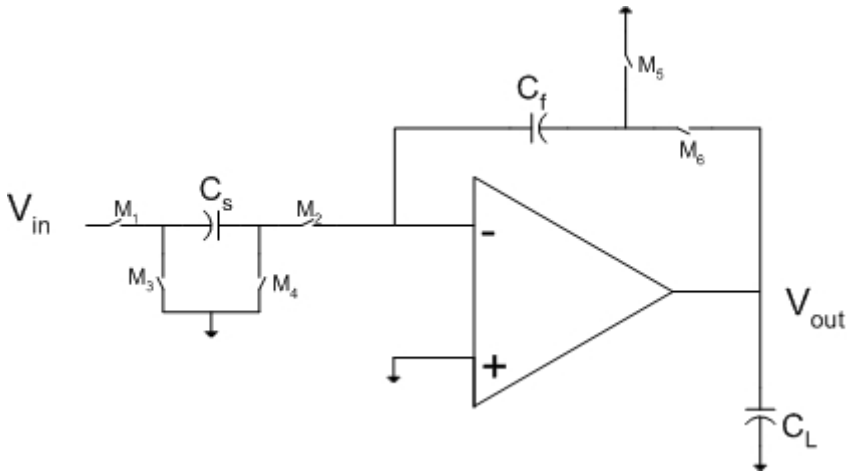


Figure 3.11 Switched-capacitor gain stage schematic.

$$V_{n1}^2 = \frac{2kT}{C_s} \tag{3.21}$$

$$V_{n2}^2 = \frac{kT}{C_f} \tag{3.22}$$

$$V_{n3}^2 = \frac{kT}{C_L} \tag{3.23}$$

where V^2_{n1} is the switches M_1-M_4, generated the noise; V^2_{n2} is the top switch M_5 generated noise, V^2_{n3} is M_6 generated noise. Ignoring 1/f noise contributions, the total noise at the output of the gain stage is

$$V_{n\_switch(rms)}^2 = \frac{2kT}{C_s} \cdot \left(\frac{C_s}{C_f}\right)^2 + \frac{kT}{C_f} + \frac{kT}{C_L} = \frac{(1-f)kT}{fC_s}\left(\frac{2-3f}{f}\right) + \frac{kT}{C_L} \tag{3.24}$$

where $f = \dfrac{C_f}{C_f + C_s}$.

If assuming $C_s = \alpha C_L$, (3.24) can be simplified to

$$V_{n\_switch(rms)}^2 = \frac{kT}{C_L}\left(\frac{(1-f)(2-3f)}{f^2\alpha} + 1\right) \tag{3.25}$$

The total noise is

$$V_{n(rms)}^2 = \frac{kT}{C_L}\left(\frac{(1-f)(2-3f)}{f^2\alpha} + 1 + \gamma\right) \tag{3.26}$$

The settling time of the gain stage is

$$t_{settle} = \frac{k \cdot C_L}{A \cdot g_m} = \frac{f \cdot k \cdot C_L}{(1-f) \cdot g_m} \tag{3.27}$$

Combining (3.14), (3.16),(3.26), and (3.27), the power settling product efficiency for the saturation range is

$$\begin{aligned} \Pi &= \frac{(1-f)(2V_{DD} - N \cdot \Delta V)^2}{mk \cdot \Delta V \cdot V_{DD} \cdot f\left(\frac{(1-f)(2-3f)}{f^2\alpha} + 1 + \gamma\right)} \\ &= \frac{(2V_{DD} - N \cdot \Delta V)^2}{mk\gamma' \cdot \Delta V \cdot V_{DD}} \end{aligned} \tag{3.28}$$

For the same feedback factor, (3.28) is almost the same as the (3.19). γ' is the same as the γ representing the configuration factor independent of process. From (3.28), similar conclusions will be derived as (3.19). Furthermore, for the same α, β, γ', different stage gains or feedback factor vs. the normalized Π is shown in Figure 3.12.

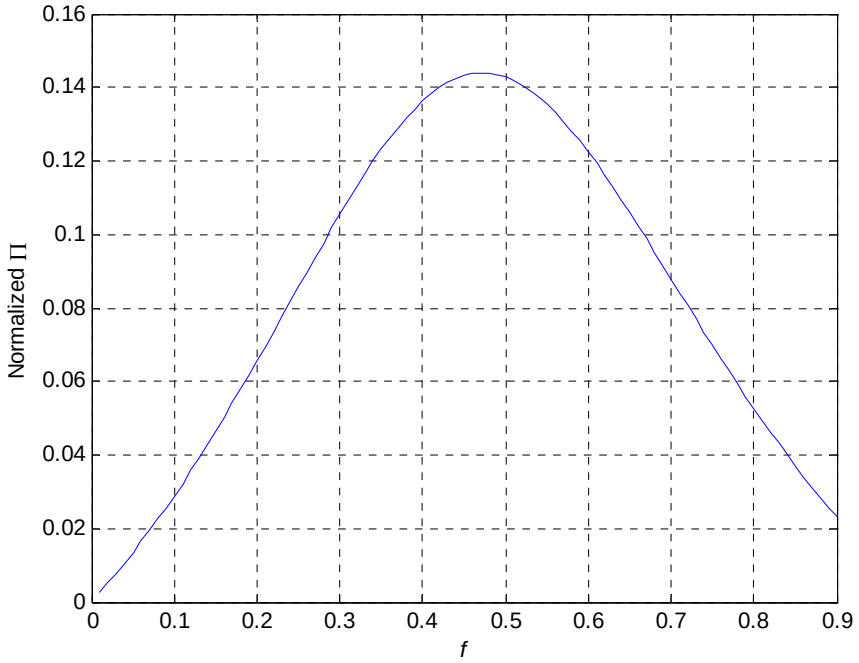


Figure 3.12 Normalized Π versus *f* for α=1/2, and γ=4.

From the Figure 3.12, we can get one important conclusion that, there is maximum power settling product around feedback factor around 1/2. This point is very important in the later pipelined ADC design.

3.3 Analog scaling rule

From the Figure 3.10, we observe that the power settling product efficiency reduces with the scaling of the supply voltage for the same overdrive voltage. We also find that the reduction of the overdrive voltage can improve the power settling product efficiency with the reduction of the supply voltage. *Here we defined the analog scaling rule as follows; the overdrive voltage should scale at the same the rate or faster than the*

supply voltage to keep the power efficiency constant or improving. This is readily observed from the (3.19) and (3.20).

If one uses a 300mV overdrive voltage and assumes a 3.3V technology as a reference, the overdrive voltage will scale down to 100mV for a 1.1V supply voltage to keep the power efficiency constant. From the above analysis and measured data, this is achievable. When the supply voltage scales below 1.1V, the transistors may have to operate in the weak-inversion range to maintain power efficiency but at a reduced bandwidth. Scaling overdrive voltage and normalized power settling product efficiency vs. supply voltage is shown in Figure 3.13. This implies a future mixed signal supply corner will be 1.1V+/-100mV.

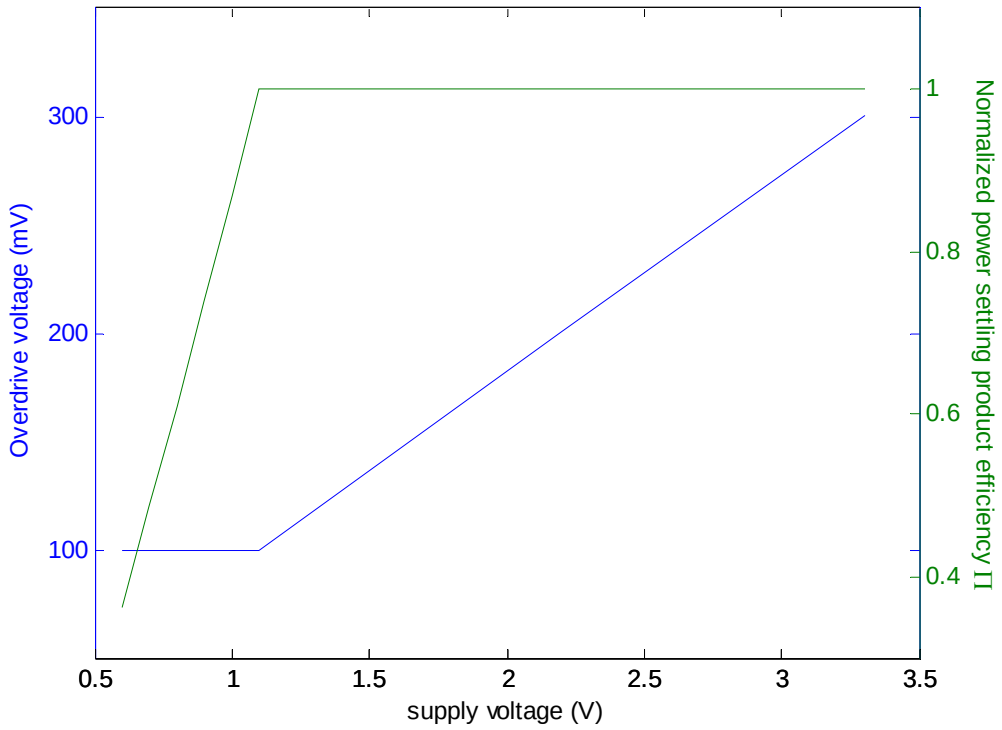


Figure 3.13 Scaling overdrive voltage and normalized power efficiency factor vs. supply voltage.

3.4 Conclusion

CMOS scaling down to sub-100nm device not only affects single device performance, but also impacts on the topology choice of analog design. In this chapter, we determined that the final scaled value of the supply voltage when considering saturation voltage, threshold voltage, subthreshold, velocity saturation and know process technologies for analog design is 1.1V, and that threshold voltages must be greater than 300mV. In order to give guidance to the topology and parameters choice in the analog design, power settling product efficiency concept is proposed as criteria and analog scaling rule is proposed. In the context of the following conclusions the "low" supply voltage is assumed to be less than 1.1V. Supply voltage and overdrive voltage scaling effects on an analog circuits' performance were presented with following conclusions. 1) With the supply voltage scaling, lower overdrive voltage topologies will have greater power settling product efficiency. 2) Cascode topology may not be suitable for the low supply voltage design as a result of the poor power settling product efficiency. 3) N=5 topology is unsuitable for DSM scaling. 4) Overdrive voltage should always scale at the same or greater rate than the supply voltage to keep the same or better power settling product efficiency. 5) Supply voltages less than 1.1V supply voltage are not suitable for the portable analog applications. 6) Analog circuits are quite feasible supply levels less than 1.1V, however a dramatic power consumption penalty must be paid.

Chapter 4

Power Efficiency Analysis of OTA Architecture

4.0 Introduction

Bandwidth (or settling time), accuracy (offset, noise, and gain), and power consumption are three of the most important properties of the analog design. However, they cannot be satisfied simultaneously. In order to achieve higher bandwidth (or less settling time), accuracy is often sacrificed and more power consumed. It is the analog designer's task to make the trade offs between bandwidth and accuracy at minimum power consumption. Less power is especially important in this "portable" era.

As mentioned in the Chapter 3, CMOS device scaling has continued to pursue higher bandwidth while at the same time not sacrifice transistor performance. One direct result is that supply voltage scales from 3V to sub-1V [12] for the stability considerations due to the thin gate oxide even for the long channel devices. More power is needed to obtain the same dynamic range. The other result of the scaling is the reduction of the transistor output impedance which results in less gain ($g_m \cdot r_o$ or $2V_A/\Delta V = \mu$)[46], where μ is the transistor intrinsic gain. The cascode or miller compensated two-stage topologies,

typically used in the high gain applications with long channel devices, are no longer suitable in many applications especially with the sub-100nm transistors. On the other hand, end users typically demand higher and higher quality electronics products, with lower distortion for music, motion pictures, and cell phone voice etc. These applications typically demand greater than 16-bit resolution ADCs for the music compression and decompression and/or 10-12 bits 20-40Msps ADC for the cell phone IF conversion [47]. To achieve 12-bit resolution, an OTA with 75-80dB of DC gain is required for the pipelined ADC topology. This requires μ^4 or μ^5 to achieve the desired gain from a device with an intrinsic gain around 20dB. Adequate gain can only be realized by multi-stage topologies.

Multi-stage topologies are realized by "vertical" boosting via boosted OTAs [48] or nested boosted OTAs [49]. On the other hand, multi-stage in the "horizontal" direction includes miller compensated OTAs [50], or NGCC (Nested Gm-C Compensation) OTA [51, 52], in which each stage can be a cascode or a simple common-source stage.

As mentioned at the beginning of this chapter, power consumption must be considered when selecting the OTA topology. However, in people's analysis of OTAs, most considerations are limited to bandwidth, or settling time and stability. The worse is that lots of works [48, 50-52] have only analyzed the open loop transfer function of the above topologies. Although several[53, 54] have analyzed the close loop, the effects of zeros are ignored due to the feed-forward signal through the bilateral feedback path.

In this chapter, as far as the author knows, it is the first time that each of the OTAs topologies contained here have been analyzed and compared in the real close loop form (not via the feedback factor) in order to determined the scaling effect on OTAs performance. The power settling product efficiency concept proposed in Chapter 2 is used as a criteria to find an optimumal topology. For convenient, the power settling power efficiency expression for operation in strong-inversion is rewrite as (4.1), in which N and m are determined by the topology. *In this chapter, the aim of the analysis is to find the m — the total g_m normalized to the specific g_{m1}, which drives the g_{m1} / C_s to 0.1 ω_T (the unit current gain)*. This assumption is based on following reasons. Due to the internal parasitic capacitances, power consumption is not linear proportional to the bandwidth. Too large a g_{m1}/C_s means a dramatic increase in power consumption and the comparison basis with other topology is lost; too small g_{m1}/C_s will not be used in real applications. This chapter is organized as follows: First, the folded-cascode OTA topology is presented with the necessary analysis in the close loop form, and then compared with a two-stage miller compensated OTA. Secondly, high order OTA topologies will be analyzed to determine their power settling product efficiencies. Finally, the optimum OTAs' topology choice is given in the context of a specific application. The conclusions drawn will be used in a later design application to validate of the approach.

$$\Pi = \frac{A(2V_{DD} - N \cdot \Delta V)^2}{km\gamma \cdot \Delta V \cdot V_{DD}} \tag{4.1}$$

Throughout this chapter, the OTAs are assumed to operate in a switched capacitor (SC) environment. However, the approach is universally applicable to all OTA and

amplifier topologies and circuit applications. What should be noted is that only a linear settling analysis is performed here. Slew rate effects are neglected in this study.

4.1 Introduction of the SC circuit

Due to the fact that it is difficult to obtain accurate high valued resistors in CMOS integrated circuit processes, SC techniques emerged to replace the resistor. As shown in the Figure 4.1 (a), ϕ_1 and ϕ_2 are non-overlapping clocks, and V_1 and V_2 are voltage sources. From the charges transfer point of the view, 4.1 (a) is the same as 4.1 (b) if

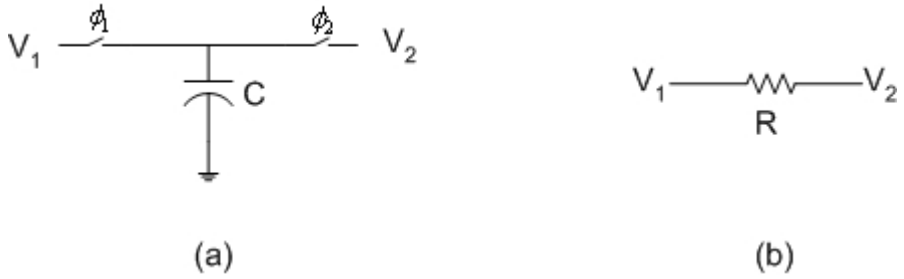


Figure 4.1 Switched capacitor equivalence of a resistor. (a) Switched-capacitor circuit; (b) Resistor equivalent.

$$R = \frac{1}{Cf_s} \tag{4.2}$$

f_s, the sampling frequency is sufficiently faster than the application.

SC techniques can be applied to many applications, including: integrators, sample and holds, and gain stages in signal condition which includes filters, DACs, ADCs etc. Figure 4.2 is a fully-differential SC gain stage. The results will be used repeatedly in later analysis for it is the basic cell configuration used in the SC integrators, MDACs, gain

stages, etc. The resulting analysis method provides bandwidth, gain, and settling response for a general category of SC and capacitor loaded OTA circuits.

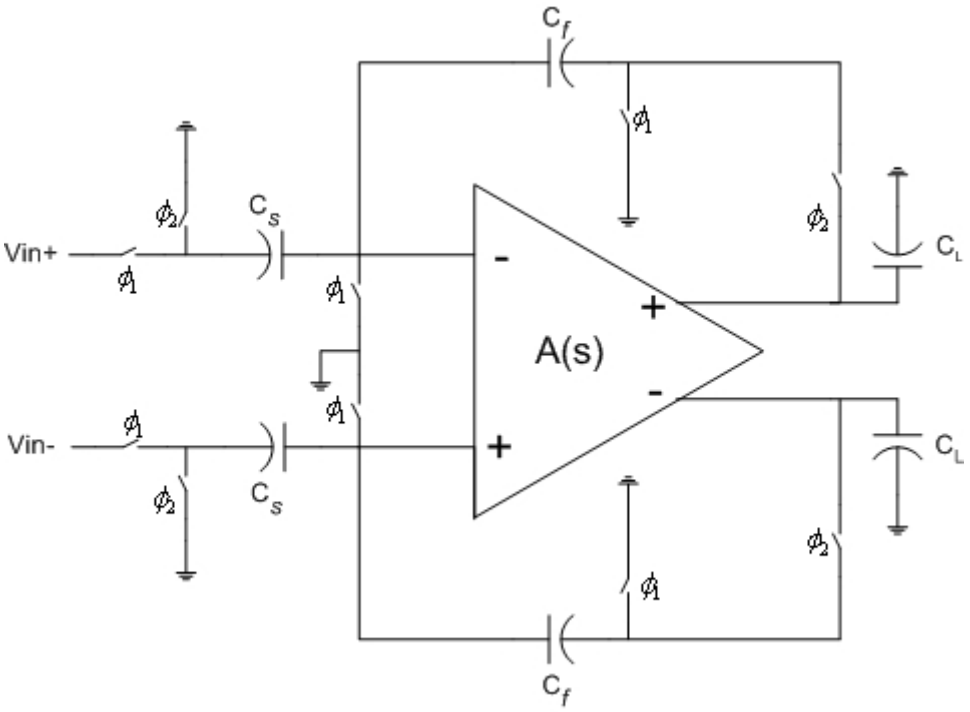


Figure 4.2 Fully-differential SC gain stage.

In Figure 4.2, A(s) represents the transfer function of an OTA with capacitor feedback that can be implemented with different OTA topologies Where C_s is the sampling capacitor, C_f is the feedback capacitor, and C_L is the load capacitor. During ϕ_1, the input voltage is sampled on C_s, and circuit settles during ϕ_2. The equivalent circuit during the settling period is shown in Figure 4.3 where only halved circuit is shown for simplicity.

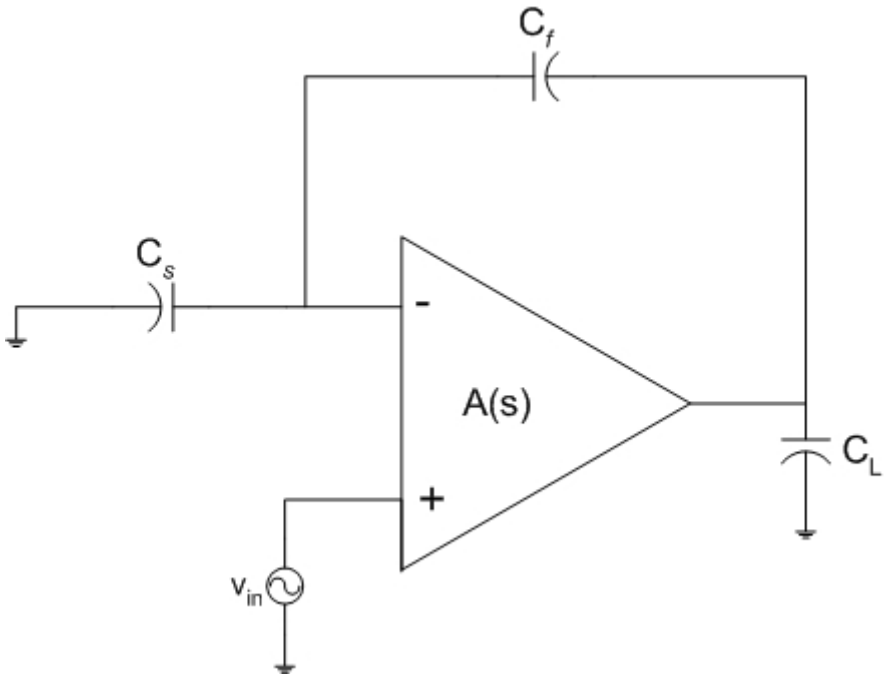


Figure 4.3 The equivalent schematic of an SC gain stage.

In order to compare the power settling product efficiency, it is assumed that the transistors are biased at the same gate overdrive voltage, which means ω_T is fixed. For simplicity, the bias generator's effect on power efficiency is ignored for each realization.

4.2 Folded-cascode OTA

The folded-cascode OTA is one topology to realize the open loop gain μ^2. In order to demonstrate the necessity of using close loop form circuit to analyze the capacitor loaded OTAs, three kinds of analysis — open loop, close loop using feedback factor, and close loop, are analyzed and compared. First, the schematic of the open loop cascode OTA is shown in Figure 4.4.

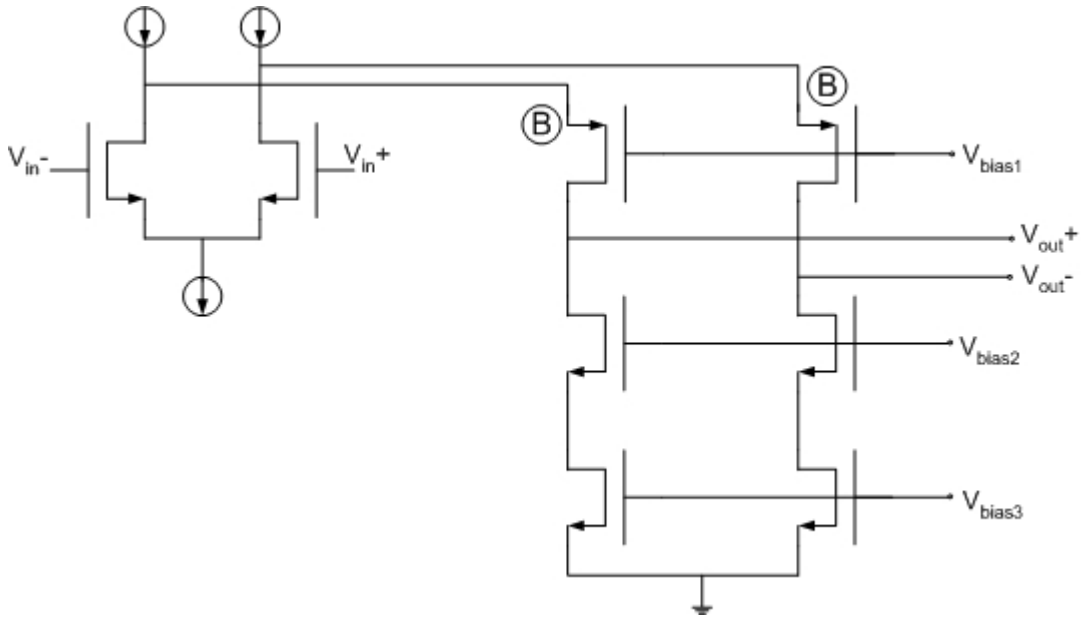


Figure 4.4 The schematic of the classical folded-cascode OTA.

The small signal model of the Figure 4.4 is shown in Figure 4.5.

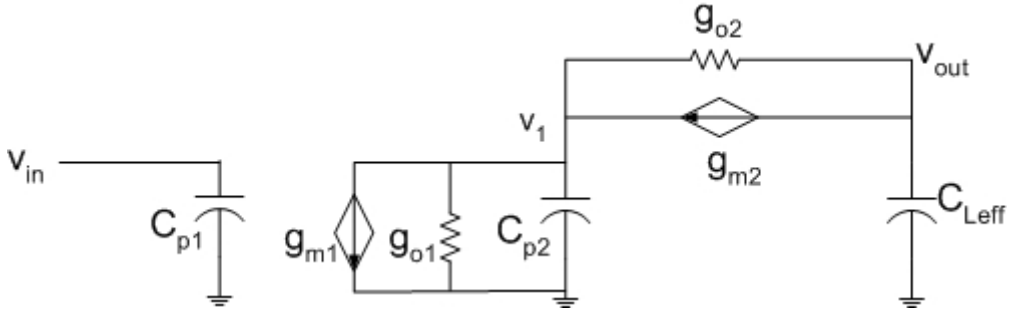


Figure 4.5 The small signal model of the folded-cascode OTA.

Where C_{p1} is the total parasitic capacitances at the input node, the C_{p2} is the total parasitic capacitances at the B node, and the C_{Leff} is the total effective capacitance load at the output node. The open loop transfer function is

$$\frac{v_{out}}{v_{in}} = -\frac{g_{m1}g_{m2}}{C_{Leff}C_{p2}s^2 + C_{Leff}g_{m2}s + g_{o1}g_{o2}} \tag{4.3}$$

Apparently, there are two poles in the (4.3), in which the dominant pole is

$$p_{dom} = \frac{g_{o1}g_{o2}}{C_{Leff}g_{m2}} = \frac{g_{o1}}{C_{Leff}A_2} \tag{4.4}$$

where the A_2 is the cascode transistor self-gain, and the non-dominant pole is

$$p_{non} = \frac{g_{m2}}{C_{p2}} \tag{4.5}$$

The open loop GBP is g_{m1}/C_{Leff}.

In order to get minimum settling time and maintain the close loop stability, the phase margin should be approximately 70º-75º for settling errors from 1% to 0.001% [55]. For the folded-cascode, the phase margin is determined by the relative location of non-dominant pole and GBP.

$$PM = 90° - \tan^{-1}\left(\frac{\omega}{p_{non1}}\right)\Big|_{\omega=GBP} \tag{4.6}$$

From (4.6), non-dominant pole should be 2.75-3.7 times of GBP in order to achieve 70º-75º phase margin. C_{p2} includes the C_{gs} of the cascode transistor and C_{gd} of the current source transistor and input transistor. Assuming $g_{m2} = k \cdot g_{m1}$, and $C_{gd} = 0.25C_{gs}$, (4.5) can be simplified to

$$\frac{g_{m2}}{C_{p2}} = \frac{k \cdot g_{m1}}{kC_{gs1} + (2+k)C_{gd1}} = \frac{k \cdot g_{m1}}{kC_{gs1} + \frac{(2+k)}{4}C_{gs1}} = \frac{4k}{5k+2}\omega_T \tag{4.7}$$

Which is k=0.21 for 70º phase margin and 0.35 for 75º.

Secondly, the small signal model of the closed loop gain stage of the folded-cascode OTA is shown in Figure 4.6

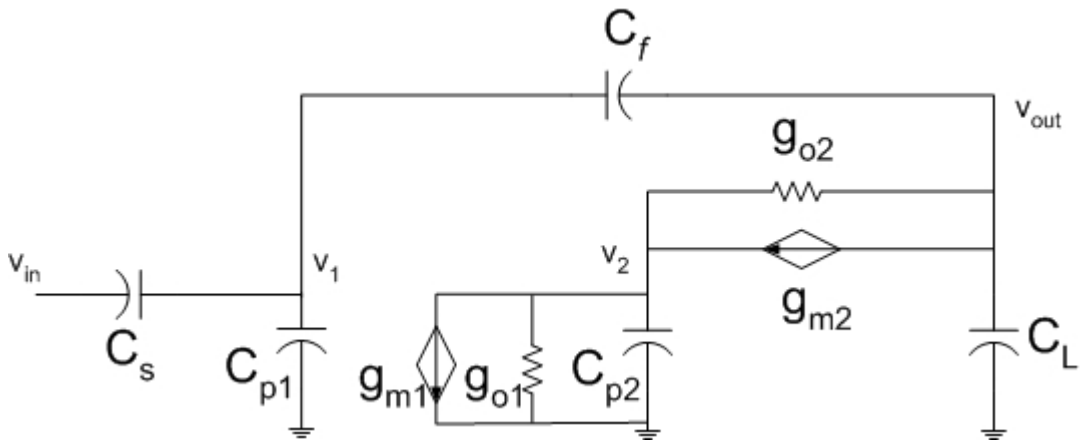


Figure 4.6 The small signal model of the close loop folded-cascade gain stage.

The nodal analysis yields the following equations.

$$(v_{in} - v_1) \cdot sC_s = v_1 \cdot sC_{p1} + (v_1 - v_{out}) \cdot sC_f \tag{4.8}$$

$$g_{m1}v_s + v_2(g_{o1} + sC_{p2}) = (v_{out} - v_2) \cdot g_{o2} - g_{m2}v_2 \tag{4.9}$$

$$(v_1 - v_{out}) \cdot sC_f = (v_{out} - v_2) \cdot g_{o2} - g_{m2}v_2 + v_{out} \cdot sC_L \tag{4.10}$$

Assuming C_{p1}, g_{o1}, and g_{o2} are small enough to be ignored and $C_f/(C_f + C_s) = f$, (4.8) – (4.10) can be solved to yield the close loop transfer function as (4.11).

$$\frac{v_{out}}{v_{in}} = \frac{C_{p2}C_s f s^2 + C_s f g_{m2} s - (1-f) g_{m1} g_{m2}}{C_{p2}(C_L + fC_s)s^2 + g_{m2}(C_L + fC_s)s + f g_{m1} g_{m2}} \tag{4.11}$$

The dominant pole of the (3.21) is

$$\omega_{3dB\_CL} = \frac{g_{m1} \cdot f}{C_L + fC_s} = \frac{g_{m1} \cdot f}{C_{Leff}} \tag{4.12}$$

where $C_{Leff} = C_L + \frac{C_f C_s}{C_T}$.

and the non-dominant pole is

$$p_{non} \approx -\frac{g_{m2}}{C_{p2}} \tag{4.13}$$

In this case, The GBP is

$$GBP = \frac{g_{m1}}{C_L + C_f + \dfrac{C_L C_f}{C_s}} = \frac{g_{m1}}{C_{Leff}}(1-f) \tag{4.14}$$

in which close loop GBP is not g_{m1}/C_{Leff} any more, unless f is sufficiently small enough. The zeros are

$$z_1 \approx -\frac{g_{m2}}{C_{p2}} \tag{4.15}$$

$$z_2 \approx \frac{g_{m1}}{C_f} \tag{4.16}$$

From (4.13) and (4.15), there is a single pole-zero doublet in the transfer function. The doublet is around ω_T when C_{p2} is almost equal to C_{gs2}. The doublet moves towards the unit gain bandwidth as the cascode transistor size is reduced as a result of the other additional parasitic capacitors becomes significant in addition to C_{gs2}. The cascode transistor cannot be allowed to become too small as this result in the doublet coming too close to the ω_{3dB} and as a result harming the settling time[48]. If slew rate is not considered, the doublet can be closer to the GBP, which means g_{m2} can be as small as 0.02 g_{m1}, when the other parasitic capacitance of the C_{p2} besides the C_{gs2} is $0.25C_{gs1}$. In practice, g_{m2} is usually no less than one fourth of the g_{m1}[56] for slew rate consideration. In order to calculate the close loop bandwidth, we assume C_L is equal to ½ of the C_s result in $C_{Leff} = C_s$ for simplicity. In the typical circuits, C_L is in the range of ¼ ~ ½ of the C_s. This assumption will not affect later conclusions. The topology factor, m, is $(g_{m1}+g_{m2}) / g_{m1} = 1.25$. Normalizing ω_T to 1, the settling time for 0.1% settling error is 146. To place

this in perspective gm1/C_s was conservatively selected to 0.1 ω_T to accommodate bulk processes, a 0.1% settling 6.9 time constants so the baseline is essentially 69 with the choice of circuit topology taking and additional factor of 2 in bandwidth and with consideration for the effect of the zero. Taking a less conservative approach e.g. gm1/C_s selected to 0.2 ω_T the resulting figure of merit would be 73.

When the feedback factor is utilized to do the close loop analysis, the small signal model is as shown in Figure 4.7.

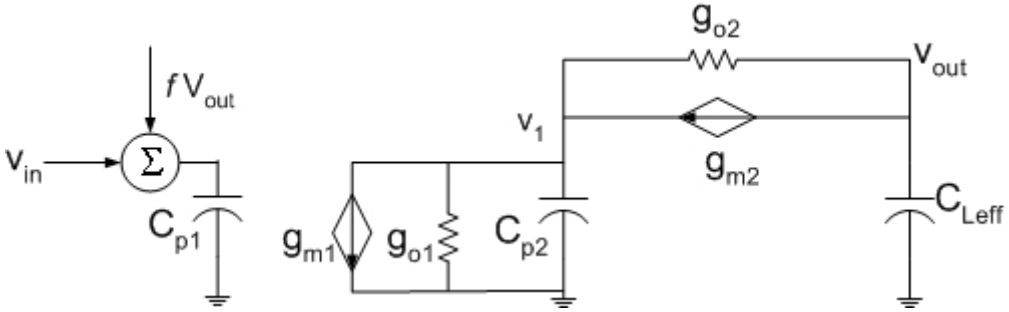


Figure 4.7 The small signal model of close loop with feedback factor analysis

The transfer function is

$$\frac{v_{out}}{v_{in}} = \frac{-g_{m1}g_{m2}}{C_{Leff}C_{p2}s^2 + C_{Leff}g_{m2}s + fg_{m1}g_{m2}} \tag{4.17}$$

Compare (4.11) and (4.17), feedback factor method only gives the poles of the close loop. It does not give zeros information. Furthermore, from (4.15), the smaller the second stage, the closer the two poles resulting in the damping of the system, which is undesired in fast settling applications. In addition, we can not determine the correct settling time. This type of analysis is only valid for small f or large close loop gain conditions. For small close loop gain, the resulting optimized configuration resulting from the feedback factor analysis will waste power by pushing the non-dominant pole unnecessary beyond GBP.

When compared above three methods to calculate the settling time, the close loop form gives the most accurate result. From this point of view, from now on, close loop form is used to analyze the power settling product efficiency.

4.3 Two stage miller compensation OTA

The schematic of the two stage miller compensation OTA, in which each stage is a common source configuration achieving an open loop gain of μ^2, is shown in Figure 4.8. The small signal model of Figure 4.8 is shown in Figure 4.9.

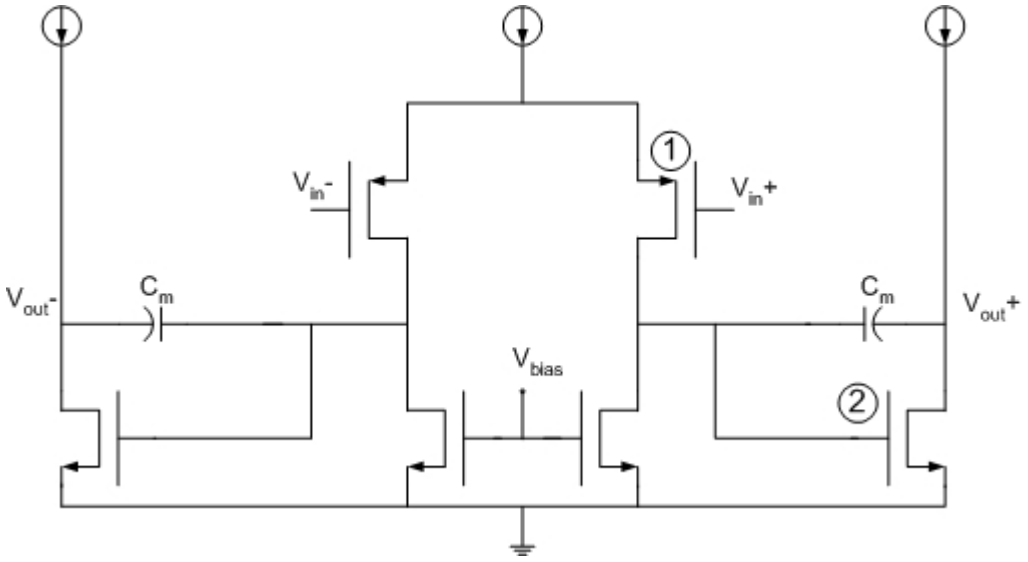


Figure 4.8 Schematic of the two stage miller compensated OTA.

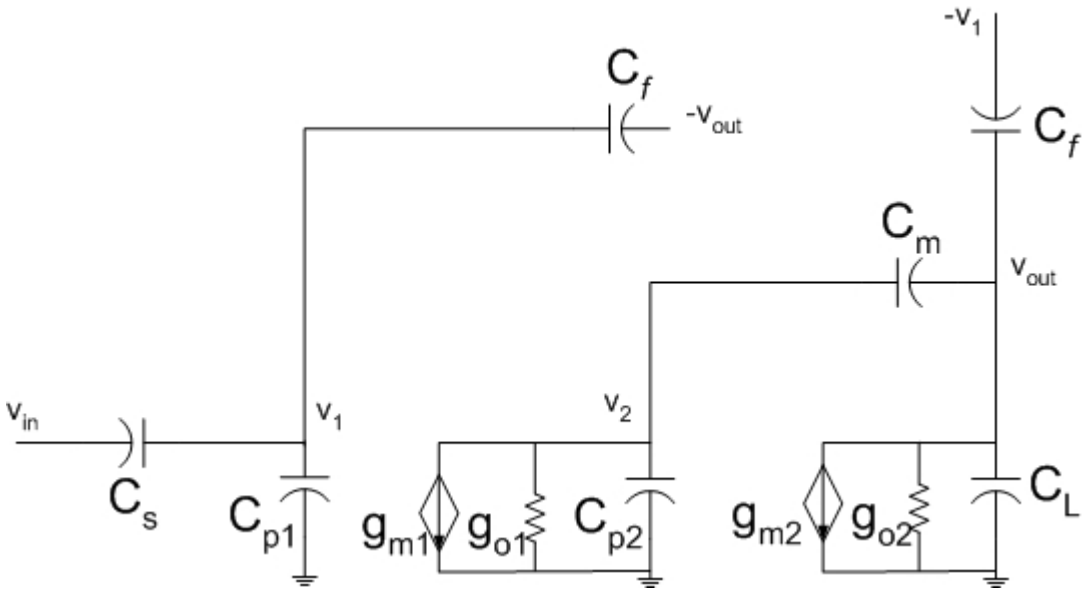


Figure 4.9 The small signal model of the two stage miller compensated OTA.

Nodal analysis yields following equations.

$$v_{in} \cdot sC_s = v_1 \cdot s(C_s + C_{p1}) + (v_1 + v_{out}) \cdot sC_f \tag{4.18}$$

$$g_{m1}v_1 + v_2(g_{o1} + sC_{p2}) = \frac{v_{out} - v_2}{1/(sC_m)} \tag{4.19}$$

$$-(v_1 + v_{out}) \cdot sC_f = \frac{v_{out} - v_2}{1/(sC_m)} + g_{m2}v_2 + v_{out}(sC_L + g_{o2}) \tag{4.20}$$

where C_{p1}, C_{p2} are the total parasitic capacitors at the nodes 1 and 2 respectively; C_m is the miller capacitor; g_{o1} and g_{o2} are the output conductance of each stage respectively. (4.18) – (4.20) can be solved to yield the close loop transfer function as (4.21), where C_{p1}, g_{o1} and g_{o2} are small enough to be ignored when compared to the g_{m1} and g_{m2}.

$$\frac{v_{out}}{v_{in}} = \frac{-C_s(C_f(C_m + C_{p2})s^2 + C_m g_{m1}s - g_{m1}g_{m2})}{(C_f + C_s) \cdot (as^2 + bs + c)} \tag{4.21}$$

$$a = C_{Leff}(C_m + C_{p2}) + C_m C_{p2}$$

$$b = C_m(g_{m2} - fg_{m1})$$

$$c = fg_{m1}g_{m2}$$

In order to make (4.21) stable, b should be large than 0, which means

$$g_{m2} > \frac{C_f g_{m1}}{C_f + C_s} = fg_{m1} \tag{4.22}$$

There are two zeros in the (4.21), which are

$$z_1 = \frac{g_{m2}}{Cm} \tag{4.23}$$

and

$$z_2 = -\frac{C_m g_{m1}}{C_f(C_m + C_{p2})} \tag{4.24}$$

The two zeros, especially the RHP (right half plane) zero are not far away the GBP. Their effects to the settling time cannot be ignored. The poles of the (4.21) can be two real poles or two complex poles, depending on f and g_{m2} sizing related to g_{m1}. It can be also shown that there is not direct solution of g_{m2} scaling to the g_{m1} in order to achieve optimum power settling product efficiency. One must consider C_m, C_L and the feedback factor as well. In order to compare with cascode topology, the following assumptions are made: $C_f = C_s$, $C_L = 1/2C_s$, $g_{m1}/C_s = 0.1\omega_T$. Matlab was used to do numerical simulation and with the results are shown in Figure 4.10, in which $g_{m2}=3.25g_{m1}$, and $C_m=1.22C_s$ to achieve minimum settling time for 0.1% settling error with 70.

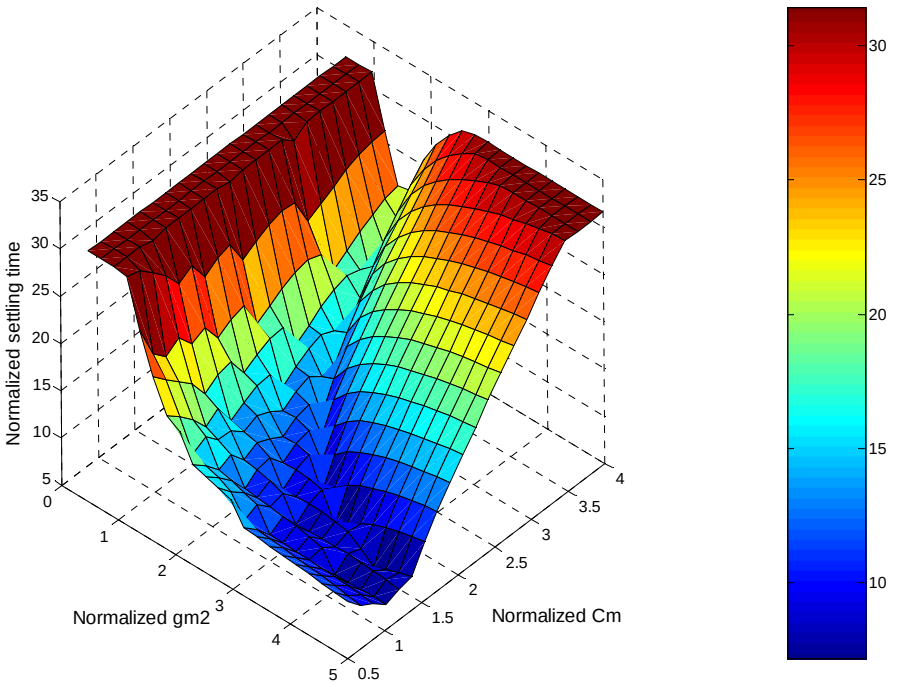


Figure 4.10 Normalized settling time versus normalized the g_{m2} and C_m, when settling error is 0.1%.

In order to vividly compared the power efficiency of the two-stage miller compensated OTA and cascode OTA, we define the normalized Π' the ratio of power settling product efficiency of the miller compensated OTA and cascode OTA.

$$\Pi' = \frac{\Pi_{cascode}}{\Pi_{miller}} \tag{4.25}$$

Π' versus the supply voltage and overdrive voltage is shown in Figure 4.11. A Π' of less than 1 means that the cascode OTA has less power settling product efficiency.

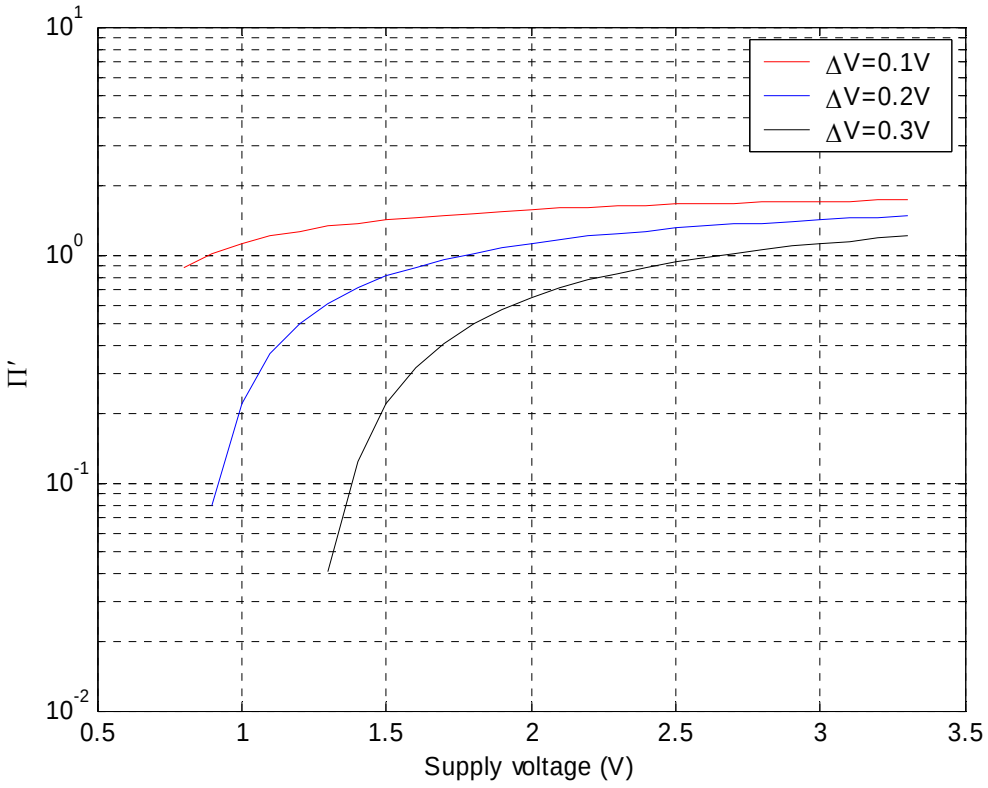


Figure 4.11 Comparison of the normalized power settling product efficiency of the miller compensated two stage and folded-cascode topologies.

From the Figure 4.11 we can find that when the overdrive voltage is 0.3V, the two-stage OTA is less power efficiency until 2.7V. When the overdrive voltage is reduced to 0.2V, the folded-cascode OTA achieves better power efficiency until 1.7V. When operating in weak inversion range or velocity saturation, such that V_{DS} is 0.1V, the folded-cascode OTA always has better power efficiency until 0.8V.

4.4 Nested gain-boosting cascode OTA

Nested gain-boosting OTA as shown in Figure 4.12 was proposed by [11], to achieve high gain. The structure can be expanded to n levels to achieve gain as high as $(g_m \cdot r_o)^{2n}$.

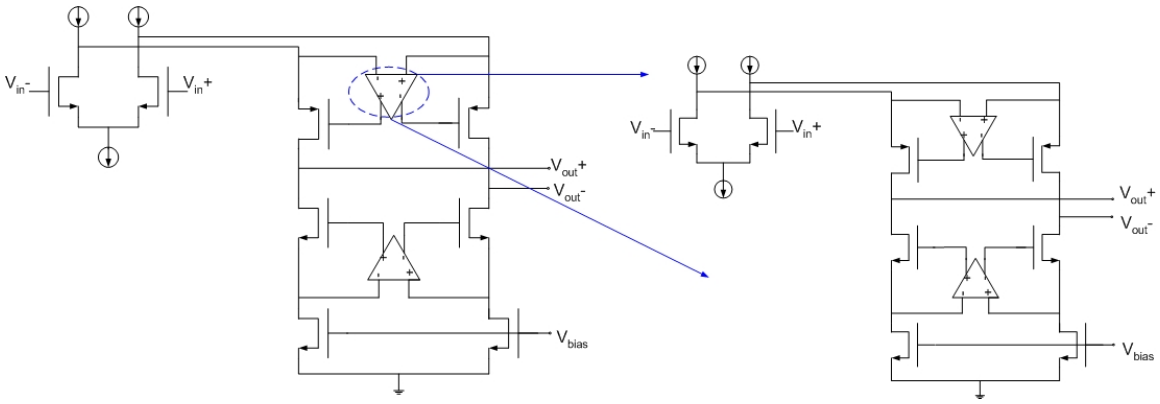


Figure 4.12 Concept of nested gain-boosting cascode OTA

As an example, 2-level nested gain-boosting OTA was used to investigate the power efficiency factor. The small signal model of the close loop of 2-level nested gain-boosting OTA is shown in Figure 4.13.

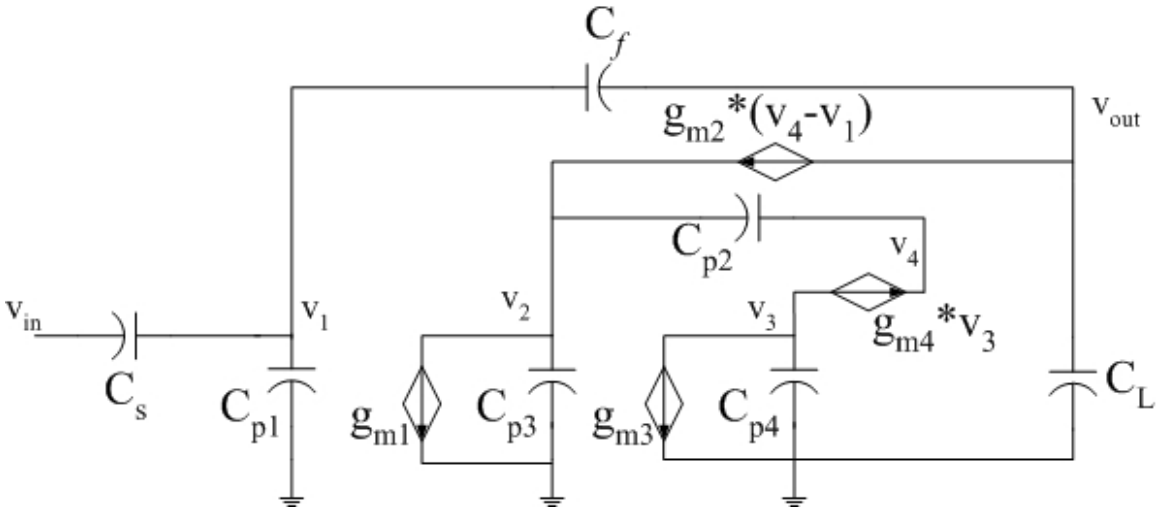


Figure 4.13 The small signal model of closed loop 2-level nested gain-boosting OTA

In Figure 4.13, all the transistor intrinsic conductances are ignored. $C_{p1} - C_{p4}$ are the total parasitic capacitors at the transistor input nodes. The nodal analysis yields following equations.

$$(v_{in} - v_1) \cdot sC_s = v_1 \cdot sC_{p1} + (v_1 - v_{out}) \cdot sC_f \tag{4.26}$$

$$g_{m1}v_1 + v_2 \cdot sC_{p3} = (v_4 - v_2) \cdot sC_{p2} + g_{m2} \cdot (v_4 - v_2) \tag{4.27}$$

$$g_{m3}v_2 + v_3 \cdot sC_{p4} + g_{m4}v_3 = 0 \tag{4.28}$$

$$(v_2 - v_4) \cdot sC_{p2} + g_{m4}v_3 = 0 \tag{4.29}$$

$$(v_1 - v_{out}) \cdot sC_f = g_{m2}(v_4 - v_2) + v_{out} \cdot sC_L \tag{4.30}$$

Assuming C_{p1} is small enough to be ignored and C_f / (C_f+C_s) = f, (4.26) – (4.30) can be solved to yield the close loop transfer function as (4.31).

$$\frac{v_{out}}{v_{in}} = \frac{C_s}{C_f + C_s} \frac{C_f}{C_{Leff}} \frac{(as^3 + bs^2 + cs^1 + 1)(s - \frac{g_{m1}}{C_f})}{(as^3 + bs^2 + cs^1 + 1)(s + \frac{fg_{m1}}{C_{Leff}})} \tag{4.31}$$

$$a = \frac{C_{p1}C_{p2}C_{p3}}{g_{m1}g_{m2}g_{m3}}$$

$$b = \frac{C_{p2}C_{p3}}{g_{m2}g_{m3}}$$

$$c = \frac{C_{p3}}{g_{m3}}$$

The dominant pole of the (4.31) is

$$\omega_{3dB\_CL} = \frac{g_{m1} \cdot f}{C_L + fC_s} = \frac{g_{m1} \cdot f}{C_{Leff}} \tag{4.32}$$

There is one RHP low frequency zero

$$z_1 \approx \frac{g_{m1}}{C_f} \tag{4.33}$$

From (4.31), there are three pole-zero doublets in the transfer function at or approaching ω_T. This is similar to the folded-cascode OTA. In order to move the doublets away from the unit-gain bandwidth, the boosted amplifier cannot be allowed to become too small, harming the settling time of the boosted cascode OTA. The boost and their cascades are typically set to be one fourth of the main OTA. For C_L=1/2C_s, C_f = C_s, m = (g_{m1} + g_{m2}+ g_{m3}+ g_{m4}) / g_{m1} = 1+0.25+0.25+0.125 = 1.625. This results a performance similar to the folded-cascode OTA topology. As stated in section 4.2, for g_{m1} / C_s = 0.1ω_T and a 0.1% setting error, the settling time is 146 with ω_T normalized to 1.

4.5 Nested G_m-C compensation OTA

One of the main challenges in designing the multi-stage OTA is the choice of the feedback path to stabilize the system. There are two main approaches to compensate, the NMC (nested Miller compensation) [57] and NGCC (nested G_m-C compensation) [51]. In [51], it is claimed that for the same power consumption, NGCC has a larger bandwidth than the NMC, or alternately stated less power consumption for the same bandwidth, which means NGCC is more power efficient than NMC. For this reason the NGCC topology is analyzed to determine its power efficiency factor. The block diagram of the NGCC is shown in Figure 4.14.

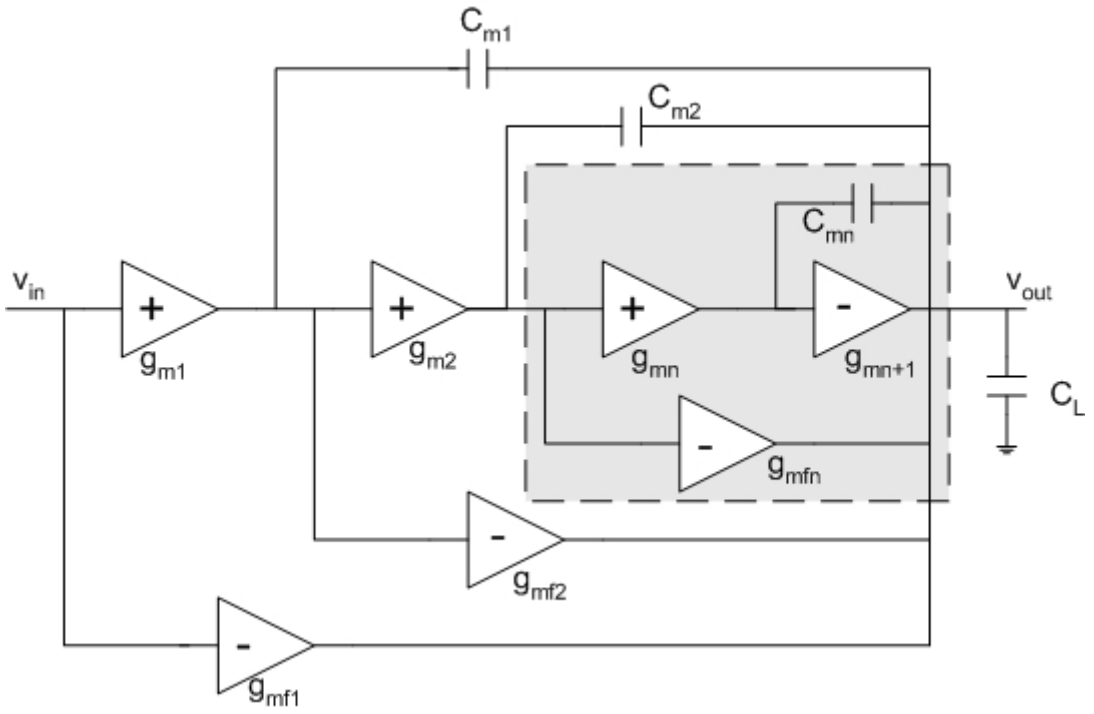


Figure 4.14 Block diagram of the NGCC concept. The shadow dashed block is the basic module of the topology.

In order to compare it with the nested boost cascode OTA topology, the small signal transfer function of the 4-stage NGCC OTA is analyzed. The small signal model of the close loop form is shown in Figure 4.15.

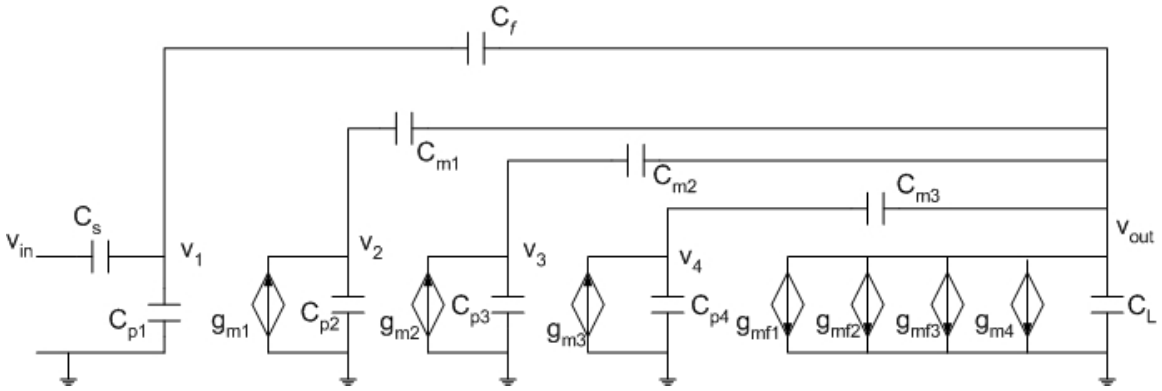


Figure 4.15 The small signal model of the close loop form of 4-stage NGCC OTA.

In Figure 4.15, all the transistor output conductances are ignored for simplicity. The node equations are shown in (4.34-4.38) and the transfer function is derived in (4.39), where all the parasitic capacitors are ignored, $g_{mf1}=g_{m1}$, $g_{mf2}=g_{m2}$, $g_{mf3}=g_{m3}$, and $C_{Leff}= C_L+f\cdot C_s$.

$$(v_{in}-v_1)s\cdot C_s = v_1\cdot s\cdot C_{p1}+(v_1-v_{out})s\cdot C_f \tag{4.34}$$

$$g_{m1}v_1 = v_2 s\cdot C_{p2}+(v_2-v_{out})s\cdot C_{m1} \tag{4.35}$$

$$g_{m2}v_2 = v_3 s\cdot C_{p3}+(v_3-v_{out})s\cdot C_{m2} \tag{4.36}$$

$$g_{m3}v_3 = v_4 s\cdot C_{p4}+(v_4-v_{out})s\cdot C_{m3} \tag{4.37}$$

$$v_1 g_{mf1}+v_2 g_{mf2}+v_3 g_{mf3}+v_4 g_{m4}+v_{out}sC_L=(v_1-v_{out})sC_f+(v_2-v_{out})sC_{m1}+(v_3-v_{out})sC_{m2}+(v_4-v_{out})sC_{m3} \tag{4.38}$$

$$\frac{v_{out}}{v_{in}}=\frac{C_s}{C_f+C_s}\frac{C_f C_{m1}C_{m2}C_{m3}s^4-g_{m1}g_{m2}g_{m3}g_{m4}}{as^4+bs^3+cs^2+ds+f\cdot g_{m1}g_{m2}g_{m3}g_{m4}} \tag{4.39}$$

$$a = C_{Leff}C_{m1}C_{m2}C_{m3}$$

$$b = C_{m1}C_{m2}C_{m3}g_{m4}$$

$$c = C_{m1}C_{m2}g_{m3}g_{m4}$$

$$d = C_{m1}g_{m2}g_{m3}g_{m4}$$

(4.39) can be simplified to

$$\frac{v_{out}}{v_{in}}=\frac{C_s}{C_f+C_s}\frac{\frac{C_f s^4}{C_{Leff}}-\omega_1\omega_2\omega_3\omega_4}{s^4+\omega_4 s^3+\omega_3\omega_4 s^2+\omega_2\omega_3\omega_4 s+f\omega_1\omega_2\omega_3\omega_4} \tag{4.40}$$

where $\omega_1=g_{m1}/C_{m1}$, $\omega_2=g_{m2}/C_{m2}$, $\omega_3=g_{m3}/C_{m3}$, $\omega_4=g_{m4}/C_{Leff}$, assuming $C_{m1}>>2C_{gs2}$, $C_{m2}>>2C_{gs3}$, $C_{m3}>>C_{gs4}$.

Using Routh's array, the stabilization condition of the (4.40) is

$$\omega_4 = \frac{g_{m4}}{C_{Leff}} > \frac{g_{m2}}{C_{m2}} \cdot \frac{1}{1 - \frac{g_{m1} f \cdot C_{m3}}{g_{m3} \cdot C_{m1}}} = \omega_2 \frac{1}{1 - f \frac{\omega_1}{\omega_3}} \tag{4.41}$$

Assuming $\omega_1 = 0.1\ \omega_T$, $\omega_2 = \alpha\ \omega_1$, $\omega_3 = \beta\ \omega_1$, $\omega_4 = \gamma\ \omega_1$, and sweeping the α, β, and γ, numerical simulation is used to find the optimum configuration. Under α = 1.208, β = 2.436, and γ=4.75 configuration, the system has minimum product of the settling time 54 for the 0.1% settling error with ω_T normalized to 1. Here m is difficult to be determined for the C_{m1}, C_{m2}, and C_{m3} can not be determined by the numerical simulation. However, we can derive the following. From above simulations, $\omega_4 = g_{m4}/C_{Leff} = g_{m4}/C_s = 0.475\ \omega_T$ result in C_{p4} approximately 0.5 C_s. In the above assumption that C_{m3} is much larger than C_{p4}, C_{m3} should be at least $2.5C_s$. From the similar approach, we can get $C_{m2} > 2.5C_s$, and $C_{m1} > C_s$. The topology factor, $m = (2 \cdot g_{m1} + 2 \cdot g_{m2} + 2 \cdot g_{m3} + g_{m4}) = (2 + 2*2.5*1.208 + 2.5 \cdot 2.436 + 4.75) = 24.97$. In actuality, this results in an underestimation of m.

From the above analysis, NGCC topology is not a suitable configuration for using multiple stages to achieve high gains. Several miller capacitors not only increase the difficulties in setting the poles of the system in order to make it stable, power consumption increase dramatically as well with the increase of the stage number — larger current needed to push the non-dominant poles out, due to large capacitor loads generated by the miller capacitors, beyond the GBP. To achieve high output swing, the nested cascode miller compensated (NCMC) OTA is recommended.

4.5 Nested cascode miller compensated (NCMC) OTA

Nested cascode miller compensated OTA is formed by combing a nested cascode in the first stage with a common source as the second stage to achieve both high gain and high output swing. The first stage can be expanded to n levels as needed to achieve total gain as high as $(g_m \cdot r_o)^{2n+1}$.

The schematic of nested cascode miller compensated OTA is shown in Figure 4.16. In order to compare with nested cascode and NGCC configurations, the boosted stage is analyzed as follows. The small signal model of the close loop form is shown in Figure 4.17. The node equations are shown from (4.42)-(4.46).

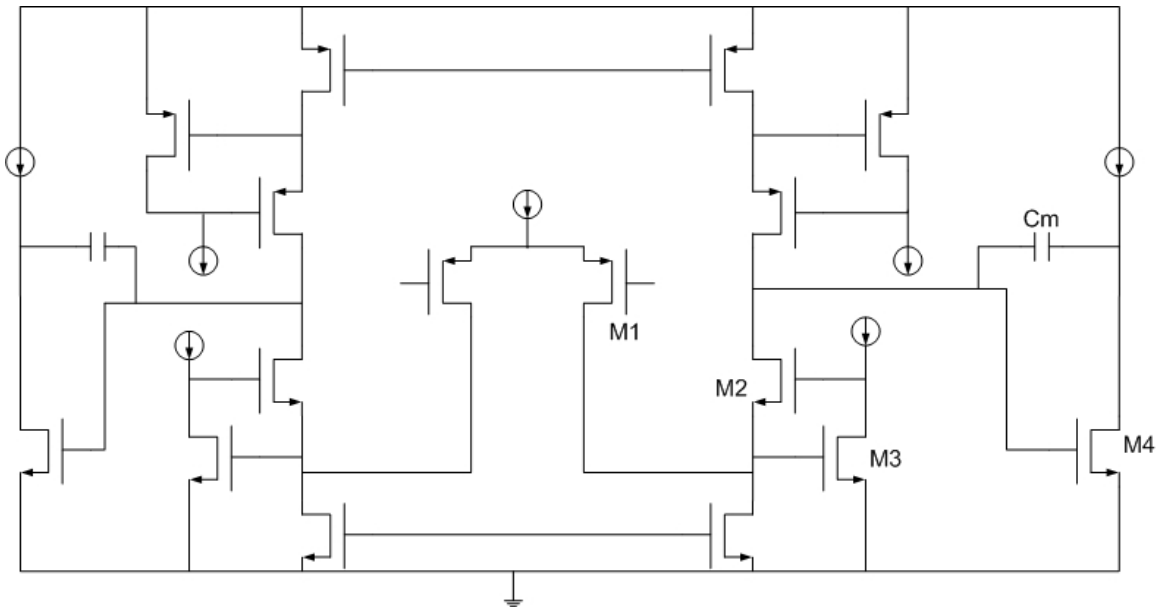


Figure 4.16 the schematic of the nested cascode miller compensated OTA

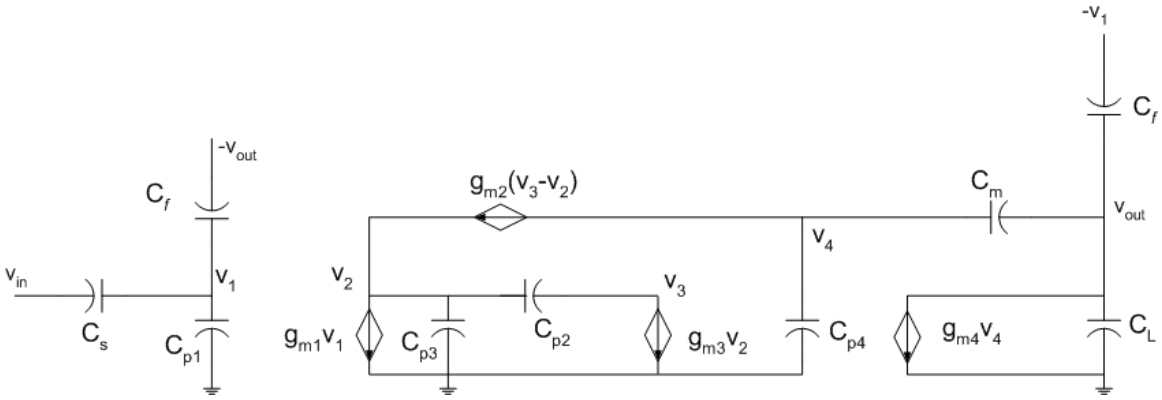


Figure 4.17 the small signal model of the nested cascode miller compensated OTA.

$$(v_{in} - v_1)s \cdot C_s = v_1 s \cdot C_{p1} + (v_1 + v_{out})s \cdot C_f \tag{4.42}$$

$$g_{m1}v_1 + v_2 \cdot s \cdot C_{p3} = (v_3 - v_2)g_{m2} + (v_3 - v_2) \cdot s \cdot C_{p2} \tag{4.43}$$

$$(v_2 - v_3) \cdot s \cdot C_{p2} = g_{m3} \cdot v_2 \tag{4.44}$$

$$g_{m2}(v_3 - v_2) + v_4 \cdot s \cdot C_{p4} = (v_{out} - v_4) \cdot s \cdot C_m \tag{4.45}$$

$$(-v_1 - v_{out}) \cdot s \cdot C_f = g_{m4}v_4 + v_{out} s \cdot C_L + (v_{out} - v_4) \cdot s \cdot C_m \tag{4.46}$$

The simplified transfer function is shown in (4.47) where C_{p2} and C_{p3} are ignored — which is quite reasonable for a cascoded transistor and boosted transistor which are both scaled to ¼ of the input stage transistor from the previous analysis in section 4.2.

$$\frac{v_{out}}{v_{in}} = \frac{-C_s(C_f(C_m + C_{p4})s^2 + C_m g_{m1} s - g_{m1}g_{m4})}{(C_f + C_{p1} + C_s) \cdot (as^2 + bs + c)} \tag{4.47}$$

$$a = C_{Leff}(C_m + C_{p4}) + C_m C_{p4}$$

$$b = C_m(g_{m4} - fg_{m1})$$

$$c = fg_{m1}g_{m4}$$

Comparing (4.47) with (4.21), we find the two equations are exactly identical except g_{m4} is replaced with g_{m2}. This is a "new" topology for the low supply voltage design, the first stage generates sufficiently high gain by using the nested cascode configuration with very little power penalty, while the second stage supplies high output swing. In the typical case, the $m = C_m / C_s \cdot (g_{m1}+g_{m2}+g_{m3}+g_{m4})$ / bandwidth = $1.22 \cdot (1+0.25+0.25+3.25) = 5.795$. For $g_{m1} / C_s = 0.1\omega_T$ and a 0.1% settling error, the settling time is 70 with ω_T normalized to 1.

(4.25) is used to compared the power efficiency of two different topologies. The nested cascode and NCMC OTAs are compared in Figure. 4.18. From it, we find that for an overdrive voltage equal 0.3V, NCMC OTA always has better power efficiency in the supply range of interest; for overdrive voltage equal 0.2V, the NCMC OTA has better power efficiency until supply voltage is greater than 2.2V; when operating in weak inversion or velocity saturation, the nested cascode OTA has better efficiency for supply voltages greater than 1V. The NGCC OTA performance comparison with the nested cascode OTA is shown in Figure 4.19. It is observed for overdrive voltages of 0.1V or 0.2V, the nested OTA has better power efficiency. However, for an overdrive voltage of 0.3V, the nested cascode only has better power efficiency until supply voltage reaches 1.6V.

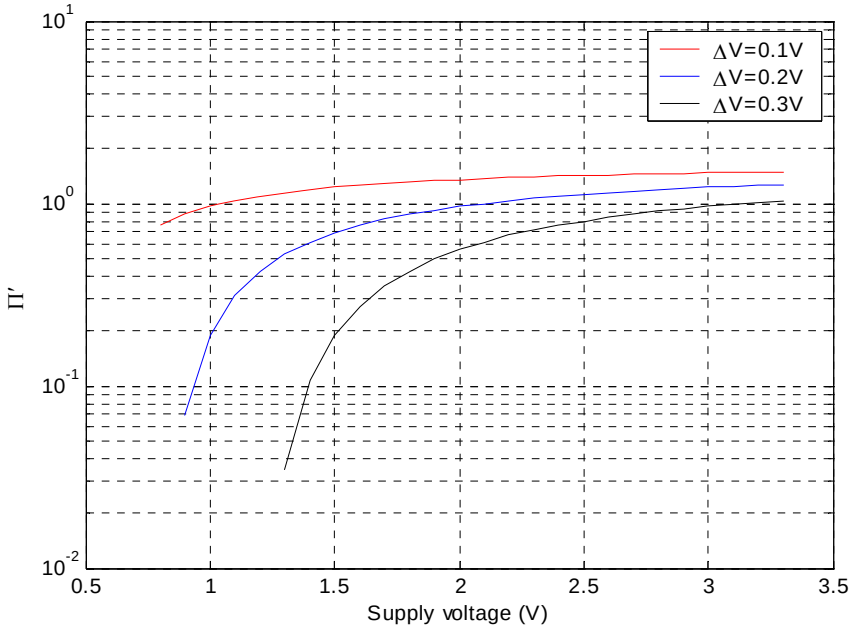


Figure 4.18 Comparison of the normalized power efficiency factor of NCMC OTA and nested gain boost cascode OTA topology, where $\Pi' = \frac{\Pi_{nested\_cascode}}{\Pi_{NCMC}}$.

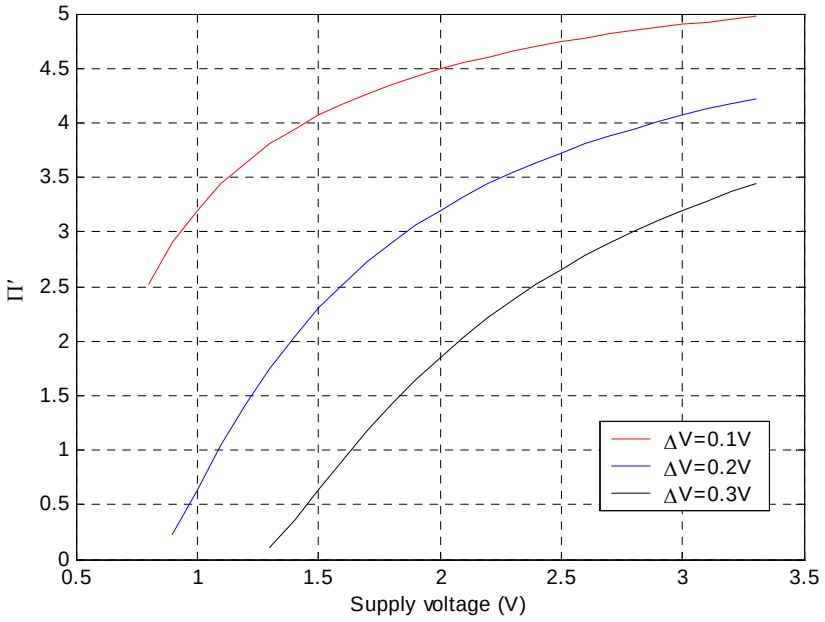


Figure 4.19 Comparison of the normalized power efficiency factor of NGCC OTA and nested gain boost cascode OTA topology, where $\Pi' = \frac{\Pi_{nested\_cascode}}{\Pi_{NGCC}}$.

4.6 Conclusion

In this chapter, the power efficiency of different OTA topologies are compared. In order to find the power efficiency of each topology, close loop analysis is adopted. Analysis shows that this method gives greater accuracy than the open loop or feedback factor method and results power savings especially when a circuit is configured for low close loop gain.

Analysis also shows that nested cascode miller compensated OTA is the first choice in the high gain applications with the scaling supply voltage. Because it has both advantages of the nested cascode and cascaded OTAs — high gain in the first stage and high output swing in the second stage. The multiply stage cascaded OTA is a poor choice for the many feedback stages introduce a large load effective at each stage resulting in heavy power penalty.

Chapter 5

Low-Voltage Low-Power Pipelined ADC Design

5.1 Introduction to pipelined ADC

The pipelined ADC consists of k cascaded stages, which resolves n bits as shown in Fig. 5.1. In each stage, the signal is first sampled and held, then quantized to n bits by sub-A/D. After the sampled data is subtracted from the quantized data, the residue is amplified to its original full-scale data range and applied to the next stage. The quantized data of each stage is applied to the digital correction logic to generate the final m bits.

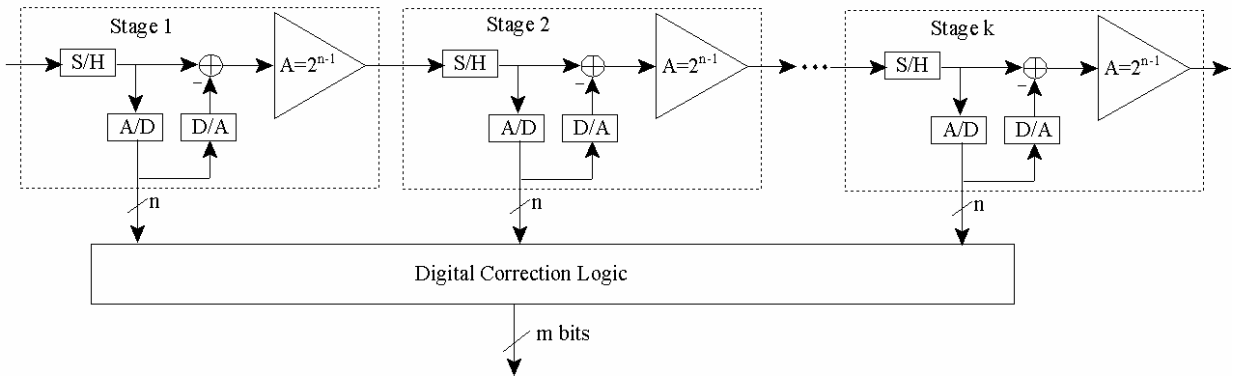


Figure 5.1 Pipelined ADC block diagram

The advantage of this architecture is that each stage only resolves a few bits and the residue is applied to the next stage resulting in high speed conversion. The disadvantage is that there is a k-1 clock cycle latency, which will limit its use in some real-time applications usually involving feedback control. The accuracy of this architecture is limited by the component matching. Currently pipelined ADC architectures typically resolve 10-14 bits sampled at 10-50MHz[58-61].

5.2 Accuracy requirements

Any non-ideal effects will hurt the overall ADC performance. Among them, OTA gain error, OTA settling error, components matching errors and the thermal noise have the most important affects.

5.2.1 OTA gain error

OTA gain error is due to the finite open loop DC gain of the OTA. The output voltage of the close loop OTA is shown as below

$$v_{out} = \frac{1}{\beta}\left(\frac{1}{1+\frac{1}{A_o\beta}}\right)v_{in} \approx \frac{1}{\beta}\left(1-\frac{1}{A_o\beta}\right)v_{in} \tag{5.1}$$

where A_o is the open loop gain of the OTA, 1/β is the close loop gain and v_{in} is the amplitude of the input voltage.

The gain error results from the finite DC gain is

$$error_{gain} = \frac{1}{A_o \beta^2} v_{in} \tag{5.2}$$

5.2.2 OTA settling error

As a result of the finite bandwidth of the OTA, the output signal may not settle to the expected value in a limited settling time period. The difference between the output final voltage and the expected value is the settling error. Assuming the OTA is a first order system and DC open loop gain is infinite, the output voltage is expressed as

$$v_{out} = \frac{1}{\beta}(1 - e^{-t/\tau}) v_{in} \tag{5.3}$$

where τ is the time constant of the first order system, t is the settling time, and 1/β is the close loop gain. The settling error is written;

$$error_{gain} = \frac{1}{\beta} e^{-t/\tau} v_{in} \tag{5.4}$$

5.2.3 Components mismatch errors

Mismatching errors mainly come from capacitors mismatch and offset voltage due to the mismatch of the comparator and OTA. The mismatch of the capacitors dominates the error in a low-to-median resolution ADC. From a thermal noise perspective, a 10-bit, 1V full scale signal only requires a 0.625pF sampling capacitance. However, the practical capacitance is much larger due to capacitor match coefficients. Furthermore, the mismatch of the capacitors is inverse proportional to the capacitor's area. To reduce the mismatch by half, four times the area and power must be consumed. Assuming the open loop gain of the OTA is infinite and the bandwidth of the OTA is also infinite, the output voltage is

$$v_{out} = \frac{1}{\beta} = \frac{C_s(1 \pm \varepsilon)}{C_f(1 \pm \varepsilon)} v_{in} \tag{5.5}$$

where ε is the capacitor mismatch.

Considering only the first terms in making the approximation, the maximum error resulting from the capacitor mismatch is written as;

$$\begin{aligned} v_{out} &= \frac{C_s(1+\varepsilon)}{C_f(1-\varepsilon)} v_{in} = \frac{1}{\beta} \frac{(1+\varepsilon)^2}{(1-\varepsilon^2)} v_{in} \approx \frac{1}{\beta}(1+2\varepsilon) v_{in} \\ error_{mis} &= \frac{2\varepsilon}{\beta} v_{in}. \end{aligned} \tag{5.6}$$

Offset voltage is mainly due to the Pellegrom threshold voltage and β mismatch, which coefficients can be represent by [62]

$$\sigma^2(\Delta V_T) = \frac{A_{1V_T}^2}{WL} + \frac{A_{21V_T}^2}{WL^2} + \frac{A_{3V_T}^2}{W^2 L} \tag{5.7}$$

$$\sigma^2\left(\frac{\Delta\beta}{\beta}\right) = \frac{A_{\beta 1}^2}{WL} + \frac{A_{\beta 2}^2}{WL^2} + \frac{A_{\beta 3}^2}{W^2 L} \tag{5.8}$$

where W is the transistor width, L is the transistor length.

Offset voltage is not of a major concern in the pipelined ADC design. The resulting conversion error can be corrected by the digital correction algorithm by utilizing the one or more extra bits of each stage.

5.2.4 Thermal noise

Thermal noise is generated by the MOSFET switch on-resistance and the transistors of the OTAs. Which one dominates the noise source is circuit dependent. This will be demonstrated in the following analysis. The input referred mean-squared thermal noise voltage of a stage from the switches is

$$v_{ni}^2 = \gamma \frac{kT}{C_{si}} \qquad i = 1,2,\cdots,k \tag{5.9}$$

where i is the stage number of the ADC, and γ is circuit topology constant.

The total input referred noise voltage of the switches is

$$v_n^2 = \gamma kT \sum_{i=1}^{k} \frac{1}{C_{si} A_c^{2(i-1)}} \tag{5.10}$$

where A_c is the closed loop gain of each stage.

The thermal noise from the transistors of the OTA can be expressed as

$$v_n^2 = \gamma_1 \frac{8kT}{3g_m} BW \tag{5.11}$$

where γ_1 is determined by the circuit topology, and BW is the noise bandwidth. In order to calculate the noise bandwidth, assuming the OTA is represented by a single dominant pole system, the OTA transfer function can be expressed as

$$A(s) = \frac{A_0}{1 + \frac{s}{2\pi f_B}} \tag{5.12}$$

where A_0 is the open loop DC gain, and f_B is the 3dB frequency. The noise power is

$$v_n^2 = \int_0^\infty \gamma_1 \frac{8kT}{3g_m} \frac{A_0}{1+\frac{f}{f_B}} df = \gamma_1 \frac{8kT}{3g_m} \frac{\pi A_0 f_B}{2} = \gamma_1 \frac{4\pi kT}{3g_m} GBP \tag{5.13}$$

(5.13) can be simplified to

$$v_n^2 = \gamma_1 \frac{4\pi kT}{3g_m} \frac{g_m}{C_{Leff}} (1-f) = \gamma_1 \frac{4\pi kT}{3C_{Leff}} (1-f) \tag{5.14}$$

Comparing (5.9) with (5.14), thermal noise from the OTA is on the same order of the noise from the switches.

5.2.5 Error tolerance

Gain error, settling error and capacitor mismatch result in DNL (Differential Non-linearity) and INL (Integral Non-linearity) of the ADC. DNL is the difference between a specified code bin width and the averaged code bin width, while INL is the maximum difference between the ideal and actual code transition levels after correcting for gain and offset. INL error is accumulated from stage to stage. Due to the non-linear and random properties of the error sources, INL is very difficult to analysis [53]. It is neglected here. DNL is required to be less than ½ LSB, which means the total output voltage error of each stage due to finite gain, finite settling time and capacitor mismatch should be less than ½ LSB of the next stage. For a N-bit resolution ADC, using 1st stage as an example, its total output voltage error should be less than

$$\begin{aligned} &error_{gain} + error_{settle} + error_{mis} < \frac{1}{2} LSB_{N-1} \quad \Rightarrow \\ &\left(\frac{1}{A_o \beta^2} + \frac{1}{\beta} e^{-t/\tau} + \frac{2\varepsilon}{\beta} \right) v_{in} < \frac{1}{2} \frac{v_{in}}{2^{N-1}} \end{aligned} \tag{5.15}$$

How to allocate the gain error, settling error and capacitor mismatch error in the total tolerated error? We will analysis it from the perspective of the required power consumption needed to minimize the each error. For gain error, the DC gain needed is

$$A_o > \frac{2^N}{v_{in}\beta^2} \tag{5.16}$$

Assuming a nested cascode OTA is used, the normalized power consumption vs. the resolution bits is shown in Figure 5.2, where vin full scaling voltage is assumed to be 1V, and β is 1/2. From Figure 5.2, with the resolution increase, the power consumption increases less and less. In theory, as resolution goes to the infinity, the power consumption is a constant and independent of the resolution. From this point of view, the penalty induced by the gain error is quite minimumal.

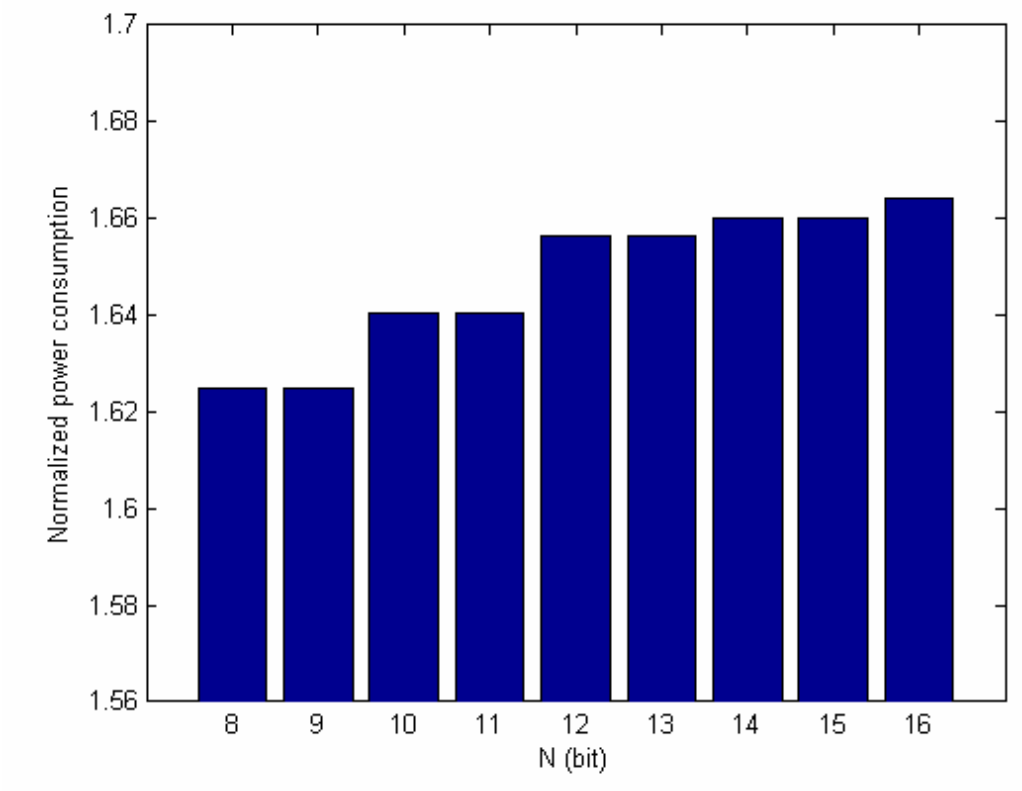


Figure 5.2 The normalized power consumption vs. resolution.

As to the settling time error, there are two potential solutions: 1) increase the sampling time, which will not increase the power consumption, however, this usually is not desired or permitted; 2) decrease the settling time constant, which means an increase in the power consumption. Still assuming the unit gain feedback that v_{in} full scaling voltage is 1V, β is 1/2, the settling error can be rewritten as

$$e^{-t_s/\tau} < \frac{1}{2^N} \Rightarrow \tau < \frac{t_s}{N \ln 2} \tag{5. 17}$$

Assuming a single pole system, (5.17) can be written

$$\frac{A_o C_{Leff}}{g_m} < \frac{t_s}{N \ln 2} \Rightarrow g_m > \frac{A_o C_{Leff} N \ln 2}{t_s} \tag{5.18}$$

From (5.18), it clearly shows that the power consumption is linearly proportional to the resolution.

Capacitor mismatch error will need considerably more power consumption to overcome its resulting error, for the mismatch is inversely proportional to the capacitor area [63], or in summary a four times larger capacitor (resulting in 4x greater power) is required to reduce mismatch by half. In order to maintain constant bandwidth, g_m and power consumption must increase exponentially. So the design strategy is to allocate the error is to minimize the gain error and settling error, leaving more tolerate error to the capacitor mismatch. The calibration algorithm at the back end of the ADC should also be considered in the high resolution application.

5.3 Stage scaling

As mentioned in [54, 64], the inter-stages beyond the first stage can be scaled because their gain, settling error, and noise requirement is less restricted than that of the first stage. Scaling later stages is one of the most efficient ways to save power. The pipelined ADC block diagram after the scaling is shown as in Figure 5.3. Inter-stage scaling is heavily dependent on the number of bits extracted per stage.

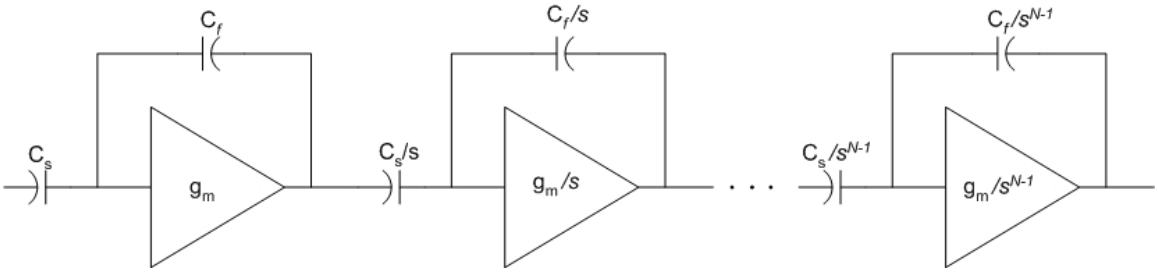


Figure 5.3 Pipelined ADC block diagram after scaling.

In order to have a clear look at the scaling effect on the power consumption, assume the capacitor matching is perfect, that is there is no extra power consumption needed to compensate for the capacitor mismatch, and that the OTA is a single pole system. In other words the design is thermal noise limited. Power consumption is proportional to total g_m, which is also proportional to the close loop gain times the effective capacitance load at the fixed bandwidth summed across the total number of stages. From (4.12), the power consumption of each stage is shown as below.

$$\begin{aligned} P_i &\propto C_{Leffi} / f = (C_{Li} + fC_{si}) / f = (C_{s(i+1)} + fC_{si}) / f \\ &= C_{si}(\frac{1}{sf} + 1) \quad i = 1,2,\cdots k \end{aligned} \tag{5.19}$$

where s is the scaling factor, and f is the feedback factor. The total power consumption is proportional to

$$P \propto (\frac{1}{sf}+1)\sum_{i=0}^{k-1} C_{si} = C_{s1}(\frac{1}{sf}+1)\frac{1-\frac{1}{s^k}}{1-\frac{1}{s}} \tag{5.20}$$

In [54, 64], however, one item neglected is that for the same desired resolution in bits of the ADC, a different number of output bits of per stage will result a different number of stages. For example, for a 12-bit ADC, the number of stages k for 1-bit, 2-bit, and 3-bit per stage is 11, 6, and 4 respectively. Therefore, for 2-bit, and 3-bit per stage, (5.20) cannot be simplified to (5.21) by assuming k is simple large enough.

$$P \propto C_{s1}(\frac{1}{sf}+1)\frac{1}{1-\frac{1}{s}} \tag{5.21}$$

Combining (5.9) and (5.14), the total thermal noise of each stage is

$$v_{ni}^2 = \frac{\rho kT}{C_{si}} i = 1,2,\cdots k, where\ \rho = \frac{\gamma_1 4\pi(1-f)}{3(1/s+f)} + \gamma \tag{5.22}$$

With scaling s, the total noise can be simplified to

$$v_n^2 = \frac{\rho kT}{C_{s1}}\sum_{i=0}^{k-1}\left(\frac{s}{A_c^2}\right)^i = \frac{\rho kT}{C_{s1}}\frac{1-\left(\frac{s}{A_c^2}\right)^k}{1-\frac{s}{A_c^2}} \tag{5.23}$$

where A_c is the close loop gain. In order to achieve N-bit resolution, the total noise should be

$$v_n^2 < \left(\frac{LSB}{2}\right)^2 = \left(\frac{v_{in}}{2\cdot 2^N}\right)^2 = \frac{v_{in}^2}{2^{2N+2}} \tag{5.24}$$

Combining (5.20), (5.23) and (5.24), the normalized total power is written;

$$P_{nor} = \rho\left(\frac{1}{sf}+1\right)\frac{1-\frac{1}{s^k}}{1-\frac{1}{s}}\cdot\frac{1-\left(\frac{s}{A_c^2}\right)^k}{1-\frac{s}{A_c^2}} \tag{5.25}$$

Assuming γ and γ_1 are 4 and 6 respectively, the normalized the power consumption versus the scaling factor is plotted in Figure 5.4. For n, the number of bits per stage equals 1. Repeating for n equals 2, and 3 results in Table 5.1.

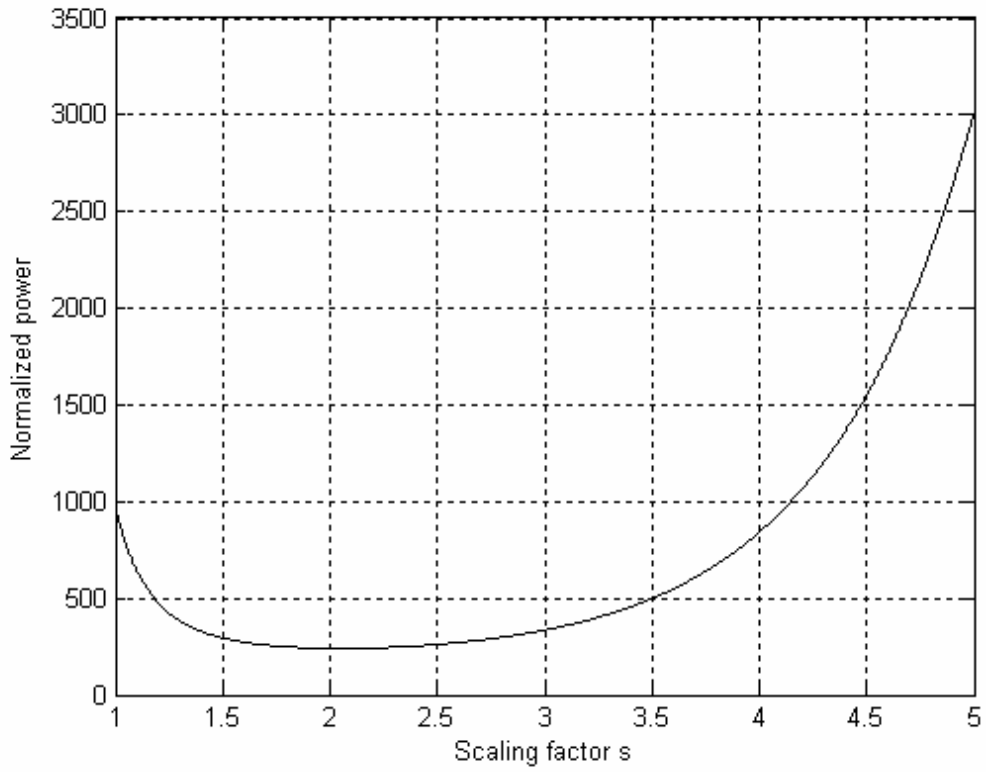


Figure 5.4 Normalized power versus scaling factor s for 12-bit pipelined ADC, where n, which equals to 1, is the output bits per stage.

Table 5.1 Optimized scaling factor s versus output bits per stage

Output bits per stage n	Scaling factor s
1	2.06
2	4.20
3	8.52

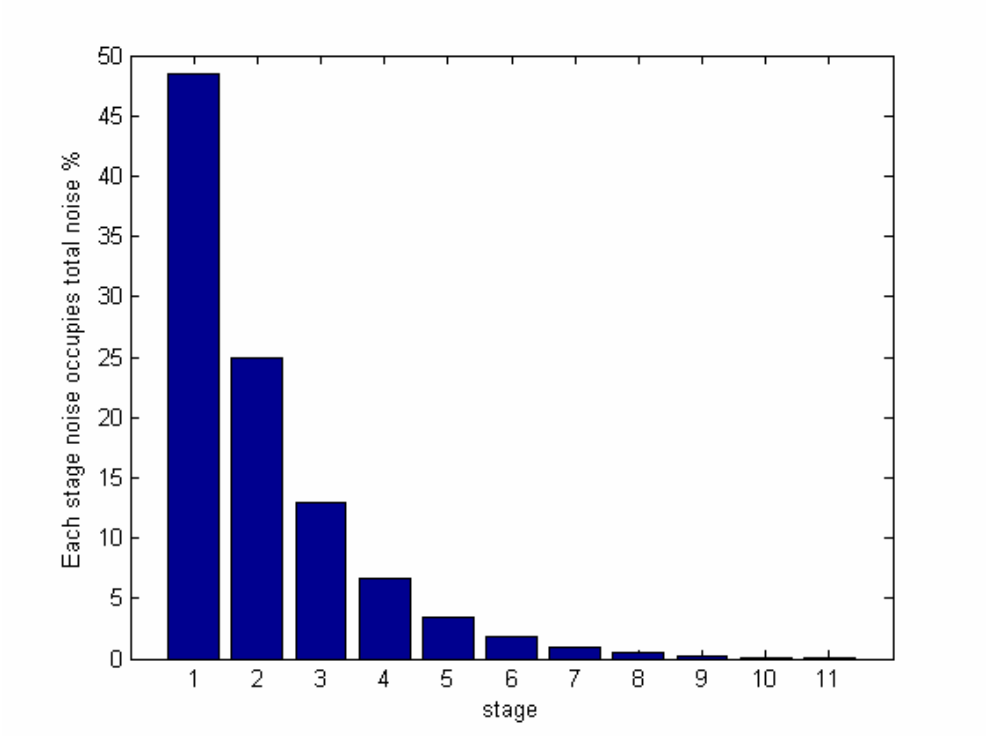


Figure 5.5 Each stage noise occupies the total noise for 12-bit pipelined ADC.

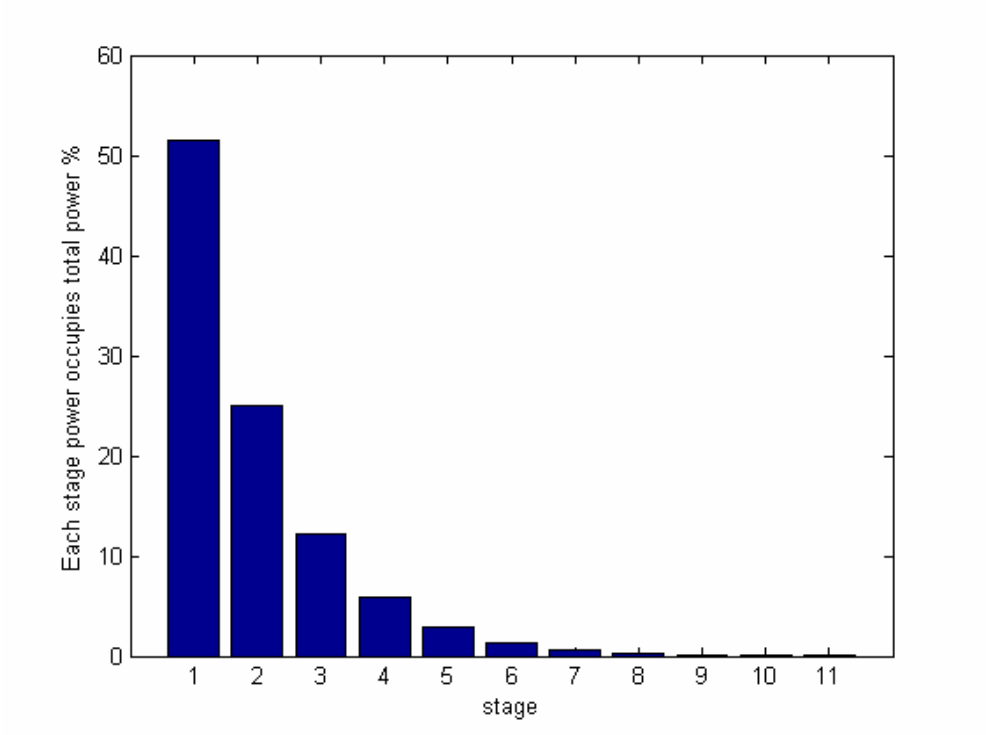


Figure 5.6 Each stage power occupies the total power for 12-bit pipelined ADC.

Figure 5.5 and 5.6 are bar graphics of the each stage noise and power allocation for the optimum scaling s. Contrary to the common thought, no average noise allocation nor average power allocation results in the optimum scaling. Note, although γ and γ_1 are different for both the different topologies and processes, it results in a near optimum scaling number. From Table 5.1, we can find that the optimum scaling s increases with the each stage's output number of bits. In practice there will a limit to the practice lower scaling bond on OTAs the bar graph of Figures 5.5 will saturate. The reader should note that capacitor mismatch results in all capacitors being scaled at the same rate in all stages and as a result only pushes up the power requirement having no effect on the scaling. *The end result is that (5.25) is independent of capacitor mismatch.*

5.4 Modeling pipelined ADC in Simulink

In order to achieve the design objectives, extensive circuit analysis and simulations need to be performed to guide the sizing of transistors in each of the ADC components. Classical transient simulation methods in SPICE to determine of ADC design specifications such as INL, DNL, and SFDR are time prohibitive when trying to explore the tradeoffs in the design space. A top down design approach described in this section is to model the pipeline ADC in MATLAB Simulink environment, which speeds up in simulation time compared to the SPICE and as a result facilitates analysis and optimization of system performance metrics. Once the optimized architecture is known, the designer can design each component more efficiently as one knows the architecture and specifications to be met.

A key feature of MATLAB Simulink is the ability to describe circuit blocks in discrete time operation, where circuit responses to inputs are only calculated at clock edges therefore increasing the speed of the simulation. Other relevant features of Simulink include the creation of hierarchical block diagrams, an extensive collection of functions and data functions, and flexible simulation time step control.

5.4.1 Pipelined ADC Simulink library overview

Pipelined ADC is mainly composed by several stages with digital correction circuits (Figure 5.7). Each stage is composed by the sub-ADCs, and MDACs (Figure 5.8) — besides first stage, MDAC of each stage includes S/H function. In order to capture the

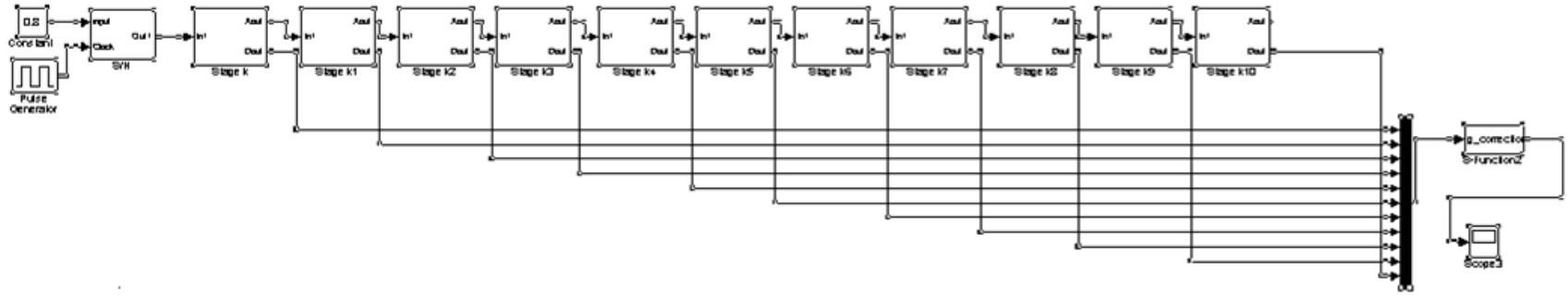


Figure 5.7 Pipelined ADC in Matlab Simulink

impact of non-idealities' on the ADC system performance, gain error, settling error, components mismatch, slew-rate, and switch thermal noise are all considered in the behavior simulation.

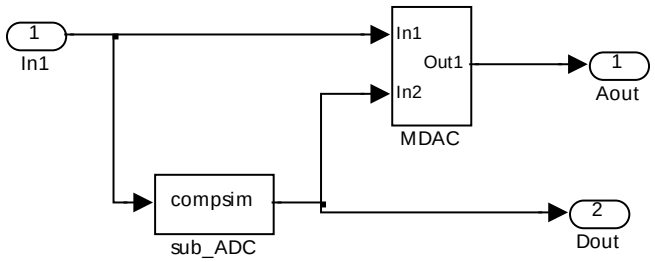


Figure 5.8 Simulink block diagram of each stage

4.2 Pipelined ADC components

5.4.2.1 S/H circuit

Sample and hold circuit is shown as below

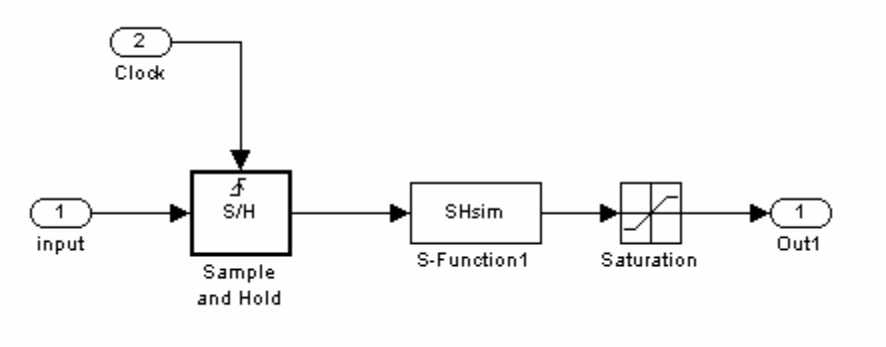


Figure 5.9 Sample and hold circuit.

The sample-and-hold is a critical component of the pipelined ADC. As mentioned in section 5.2, it must have a high open-loop gain and GBP sufficient to maintain adequate precision.

The sample-and-hold model shown in Figure 5.9 has three blocks. The sample-and-hold block performs an ideal sample and hold on each clock edge. The saturation block simply limits the signal output to $\pm V_{fs}/2$, in which V_{fs} is the full-scale voltage range of the ADC. The SHsim block is a gain of unity amplifier. The amplifier block calls a function which takes into consideration the sampling time, the gain bandwidth, the finite open loop gain, offset voltage and the slew rate of the amplifier. It initially checks to see if the amplifier is going to slew-rate limit by checking whether or not the input step is greater than $\sqrt{2}\Delta V$ (where ΔV is the overdrive voltage). If the input is less than this value, simple linear settling occurs and the voltage endpoint can be calculated by using the following equation:

$$Vout = Vin \cdot A_{DC}(1 - e^{-\delta T/\tau}) \tag{5.26}$$

where δ is the fraction of the sampling time (or the duty cycle) that the circuit takes a sample, T is the inverse of the sampling frequency, f_s, and τ is the time constant, $A_{DC}/(2\pi GBP)$.

If the input is greater than $\sqrt{2}\Delta V$ (Figure 5.10), then the function calculates how long the circuit will slew-rate limit. This is determined by the following equation:

$$t_0 = \frac{Vin \cdot A_{DC} - \sqrt{2}\Delta V}{SR} \tag{5.27}$$

where A_{DC} is the closed-loop DC gain, which is unity in the case of the sample-and-hold. Once t_0 is known, it is compared with the sampling window, δT. If it is greater than the sampling window, the amplifier slew-rate limits the entire time and the final voltage is

simply SR·δT. However, if t_0 is smaller than the sampling window, this means the final voltage of the amplifier output results from a combination of slew-rate limiting and linear settling. In summary the amplifier slews to t_0 and linear settles up until δT. The equation for this case is as follows:

$$Vout = SR \cdot t_0 + (Vin \cdot A_{DC} - SR \cdot t_0)(1 - e^{-(\delta T - t_0)/\tau}). \tag{5.28}$$

The worst case scenario for the sample-and-hold is when the input signal is + or – $V_{fs}/2$ since this produces the largest signal swing on the amplifier during the sampling window. The sample-and-hold needs to settle to an accuracy of less than ½ LSB of the N-bit pipelined ADC with the worst-case input value to be adequate [65].

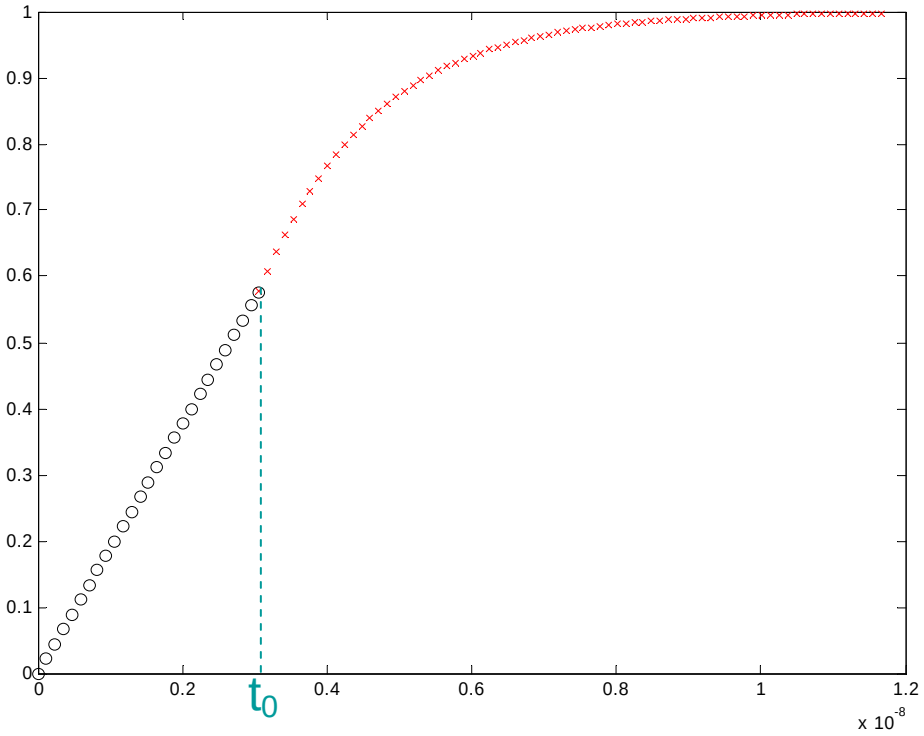


Figure 5.10The conceptual of the settling combined with the occurrence of OTA signal slewing. Note the o's represent slewing while the +'s represent linear settling.

5.4.2.2 Sub-ADC

The sub-ADC's precision is affected by resistor mismatch, thermal noise, and systematic and random comparator offsets. The composite of these non-idealities again needs to remain small enough to produce an error of less than ½ LSB of the stage.

The sub-ADC model is an S-function, which is based on an architecture using a reference resistor string to provide 2^N-1 reference voltages to the reference input of the comparators. The input voltage signal is then connected to the remaining inputs of the comparators. The resulting output of this circuit is a thermometer-coded digital output. As shown in Figure 5.11, it has N, Res, deltaR, Vref and offsetV input parameters, which are define resolution bits, reference resistor value, resistor mismatch standard deviation, reference voltage and offset voltage respectively. Therefore, the non-idealities of a sub-ADC are included in the modeled functions. The function, *mismatch_R* calculates a vector of resistor values given the number of bits, the mean, and the standard deviation percentage mismatch. The function, *setup_refer* calculates the reference voltages of the flash given the number of bits, the resistance vector previously calculated, and the comparator offset voltage. The user is able to input both the percentage accuracy of each resistor as well as the comparator offset voltage as a standard deviation. Thermal noise is also added later to the voltage levels as a random variable which constantly changes with each time step.

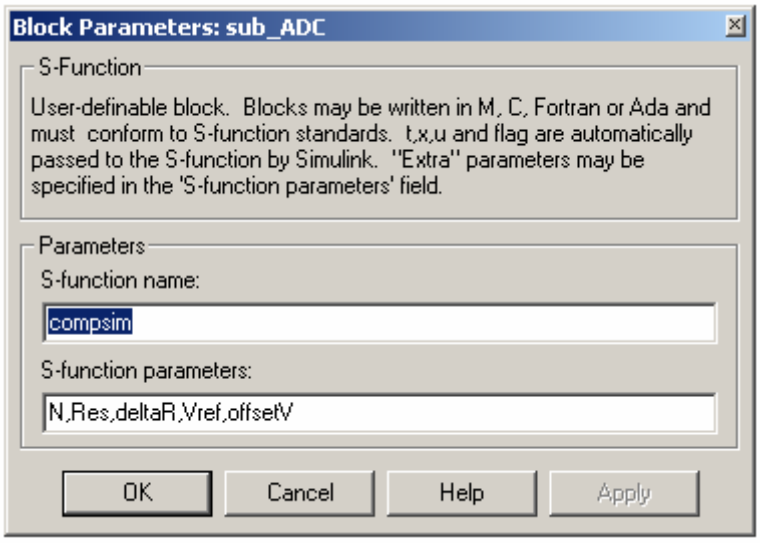


Figure 5.11 Definition of sub-ADC function.

5.4.2.3 MDAC

The schematic of a MDAC, whose gain is 4, is shown in Figure 5.12. Its Simulink model is shown in Figure 5.13. The non-idealities of a MDAC model include thermal noise voltage, capacitor mismatch, finite gain and finite GBP. These non-idealities can be set as input parameters before the start of each simulation (Figure 5.14). The amplifier is modeled the same as the S/H circuit OTA. The difference between the two is that the closed loop gain of the MDAC amplifier is not unity, meaning that more bandwidth is required of the MDAC OTA. The thermal noise voltage is model as shown in Figure 5.15. It represents equation as (5.9).

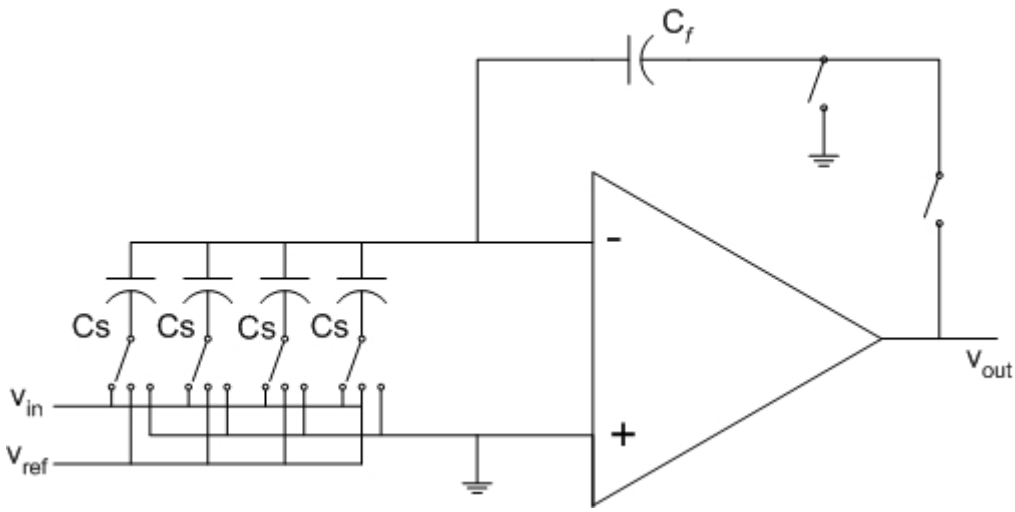


Figure 5.12 Schematic of a MDAC.

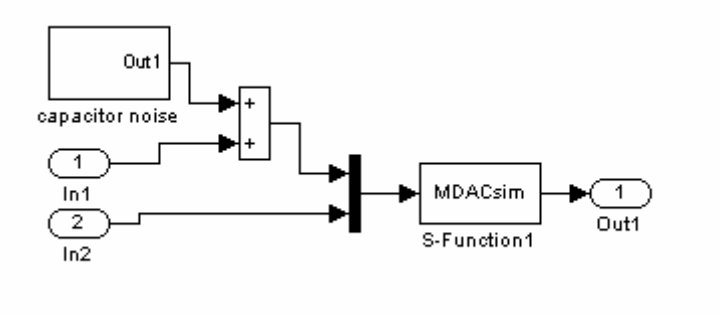


Figure 5.13 Simulink model of MDAC.

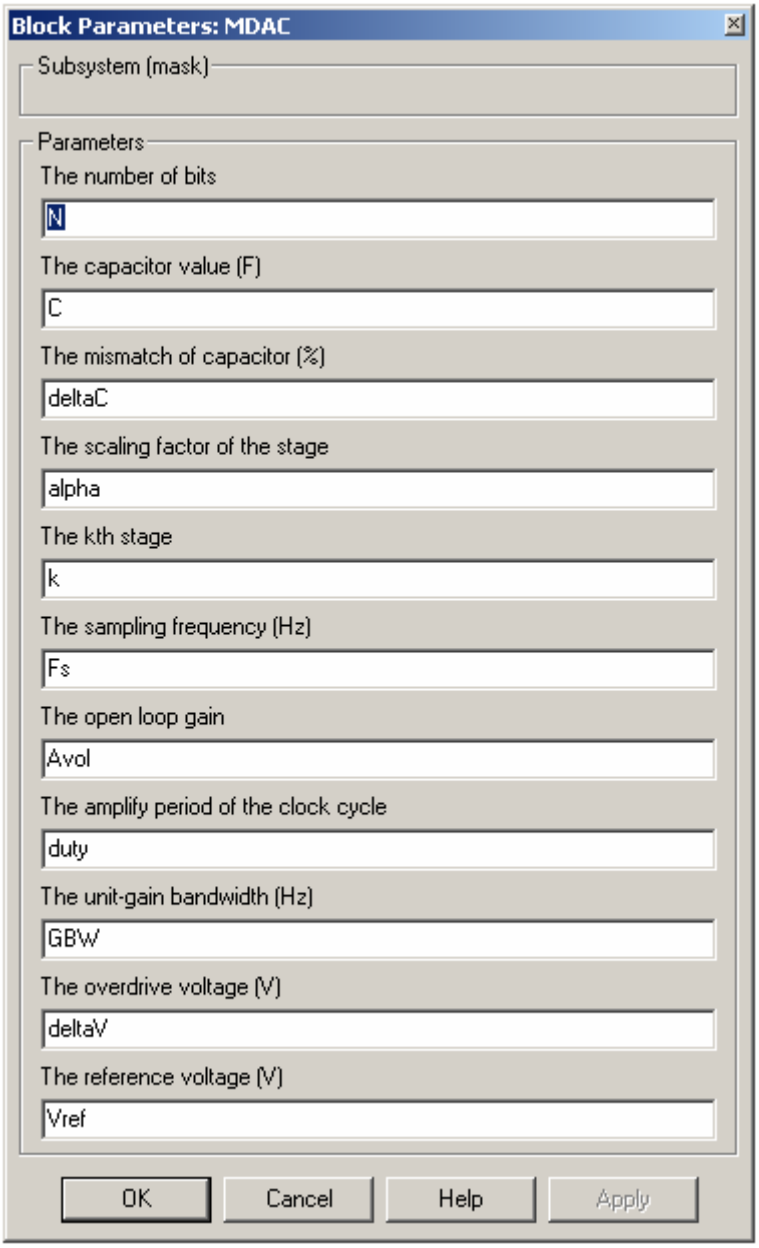


Figure 5.14 Input parameters window of MDAC.

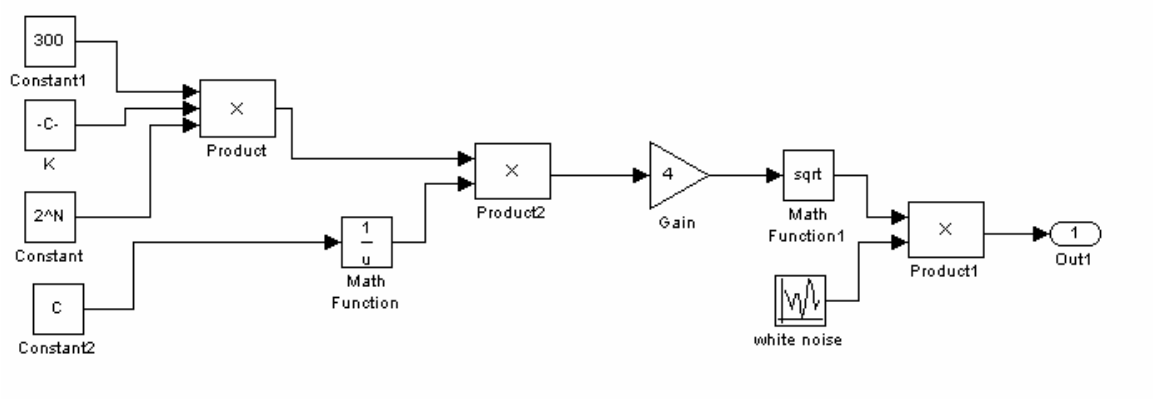


Figure 5.15 Simulink model of thermal noise voltage.

5.4.2.4 Digital correction circuit

The digital circuitry in the pipelined converter performs the functions of digital error correction combining the binary results of each stage into a final N-bit binary number. In order to demonstrate how the digital correction circuit works, a 4-bit pipelined ADC is used as an example. Assuming a bipolar reference voltage is used, the input signal of each stage can be express as

$$v_{in} = (2 \cdot B_{in} - 1)v_{ref} + v_{res} \tag{5.29}$$

where Bin=$b_1 \cdot 2^{-1} + b_2 \cdot 2^{-2}$, and v_{res} is the residue after the 2-bit sub-ADC. While the output of the 2-bit DAC is

$$v_{DAC} = (2 \cdot B_{in} - 3/4)v_{ref} \tag{5.30}$$

With the MDAC gain is 2, the output voltage of each stage is

$$v_{out} = 2 \cdot (-1/4v_{ref} + v_{res}) = -1/2v_{ref} + 2v_{res} \tag{5.31}$$

If offset voltage present in the comparator, the (5.31) will be modified to

$$v_{out} = -1/2v_{ref} + 2v_{res} - 2v_{off} \tag{5.32}$$

where v_{off} is the offset voltage of the comparator. The absolute value of the output voltage of each stage should be less than the reference voltage in order to be quantized. This will lead to

$$\left|-1/2v_{ref}+2v_{res}-2v_{off}\right|<v_{ref}\Rightarrow -1/4v_{ref}<v_{res}-v_{off}<3/4v_{ref} \tag{5.33}$$

From (5.29), we know

$$0<v_{res}<1/2v_{ref} \tag{5.34}$$

Combining (5.33) and (5.34), the maximum tolerate offset voltage of the comparator is

$$\left|v_{ref}\right|<1/4v_{ref} \tag{5.35}$$

or an LSB/2 of each stage. What should be noted is that the offset voltage of the OTA cannot be corrected by the digital correction circuit for it is added offset voltage on the input signal of the next stage.

The digital circuitry converts each of the flash outputs to a 2's complement form by subtracting ½ LSB from each digital word. Once each digital word is in 2's complement form, they are added together by overlapping the first bit of the following stage with the last bit of the previous stage and zero padded everywhere else. One bit per MDAC is lost due to the error correction. Hence, at the output of each stage, one less bit is resolved than the sub-ADC resolution to achieve error correction.

5.4.2.5 Simulink simulation result

The pipelined ADC can be tested using both static and dynamic methods. Two common tests for this are the ramp test and the FFT test respectively. The ramp test is performed by simply inputting an analog ramp having a slope which covers the full scale

range of the ADC. In practical ADCs, the output is not perfect and results in both DNL and INL.

The FFT test is performed by sending a pure analog sine-wave into the ADC and analyzing the power spectrum of the output of the ADC. Performance metrics that arise from this test are the effective number of bits (ENOB), signal to noise and distortion ratio (SINAD), and the spurious free dynamic range (SFDR).

The DNL and INL of the static test results are shown in Figure 5.16. The spectrum of the FFT test results is shown in Figure 5.17. The DNL and INL are less than 1.1LSB and 1.2LSB respectively. SFDR is 71.2dB.

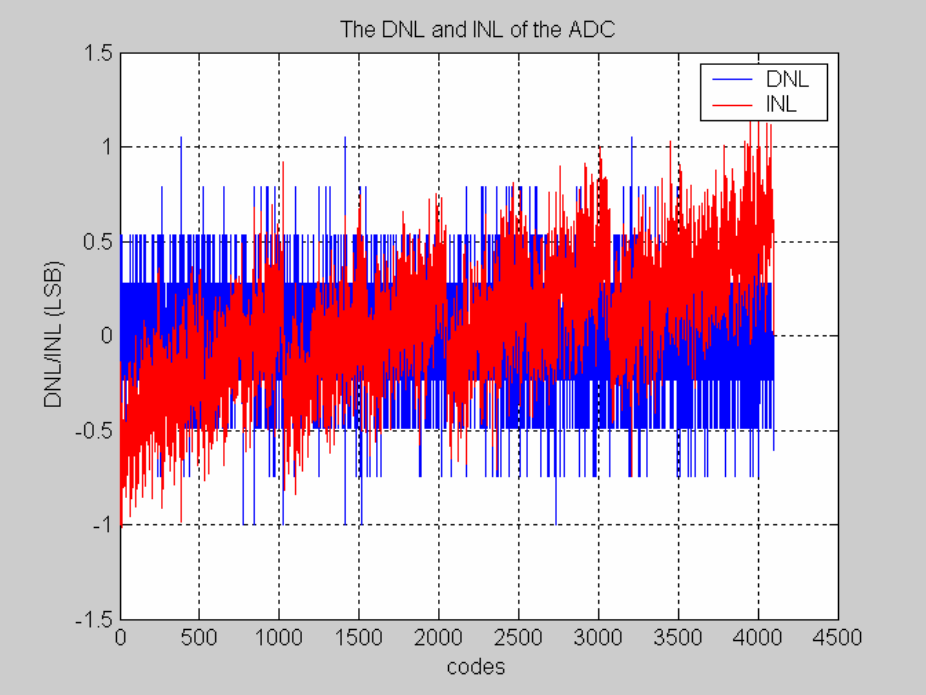


Figure 5.16 The DNL and INL of a pipelined ADC Simulink simulation.

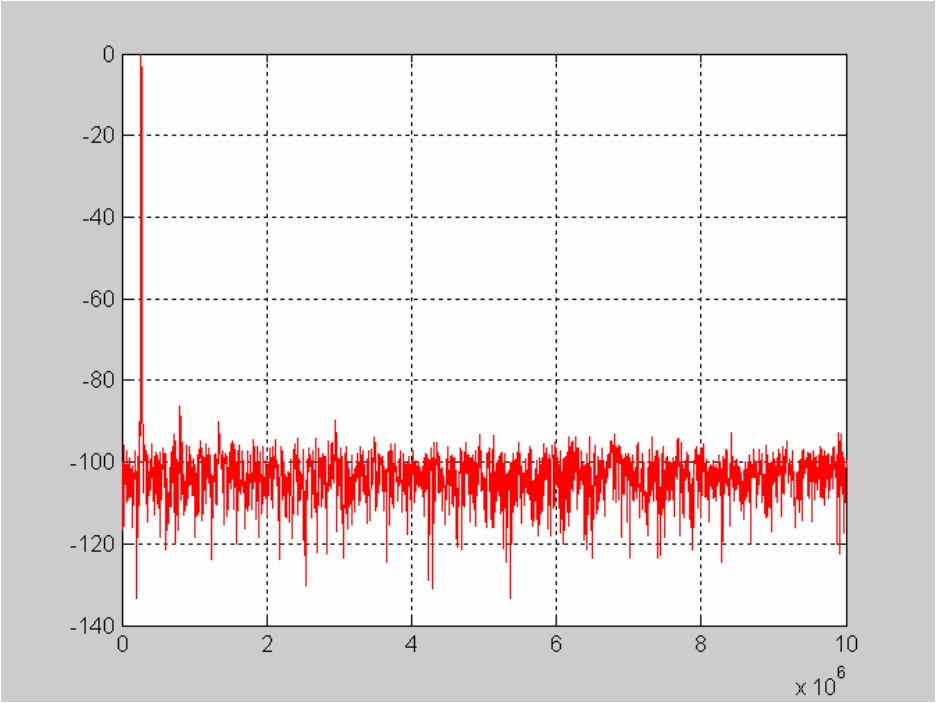


Figure 5.17 The spectrum of the ADC FFT test Simulink simulation results.

5.5 Design example

5.5.1 Specifications of the pipelined ADC

In this section, low voltage, 11-bit, 20Msample/s pipelined ADC is designed as a practice to demonstrate above discussed techniques. The process assumes the Peregrine 0.5um SOI CMOS technology. The desired specification of the ADC are shown in Table 5.2.

Table 5.2 Specifications of the object ADC

Voltage supply (V)	±1
Resolution (bits)	11
Sampling frequency (MS/s)	20
Bits/stage	2
Sampling capacitor (pF)	2
Open loop gain	80dB
Reference voltage (V)	±1

5.5.2 2-bit/stage architecture

A typical block diagram of 2-bit/stage is shown as Figure 5.18. The reason for selecting 2-bit resolution/stage is mainly a bandwidth consideration. The more bits of resolution per stage means less bandwidth. Each stage of the pipelined ADC is composed of a 2-bit sub-ADC and a MDAC. After the previous stage output is subtracted by the quantized analog voltage from the sub-ADC, the resulting residue is amplified by the close loop gain of 2 of each amplifier, and then fed to the next stage. During the ϕ_1, feedback capacitor is connected with the input signal together with the sampling capacitor. During the ϕ_2, feedback capacitor is connected to the output to perform gain of 2 close loop. In this way, the feedback factor increases from 1/3 to 1/2 and resulting in larger bandwidth [53] and improved settling.

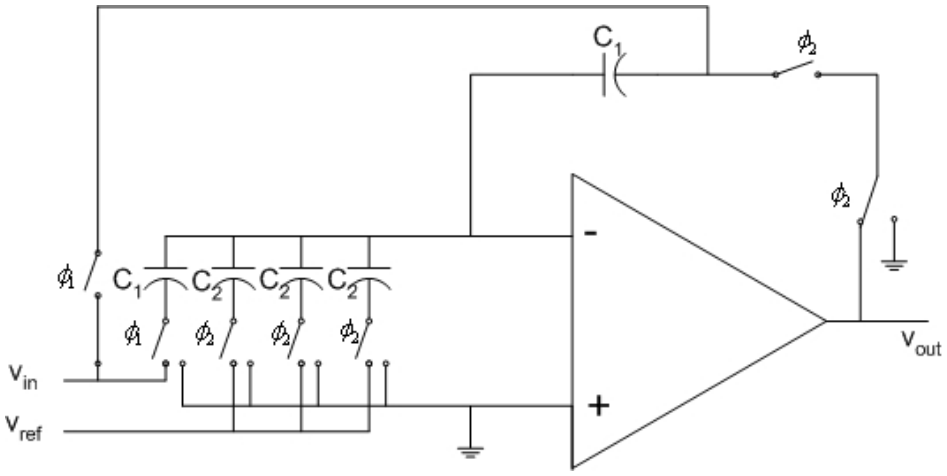


Figure 5.18 The block diagram of 2-bit MDAC, C_2=1/2C_1.

5.5.3 Nested cascode OTA

As to shown in Chapter 4, the nested cascode OTA has better power efficiency than the NGCC OTA and Nested cascode miller compensated OTA at a 2V voltage supply and 0.15V overdrive voltage. Therefore, a nested cascode OTA was used in the

gain stage. The schematic of the fully differential OTA without the common mode feedback circuit is shown in Figure 5.19.

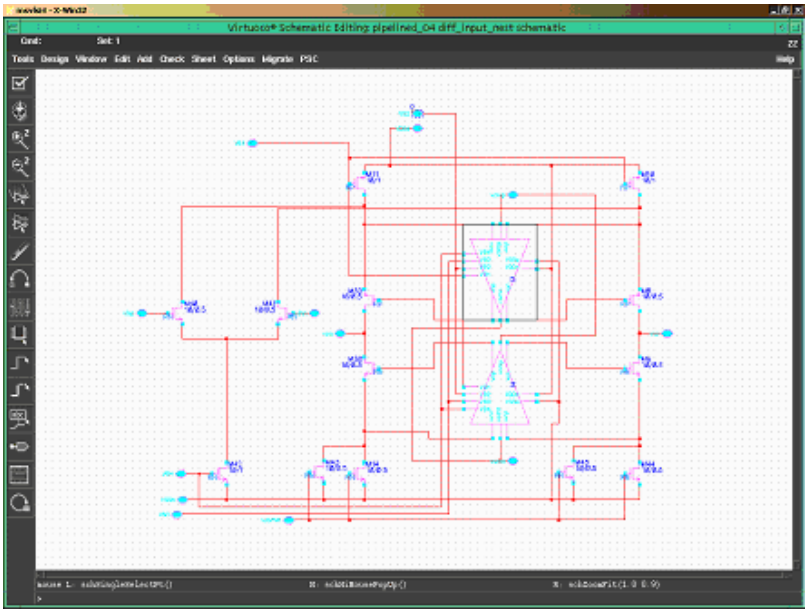


Figure 5.19 The schematic of the nested cascode OTA.

The boosted stage is shown in Figure 5.20. The right half part is the common mode feedback circuit.

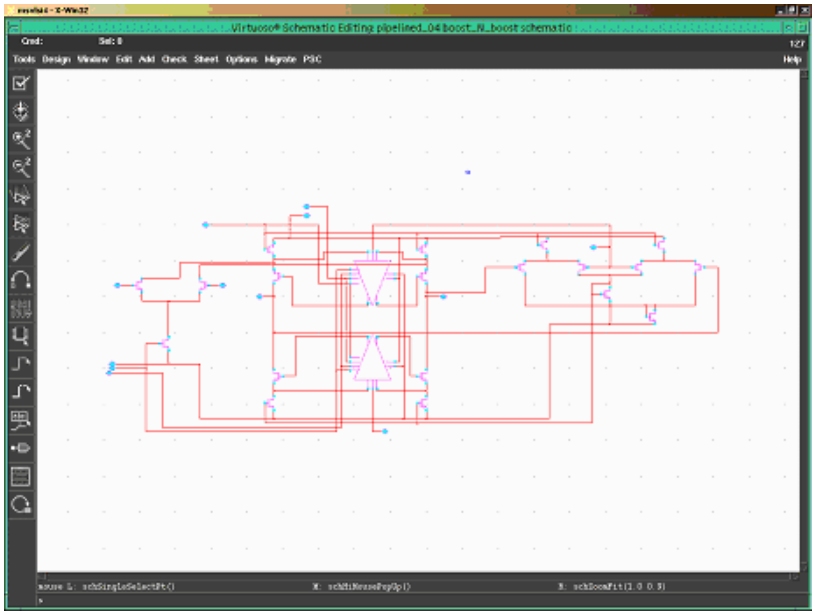


Figure 5.20 The schematic of the boosted stage.

The frequency response of the open loop OTA is shown in Figure 5.21, in which the DC gain is 107dB. The settling time of the gain stage is 18ns for 0.01% settling error.

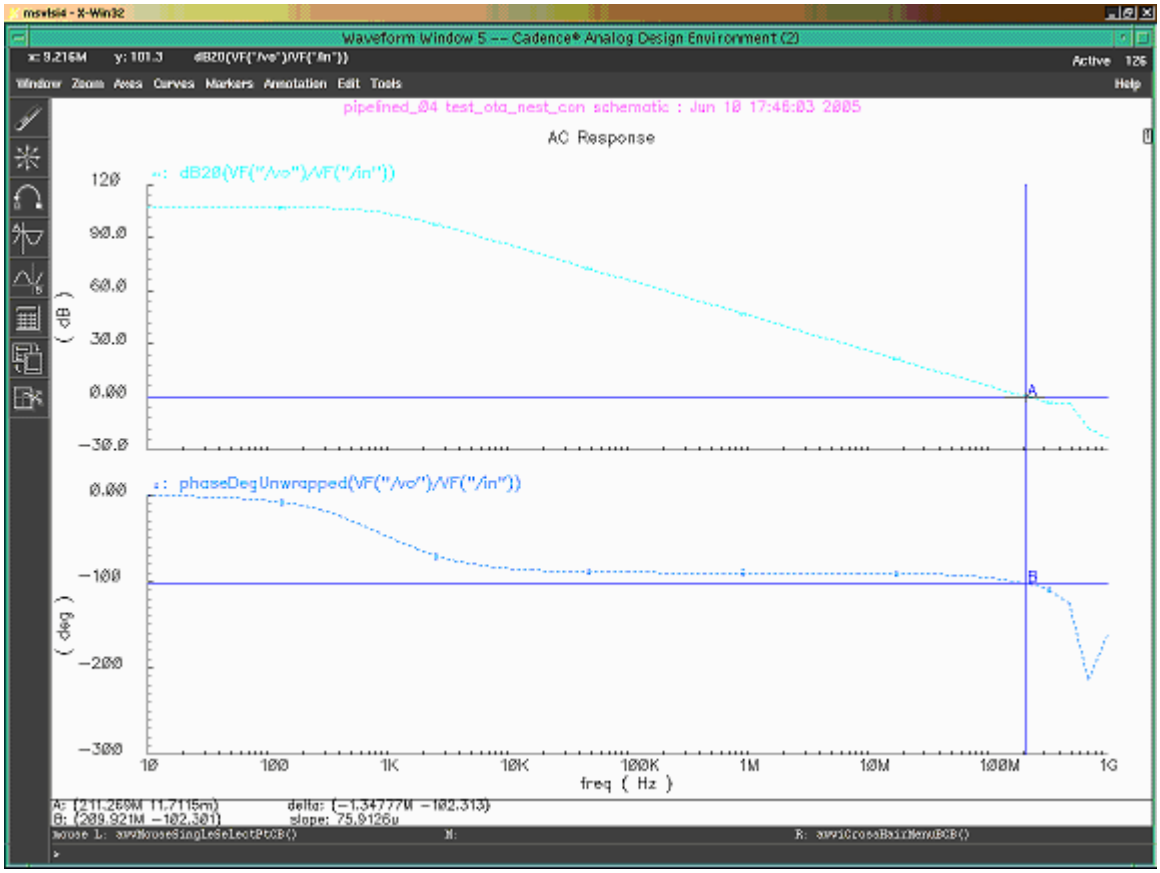


Figure 5.21 Frequency response of the open loop OTA, with DC gain = 107dB, Phase margin = 78º, GBP = 211MHz.

5.5.4 boot-strap clock generator

In conventional CMOS process, threshold voltage V_T is around 0.7 to prevent leakage current. When supply voltage scales below 2V, or even 1V, on-resistance of the switches is a problem resulting in limited charge transfer time constant in SC application, which can be shown graphically in Figure 5.22, even if in the deep sub-micro process the threshold voltage would be 0.3V as discussed in Chapter 3.

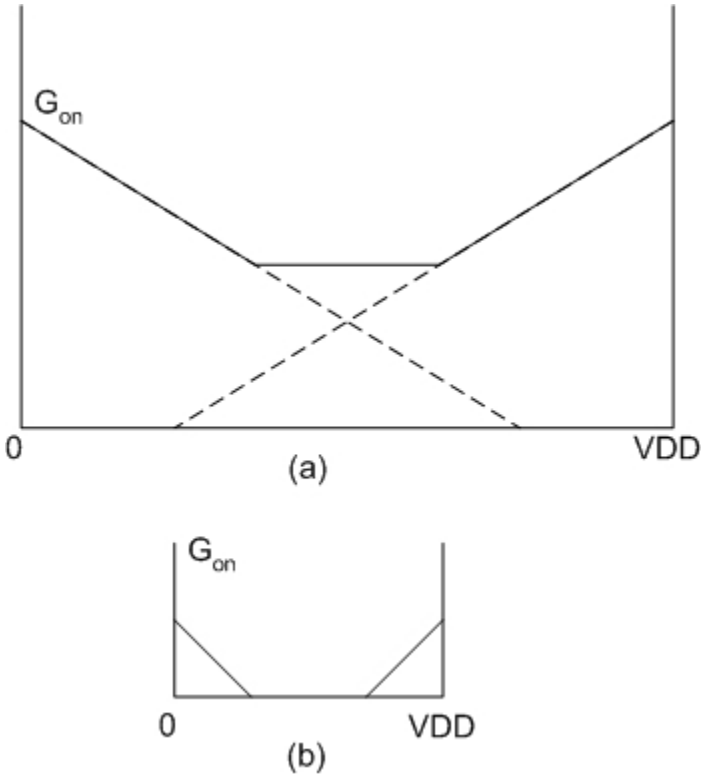


Figure 5.22 Switch on-conductance (a) under high voltage supply; (b) under low voltage supply.

In Figure 5.22, the dashed lines are the PMOS and NMOS on-conductance respectively. The solid line is the total on-conductance. When the voltage supply is less than the sum of the threshold voltage of PMOS and NMOS, the switch will not completely turn on as shown in Figure 5.22 (b).

Usually, there are three ways to solve this problem. One is using the low threshold voltage devices [66]. However, it will induce larger leakage currents and limit the resolution of the ADC (refer to Chapter 1). Secondly, one is using a switched OTA or Opamp reset switch[60, 67]. However, the switching time of the OTA limits its usage in the high speed applications. An Opamp reset switch overcomes the long setting time problems of a switched OTA by keeping the OTA active and on at all times. An

additional power penalty is paid by adding two more OTAs for each stage. The third is using a boot-strapping clock generator [61]. Although it has an implied a long term stability problem resulting from gate oxide stressing (This is s strongly process dependent), the adoption of high κ material in the gate may alleviate this problem.

The schematic of the boot-strapping clock generator is shown in Figure 5.23. The simulation result of boot strapping clock versus input signal is shown in Figure 5.24.

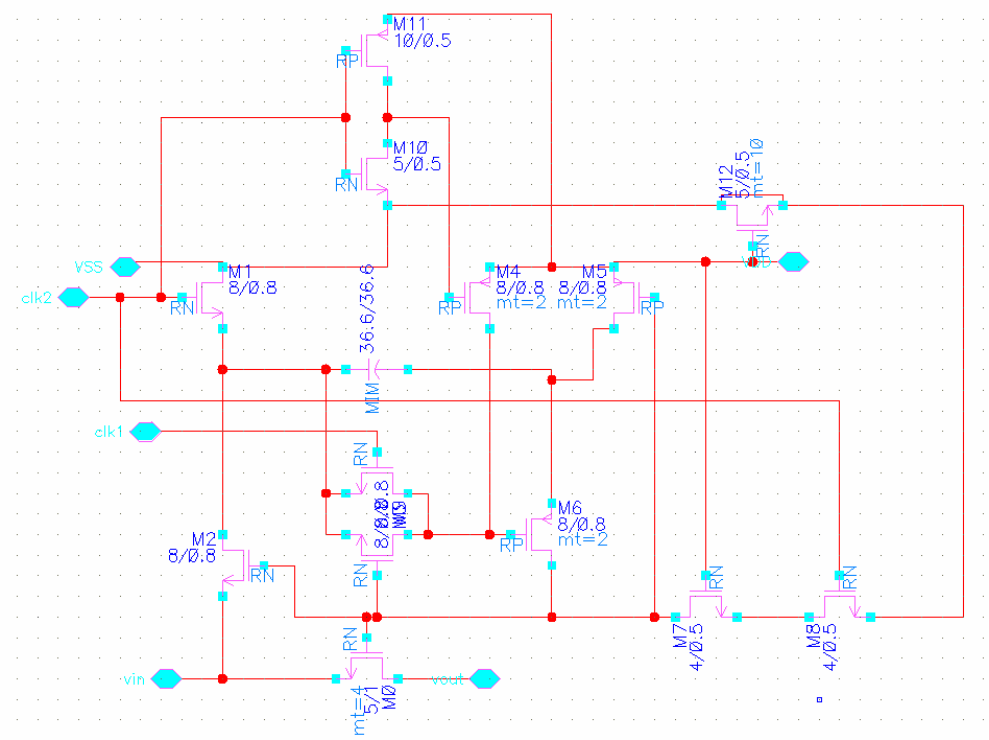


Figure 5.23 The schematic of boot-strapping circuit.

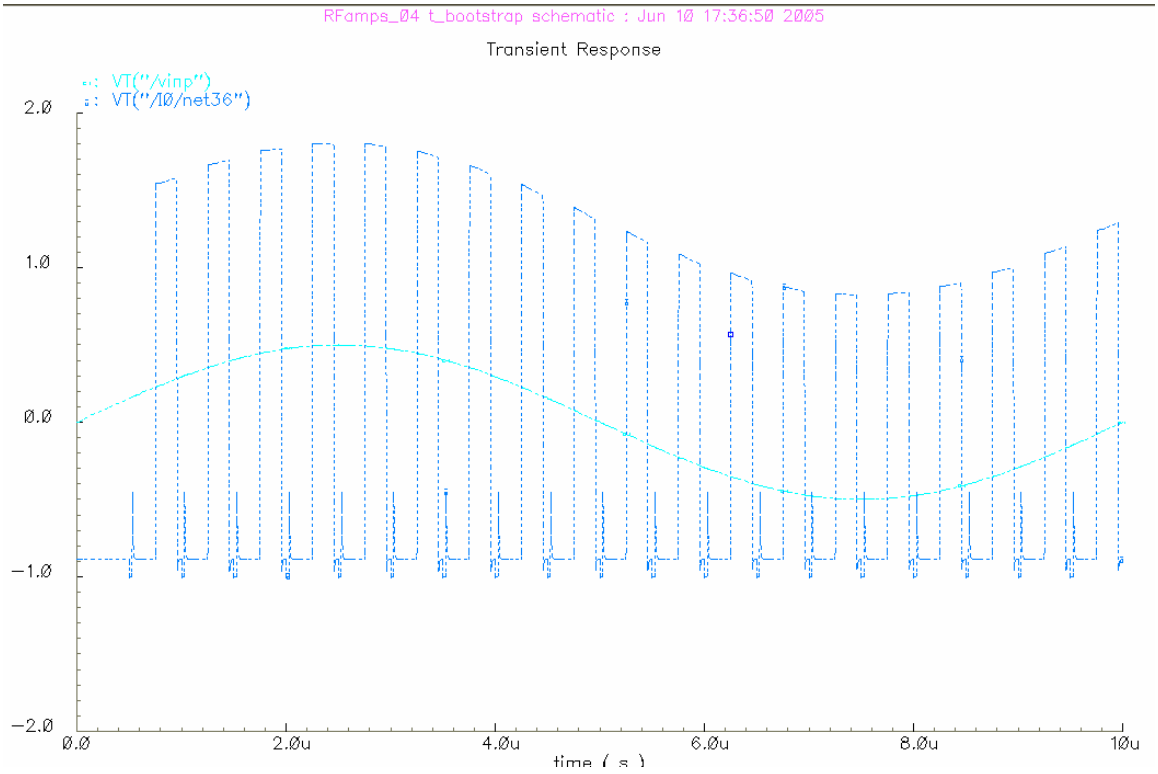


Figure 5.24 Simulation of boot-strapping clock versus input signal.

5.5.5 Comparator

The sub-ADC is a 2-bit quantizer, whose output and input characteristic is shown in Figure 5.25. It consists of three dynamic comparators shown in Figure 5.26. During phase 2, voltage V_{ref}-V_{cm} is sampled on the capacitor C. During phase 1, voltage V_{in}-(V_{ref}-V_{cm})/2 is fed to the amplifier input. V_{cm} is the common mode voltage used to raise input signal level. As shown in the Figure 5.27, a regenerative amplifier is used to make the comparison. A low offset comparator is not required here. First, the digital correction logic circuit can detect the error and correct it as long as the offset voltage is less than ±¼ of reference voltage (5.35). Second, the low resolution of the sub-ADC does not require low offset comparators. The offset voltage can be ±¼ of reference voltage without degradation of the overall performance. In our case, the reference voltage is ±1V, which

means the offset voltage should be less than ±250mV. The offset voltage is determined by

$$V_{off} = \Delta V_T + \left(\frac{\Delta Kp}{Kp}\right)\cdot\left(V_{gs} - V_T\right) + \left(\frac{\Delta W}{W}\right)\cdot\left(V_{gs} - V_T\right) + \left(\frac{\Delta L}{L}\right)\cdot\left(V_{gs} - V_T\right) \quad (5.36)$$

where ΔV_T is the threshold voltage mismatch, ΔKp is the mismatch of transconductance parameter, ΔW is the mismatch of the width, ΔL is the mismatch of the length.

As a design example based on measured data from the Peregrine process [68], ΔV_T is 10mV/um^2, $\Delta Kp/Kp$ is 2.5%, ΔW is 0.01um and ΔL is 0.01um. Offset voltage requirement can be easily satisfied.

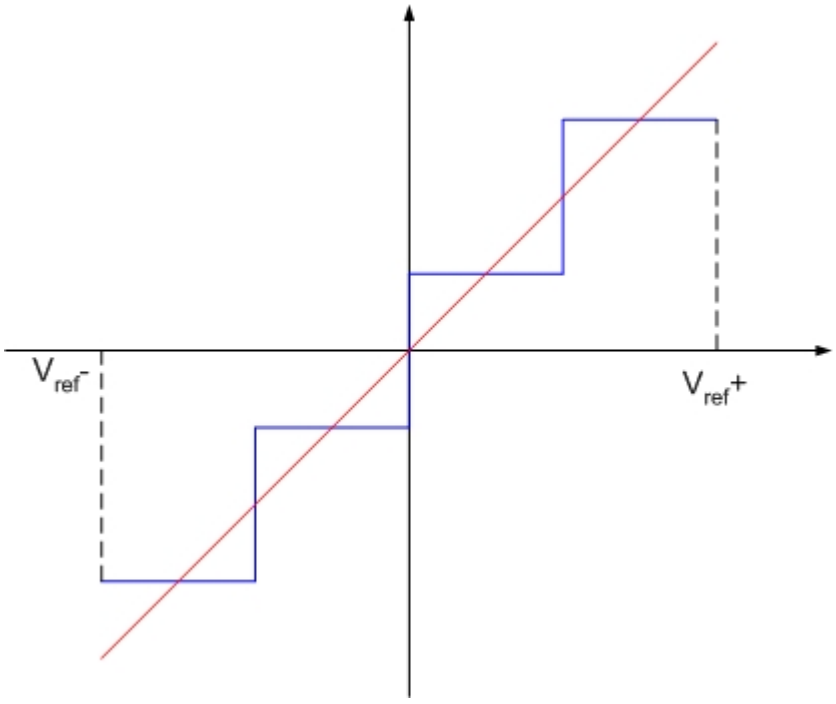


Figure 5.25 Input-output transfer curve of 2-bit quantizer.

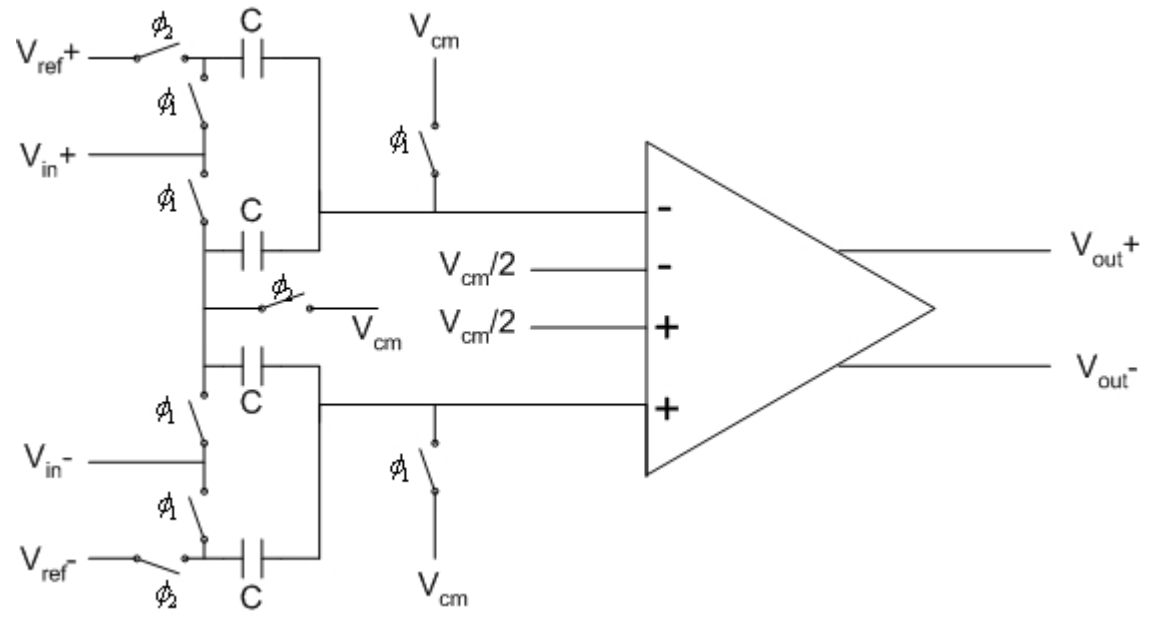


Figure 5.26 Conceptual comparator.

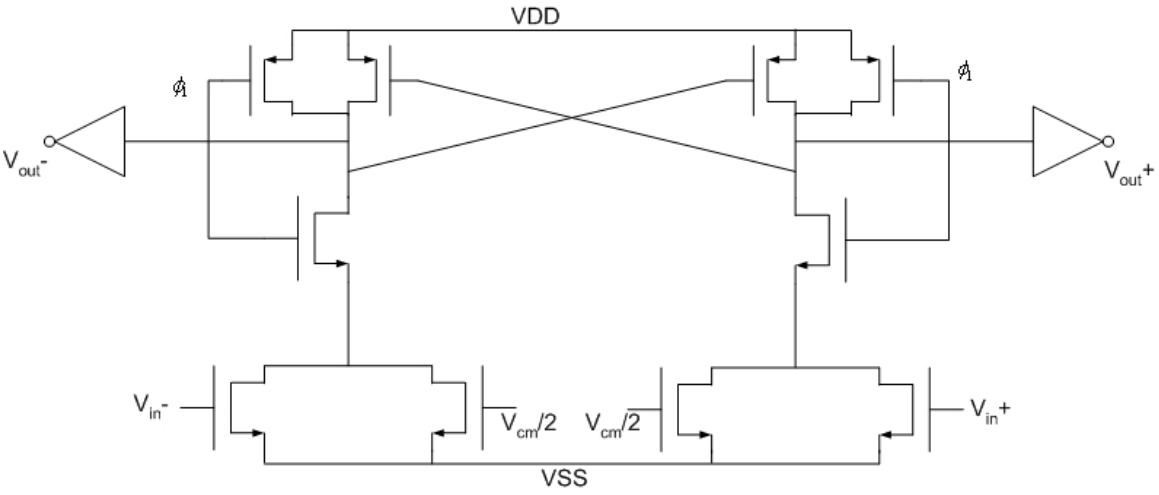


Figure 5.27 The schematic of the regenerative amplifier.

5.6 Summary

The schematic of the nested cascade OTA, MDAC, comparator, and bootstrapped clock circuits along with the require design and circuit simulation are complete. DC, AC,

and transient simulations have been done on these components. All block have been simulated with the exception of the S/H. The simulation results have all met the design specifications demanded of the Matlab model. This verifies the feasibility of a low voltage (2V), 11-bit, 20Msample/s, $1V_{FS}$ pipelined ADC based on the Peregrine 0.5um SOI CMOS technology. The projected power dissipation of the optimized ADC is 62mW.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This research explored the impact of the CMOS scaling on the analog design. Key research contributions and results are summarized below:

- Demonstrate the minimum threshold voltage V_T for the analog design is larger than 300mV; the minimum V_{DSsatQ} is approximate 100mV; and the minimum supply voltage is around 1.1V.
- Power settling product efficiency (Dynamic range / (Power · Settling time) is proposed as a guide to choose an OTA in the approach to analog design in the presents of device scaling.
- All the common OTA topologies are analyzed in the close loop form and compared by using the power settling product efficiency as criteria as the supply voltage scales. Nested gain boost topology is demonstrated as an optimum topology especially operation in the weak inversion and any time the supply voltage to V_{DSsat} ratio is greater than approximately 11.

- Designed a low-voltage, low-power pipelined ADC as a vehicle to demonstrate the utility of power settling product efficiency and closed loop analysis. New stage to stage scaling factor for pipelined ADCs was developed based on real stages number and close loop analysis. Spice and Simulink simulations were combined to demonstrate its feasibility in the Peregrine SOS process.

6.2 Recommended future work

This work served as a proof of concept to demonstrate that operation in the weak-inversion or velocity saturation is only choice for the future analogy design. The next step is to integrate a demonstration circuit on the chip and validate its power efficiency. Issues such as threshold mismatch may be a much greater problem in future circuit design.

Capacitor mismatch is one of the major problems in the high resolution ADC design. Novel structure or correction algorithm would be a good topic in the future work.

Switch-on resistance is another concern in the future high speed SC circuit design. Boot-strap clock structure may be not suitable for the future ultra-thin gate oxide transistors. Issues considering reliability, the type of the switch-OTA structure need further the resolution.

REFERENCE

[1] Asai, S. and Wada, Y., "Technology challenges for Integration Near and Below 0.1um," *Proc. IEEE*, vol. 85, Apr. 1997.

[2] Komurasaki, H., Sano, T., Heima, T., Yamamoto, K., Wakada, H., Yasui, I., Ono, M., Miwa, T., Sato, H., Miki, T., and Kato, N., "A 1.8-V operation RF CMOS transceiver for 2.4-GHz-band GFSK applications," *Solid-State Circuits, IEEE J.*, vol. 38, May 2003.

[3] Deiss, A. and Huang, Q., "A low-power 200-MHz receiver for wireless hearing aid devices," *Solid-State Circuits, IEEE J.*, vol. 38, May 2003.

[4] Floyd, B. A., Shi, L., Taur, Y., Lagnado, I., and Kenneth, K. O., "A 23.8-GHz SOI CMOS Tuned Amplifier," *Microwave Theory and Techniques, IEEE trans.*, vol. 50, Sept. 2002.

[5] Wann, C. and Su, L., "RF perspective of sub-tenth-micron CMOS," *Proc. IEEE ISSC Conf.*, 1998.

[6] Viviani, A., Flandre, D., and Jespers, P., "High-temperature sigma-delta modulator in thin-film fully-depleted SOI technology," *Electronics Letters, IEEE*, vol. 35, Apr. 1999.

[7] Hisamoto, D., Tanaka, S., Tanimoto, T., and Kimura, S., "Suspended SOI structure for advanced 0.1-um CMOS RF devices," *IEEE Trans. on Electron Devices*, vol. 45, May 1998.

[8] Chang, L., Choi, Y.-K., Kedzierski, J., Lindert, N., Xuan, P., Bokor, J., Hu, C., and King, T.-J., "Moore's law lives on [CMOS transistors]," *IEEE Circuits and Devices Magazine*, vol. 19, Jan 2003.

[9] Colinge, J. P., *Silicon-on-Insulator Technology: Materials to VLSI*: Kluwer, 1997.

[10] Thompson, S., Pachan, P., and Bohr, M., "MOS Scaling: Transistor Challenges for the 21st Century," *Intel Technology J.*, 1998.

[11] Dennard, R. H., Gaensslen, F. H., Ridiout, V. L., Bassous, E., and LeBlanc, A. R., "Design of ion-implanted MOSFET's with very small physical dimensions," *Solid-State Circuits, IEEE J.*, vol. 9, Oct. 1974.

[12] "ITRS roadmap," http://public.itrs.net/

[13] Taur, Y., Buchanan, D. A., chen, W., Frank, D. J., Ismail, K. E., Lo, S.-H., Sai-Halasz, G. A., Viswanathan, R. G., C. Wann, H.-J., Wind, S. J., and Wong, H.-S., "CMOS Scaling into the Nanometer Regime," *Proceedings of the IEEE*, vol. 85, April 1997.

[14] Hu, C., "Future CMOS Scaling and Reliability," *Proce. IEEE*, vol. 81, May 1993.

[15] Pelgrom, M., "Transistor matching in analog CMOS applications," *IEDM*, 1998.

[16] Wong, H.-S. P., "Discrete random dopant distribution effects in nanometer-scale MOSFETs," *Mircroelectronics Reliability*, 1998.

[17] Trivedi, V. P., Forssum, J. G., and Vandooren, A., "Non-Classical CMOS Device Design," *2003 SOI Conference, IEEE international*, 29 Sept.-2 Oct., 2003.

[18] Wong, H., "Nanoscale CMOS," *Proce. IEEE*, vol. 87, Apr. 1999.

[19] Yang, N., "A Comparative Study of Gate Direct Tunneling and Drain Leakage Currents in N-MOSFET's with Sub-2nm Gate Oxides," *IEEE Trans. on Electron Devices*, vol. 47, Aug. 2000.

[20] Mukhopadhyay, S., "Gate Leakage Reduction for Scaled Devices Using Transistor Stacking," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 11, Aug. 2003.
[21] Choi, C.-H., Nam, K.-Y., Yu, Z., and Dutton, R. W., "Impact of gate direct tunneling current on circuit performance: a simulation study," *IEEE Trans. on Electron Devices*, vol. 48, Dec. 2001.
[22] S.-H., L., Buchanan, D. A., Taur, Y., and Wang, W., "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's," *Electron Device Letters, IEEE*, vol. 18, May 1997.
[23] Chen, M., "Back-Gate Bias Enhanced Band-to-Band Tunneling Leakage in Scaled MOSFET's," *Electron Device Letters, IEEE*, vol. 19, Apr. 1998.
[24] Chen, J., "Subbreakdown Drain Leakage Current in MOSFET," *Electron Device Letters, IEEE*, vol. EDL-8, Nov. 1987.
[25] Fair, R. B. and Wivell, H. W., "Zener and Avalanche Breakdown in As-Implanted Low-Voltage Si n-p Junctions," *IEEE Trans. on Electron Devices*, vol. ED-23, May 1976.
[26] Mead, C. A., "Scaling of MOS technology to submicrometer feature sizes," *J. VLSI Signal Processing*, vol. 8, pp. 9-25, 1994.
[27] Fu, K.-Y. and Tsang, Y. L., "On the punchthrough phenomenon in submicron MOS transistor," *IEEE Trans. on Electron Devices*, vol. 44, 1997.
[28] Henson, W. K., "Analysis of Leakage Currents and Impact on Off-State Power Consumption for CMOS Technology in the 100-nm Regime," *IEEE Trans. on Electron Devices*, vol. 47, July 2000.
[29] Taur, Y. and Ning, T. H., *Fundamentals of Modern VLSI Devices*: Cambridge, 1998.
[30] Lee, T., *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, U. K.: Cambridge University Press, 1998.
[31] Tsividis, Y., *Operation and Modeling of the MOS Transistor*. New York: McGraw-Hill, 1999.
[32] Scholten, A. J. and Tromp, H. J., "Accurate thermal noise model for deep-submicron CMOS," *Proc. Internaltional Electron Devices Meeting*, Dec., 1999.
[33] Knoblinger, G., Klein, P., and Tiebout, H., "A new model for thermal channel noise of deep-submicron MOSFETs and its application in RF-CMOS design," *Solid-State Circuits, IEEE J.*, vol. 36, May 2001.
[34] Goo, J., Choi, C., Abramo, A., and Ahn, J., "Physical origin of the excess thermal noise in short channel MOSFETs," *Electron Device Letters, IEEE*, vol. 22, Feb. 2001.
[35] Enz, C. C., "An MOS Transistor Model for RF IC Design Valid in All Regions of Operation," *IEEE Trans. on Microwave theory and techniques*, vol. 50, Jan. 2002.
[36] Abidi, A., "High-frequency noise measurements on FET's with small dimensions," *IEEE Trans. on Electron Devices*, vol. ED-33, Nov. 1986.
[37] Enz, C. and Temes, G. C., "Circuit Techniques for Reducing the Effects of Op-Amp Imperfections: Autozeroing, Correlated Double Sampling, and Chopper Stabilization," *Proc. IEEE*, vol. 84, Nov. 1996.

[38] Chew, K. W. and Yeo, K. S., "Impact of 0.25um dual gate oxide thickness CMOS process on flicker noise performance of multifingered deep-submicron MOS devices," *IEE Proc. Circuits Devices Syst.*, vol. 148, Dec. 2001.
[39] Hung, K. K. and Ko, P. K., "A Unified Model for the Flicker Noise in Metal-Oxide-Semiconductor Field-Effect Transistors," *IEEE Trans. on Electron Devices*, vol. 37, Mar. 1990.
[40] Simoen, E., "Empirical Model for the Low-Frequency Noise of Hot-Carrier Degraded Submicron LDD MOSFET's," *Electron Device Letters, IEEE*, vol. 18, Oct. 1997.
[41] Celik-Butler, Z., "Low-frequency noise in deep-submicron metal-oxide-semiconductor field-effect transistors," *IEE Proc. Circuits Devices Syst.*, vol. 149, Feb. 2002.
[42] Tseng, Y.-C., "Device Design Methodology to Optimize Low-Frequency Noise in Advanced SOI CMOS Technology for RF IC's," *IEDM*, 1998.
[43] Lee, J.-S., Choi, Y.-K., Ha, D., Balasubramanian, S., King, T.-J., and Bokor, J., "Hydrogen annealing effect on DC and Low-frequency noise characteristics in CMOS FinFETs," *Electron Device Letters, IEEE*, vol. 24, pp. 186-188, March 2003.
[44] Vittoz, E. A. and Tsividis, Y., "Frequency-Dynamic Range-Power," in *Trade-offs in Analog Cirucuit Design*, C. Toumazou, B. Gilbert, and G. Moschytz, Eds. Boston, MA: Kluwer, 2002.
[45] Castello, R. and Gray, P., "Performance Limitations in Switched-Capacitor Filters," *IEEE Trans. on Circuits and Systems*, vol. CAS-32, Sept. 1985.
[46] S. Wong and Salama, C. A. T., "Impact of Scaling on MOS Analog Performance," *Solid-State Circuits, IEEE J.*, vol. 18, Feb. 1983.
[47] J. Arias, V. Boccuzzi, L. Quintanilla, L. Enriquez, D. Bisbal, M. Banu, and Barbolla, J., "Low-Power Pipeline ADC for Wireless LANs," *Solid-State Circuits, IEEE J.*, vol. 39, Aug. 2004.
[48] K. Bult and Geelen, G. G. M., "A Fast-Settling CMOS Op Amp for SC Circuits with 90-dB DC Gain," *Solid-State Circuits, IEEE J.*, vol. 25, Dec. 1990.
[49] Y. Chiu, P. R. Gray, and Nikolic, B., "A 1.8-V, 14-b, 10-MS/s Pipelined ADC in 0.18-um CMOS with 99-dB SFDR," *2004 IEEE International Solid-State Circuits Conference*, Feb., 2004.
[50] P. Hurst, S. Lewis, J. Keane, F. Aram, and Dyer, K., "Miller Compensation Using Current Buffers in Fully Differential CMOS Two-Stage Operational Amplifiers," *Circuits and Systems I, IEEE trans.*, pp. 275-285, Feb. 2004.
[51] F. You, S. H. K. Embabi, and Sanchez-Sinencio, E., "Multistage Amplifier Topologies with Nested Gm-C Compensation," *Solid-State Circuits, IEEE J.*, vol. 32, Dec. 1997.
[52] H.-T. Ng, R. M. Ziazadeh, and Allstot, D. J., "A Multistage Amplifier Technique with Embedded Frequency Compensation," *Solid-State Circuits, IEEE J.*, vol. 34, Mar. 1999.
[53] Abo, A. M., *Design for Reliability of Low-Voltage, Switched-Capacitor Circuits*, PhD thesis, University of California, Berkeley, 1999.
[54] Cline, D. W., *Noise, Speed, and Power Trade-offs in Pipelined Analog to Digital Converters*, PhD thesis, University of California, Berkeley, 1995.

[55] Howard C. Yang and Allstot, D. J., "Considerations for Fast Settling Operational Amplifiers," *IEEE Transactions on Circuits and Systems*, vol. 37, pp. 326-334, March 1990.
[56] D.Johns and Martin, K., *Analog Integrated Circuit Design*: John Wiley, 1997.
[57] R. Eschauzier and Huijsing, J., *Frequency Compensation Techniques for Low-Power Operational Amplifiers*. Boston: Kluwer, 1995.
[58] Chiu, Y., Gray, P. R., and Nikolic, B., "A 1.8-V, 14-b, 10-MS/s Pipelined ADC in 0.18-um CMOS with 99-dB SFDR," *2004 IEEE Internaltional Solid-State Circuits Conference*, February, 2004.
[59] Chen, H.-S., Song, B.-S., and Bacrania, K., "A 14-b 20-Msamples/s CMOS pipelined ADC," *Solid-State Circuits, IEEE J.*, vol. 36, June 2001.
[60] Chang, D.-Y. and Moon, U.-K., "A 1.4-V 10-bit 25-MS/s pipelined ADC using opamp-reset switching technique," *Solid-State Circuits, IEEE J.*, vol. 38, pp. 1401-1404, Aug 2003.
[61] Abo, A. M. and Gray, P. R., "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipelined analog-to-digital converter," *Solid-State Circuits, IEEE J.*, vol. 34, May 1999.
[62] Bastos, J., Steyaert, M., Pergoot, A., and Sansen, W., "Mismatch Characterization of Submicron MOS Transistors," *Analog Integrated Circuits and Signal Processing*, vol. 12, pp. 95-106, Feb. 1997.
[63] Shyu, J.-B., Temes, G. C., and Yao, K., "Random Errors in MOS Capacitors," *Solid-State Circuits, IEEE J.*, vol. SC-17, Dec. 1982.
[64] Cline, D. W. and Gray, P. R., "A power optimized 13-b 5Msamples/s pipelined analog to digital converter in 1.2um CMOS," *Solid-State Circuits, IEEE J.*, vol. 31, March 1996.
[65] Brigati, S., Francesconi, F., Malcovati, P., Tonietto, D., Baschirotto, A., and Aloberti, F., "Modeling Sigma-Delta Modulator Non-idealities in Simulink," *Proceedings of the 1999 IEEE International Symposium on.*, 1999.
[66] Adachi, T., Ishinawa, A., Barlow, A., and Takasuka, K., "A 1.4V switched capacitor filter," *Proc. CICC*, May, 1990.
[67] Crols, J. and Steyaert, M., "Switched-Opamp: An Approach to Realize Full CMOS Switched-Capacitor Circuits at Very Low Power Supply Voltages," *Solid-State Circuits, IEEE J.*, vol. 29, Aug. 1994.
[68] Kayathi, N., *Extended Models of Silicon On Sapphire transistors for Analog and Digital Design at elevated temperatures (200°C)*, Master thesis, Oklahoma State University, 2005.

VITA

Xunyu Zhu

Candidate for the Degree of

Doctor of Philosophy

Thesis: IMPACTS OF CMOS SCALING ON THE ANALOG DESIGN

Major Field: Electrical Engineering.

Biographical Information:

Education: Received Bachelor of Engineering degree and Maser of Engineering in Electrical and Computer Engineering department from University of Science and Technology Beijing, Beijing, China P.R. in July, 1995 and March, 1998 respectively. Completed the requirements for the Doctor of Philosophy with a major in Electrical and Computer Engineering at Oklahoma State University in July 2005.

Experience: Employed by Oklahoma State University, Department of Electrical and Computer Engineering as a Graduate Research Assistant from August 1999 to present, including research work for Space and Naval Warfare Systems Command (SPAWAR), San Diego, California.