

LOW POWER DESIGN METHODOLOGIES IN ANALOG BLOCKS  
OF CMOS IMAGE SENSORS

WEI GAO

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE  
STUDIES  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMPUTER SCIENCE AND ENGINEERING  
DEPARTMENT  
YORK UNIVERSITY  
TORONTO, ONTARIO  
MARCH 2011

**LOW POWER DESIGN METHODOLOGIES  
IN ANALOG BLOCKS OF CMOS IMAGE  
SENSORS**

by **Wei Gao**

a dissertation submitted to the Faculty of Graduate Studies of York University in partial fulfilment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

© 2011

Permission has been granted to: a) YORK UNIVERSITY LIBRARIES to lend or sell copies of this dissertation in paper, microform or electronic formats, and b) LIBRARY AND ARCHIVES CANADA to reproduce, lend, distribute, or sell copies of this dissertation anywhere in the world in microform, paper or electronic formats *and* to authorise or procure the reproduction, loan, distribution or sale of copies of this dissertation anywhere in the world in microform, paper or electronic formats.

The author reserves other publication rights, and neither the dissertation nor extensive extracts for it may be printed or otherwise reproduced without the author's written permission.

# LOW POWER DESIGN METHODOLOGIES IN ANALOG BLOCKS OF CMOS IMAGE SENSORS

by **Wei Gao**

By virtue of submitting this document electronically, the author certifies that this is a true electronic equivalent of the copy of the dissertation approved by York University for the award of the degree. No alteration of the content has occurred and if there are any minor variations in formatting, they are as a result of the conversion to Adobe Acrobat format (or similar software application).

Examination Committee Members:

1. Prof. Richard Hornsey
2. Prof. Wolfgang Stuerzlinger
3. Prof. James Elder
4. Prof. Regina Lee
5. Prof. Minas Spetsakis
6. Prof. Orly Yadid-Pecht

## **Abstract**

Complementary metal-oxide-semiconductor (CMOS) active pixel image sensors (APS) have been widely used in many portable devices, such as digital cameras, mobile phones, laptop computers, and bio-sensors. Power consumption is a critical issue in portable applications due to the limitation of the battery lifetime. This thesis describes research on low-power design for CMOS image sensors, in which low-power design techniques at the circuit level are addressed. Power reduction in the analog blocks of these mixed-signal devices is much more difficult than the digital blocks, because simple power supply scaling can significantly affect their performance. Two major issues related to low-power design in the analog blocks of CMOS image sensors using modern sub-micron processing technologies are studied in this research: maintaining acceptable dynamic range at lowered power supplies, and power minimization for a required performance.

Dynamic range enhancement is a practical technique to overcome the output swing reduction caused by lowered power supplies. As an example of an application

requiring a high dynamic range, a CMOS image sensor using a new predictive integration technique was designed for laser rangefinding. This design achieved  $\sim 100$ -dB dynamic range when the power supply was kept as low as 1.8 V.

Low-power design is facilitated by using automatic analog synthesis power optimization, with which a power optimal design can be achieved efficiently for a required performance. Here, a new sub-space-based posynomial modeling method is shown to increase the MOS transistor modeling performance for sub-micron technologies. Geometric programming (GP) based power optimization using sub-space-based posynomial models has been demonstrated to be an efficient and reliable low-power design technique in analog circuits. New techniques for solving speed-driven problems using GP in basic analog building blocks of CMOS image sensors were also shown. The sub-space modeling based GP synthesis was successfully applied to three major analog blocks (in-pixel source follower, correlated double sampling circuit, and analog-to-digital converter) of CMOS image sensors to achieve power-optimal designs. Synthesized circuits for these three blocks were verified with HSPICE simulations, and op-amps for a correlated double sampling circuit were verified experimentally with a fabricated chip. Both simulation and test results show that sub-space modeling based GP synthesis is an efficient and reliable power reduction technology in sub-micron technologies.

## Acknowledgements

I would like to express my deepest gratitude towards my supervisor, Dr. Richard Hornsey, for giving me opportunities in doing research on CMOS image sensors in a disciplined but relaxed environment, for his valuable guidance and support in solving difficult problems and coping with critical situations, for his understanding in my non-academic roles as a parent, and for his patience and encouragement throughout my doctoral research.

I am greatly appreciative of Dr. Paul Thomas at Topaz Technology Inc. for his support in my NSERC Industrial Postgraduate Scholarship, and his precious guidance in my early research projects.

I am very thankful to Mr. Angelo Beraldin at NRC Canada for providing me with a chance to acquire and practice my design and test skills for CMOS image sensors in an industrial environment.

I would like to acknowledge all my colleagues at the VISOR Lab at York University for their helpful advice in chip design and testing, and jokes to relieve my tensions

at tape-out. In particular: Edward Shen, for sharing his knowledge in CMOS circuit design, for his encouragement to pursue academic career, and for helping to tackle down Latex bugs; Winnie Wong, for taking me to her spinning class to rejuvenate my mind; Mike Liscombe, for his valuable help in making a PCB test board and proof-reading my thesis; Cyrus Minwalla, for his knowledge and help in optical setup in my chip test; Elliott Tsai, for suggestions in chip testing and interesting technical discussions. I also like to thank technical staffs of the Computer Science and Engineering department at York University for helping to solve technical problems in design tools.

I would like to thank the Canadian Microelectronics Corporation for fabrication of chips, lending test experiments, and providing technical support. I would like to acknowledge the support of NSERC in my postgraduate scholarship.

Special thanks to my loved ones: my parents and Xia's parents for encouraging and helping to take care of Helen; my husband, Xia, for his love, understanding, support, and sacrifice; My daughter, Helen, for her charming smile and eternal love.

# Table of Contents

<b>Abstract</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Table of Contents</b>	<b>viii</b>
<b>List of Tables</b>	<b>xii</b>
<b>List of Figures</b>	<b>xiv</b>
<b>Abbreviations</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Understanding low-power design . . . . .	1
1.2 Low-power design in the CMOS APS image sensor . . . . .	6
1.3 Low-power design challenges in analog blocks of CMOS APS image sensors . . . . .	11

1.4	Contributions . . . . .	14
1.5	Thesis organization . . . . .	16
<b>2</b>	<b>Algorithms for global optimization</b>	<b>18</b>
2.1	Genetic algorithm . . . . .	19
2.1.1	Analog circuit synthesis using genetic algorithms . . . . .	21
2.2	Simulated annealing . . . . .	23
2.2.1	Analog circuit synthesis using simulated annealing . . . . .	27
2.3	Geometric programming . . . . .	28
2.3.1	Analog synthesis using geometric programming . . . . .	31
2.4	Conclusion . . . . .	35
<b>3</b>	<b>CMOS transistor modeling</b>	<b>36</b>
3.1	PWL modeling . . . . .	38
3.2	Sub-space-based max-monomial models . . . . .	48
<b>4</b>	<b>Low power design in in-pixel source follower</b>	<b>52</b>
4.1	Operation dependent low power design in in-pixel source follower . . .	52
4.2	Power optimization in in-pixel source follower . . . . .	56
4.3	Synthesis Evaluation . . . . .	60
<b>5</b>	<b>Power optimization in CDS op-amp</b>	<b>63</b>

5.1	Op-amp power optimization with geometric programming . . . . .	64
5.1.1	Two-stage Miller op-amp power optimization . . . . .	64
5.1.2	Symmetrical OTA power optimization . . . . .	74
5.2	Experimental verification of synthesized results . . . . .	82
5.3	CDS op-amp design . . . . .	87
<b>6</b>	<b>Power optimization in successive-approximation A/D converters</b>	<b>97</b>
6.1	Power optimization in a two-stage open-loop op-amp . . . . .	99
6.2	SAR ADC design . . . . .	106
<b>7</b>	<b>System optimization</b>	<b>110</b>
<b>8</b>	<b>Dynamic range enhancement for a laser rangefinding image sensor with a lowered power supply</b>	<b>112</b>
8.1	Dynamic range enhancement technologies . . . . .	114
8.2	Predictive integration . . . . .	119
8.3	Sensor architecture and operation . . . . .	124
8.4	Chip test and characterization . . . . .	130
8.5	Summary . . . . .	146
<b>9</b>	<b>Conclusion and suggestions for future work</b>	<b>149</b>
9.1	Future work . . . . .	155

9.1.1	Model . . . . .	155
9.1.2	Chip . . . . .	156
	<b>Bibliography</b>	<b>158</b>

## List of Tables

3.1	Fitting comparison among three algorithms . . . . .	47
3.2	Sub-space map for all parameters but $1/g_m$ . . . . .	49
3.3	Comparison of MRE (%) for different models . . . . .	51
4.1	Mean, maximum, and minimum relative errors in settling time model of in-pixel source follower . . . . .	57
4.2	Synthesized solutions of the source follower with the minimum bias current . . . . .	59
5.1	Two-stage op-amp models and constraints . . . . .	66
5.2	Op-amp design specifications . . . . .	74
5.3	Two-stage Op-amp design verification with HSPICE . . . . .	74
5.4	GP-based design comparisons . . . . .	76
5.5	Conventional symmetrical OTA models and constraints . . . . .	77
5.6	CMFB OTA models and constraints . . . . .	78

5.7	The performance of the optimized design of the conventional symmetrical OTA . . . . .	82
5.8	The performance of the optimized design of the CMFB OTA . . . . .	82
5.9	Static power measurements comparison . . . . .	84
5.10	AC Performance measurements comparison . . . . .	86
5.11	Slew rate measurements comparison . . . . .	86
5.12	CDS op-amp design verification with HSPICE . . . . .	95
6.1	Performance measures of the optimized two-stage open-loop comparator	105
7.1	Bias current comparison between a referenced and an optimized design	111
8.1	Power consumption with 3500 frames/s at 1.8V power supply . . . . .	147
8.2	Specifications and performance of the laser rangefinding chip . . . . .	148

## List of Figures

1.1	Design efforts (a) and power reduction (b) along with design levels . . .	2
1.2	A typical architecture of voltage-mode CMOS APS image sensors . . .	11
2.1	Generic algorithm flow of a genetic optimization process . . . . .	20
2.2	Flow chart of the simulated annealing algorithm [Maji 06] . . . . .	24
2.3	A single stage common source amplifier . . . . .	32
3.1	Non-convex fitting case . . . . .	39
3.2	Convergency distribution . . . . .	43
3.3	Fitting results of four constructed functions . . . . .	44
3.4	Effect of trial number on fitting accuracy . . . . .	45
3.5	Effect of partition on fitting accuracy . . . . .	46
3.6	$g_m$ versus $V_{gs}$ for an NMOS transistor in 0.18 $\mu m$ technology with $L = 1.02\mu m$ , $W : 0.4 \sim 100\mu m$ , $V_{ds} : 0.2 \sim 1.8V$ , working in (a) weak inversion (b)moderate inversion (c)strong inversion. . . . .	48

4.1	Settling time of the source follower during charging and discharging . . . . .	53
4.2	Column-wise signal readout in a CMOS image sensor . . . . .	54
4.3	Modified column-wise signal readout . . . . .	55
4.4	(a) ideal source follower (b) 3T source follower (c) 2T source follower . . . . .	56
4.5	Relative settling error versus bias current . . . . .	60
4.6	Source follower synthesis evaluation with six random requirements . . . . .	61
5.1	Switched-capacitor-based CDS . . . . .	63
5.2	Two-stage Miller compensated op-amp . . . . .	65
5.3	Relative performance errors in the two-stage Miller op-amp . . . . .	73
5.4	Conventional symmetrical OTA (a) and CMFB OTA (b) . . . . .	75
5.5	Relative performance errors in the CS OTA . . . . .	80
5.6	Relative performance errors in the CMFB OTA . . . . .	81
5.7	A micro-photograph of the op-amp chip . . . . .	83
5.8	Frequency response of the six studied op-amps . . . . .	85
5.9	Settling of an op-amp responding to a large step input . . . . .	88
5.10	CDS op-amp operation modes: (a) reset mode (b) data transfer mode . . . . .	88
5.11	CDS op-amp settling during reset . . . . .	89
5.12	CDS op-amp settling during data transfer . . . . .	91
5.13	A hierarchical GP algorithm for a CDS op-amp . . . . .	94

6.1	A typical SAR ADC . . . . .	97
6.2	Two-stage open-loop comparator . . . . .	99
6.3	Optimization algorithm for a two-stage open-loop comparator . . . . .	102
6.4	The normalized sum of transconductance at different bias conditions .	104
6.5	Low-power 8-b successive approximation ADC . . . . .	106
6.6	SAR ADC onversion waveform . . . . .	107
6.7	Differential non-linearity plot . . . . .	108
6.8	Integral non-linearity plot . . . . .	109
8.1	Response of a pixel in an adaptive integration scheme. . . . .	117
8.2	Response of a pixel to three illuminations for various sub-periods $T_n$ 's.	120
8.3	Effect of the reference value on the detectable light intensity. The letter "a" represents the value of $V_{reset}$ at $T = 0$ . Lines $ab$ , $ac$ , and $ad$ are the integration curves, in which signal value drops to $V_{ref2}$ at $T_1$ , $T_2$ , and $T_3$ , respectively. . . . .	123
8.4	Laser rangefinding chip architecture . . . . .	125
8.5	Schematics for (a) HDR and (b) WTA pixel arrays . . . . .	127
8.6	Timing diagram for the HDR and WTA pixel arrays . . . . .	128
8.7	Laser rangefinding chip micrograph . . . . .	131
8.8	Measured spectral responsivity . . . . .	133

8.9	Measured average spectral response . . . . .	134
8.10	Predictive integration with multiple sampling. (a) Analog output at a light intensity of 49.1 lux. (b) Analog output at a light intensity of 76.5 lux. (c) Sampling control signal . . . . .	136
8.11	Multiple sampling and single integration . . . . .	137
8.12	Photoconversion characteristics for single and predictive integrations.	138
8.13	Conversion slope as a function of the incident light intensity. . . . .	139
8.14	Laser spots captured with the single and predictive integration schemes.	142
8.15	Laser spot centroid detection setup. . . . .	143
8.16	Laser spots measured at 15 consecutive movements. . . . .	144
8.17	Settling time at different source-follower bias currents. . . . .	145

## Abbreviations

ADC	- analog-to-digital converter
ALU	- arithmetic logic unit
APE	- analog processing element
APS	- active pixel image sensor
$A_V$	- open loop gain
CAD	- computer-aided-design
CDS	- correlated double sampling
CLS	- constrained least-squares
CMFB OTA	- common mode feedback operational transconductance amplifier
CMOS	- complementary metal-oxide-semiconductor
CS OTA	- conventional symmetrical operational transconductance amplifier
DAC	- digital-to-analog converter
DNL	- differential non-linearity
DR	- dynamic range

FPN	- fixed pattern noise
GA	- genetic algorithm
GAP	- genetic-algorithm-based posynomial modeling
GBW	- gain-bandwidth product
GP	- geometric programming
HDR	- high dynamic range
INL	- integral non-linearity
I/Os	- inputs/outputs
LSB	- least significant bit
MARE	- Maximum relative error
MNRE	- minimum relative error
MRE	- mean relative error
NCLS	- non-constrained least-squares
PM	- phase margin
PSH	- phase shift
PWL	- piecewise linear fitting
$PWL_{sub}$	- sub-space-based piecewise linear fitting
PWM	- pulse-width-modulation
RE	- relative error

RMRSE - root mean relative square error

ROI - region of interest

SA - simulated annealing

SAR - successive-approximation register

SNR - signal-to-noise ratio

SOI - silicon-on-insulator

SR - slew rate

WTA - winner-take-all

QE - quantum efficiency

1P6M - one poly-silicon and six metal

4T - four-transistor

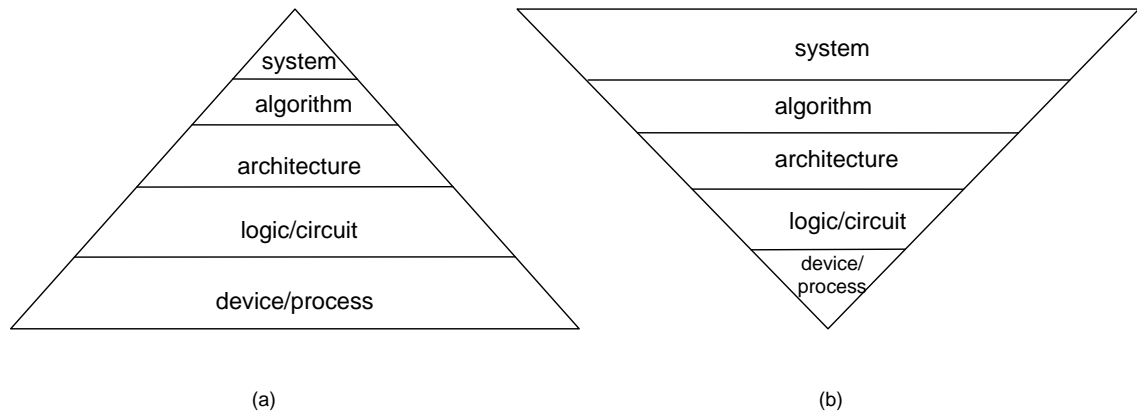
# 1 Introduction

## 1.1 Understanding low-power design

Portable electronic devices, such as cell phones, laptops, and digital cameras have benefited from the scaling of processing technologies. The use of advanced sub-micron technology can allow more functional blocks to be embedded into one chip due to smaller device sizes, and can improve the processing speed due to the reduction of inter-chip inputs/outputs (I/Os). Since these portable devices are battery powered, battery capacity is a significant constraint. To improve the performance of such a system and prolong the battery lifetime without increasing the battery capacity, low-power design technologies must be employed. Even for non-portable devices, low-power design is an important design factor because excessive power dissipation can result in increased packaging and cooling costs, and can potentially affect reliability [Pedr 02].

Low-power design technologies in electronic devices have been extensively studied

and applied at different design levels. It has been noticed that more power can be reduced at a higher design level while simultaneously less design effort is required [Chan 95,Raba 96,Beni 97,Ragh 98,Sand 02], which is illustrated in Figure 1.1. Low-



**Figure 1.1:** Design efforts (a) and power reduction (b) along with design levels

power design at the system level can achieve power-efficient functionality by using power management and power optimization technologies, and it can help to estimate the power consumption of the system by using system power models. The goal of low-power design at the algorithm level is to implement power-efficient system functions. A low-power architecture can be achieved at the architecture level by performing architecture space searching. Low-power design at the logic/circuit design level can allow designers to find low-power structures for a circuit. Power consumption can be reduced at the device/process design level by changing the structure of the electronic

devices, or modifying and inventing new device processing technologies. Usually, the design at the system, algorithm, and architecture level is referred to as a high-level design, and the design on the logic/circuit and device/process level is referred to as a low-level design.

Low-power design techniques are usually different in digital and analog systems due to the nature of their operations. In a digital integrated system, the power consumption is usually contributed by the following sources:

$$P_{avg} = P_{sw\_cap} + P_{short} + P_{leakage} + P_{static} \quad (1.1)$$

where  $P_{sw\_cap}$  is capacitive-switching power or dynamic power,  $P_{short}$  is the short circuit power,  $P_{leakage}$  is the power consumption due to leakage currents, and  $P_{static}$  is the static power consumption. The capacitive-switching power contributes most of the power consumption in a digital electronic system, which can be reduced by reducing the switching activity and parasitic capacitance at the internal nodes, and by decreasing the power supply voltage by using scaled technologies [Ragh 98, Maci 98, Beni 97]. Short circuit power can be reduced by improving the skew of clock signals, reducing the transition time of signals, and reducing the supply voltage [Wilk 08, Ko 95, Vrat 06]. The leakage current can be reduced by using oxide thickness modulation [Agar 06], gate-induced drain leakage reduction [Chen 06] gate-oxide tunneling reduction [Roy 03], and drain-induced barrier-lowering effect reduction [Roy 03].

The static power can be eliminated by replacing pseudo-nMOS logic with standard CMOS logic [Sand 02]. There are many design tools available for implementing low-power design in digital electronic systems, and they can provide very good accuracy [Pigu 05].

Low-power design in analog systems is not as straightforward as in digital systems, because analog design is perceived as less systematic and more heuristic and knowledge-intensive in nature [Giel 00]. The power consumed in an analog circuit is directly dependent on the performance of the circuit, and can not be reduced simply by decreasing the operating power supply voltage as in a digital circuit [Enz 96]. For example, the product of signal-to-noise ratio (SNR) and the bandwidth (B) in an integrator is proportional to the transconductance ( $g_m$ ) of the amplifier [Enz 96]:

$$SNR \bullet B = V_{pp}^2 \frac{g_m}{8kT} \quad (1.2)$$

where  $V_{pp}$  is the peak-peak output amplitude. In most cases, scaling the supply voltage  $V_{dd}$  by a factor  $K$  requires a proportional reduction of the signal swing  $V_{pp}$ . It is only possible to maintain the bandwidth and the SNR if the transconductance is increased by a factor of  $K^2$ , which in turn increases the power consumption. It is concluded that the reduction of the power supply may not directly lead to power reduction in analog circuit designs. The key to reducing the power in analog circuits is to reduce the quiescent bias current and increase the current efficiency. The most

popular way to reduce the quiescent bias current is to make the circuit work in the weak or moderate inversion region [Enz 96, Yan 00, Lope 05, Cama 05]. Charge-sheet-based compact models (EKV [Enz 95] and ACM [Cunh 98] models) for CMOS transistors use a single equation to describe the behavior of the transistor in all regions (weak, moderate, and strong inversion). They can provide good accuracy with hand- and computer-aided-design, and have been successfully applied in low-power analog circuit designs [Silv 02, Gao 05, Tang 09, Sham 06]. To compensate for the speed reduction induced by decreasing the quiescent bias current, an analog circuit must be designed with a high current efficiency, which can be achieved by using super-class AB design methodology in amplifiers [Lope 05]. Transistors in super-class AB designs operate in the weak or moderate inversion region with very low bias currents. By sensing the variation of the input signals, the circuit can automatically boost a large pulse current to charge the load, which can result in a high slew rate and small power penalty. A much higher power efficiency strategy is proposed by Kwon *et al.* [Kwon 05] by using class C inverters in switched-capacitor circuits. This scheme only consumes a very small quiescent current due to the weak inversion operation, and reaches approximately 100% current efficiency because only one transistor is on during each transition. Dynamic components, such as the dynamic CMOS amplifier [Wang 95] and dynamic comparator [Utha 03, Wulf 05], are considered to be another

solution to reducing the static current while keeping the speed requirement.

Design with low supply voltages in deep sub-micron technology is a critical challenge for low-power analog circuits as well, which can usually be implemented by using bulk-driven MOSFETs [Agga 09], floating-gate MOSFETs [Rami 04], self-cascode MOSFETs [Enz 96], floating voltage sources [Rami 04], flipped voltage followers [Carv 05], and current mode technology [Inou 02].

## **1.2 Low-power design in the CMOS APS image sensor**

Low-power design in CMOS APS image sensors has become very important due to increasing demands for mobile applications. CMOS APS image sensors have benefited from technology scaling by using scaled power supplies, but the power consumption in the start-of-the-art imaging systems are expected to increase or, at least, not to decrease greatly, due to increased complexity and resolution. As a special mixed signal system, the CMOS APS image sensor can adopt some existing low-power design techniques for digital and analog blocks. In this section, some specific low-power design technologies applied to CMOS APS image sensors will be reviewed.

Signal processing in a parallel form (parallelism) is an efficient way to reduce the power consumption in CMOS image sensors [Zhon 01, Taka 05, McIl 01]. By using

parallelism in digital blocks, the overall power is reduced by using low-power supply while the speed is maintained. In analog blocks, the signal settling requirement can be reduced by applying parallelism, and therefore, the bias current can be reduced. Parallelism in CMOS image sensors includes parallel readout for pixels in even and odd rows, column-wise analog-to-digital converters (ADCs), and pixel-wise ADCs.

The weak inversion design approach has been widely applied in analog blocks of CMOS image sensors to reduce the power consumption. McIlrath [McIl 01] reported a low quiescent power of 40 nW/pixel by letting transistors operating in the weak inversion region, and Bandyopadhyay *et al.* [Band 06] reported a low-power of 60 nW/pixel, including a pre-image processing operation on the pixel level.

Power management is very attractive in CMOS image sensor design for low-power consumption [Fish 05a, Sohn 03, Dege 03, Cho 03]. By allowing the sensor system to operate in active, stand-by, and sleep modes, the power consumption can be controlled by selecting its operating mode. A active mode consumes the most power, stand-by mode requires intermediate power consumption, and sleep mode uses the least power; however using sleep mode can cause a speed penalty while the circuit wakes up. In fact, this wake-up can cause extra power dissipation, and may offset the benefit from the sleep mode. Sleep mode must therefore be used carefully.

Low-power design can also benefit from silicon-on-insulator (SOI) design [Afza 03,

Sunt 07]. In the SOI design, the body effect is smaller compared to the bulk device, which can reduce the threshold voltage. The parasitic capacitance in the SOI device is also smaller than the one in the bulk device, which can reduce the column capacitance. As a result, a smaller column bias current is needed for SOI to achieve the same speed.

The current-mode-based design (photogenerated current is directly used for signal processing) is potentially a very power efficient scheme when pre-image processing functions are required at the low level. Signal processing in the current mode is very simple, for instance, summation and subtraction can easily be achieved by simple wire connections [Bous 04, Grue 10]. A vision chip, SCAMP, based on current mode operation has been developed by Dudek [Dude 05] in the University of Manchester, in which an analog equivalent arithmetic logic unit (ALU), and an analog processing element (APE), are integrated into each pixel to perform some image processing at a low level with low-power consumption and high computational performance. Since the APE sits right next to the pixel, the power consumption and I/O bandwidth requirements are reduced significantly. To perform the same edge detection, this current-mode-based chip required only 1/27 of the power of the equivalent in voltage-mode chip (photogenerated current is converted to a voltage signal for signal processing).

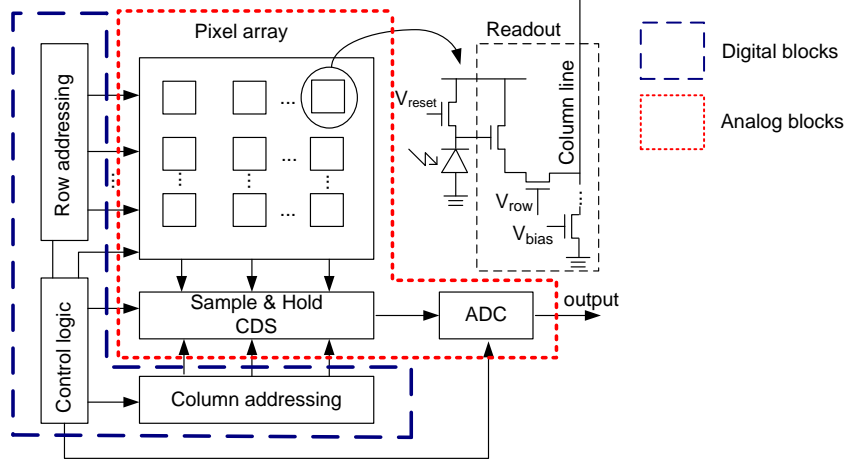
Low-voltage design is also a design trend in low-power CMOS image sensors. By reducing the power supply voltage, the power dissipated in digital parts can be reduced significantly. A low-power supply voltage can increase the difficulty in analog design, and even more so in the voltage-mode design because the output swing of the pixel will be reduced accordingly. This will degrade the dynamic range (DR) of the pixel, and add extra requirements on the following A/D converter to keep the same resolution. To overcome this problem, several technologies have been suggested. The first one is to use complementary pixel design to achieve the rail-to-rail output swing [Xu 02], in which a pair of complementary source followers is designed to remove the threshold voltage reduction. Another attractive low-voltage pixel design is the self-reset pixel design [McIl 01, Wang 06], which is sometimes called pulse-width-modulation (PWM) pixel design. In this scheme, the photo-current is directly encoded into free running pulses. By measuring the frequency of pulses, or the pulse width, the light intensity can be derived. The minimum power supply can be applied to this scheme because the voltage swing is not important. In fact, the current-mode design can allow low-voltage design as well, and is compatible with the digital process [Wang 03].

Energy harvesting technology has become a new low-power design trend in CMOS image sensors recently for ultra low-power applications. By using an energy generat-

ing component (an extra photodiode [Fish 05b] or in-pixel photodiode itself [Shi 09]) in a pixel, the pixel can be self-powered. As a result, the energy requirement for the battery can be reduced dramatically at the expense of resolution.

Other technologies involved in low-power CMOS image sensor design include global shutter design for reducing the pixel frequency [Tann 01], leakage control by using stacked transistors [Fish 05a], using a digital inverter as an analog comparator [Fish 05a], and dynamic circuit design, such as the dynamic comparator [Zhan 03] and dynamic logic for a low-power address shift register [Jin 03].

In voltage/current-mode CMOS image sensors, the analog blocks usually play a more important role compared to digital blocks. From the front end where a pixel converts light intensity to an electrical signal (voltage/current) to the back end where an analog signal is converted to digital signals, the photo-signal is amplified and processed through an analog processing chain, which is shown in Figure 1.2. The analog processing chain determines the performance of the output signal, and it should be properly designed to minimize the noise. Digital blocks in CMOS image sensors are usually used for addressing, generating control signals, and post-processing. As we have mentioned before, the low-power design in digital circuits is relatively simple compared to analog circuits, and it can be achieved by using available commercial tools. Special design efforts have to be taken in designing analog blocks of CMOS



**Figure 1.2:** A typical architecture of voltage-mode CMOS APS image sensors

APS image sensors to reduce the power consumption.

### 1.3 Low-power design challenges in analog blocks of CMOS APS image sensors

Designers are facing challenges to improve performances of CMOS APS image sensors when the deep sub-micron processing technologies are employed. A 3T (three-transistor) pixel structure will suffer from spectral response reduction with modern deep sub-micron standard technologies [Duri 10]. Both 3T and 4T (four-transistor) pixel structures will suffer from output swing reduction with reduced power supply, and performance deviation with simply scaling transistors in analog blocks. From a low-power point of view, the last two issues become more important. To minimize the

time to market, a short design turnaround time is desired. It is desirable for a design methodology to have a good scalability with respect to processing technologies.

The major issue caused by the output swing reduction in a pixel is the overall dynamic range reduction of the image sensor. Complementary pixel design [Xu 02] can only eliminate the threshold voltage influence, and it may not be good enough to achieve the required dynamic range with a scaled power supply in deep sub-micron technologies. In addition, the linearity of the pixel may be degraded due to the different transfer functions in NMOS and PMOS source followers. Dynamic range enhancement technologies become more attractive compared to other low-voltage analog design technologies, because they usually have a high efficiency in enhancing dynamic range, and have a good scalability with advancing processing technologies. Some dynamic range enhancement technologies are compatible with low-voltage design [Wang 06]. In this study, a high dynamic range enhancement technology used for low-voltage design is proposed.

When the performance specification or processing technology changes, transistors in existing analog blocks have to be resized and properly biased, and it usually takes a long time to meet specifications using manual adjustment. For low-power design, there is usually a trade-off between the power consumption and other performance figures of merit. Analog-synthesis-based power optimization is considered

to be an efficient low-power design methodology for existing architectures. During the synthesis, a power-optimal design, constrained by other performance measures, can be simply synthesized by including power consumption in the objective function. Among three popular analog synthesis methods (knowledge-based [Degr 87, Harj 89], simulation-based [Nye 88, Ocho 96], and analytical-equation-based synthesis [Wim 95, Giel 90, Gira 06, Mand 01, Hers 01]), the analytical-equation-based method is the most computationally efficient and most likely to provide the global optimization [Mand 01]. The equation-based synthesis method requires designers to convert the circuit performance measures and constraints to analytical equations in terms of the design parameters (sizes and biases of transistors), which is usually called performance modeling. The synthesis efficiency depends on the selected optimization algorithm. Among all available global optimization algorithms, the geometric programming algorithm is considered to be the most efficient, because the computational time can be reduced to literally a few seconds for solving optimization problems with hundreds of variables and thousands of constraints [Hers 01, Agui 08]. In addition, geometric programming problems can be reformulated to be convex optimization problems and, therefore, a truly global optimal solution can be guaranteed [Boyd 04]. Performance models used for geometric programming in a specific analog circuit stay the same for different processing technologies, and models for electrical parameters of

MOS transistors can be derived in the same way for different processing technologies. As a result, a geometric-programming-based low-power design methodology should have good scalability. In this study, the feasibility of geometric-programming-based low-power design in analog blocks of CMOS APS image sensors will be investigated.

## 1.4 Contributions

By exploring a new predictive-integration scheme to overcome the output swing reduction problem, and by performing geometric-programming-based power optimization on all basic analog building blocks required for CMOS image sensors, a number of contributions have been achieved:

**Predictive-integration for low-voltage operation:** A new predictive integration scheme for enhancing dynamic range of CMOS image sensor at a low operating supply voltage was proposed and implemented in a laser rangefinding sensor chip [Gao 09]. The chip was demonstrated to achieve  $\sim 100$ -dB DR by using this new predictive-integration scheme when it was powered using 1.8 V. The predictive-integration can automatically adjust the integration time to the local illumination level. It is not dependent on a particular processing technology, and therefore, should have good scalability and can be applied to other deep sub-micron processing technologies.

**New methodology for a feasible and robust analog synthesis tool:** A new sub-spaced max-monomial modeling scheme for CMOS transistors in sub-micron technologies was proposed to improve the modeling accuracy [Gao 10]. By improving the device model accuracy, the synthesis accuracy was improved. The method used for device modeling is not technology-dependent, and can be applied to other sub-micron technologies than the 0.18  $\mu\text{m}$  technology used here; therefore, a robust synthesis tool can be achieved.

**Efficient low-power design in basic analog blocks of CMOS APS image sensors:** Geometric-programming-based power optimization was applied to the design of basic analog building blocks of CMOS APS image sensors. It was demonstrated to give an acceptable accuracy when synthesizing an in-pixel source follower, a column readout circuit, and an analog-to-digital converter. It is an efficient way for designers to reduce power consumption for an existing architecture or explore new low-power architectures [Gao 10]. An optimal design based on a published design was implemented, and it was noticed that a significant improvement of the power consumption in in-pixel source follower and ADC comparator blocks is possible.

**Speed-driven power optimal design in op-amps and comparators:** Specifications associated with op-amps and comparators in analog blocks of CMOS APS image sensors are usually speed-related, such as settling time for op-amps and prop-

agation delay time for comparators. These speed-driven power optimal designs can not be simply solved without knowing other regular performance requirements, such as DC gain, gain-bandwidth product, slew rate, phase margin, and others. In this study, some new methodologies are suggested to convert speed-related requirements to regular performance requirements, and then a power optimal design can be synthesized with geometric programming.

## 1.5 Thesis organization

Some typical optimization algorithms are reviewed and compared in Chapter 2. A proposed new sub-space-based modeling method for electrical parameters of CMOS transistors is presented in Chapter 3. A convex piecewise-linear fitting algorithm used for device modeling is discussed as well. In Chapter 4, issues related to low-power design in in-pixel source followers are discussed, including the operation-dependent low-power design in an in-pixel source follower, and geometric-programming-based power optimization in an in-pixel source follower. The synthesis evaluation is presented at the end of the chapter. In Chapter 5, a detailed power optimization design methodology for op-amps used for column readout block is discussed, synthesized results are verified with experimental data obtained from a tested chip, and a design for a low-power correlated double sampling (CDS) op-amp according to a

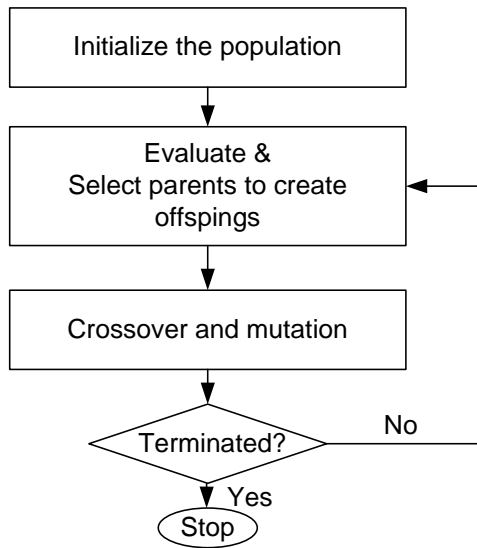
practical requirement is presented. Low-power design methodologies for successive-approximation (SAR) ADC are discussed in Chapter 6, including power optimization in two-stage open-loop op-amp and SAR ADC design. An example of power optimization on an existing CMOS image sensor is given in Chapter 7. In Chapter 8, a brief review of dynamic range enhancement technologies is presented, followed by the principle of the proposed predictive-integration scheme. The design for a laser rangefinding CMOS image sensor using predictive-integration is also demonstrated. Chapter 9 presents the conclusion and future work.

## 2 Algorithms for global optimization

As it has been mentioned in the last chapter, choice of optimization algorithm plays an important role in determining the efficiency of the synthesis-based low-power design. The classical general purpose optimization methods, such as steepest descent, sequential quadratic programming, and Lagrange multiplier methods, have been widely used in analog-circuit computer-aided-design (CAD) tools [Hers 01]. They can handle a wide variety of problems, but can usually only find locally optimal designs. A global optimization algorithm is needed to implement low-power design by finding the minimum power consumption. In this chapter, three typical global optimization algorithms will be reviewed, and a comparison will be made according to their implementation complexity and computation efficiency.

## 2.1 Genetic algorithm

The Genetic algorithm (GA) was created by John Holland in 1975 [Holl 75]. It is a popular search technique used in computer science to find approximate solutions to optimization problems. The Genetic algorithm is an evolutionary optimization method whose mechanism is analogous to biological evolution. Over successive generations, individuals who are best suited to survive in an environment live on and reproduce, while other individuals die off. For a complicated problem, the genetic algorithm uses a simple representation (bit strings) to encode each possible state. Traditionally, the bit string is a binary string, but a different encoding scheme can be used. For example, an integer number can be used in the transistor size optimization. Each encoding state is called an individual or chromosome, and a set of individuals are called a population or chromosomes. A general optimization procedure is shown in Figure 2.1. The optimization starts by creating an initial population, in which a set of individuals is randomly generated. A cost function (fitness function) is used to evaluate each individual. To generate the next population, a proper stochastic selection scheme must be applied to determine which individual will be copied to the next population or which individuals will be chosen as the parents to reproduce two new children for the next population. A detailed description of different selection methods can be found in [Marc 06]. Among them, roulette wheel



**Figure 2.1:** Generic algorithm flow of a genetic optimization process

selection and tournament selection are the most popular and well studied. Generally, an individual with a high fitness score has a high probability of being copied to the next population. Crossovers and mutations are used to generate two new individuals from two selected parents. Newly generated individuals will fill up vacant places in the population. After the new population is generated, the process will go back to the evaluation and repeat the successive steps. The genetic algorithm optimization can be stopped when convergence occurs, and the individual with the best fitness score is considered to be the solution to the optimization problem.

### 2.1.1 Analog circuit synthesis using genetic algorithms

The GA technique has been successfully applied in analog design for multi-objective optimization ranging from the system level to the transistor level [Dhan 06, Zebu 98, Wim 95]. Dhanwada *et al.* [Dhan 06] used GA to perform a hierarchical constraint transformation in a top-down analog system design methodology. In this hierarchical constraint transformation, the system-level performance specifications and constraints were first translated into block-level performance specifications, and then the transistor sizes in each functional block were determined according to the block-level specifications. The constraint transformation was treated as an independent optimization process. Zebulum *et al.* [Zebu 98] used GA to find the transistor sizes in an operational amplifier for micro-power consumption. Kruiskamp *et al.* [Wim 95] developed a GA based tool called Darwin, which optimizes operational amplifier designs by selecting their constituent blocks from standard amplifier stages. Transistor sizes have also been optimized using performance constraints.

A conventional way to implement automated analog circuit design with GA includes defining circuit representation, creating a fitness function, generating an initial individual population, and performing evaluation, crossover, and mutation. A circuit is usually represented by a string (also called an individual) which encodes parameters of the circuit, such as transistor sizes, bias current, compensation ca-

pacitor, or load. The string can be either made up of binary bits [Wim 95] or integer numbers [Zebu 98]. A fitness function has to be created to guide the searching algorithm towards the design space containing the optimal solution. A common way to construct a fitness function is to use a weighted-sum approach to consider multi-performance requirements, as shown in the following equation:

$$Fitness = \sum_{i=1}^n w_i f_i \quad (2.1)$$

where  $f_i$  is the fitness component of each objective,  $w_i$  is the respective weight, and  $n$  is the total number of objectives. The main problem of this approach is the specification of the weights. A robust method used for assigning weight to each objective is suggested in [Zebu 98], in which weights are generated adaptively as the optimization process progresses. A large weight will be assigned to the objective for which the average fitness is far from the target value. At the beginning of the evolution process, an initial population made up of a set of individuals should be generated randomly. This initial population will affect the convergence; if the initial population has individuals that are close to the desired solution, it may lead to a quicker convergence by reducing the search space. Solutions are usually evaluated using results from performance models or circuit simulators (e.g. HSPICE). Performance model based evaluation can lead to a quick computation compared to the simulator based method, but it depends heavily on the model accuracy. Fitness scores can be calculated from

evaluation results, by which selection for survived individuals is performed. A new population is generated by further executing crossover and mutation. Operations of evaluation, selection, crossover, and mutation will repeat until the optimal solution is found.

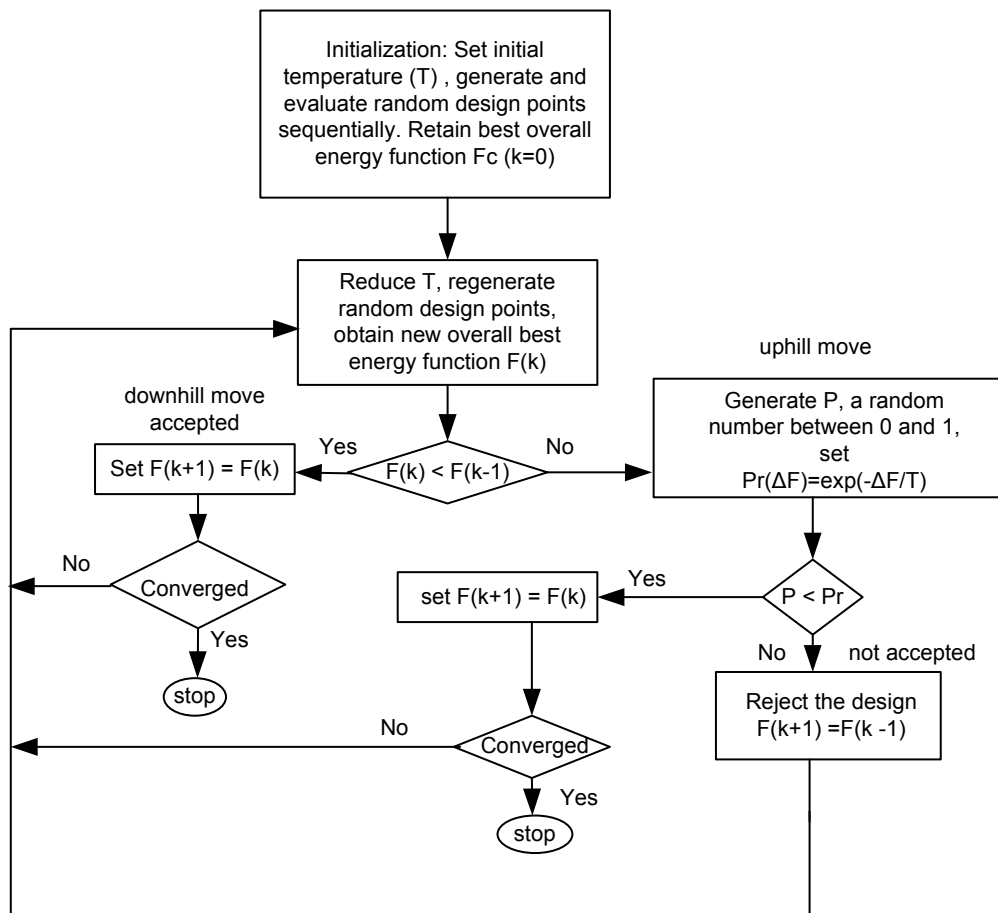
The GA technique has demonstrated a good stability for large circuit synthesis, and flexibility for different circuit structures [Zebu 98]. It also has some issues related to global optimization. To find the globally optimal solution, a large number of individuals are needed in a population and, therefore, the computation time will increase. The initial population may also affect the final solution, which may prevent the GA from finding a true globally optimal solution. An improper fitness function will over-address some objectives and under-estimate others, which in turn may not lead to a true globally optimal solution. If a circuit simulator is used for evaluation, the computation time will increase dramatically with increasing the complexity of the circuit.

## **2.2 Simulated annealing**

An alternative to genetic algorithms is simulated annealing (SA), which is a stochastic computational technique derived from statistical mechanics for finding near globally-minimum-cost solutions to large optimization problems. It was originally inspired by

the formation of crystals in solids during cooling, in which the solid is first heated to a high temperature and then cooled slowly down to the original temperature. The slower the cooling speed, the more ideal the resulting crystal structure. After the annealing, all particles in the solid have minimal bonding energy.

A general simulated annealing algorithm is shown in Figure 2.2. The simulated



**Figure 2.2:** Flow chart of the simulated annealing algorithm [Maji 06]

annealing starts with setting an initial temperature ( $T$ ), randomly generating a set of solutions in terms of input variables, and evaluating initial solutions with an energy function. The solution with the minimum energy will be selected for generating a new set of solutions, which is usually implemented with a procedure called perturbation. The perturbation will let the solution move in a random direction. During the cooling process, the temperature  $T$  is reduced according to the cooling schedule. The new set of solutions is evaluated, and the solution with the minimum energy is compared with the initial solution. If the energy of the new solution is less than that of the initial solution, the new solution will replace the initial one, and this downhill move is accepted. Otherwise, the solution is accepted with finite probability according to a function of  $T$ :

$$P(\Delta F) = \exp(-\Delta F/kT) \quad (2.2)$$

where  $\Delta F$  denotes the change in energy. A random number uniformly distributed in the interval (0,1) is used to compare with this probability. If the random number is less than  $P(\Delta F)$ , the new solution is retained, otherwise, the original solution is used to start the next step. This process allows simulated annealing to search in both downhill and uphill directions and, as a result, the probability of getting trapped in a local optimum is minimized. The annealing process will continue with a cooling schedule until the solution converges.

It is noticed that the key components in simulated annealing are the energy function, perturbation mechanism, acceptance function, and cooling schedule. The energy function depends on the properties of the design problem, which should consider effects of objectives and constraints. Generally, a weighted-sum approach similar to the method of constructing a fitness function in a GA problem can be used to construct an energy function. The perturbation mechanism will affect the computation speed, and an efficient perturbation mechanism should allow the solver to automatically select the perturbation step and range according to the current annealing state [Ocho 96]. The acceptance function, as defined in Equation 2.2, is used to determine whether the worse candidate generated by the current perturbation move is accepted or not. The acceptance function is controlled by the temperature and descends asymptotically to zero in the course of the optimization process. The cooling schedule determines the temperature change at each annealing stage, which must be designed properly to lead the process toward the global optimization. Traditionally, the cooling process is very time-consuming, because care must be taken to proceed close to the equilibrium so as not to be caught in the local minimum. An adaptive cooling schedule is the most efficient cooling scheme, which can adaptively change the temperature step according to the energy difference between two subsequent states [Nour 98, Shen 07].

### 2.2.1 Analog circuit synthesis using simulated annealing

The simulated annealing approach has been successfully applied to analog syntheses for ADCs [Maji 06], sensor interface architecture [Donn 98], operational amplifiers [Yuan 05, Gira 06], and the switched capacitor filter [Alpa 00]. At the circuit level, SA was used to size transistors in an analog circuit to meet specification requirements [Yuan 05, Gira 06, Alpa 00]. At the system level, SA was demonstrated to be an powerful optimization algorithm for implementing architecture optimization by finding architecture parameters and converting the system specifications into performance specifications for each building blocks [Maji 06, Donn 98, Alpa 00]. SA allows a wide variety of performance measures and objectives to be handled, and is extremely effective for problems involving a mixture of continuous and discrete variables.

SA usually requires a large computation time due to its annealing nature. To overcome this problem and keep the advantage of global optimization in SA, some researchers proposed an efficient way to limit the searching space by synthesizing an initial design close to the global optimum based on equations and heuristics. The final globally optimal design is then synthesized with SA within a small searching space [Yuan 05]. To reduce the computation time required by simulation-based evaluation, some researchers proposed to use equation-based evaluation [Donn 98, Alpa 00];

however, a great effort is needed to implement accurate behavioral models of circuits. A CMOS compact model [Cunh 98] was used in a SA process for synthesizing a low-power op-amp, in which the searching efficiency is improved because the compact model uses a single equation to combine operations from weak to strong inversion regions [Gira 06]. An efficient perturbation scheme was proposed in [Alpa 00] to improve computation efficiency, in which an evolution strategy was embedded in SA to select variable step size adaptively.

In principle, the SA can lead to a globally optimal solution, but it is not guaranteed in practice because the optimization process is affected by many factors, such as the perturbation mechanism, cooling schedule, and evaluation method, for which there is no precise formula. Moreover, no real-time lower bound is available, so termination is heuristic.

### **2.3 Geometric programming**

Geometric programming (GP) has been well studied for optimization designs in engineering applications since the late 1970s [Duff 67]. It is now regaining the attention of engineering designers due to two reasons: 1) new solving methods can solve very large-scale GPs extremely efficiently and reliably; 2) a number of practical problems, particularly in electrical circuit design, have recently been found to be equivalent to

(or well approximated by) GPs [Boyd 07]. Geometric programming is an equation-based mathematical optimization problem, which has the following standard form:

$$\begin{aligned}
& \text{minimize} && f_0(x) \\
& \text{subject to} && f_i(x) \leq 1, && i = 1, \dots, m \\
& && g_j(x) = 1, && j = 1, \dots, p
\end{aligned} \tag{2.3}$$

where  $x \in R_n^+$  is a vector of  $n$  real positive variables,  $f_i$  are posynomial functions (or called posynomials), and  $g_i$  are monomial functions (or called monomials). A monomial function has the form:

$$g(x) = cx_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \tag{2.4}$$

where  $c > 0$  and  $a_i \in R$ . A posynomial function is a sum of one or more monomials, which has a form:

$$f(x) = \sum_{k=1}^t c_k \prod_j x_j^{a_{kj}} \tag{2.5}$$

where  $c_k > 0$ , and  $a_{kj} \in R$ .

The reason that GP is efficient in solving optimization problem is that a posynomial/monomial function can be converted into a convex function by performing a logarithmic transformation on the function and its variables. For example, we can define new variables  $y_i = \log x_i$ , and take a logarithm of a posynomial  $f(x)$  to get

$$h(y) = \log(f(x)) = \log(f(e^{y_1}, \dots, e^{y_n})) = \log\left(\sum_k e^{a_k^T y + b_k}\right) \tag{2.6}$$

where  $a_k^T = [a_{1k}, \dots, a_{nk}]$  and  $b_k = \log c_k$ . For a monomial function, the transformed function is

$$h(y) = \log(g(x)) = \log(g(e^{y_1}, \dots, e^{y_n})) = a^T y + b \quad (2.7)$$

where  $a^T = [a_1, \dots, a_n]$ . It can be proven that  $h$  is a convex function of the new variable  $y$  [Hers 01], and the geometric programming can be converted into a convex optimization by expressing it as:

$$\begin{aligned} & \text{minimize} && f_0(e^{y_1}, \dots, e^{y_n}) \\ & \text{subject to} && f_i(e^{y_1}, \dots, e^{y_n}) \leq 0, && i = 1, \dots, m \\ & && g_j(e^{y_1}, \dots, e^{y_n}) = 0, && j = 1, \dots, p \end{aligned} \quad (2.8)$$

The convex optimization can lead to a truly global solution without being affected by the start point, and it can be solved efficiently with interior-point methods [Boyd 04]. It was reported that interior-point methods can solve large convex optimization problems, with thousands of variables and tens of thousands of constraints, in a few minutes on a small workstation [Hers 01]. Also, the stopping criteria are not heuristic because a lower bound on the achievable performance is given at each iteration.

In addition to efficient optimization, GP has advantages of facilitating feasibility, sensitivity and trade-off analyses [Vand 04, Hers 01]. During the GP solving process, the constraints will be checked to see if they are mutually consistent. If constraints are not mutually consistent, the problem is indicated as infeasible, and adjustments

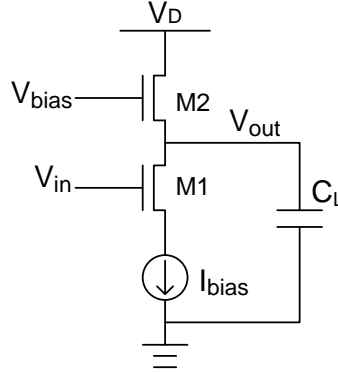
are needed on some constraints or performance specifications. Sensitivity analysis is used to investigate the effect of variable variation on the objective function. In the GP algorithm, the sensitivity analysis is achieved simultaneously when solving GPs, therefore there is no extra cost needed. A trade-off analysis can be performed easily in GP by replacing the unity value in constraint expressions (see Equation 2.3) with a non-unity value, which is called a perturbed GP.

### **2.3.1 Analog synthesis using geometric programming**

Analog circuit synthesis based on geometric programming has become an active research subject over the last decade. Its unique features of efficient computation and reliable global optimum have allowed researchers to successfully explore the automated design of CMOS op-amps [Mand 01, Hers 01, Vand 04, Agui 08], pipelined ADC [Hers 02], and CMOS DC-DC buck converters [Lee 03].

According to the definition of GP in Equation 2.3, the objective function and constraints for analog circuit synthesis have to be properly expressed in posynomial/monomial forms. The objective function is the goal of the optimization, which comprises cost metrics to be minimized. For example, the power consumption will be included in the objective function in low-power design. In GP-based analog circuit synthesis, information about transistor sizes, circuit topology, bias condi-

tions, and performance requirements have to be properly represented by convex constraints. For example, a transistor size is defined within a range:  $L_{min} \leq L \leq L_{max}$ ,  $W_{min} \leq W \leq W_{max}$ ; transistors in a current mirror are defined as:  $L_1 = L_2$ ,  $\frac{W_2}{L_2} \leq \frac{W_1}{L_1}$ ; bias for a transistor can be constrained as:  $V_{min} \leq V_{gs} \leq V_{max}$ ; and gain for an op-amp can be derived as:  $\frac{1}{gain} \leq \frac{1}{gain_{min}}$ . Constraints for transistor sizes, topology, and bias conditions are obvious, and can be derived easily. Converting performance requirements into constraints in posynomial/monomial forms usually takes more effort, in which multi-level models (at circuit and device levels) are needed. For example, the model for the DC gain at the circuit level in a simple single stage common source amplifier (as shown in Figure 2.3) is  $A_V = \frac{g_{m1}}{g_{ds1} + g_{ds2}}$ , where  $g_{m1}$  and  $g_{ds1}$  are transconductances of transistor  $M1$ , and  $g_{ds2}$  is the transconductance of transistor  $M2$ . It is noticed that the model of gain is not a posynomial,



**Figure 2.3:** A single stage common source amplifier

because the denominator is a posynomial. To convert the *gain* into an appropriate constraint required by the geometric programming, an inverted form of the gain should be used,  $\frac{1}{A_V} = \frac{g_{ds1}+g_{ds2}}{g_{m1}}$ , in which  $\frac{1}{g_{m1}}$ ,  $g_{ds1}$ , and  $g_{ds2}$  should be modeled in posynomial forms in terms of design variables at the device level, such as transistor sizes and bias voltages.

GP is a true equation-based optimization, and the evaluation completely depends on the derived models for circuits and transistors. The major error source comes from the transistor modeling. In sub-micron technologies, transistors cannot be modeled as accurately by monomials as those in long-channel technologies. Improving posynomial models for CMOS transistors in sub-micron technologies has become an active research subject in recent years. A convex piecewise-linear fitting method was proposed in [Kim 04] to generate the short-channel transistor model. The result shows a dramatic improvement in accuracy compared to the monomial modeling method used for long-channel technologies. Aggarwal *et al.* [Agga 07b] proposed a genetic-algorithm-based method to derive posynomials for CMOS transistors. It was reported to achieve a better performance than the method used in [Kim 04]. These two methods model CMOS transistor parameters in a single design space ranging from weak- to strong-inversion regions. In fact, some parameters are not convex across the entire design space, so a big modeling error will be gener-

ated if a single posynomial model is used for the entire design space. To eliminate models for transistors, Daems *et al.* [Daem 03] proposed a simulation-based method to generate posynomial models for an analog circuit, in which performance measures are modeled in terms of design variables by using circuit SPICE simulation data. Posynomial models derived from this method comprise all design variables of the circuit, resulting in a very high dimensional fitting problem. As a result, the model accuracy becomes a problem as the circuit complexity increases.

If a design problem can be formulated hierarchically within a geometric programming framework (i.e., constraints from the system level down to the circuit level are formulated in posynomial/monomial forms with a hierarchical organization) it can be solved in a flat manner and can generate solutions for parameters at different design levels at the same time. A good example was shown in [Hers 02] for a pipeline ADC design.

The requirement of representing analog circuits in a special convex form is the major disadvantage of geometric programming. Sometimes, a constraint cannot be formulated in a convex form, and more effort is needed to approximate this non-convex constraint with a set of convex constraints [Vand 04] or to force the solution to converge to the non-convex constraint [Agui 08].

## 2.4 Conclusion

Three leading global optimization algorithms used for analog automated designs have been briefly reviewed above. GA and SA have lower implementation complexity compared to GP, because they are not required to be formulated in a special form. In GA and SA, the evaluation can be achieved by using a simulator. Theoretically, GA and SA can lead to a global optimization solution, but it is not guaranteed in a practical design due to their heuristic nature. Unlike GA and SA, GP does not require a starting point or initial guess of the optimal solution, and does not need parameter tuning. A truly global optimal solution can be guaranteed by the GP when the problem is feasible. GP can be efficiently solved with a numerical method, while GA and SA are solved with time-consuming heuristic searching methods. GA and SA solve a hierarchical problem in a hierarchical way, but GP can solve it in a flat manner; therefore, the computation efficiency is improved.

GP will be selected here for aiding low-power design in analog blocks of CMOS APS image sensors due to its computation efficiency and true global property. The details of using GP for power optimization in analog blocks of CMOS APS image sensors will be presented in Chapter 4, 5, and 6.

### 3 CMOS transistor modeling

The goal of modeling a CMOS transistor is to map the input MOSFET parameters (e.g. channel width  $W$ , channel length  $L$ , gate-to-source voltage  $V_{gs}$ , drain-to-source voltage  $V_{ds}$ ) into the output electrical parameters, such as transconductance  $g_m$ , drain current  $I_{ds}$ , and parasitic capacitance  $C_{gd}$ . Posynomial modeling uses a posynomial form for this mapping in order to facilitate GP analysis. A key challenge faced by geometric-programming-based analog circuit synthesis in modern sub-micron technologies is to improve the accuracy of the CMOS transistor parameter models. To my knowledge, no posynomial models for CMOS transistors with a minimum feature size less than  $0.18 \mu\text{m}$  has been reported in the literature. The published posynomial modeling methods in  $0.18 \mu\text{m}$  CMOS technology and above include convex piecewise-linear fitting (PWL) [Kim 04], genetic-algorithm-based posynomial modeling (GAP) [Agga 07b], and monomial modeling for long-channel length technologies [Hers 01, Mand 01]. The monomial modeling can lead to a large modeling error in short-channel length technologies, and hence will not be considered in this study.

The PWL uses a max-affine function to fit the experimental data in the log-scale space, in which the fitting problem becomes an ordinary linear least-squares problem. The max-affine function can then be converted to a max-monomial function in real space. In contrast, GAP combines a genetic algorithm with quadratic programming to directly synthesize a posynomial model for the experimental data in real space. GAP was reported to have a better performance than PWL [Agga 07b] in CMOS transistor models in 0.18  $\mu\text{m}$  technology. These two schemes model CMOS transistor parameters in a single design space (ranging from weak- to strong-inversion regions). Although the accuracy of the transistor models in short-channel technologies is improved in both methods, large errors in some parameters, particularly transconductances  $g_{ds}$  [Kim 04] and  $g_m$  [Agga 07b], can lead to significant prediction errors in some circuit performance measures. In this chapter, a new sub-space modeling method is proposed to further improve the accuracy of posynomial models of CMOS transistors in sub-micron technologies. The PWL modeling method is selected as the basis for this work due to its simple implementation. In the following section, the PWL modeling method will be first discussed, and then sub-space-based max-monomial models for electrical parameters of CMOS transistors will be presented.

### 3.1 PWL modeling

A max-monomial is a monomial selected from a set of monomials, which is defined as  $f(x) = \max_i \{g_i(x)\}$  where  $g_i$  is a monomial function of  $x$ . Thus,

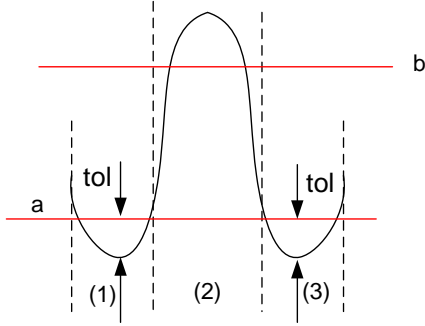
$$f(x) = \max_i \{\beta_i x_1^{a_{1,i}} x_2^{a_{2,i}} \dots x_n^{a_{n,i}}\}, \quad i = 1, \dots, k \quad (3.1)$$

where  $x \in R_n^{(+)}$ ,  $\beta_i > 0$ , and  $a_{j,i} \in R$ . For each data point, a monomial producing the maximum value will be considered as the best fit. Let  $u_i = \log(x_i)$ ,  $c_i = \log(\beta_i)$ , and  $f(u) = \log(f(x))$  is equivalent to approximating the function  $f(u)$  by a piecewise linear function

$$f(u) = \max_i \{c_i + a_{1,i}u_1 + a_{2,i}u_2 + \dots + a_{n,i}u_n\} \quad (3.2)$$

This is also called a max-affine function. As a result, a max-monomial model can be converted to a PWL model. A simple PWL modeling method was proposed by Magnani and Boyd [Magn 09], in which the data is partitioned and a least-squares fit is performed in each partitioned region. Magnani and Boyd assumed that data to be fitted is in convex form and therefore a non-constrained least-squares fit can be applied without causing a large error. By observation, we notice that some electrical parameters of CMOS transistors are not fully convex across the entire design space. A large fitting error will be generated if a simple non-constrained least-squares (NCLS) fit is used. For example, in the case shown in Figure 3.1, data in partition region (2)

demonstrate a concave property. If a non-constrained least-squares fit is used, fitting



**Figure 3.1:** Non-convex fitting case

plane  $b$  will be selected as the the fitting plane in regions (1) and (3) as well, because it generates the maximum value in regions (1) and (3) compared to the fitting plane  $a$ . The plane  $b$  is the best fit in partition region (2), but leads to big fitting errors in regions (1) and (3). A constrained least-squares (CLS) fit can be performed by adding an extra limitation as in [Crus 04]:

$$g(x_i) \leq f(x_i) + tol, \quad \text{outside the local fitting region} \quad (3.3)$$

where  $g(x_i)$  is the fitting function,  $f(x_i)$  is data to be fitted, and  $tol$  is a predefined tolerance. Actually, the value of  $tol$  will affect the model performance. In this study,  $tol$  is initially set to  $\theta$ , and a model is derived at this condition. Then a new  $tol$  value is assigned by adding a predefined step, and a new model is derived at this new condition. If the model accuracy from the new condition is improved

compared to the old one, the increment in  $tol$  is allowed and a new increment will be performed, otherwise the new  $tol$  is abandoned and the old  $tol$  is used as the final  $tol$ . This limitation will force a local fitting to generate small errors in non-fitting regions. In this case, plane  $a$  is the final fitting plane.

In this study, a constrained least-squares partition fit algorithm will be applied to model electrical parameters of CMOS transistors. A pseudo-code of the algorithm is:

```

Convert real space data to log-scale space
For  $N_{trial} = 1 : N$ 
  Generate  $k$  initial partition regions:  $P_1^{(0)}, P_2^{(0)}, \dots, P_k^{(0)}$ 
  while stop-flag & ( $l \leq l_{max}$ )
    For  $j=1$  to  $k$ 
       $a_j^{(l)}$  &  $b_j^{(l)} \leftarrow$  constrained least-squares fit on data  $P_j^{(l-1)}$ 
    end
    assign data point  $i$  to partition  $P_j^{(l)}$  if
       $f^{(l)}(x_i) = \max_w (a_w^{(l)T} x_i + b_w^{(l)}) = a_j^{(l)T} x_i + b_j^{(l)}$ ,  $w = 1, \dots, j, \dots, k$ 
    if  $P_j^{(l-1)} = P_j^{(l)}$ 
      stop-flag=0
    end
     $l=l+1$ 
  end
  Calculate RMRSE on  $P_j^{(l)}$ ,  $j=1, \dots, k$ , with  $a_j^{(l)}$  and  $b_j^{(l)}$ 
end
Find the best fit with the minimum RMRSE

```

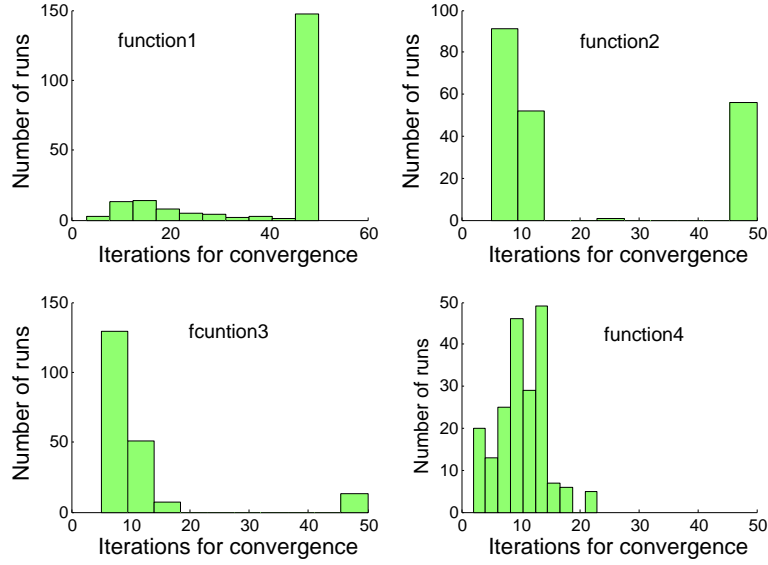
To perform CLS PWL modeling, the experimental data have to be first converted to log-scale space. This heuristic-based algorithm may lead to a poor fit with only a single run, so it is preferable to run this algorithm multiple times ( $N_{trial}$ ). During

each trial run, the experimental data will be first partitioned into  $k$  sub-regions according to their closeness by using a “kmeans” function in the MATLAB. In each trial run, seeds are randomly selected for generating sub-regions, therefore, partitions in a trial run may be different from other runs. In each sub-region, a constrained least-squares fit is performed. A set of linear fitting functions will then be derived to form a max-affine function. The max-affine function is evaluated at each data point. A data point will be assigned to a partition region if the fitting function of that region generates the maximum value at the data point. The fitting and partition operations will be repeated until new partitions are identical to the previous ones or the operation reaches the iteration limit  $l_{max}$ . For each partition operation, the root mean relative square error (RMRSE) of the max-affine function will be calculated. The best fit is the one with the minimum RMRSE. It is noticed that there are three variables ( $N_{trial}$ ,  $k$ ,  $l_{max}$ ) to be determined before running the proposed algorithm. The performance of the derived model will be affected by these variables. It is therefore worth studying the effect of each factor by running some evaluation simulations. Four constructed two-dimensional functions are used to evaluate this

algorithm as proposed by [Agga 07a]. They are:

1.  $y = x_1^{0.45} x_2^{-2.2}$
2.  $y = 10^{-7} x_1^{0.82} x_2 + 5 \times 10^{-15} x_1^{0.82} x_2 + 5 \times 10^{-15} x_1^{7.23} x_2 + 25 \times 10^{-23} x_1^{7.23} x_2^6$
3.  $y = 10^{-6} x_1 x_2^{2.1} + 10^{-2} x_1 x_2^{-1.4} + 5 \times 10^{-14} x_1^7 x_2^{2.1} + 5 \times 10^{-10} x_1^7 x_2^{-1.4}$
4.  $y = 10^{-2} x_1 + 5 \times 10^{-10} x_1^7 - 10^{-2} x_1 e^{-x_2} - 5 \times 10 x_1^7 e^{-x_2}$  (3.4)

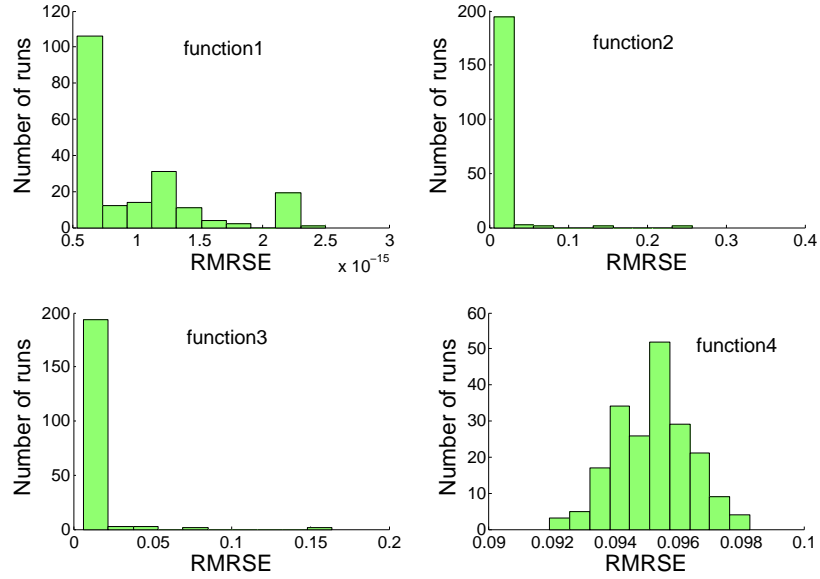
The first constructed function is a monomial and is linear in log-log space. The second one is a posynomial, which is convex and non-linear in both dimensions in log-log space. The third one is a posynomial, which is convex, non-linear in  $x_1$  and non-monotonic in  $x_2$  in log-log space. The fourth function is non-posynomial, which is non-linear in  $x_1$  and concave in  $x_2$  in log-log space. For each evaluation simulation, only one variable will be changed while other two are fixed. The iteration limit  $l_{max}$  controls the number of iterations needed for convergence. The partition number  $k$  determines the necessary partition regions for achieving an acceptable accuracy. The trial number  $N_{trial}$  is a factor to guarantee that a good fit is found. Firstly, the convergence of the algorithm will be examined by letting  $N_{trial}=200$ ,  $k=40$ , and  $l_{max}=50$ . During each trial run, the number of iterations needed for convergence will be recorded. If convergence is not reached when the iteration reaches  $l_{max}$ , it will be counted as 50. Figure 3.2 demonstrates a distribution of convergence for aforementioned functions. We notice that convergence does not occur in most of the



**Figure 3.2:** Convergency distribution

runs for function 1, but does occur in most of the runs for functions 2, 3, and 4. To find a proper iteration number for this algorithm, we should also look into the fitting result for these four functions, which is shown in Figure 3.3. For function 1, even though convergence is not reached in most of the runs, the fitting result is always good. Therefore, neither convergence nor iteration number are important. We also notice that the algorithm works pretty well for functions 2 and 3, and also gives an acceptable result for function 4. Usually, the algorithm converges within 30 iterations for functions 2, 3, and 4.

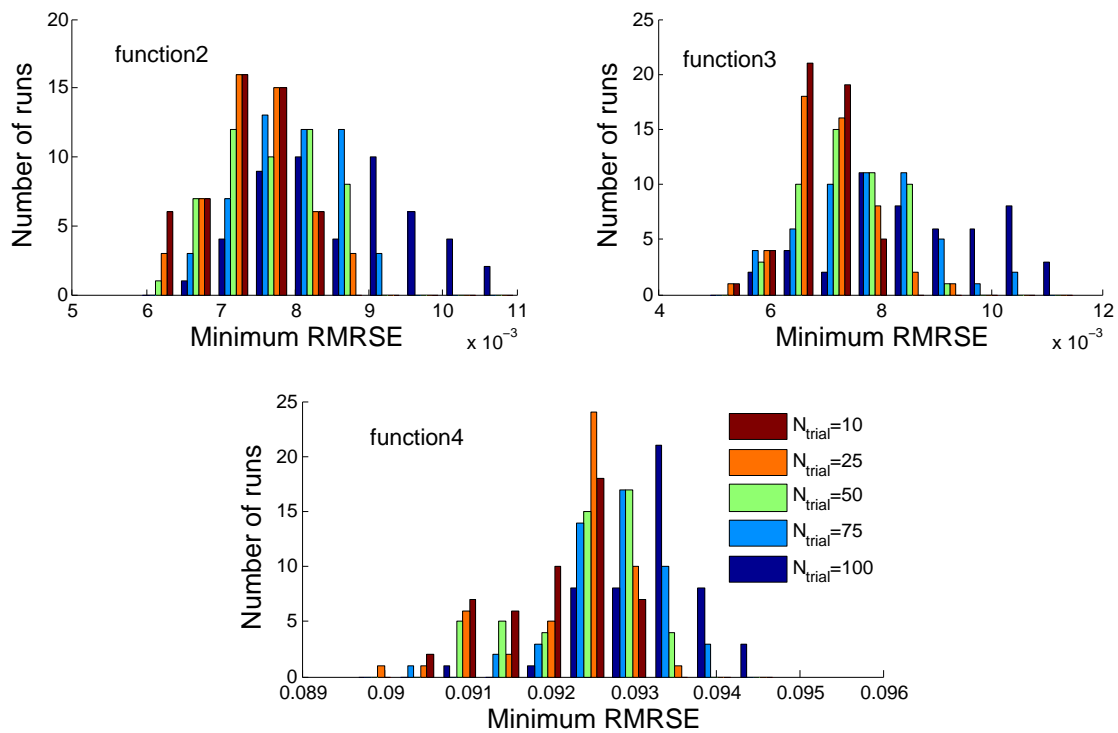
The effect of the number of trials on fitting accuracy is also investigated by setting  $k = 40$ ,  $l_{max} = 30$ . The fitting process will run 50 times for each of  $N_{trial}$  at values



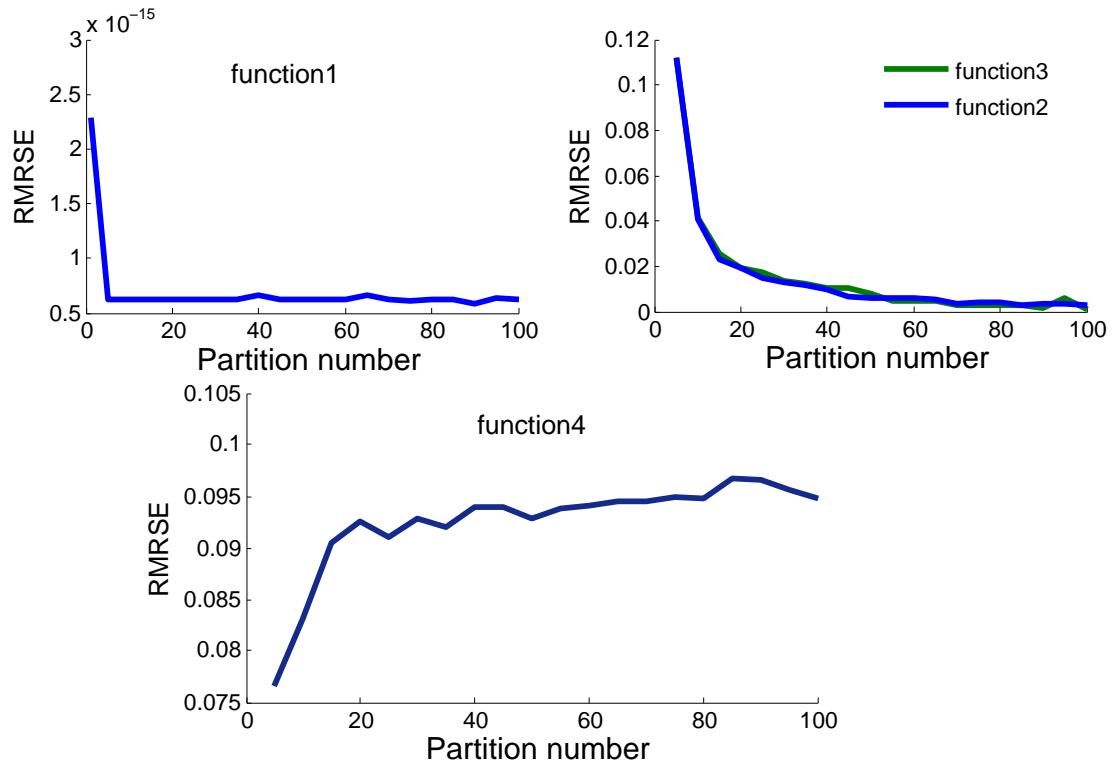
**Figure 3.3:** Fitting results of four constructed functions

of 10, 25, 50, 75, and 100. Only functions 2, 3, and 4 are evaluated here, because the function 1 always has a good fit. The simulation results are presented in Figure 3.4. It is noticed that 10 trials can provide a fit with a reasonable accuracy. In the following study of CMOS transistor modeling,  $N_{trial}=10$  will be used to achieve a trade-off between accuracy and computational time.

To evaluate the effect of the partition number on the model accuracy, we let  $l_{max} = 30$ ,  $N_{trial} = 10$ , and  $k$  vary from 1 to 100. The relationship between the minimum RMRSE and the partition number is derived and presented in Figure 3.5. It is noticed that the fit does not improve beyond five partitions for function 1. For functions 2 and 3, the fit improves as partition number increases. For function



**Figure 3.4:** Effect of trial number on fitting accuracy



**Figure 3.5:** Effect of partition on fitting accuracy

4, the fit gets worse when partition number increases. A proper partition number should be selected according to the trade-off between the modeling accuracy and the computational effort. For these four cases, 70 partitions can give a reasonable fitting accuracy. This value will be used for CMOS transistor modeling as well.

With the selected variable values, the proposed algorithm is used to fit the aforementioned functions. Two other algorithms, non-constrained least-squares fit and the algorithm in [Crus 04], are also applied for comparison. The fitting results are presented in Table 3.1. We can see that the CLS can give a better fit among these

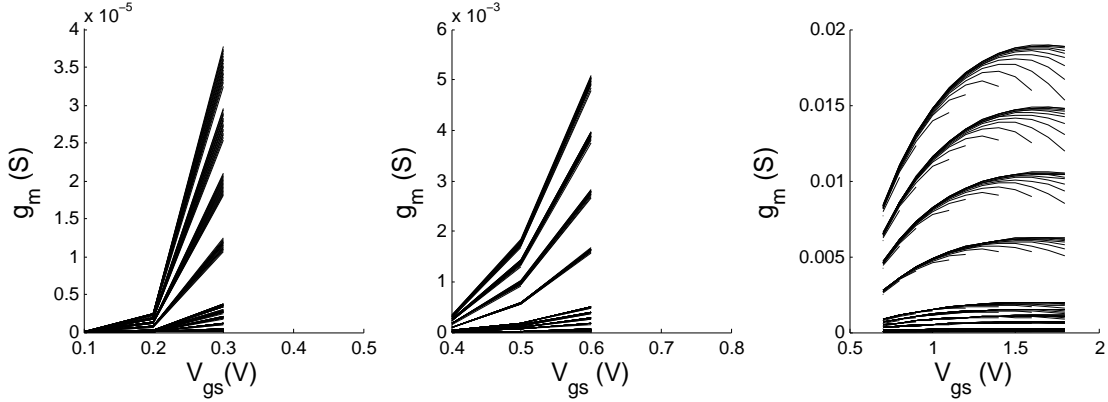
**Table 3.1:** Fitting comparison among three algorithms

<b>Function</b>	<b>RMRSE (NCLS)</b>	<b>RMRSE (algorithm of [Crus 04])</b>	<b>RMRSE (CLS)</b>
1	$1.81 \times 10^{-13}\%$	$5.9 \times 10^{-14}\%$	$6.2 \times 10^{-14}\%$
2	5.13%	1.69%	4.34%
3	6.09%	19.36%	4.34%
4	8.21%	9.69%	8.27%

three algorithms. The CLS algorithm will be used to model electrical parameters of CMOS transistors in 0.18  $\mu\text{m}$  technology, which is discussed in detail in the following section.

### 3.2 Sub-space-based max-monomial models

Unlike for long-channel length technology, parameters for a CMOS transistor in sub-micron technologies cannot be accurately modeled with a single monomial equation over the entire design space. By carefully studying the nature of MOS transistor parameters in TSMC 0.18  $\mu\text{m}$  technology, we find that some parameters exhibit very different behaviors in different operating regions. An example is illustrated in Figure 3.6. We can see that  $g_m$  has a convex property as  $V_{gs}$  varies in the weak and



**Figure 3.6:**  $g_m$  versus  $V_{gs}$  for an NMOS transistor in 0.18  $\mu\text{m}$  technology with  $L = 1.02\mu\text{m}$ ,  $W : 0.4 \sim 100\mu\text{m}$ ,  $V_{ds} : 0.2 \sim 1.8\text{V}$ , working in (a) weak inversion (b) moderate inversion (c) strong inversion.

moderate inversion regions, while it has a concave property in the strong inversion region. This implies that different models have to be applied in different operation regions.

An operation region of a CMOS transistor can be defined by its bias condition and

its size, which are associated with four independent design variables ( $W, L, V_{ds}, V_{gs}$ ). To find proper regions for accurate modeling, ten key electrical parameters of CMOS transistors have to be extracted from HSPICE simulations across the entire design space. They are  $g_m, 1/g_m, g_{ds}, I_{ds}, C_{gd}, C_{gs}, C_{db}, C_{sb},$  and  $C_{dg}$ . In order to achieve a better trade-off between the power consumption and speed, a design space defined by moderate- and strong-inversion regions will be considered, namely  $L : 0.18 - 21 \mu m, W : 0.4 - 100 \mu m, V_{ds} : 0.2 - 1.8 V, V_{gs} : 0.4 - 1.8 V$ . By observation, we notice that the model accuracy is most sensitive to  $L$  and  $V_{gs}$ . As a result, the design space of interest is divided into multiple sub-spaces by assigning different ranges to  $L$  and  $V_{gs}$ . Table 3.2 lists a detailed sub-space map for our study. There are 52

**Table 3.2:** Sub-space map for all parameters but  $1/g_m$

$V_{gs}(\text{V}) \backslash L$	$L$				
	$0.5 - 1\mu m$	$1 - 2\mu m$	$2 - 5\mu m$	$5 - 10\mu m$	$10 - 21\mu m$
0.4 - 0.5	mod1	mod4	mod7	mod10	
0.5 - 0.6	mod2	mod5	mod8	mod11	
0.6 - 0.7	mod3	mod6	mod9	mod12	
0.7 - 0.8	st1	st9	st17	st25	st33
0.8 - 1.0	st2	st10	st18	st26	st34
1.0 - 1.2	st3	st11	st19	st27	st35
1.2 - 1.4	st4	st12	st20	st28	st36
1.4 - 1.5	st5	st13	st21	st29	st37
1.5 - 1.6	st6	st14	st22	st30	st38
1.6 - 1.7	st7	st15	st23	st31	st39
1.7 - 1.8	st8	st16	st24	st32	st40

sub-spaces used in our study for all parameters except  $1/g_m$ . Only three simple sub-spaces (weak, moderate, and strong inversion regions) are needed to model  $1/g_m$  accurately. In each sub-space, a max-monomial model of an electrical parameter is derived from the CLS PWL method. Ideally, design variables for each electrical parameters should be independent. Models derived from four independent variables ( $W, L, V_{gs}, V_{ds}$ ) are compared with those derived from five dependent variables ( $W, L, V_{gs}, V_{ds}, I_{ds}$ ). A better model accuracy with using five dependent variables is observed in  $g_m, 1/g_m$ , and  $g_{ds}$ . Other parameters are modeled sufficiently accurately with 4 independent design variables. To solve an over-determined problem induced by dependent variables, an extra constraint on the current  $I_{ds}$  should be added in geometric programming for circuit synthesis, which will be explained in detail in Chapter 5. For each electrical parameter, the max-monomial model in each sub-space is evaluated with the mean relative error (MRE) and the worst case among all sub-spaces being presented in Table 3.3. Results reported in [Kim 04] and [Agga 07b] are also shown in Table 3.3 for comparison. The sub-space based PWL ( $PWL_{sub}$ ) modeling approach brings MRE error under 10%. A significant improvement in  $g_m$  and  $g_{ds}$  accuracy is observed in this modeling method compared to the other two single space schemes. Among all transistor parameters, these two play the most important roles in circuit synthesis.

**Table 3.3:** Comparison of MRE (%) for different models

<b>model</b>	<b><math>g_m</math></b>	<b><math>1/g_m</math></b>	<b><math>g_{ds}</math></b>	<b><math>I_{ds}</math></b>	<b><math>C_{gd}</math></b>	<b><math>C_{gs}</math></b>	<b><math>C_{gb}</math></b>	<b><math>C_{db}</math></b>	<b><math>C_{sb}</math></b>	<b><math>C_{dg}</math></b>
<i>PWL<sub>sub</sub></i> (this work)	2.15	0.47	3.37	9.36	3.42	4.19	4.96	2.03	4.32	2.24
<i>PWL</i> [Kim 04]	--	1.70	9.40	--	--	3.10	--	--	--	--
<i>GAP</i> [Agga 07b]	13.00	--	7.21	--	0.28	4.32	--	0.18	--	--

If a blind search is made over all 52 sub-spaces, GP circuit synthesis with sub-space max-monomial models will be of low efficiency. Compared to a single posynomial model method, the computational effort in the sub-space method will be increased by an order of  $O(m^n)$ , where  $m$  is the total number of sub-spaces to be searched, and  $n$  is the number of transistors in the circuit. Even with a simple differential op-amp (5 transistors), the computational effort with the blind search will not be acceptable. However, given a specific performance requirement and circuit topology, each transistor can only work in a few of these sub-spaces. In a differential op-amp design for example, the circuit is symmetrical in the input stage, which will allow fewer transistors to be considered. By constraining the searching space of transistors according to the performance requirements, we can make sub-space based GP circuit synthesis feasible. Some examples will be demonstrated in chapters 4, 5, and 6.

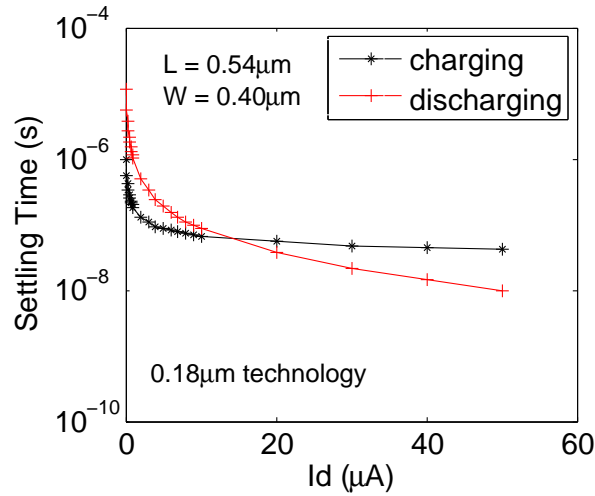
## 4 Low power design in in-pixel source follower

Major analog blocks in a typical CMOS APS image sensor include an in-pixel readout circuit (usually a source-follower), a sample-and-hold circuit (usually a CDS circuit), and an ADC, which are shown in Figure 1.2. In the following three chapters, a geometric-programming-based power optimization technology will be applied to these three analog blocks to reduce the power consumption without changing the circuit architectures.

### 4.1 Operation dependent low power design in in-pixel source follower

An in-pixel source follower in a APS image sensor is a key readout circuit and primarily affects the readout speed. The readout speed of the source follower is affected by the operation of the CDS used in the off-pixel analog signal processing circuit. In both dual-path [Mend 97] and single-path (switched capacitor based) [Cho 03]

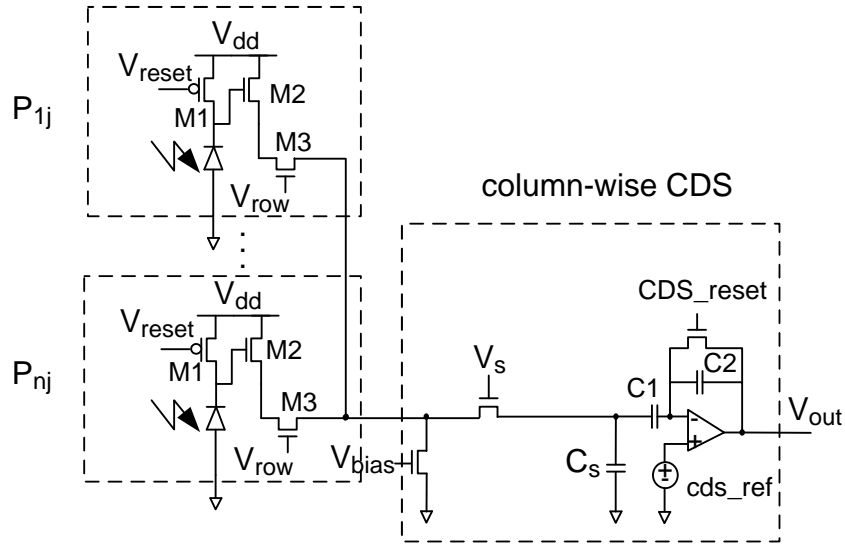
CDS circuits, the source follower will charge or discharge the sampling capacitor(s), depending on the values of the present signal and the signal already stored on the sampling capacitor(s). The settling time of the source follower is different for charging and discharging operations, as shown in Figure 4.1. The data in Figure 4.1 is



**Figure 4.1:** Settling time of the source follower during charging and discharging

extracted from the simulation using a transistor of a specific size. The worst case (the full swing of the input signal) is considered. Transistors with other sizes demonstrate a similar behavior. When the bias current is small, the readout speed is limited by the discharging operation. To increase the readout speed, a larger bias current is needed, which in turn will increase the power consumption. The settling time of the source follower decreases slightly when the bias current is beyond a specific value

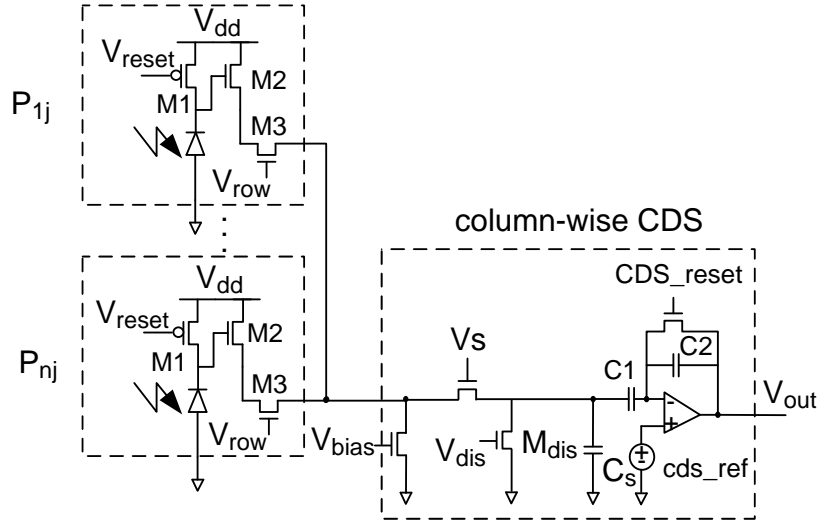
( $\sim 10 \mu A$  in this case). This implies that the readout speed can be improved without increasing the power consumption significantly if only the charging operation is needed. In many column-wise CMOS image sensor designs, a single path CDS with a switched capacitor structure is selected to reduce column fixed pattern noise (FPN). A simple single-path CDS circuit for an APS array is shown in Figure 4.2. To retrieve



**Figure 4.2:** Column-wise signal readout in a CMOS image sensor

the light intensity information, two signals (reset and photo signal) will be read out through the source follower. The source follower will charge the sampling capacitor ( $C_s$ ) during the readout of the reset signal, and will discharge the sampling capacitor during the readout of the photo signal. To reduce the power consumption induced by the discharging operation, an extra discharging transistor ( $M_{dis}$ ) can be added across

the sampling capacitor, as shown in Figure 4.3. After the reset signal is sampled to

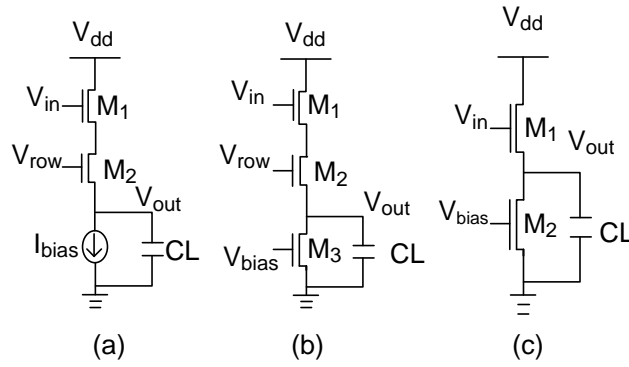


**Figure 4.3:** Modified column-wise signal readout

the sampling capacitor, the information of the reset signal is also stored on capacitor of  $C_1$ . Before the photo signal is read out, the sampling capacitor can be discharged by turning on  $M_{dis}$ . During the readout of the photo signal,  $M_{dis}$  is turned off, and the source follower will charge the sampling capacitor again. At the end of operation, the difference between the reset and photo signals will be transferred to the output of the CDS amplifier. As a result, the readout speed of the source follower is only limited by the charging operation. With only the charging operation being allowed, the bias current can be reduced significantly for a given speed requirement.

## 4.2 Power optimization in in-pixel source follower

To synthesize a source follower using geometric programming, a posynomial form of the settling time should be modeled with respect to design variables  $L$ ,  $W$ ,  $I_{ds}$ , and  $C_L$ . An ideal source follower (as shown in Figure 4.4 (a)) is used to extract experimental data. The area of the in-pixel source follower is limited by the pixel



**Figure 4.4:** (a) ideal source follower (b) 3T source follower (c) 2T source follower

size and the desired fill factor and it should be as small as possible. The design space is limited to a small range:  $L$ :  $0.54 \mu m - 10 \mu m$ ,  $W$ :  $0.4 \mu m - 10 \mu m$ ,  $I_{ds}$ :  $100 \text{ nA} - 10 \mu A$ , and  $CL$ :  $1 \text{ pF} - 10 \text{ pF}$ . The access transistor,  $M2$  has the minimum size ( $L=0.18\mu m$ ,  $W=0.4\mu m$ ). A max-monomial model of the settling time is derived in the design space defined above. The settling time calculated from the max-monomial model was evaluated with HSPICE simulation data, and has the performance shown in Table 4.1.

**Table 4.1:** Mean, maximum, and minimum relative errors in settling time model of in-pixel source follower

Root Mean Relative Square Error (RMRSE)	Maximum Relative Error (MARE)	Minimum Relative Error (MNRE)
0.76%	3.81%	$3.24^{-14}\%$

By using the settling-time model, a source-follower design problem can be treated as a geometric programming problem in which low power consumption can be easily achieved using power optimization. In practice, the ideal current source should be replaced with a transistor, as shown in Figure 4.4 (b). The access transistor usually works in the linear region, and has a very small drain-to-source voltage ( $\sim 10mV$ ). Ignoring the access transistor will not cause a big error in the final synthesis result. As a result, a 2T source follower, as shown in Figure 4.4 (c), will be used for synthesis. The objective function in this design is:

$$obj = \min(w_1 I_{bias} + w_2 (1/V_{ds\_M2}) + w_3 V_{dsat\_M2}) \quad (4.1)$$

where  $w_1$ ,  $w_2$ , and  $w_3$  are weights. In addition to the bias current, the objective function includes another two terms. The second term is used to guarantee that Kirchoff's voltage law is obeyed, and the third term is used to guarantee that  $M2$  works in the saturation region during readout. As long as  $w_1$  is much larger than both  $w_2$  and  $w_3$ , the minimum power consumption can be guaranteed. A source

follower with the same performance requirement as in [Cho 03] was synthesized for comparison. The source follower in [Cho 03] has both charging and discharging operations for reading out reset- and photo-signal, respectively. From its timing diagram and bias current, the value of the column sampling capacitor ( $\sim 6 \text{ pF}$ ) can be derived. As discussed earlier, only the charging operation is needed if a discharge transistor is added across the column sampling capacitor. The same charging readout time as in [Cho 03], i.e.  $2.6 \mu\text{s}$ , is used in this geometric programming problem. To implement the geometric programming, some practical constraints are required to limit the range of design spaces that must be searched for  $M1$  and  $M2$ . First let  $V_{gs1} = V_{ds1} \leq 0.7 \text{ V}$  (to allow a large output swing). Therefore the searching spaces for  $M1$  are mod1 – mod12 (see Table 3.2). To achieve low power consumption, the bias current should be as small as possible. As a result,  $M2$  should be limited to the moderate inversion region (mod1 – mod12). A total of 144 possible solutions is found in the defined searching spaces. We notice that the same synthesized bias current ( $I_{bias\_predict}$ ) can be achieved by combining different transistor sizes and bias voltages for  $M2$ , which can be shown in Table 4.2. Solutions listed in Table 4.2 are those with the minimum bias current. The optimal solution is the one with the minimum area cost ( $L1*W1+L2*W2$ ), which is the case listed in the second column. Compared to the current given in [Cho 03]( $1.25 \mu\text{A}$ ), the power-optimized design can reduce the

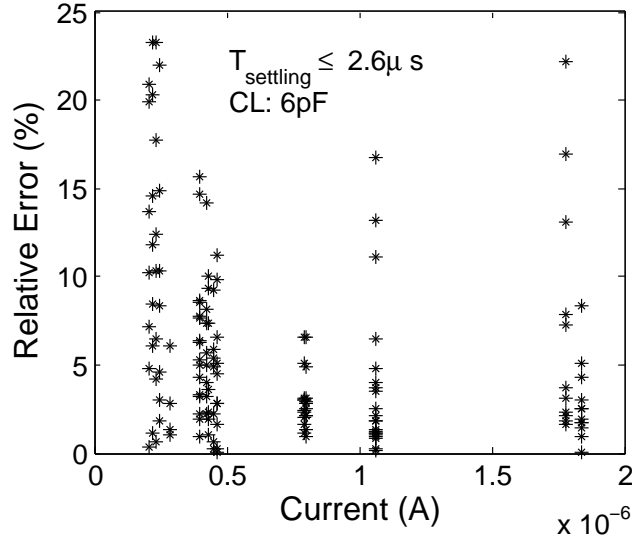
**Table 4.2:** Synthesized solutions of the source follower with the minimum bias current

Case	1	2	3	4	5	6	7
$T_{\text{settling\_predict}}(\mu\text{s})$	2.6	2.6	2.6	2.6	2.6	2.6	2.6
$I_{\text{bias\_predict}}(\text{nA})$	263	263	263	263	263	263	263
$L1(\mu\text{m})$	0.54	0.54	0.54	0.54	0.54	0.54	0.54
$L2(\mu\text{m})$	0.5	1.0	2.0	2.0	5.0	5.0	7.9
$W1(\mu\text{m})$	1.61	1.61	1.61	1.61	1.61	1.61	1.61
$W2(\mu\text{m})$	1.18	1.78	2.59	0.41	5.15	0.868	0.4
$V_{\text{gs2}}(\text{V})$	0.4	0.4	0.4	0.5	0.4	0.5	0.6
$CL(\text{pF})$	6.0	6.0	6.0	6.0	6.0	6.0	6.0
$T_{\text{settling\_HSPICE}}(\mu\text{s})$	3.17	2.84	2.68	2.39	2.54	2.66	2.48
$RE(\%)$	18.02	8.30	3.03	9.00	2.28	2.51	4.86
$I_{\text{bias\_HSPICE}}(\text{nA})$	201	231	249	288	266	251	275

bias current by a factor of 4.7.

Synthesized source followers are simulated in HSPICE for evaluation. The bias current ( $I_{\text{bias\_HSPICE}}$ ) and settling time ( $T_{\text{settling\_HSPICE}}$ ) measured from the HSPICE simulation are also listed in Table 4.2. The relative error (RE) between the predicted and simulated settling time is presented as well. We notice that the large error in the settling time is caused by the large error in the bias current. By increasing the bias voltage ( $V_{\text{gs2}}$ ) of  $M2$ , the bias current can be increased accordingly, and the settling time can be improved as well. For example, if  $V_{\text{gs2}}$  changes from 0.4 V to 0.412 V in case 1, the bias current can increase from 200.7 nA to 264.6 nA, and the settling time changes from 3.17  $\mu\text{s}$  to 2.55  $\mu\text{s}$ . The relative settling error for all 144 synthesized

circuits is presented in Figure 4.5. The maximum relative error is 21.7%. Only 19

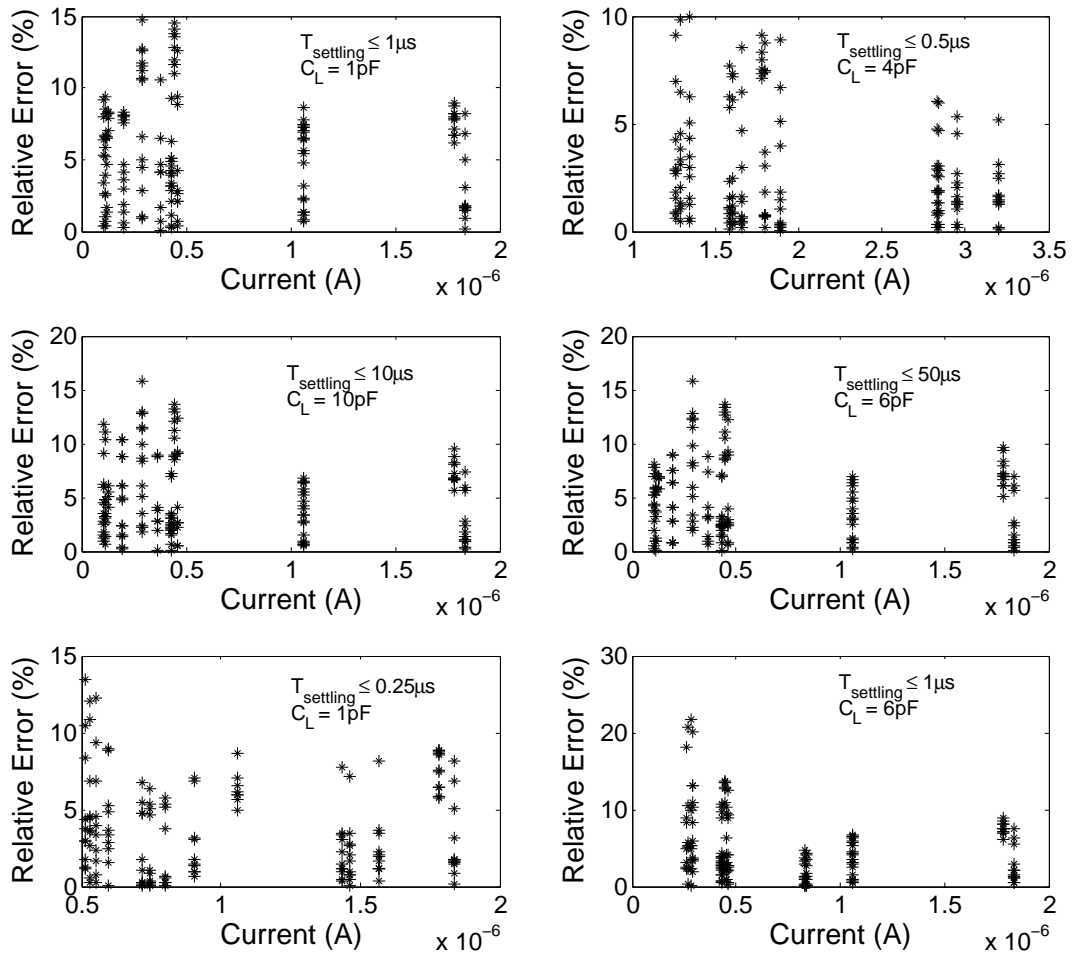


**Figure 4.5:** Relative settling error versus bias current

out of 144 synthesized results have a relative error larger than 10%. These can be improved by setting the practical bias current close to the predicted value.

### 4.3 Synthesis Evaluation

This synthesis method is tested for a variety of design requirements. Six random design requirements are used for synthesizing power optimal designs, and settling times from synthesized circuits are compared with results from the HSPICE simulation. The relative errors between the predicted and simulated settling times are depicted in Figure 4.6. We can see that the maximum relative error is less than

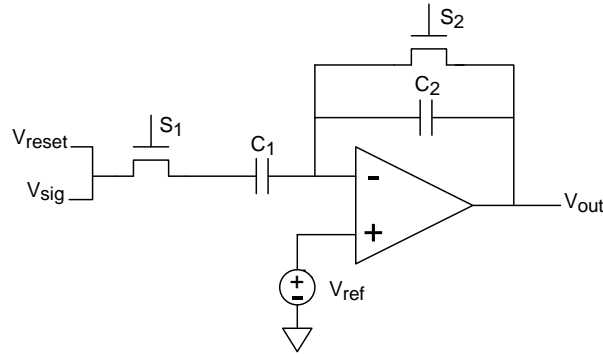


**Figure 4.6:** Source follower synthesis evaluation with six random requirements

25%, and relative errors in most cases are less than 10%. Instead of using a constant bias current, a bias transistor is used in the HSPICE test bench. The synthesized gate-to-source voltage is applied to the bias transistor, which may lead to a different bias current from the synthesized values, resulting in a larger relative error. This can be improved by changing the bias voltage of  $M2$ . This adjustment is less than 4% in the worst case. We can conclude that this synthesis method can give a reliable result in a practical design.

## 5 Power optimization in CDS op-amp

The correlated double sampling circuit is a very important block in CMOS image sensors. It is used to reduce KTC noise,  $1/f$  noise, and fixed pattern noise from the pixel [Mend 97]. Switched-capacitor-based CDS has been widely used in many applications [Deck 98, Dege 00, Taka 07, Cho 03] due to its better noise control [Dege 00]. A column-wise CDS usually gives a good trade-off between speed and power consumption. A typical switched-capacitor CDS consists of an op-amp, capacitors, and switches as shown in Figure 5.1. The key component in the CDS is the op-amp which



**Figure 5.1:** Switched-capacitor-based CDS

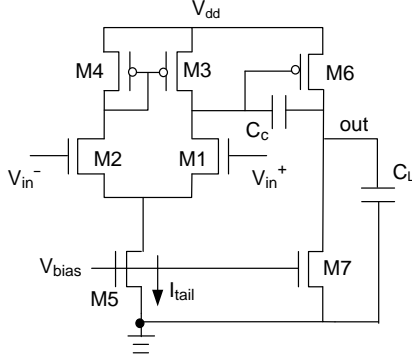
consumes most of the power. Op-amps with different architectures can lead to different power consumption for the same performance specifications. In this chapter, three different op-amps synthesized using the geometric programming method will be first examined from the power consumption perspective. A low-power CDS op-amp will then be designed according to a practical requirement and tested experimentally.

## **5.1 Op-amp power optimization with geometric programming**

To investigate the effect of op-amp type on power consumption, three different op-amps are selected for study. They are a two-stage Miller op-amp, a conventional symmetrical operational transconductance amplifier (CS OTA), and a common-mode feedback (CMFB) OTA. For a typical CMOS APS image sensor, the switched-capacitor-based CDS op-amp usually has a moderate requirement on the gain and speed. The three Op-amps selected here can easily meet this requirement.

### **5.1.1 Two-stage Miller op-amp power optimization**

The two-stage Miller op-amp in Figure 5.2 has been studied by many researchers in their GP synthesis methods [Mand 01, Hers 01, Agui 08, Kim 04]. It is revisited here in order to address sub-space-based optimization synthesis. To solve the



**Figure 5.2:** Two-stage Miller compensated op-amp

GP optimization problem, circuit performance measures have to be modeled with a posynomial/monomial form, and they must have a GP constraint. Model equations for open-loop gain ( $A_V$ ), gain-bandwidth product ( $GBW$ ), phase margin ( $PM$ ), and slew rate ( $SR$ ) are similar to those in [Hers 01]. Table 5.1 lists models and constraints for these four major performance measures. If  $1/g_m$  is used in models, the gain  $A_V$  and  $GBW$  are inverse posynomial (i.e.,  $1/A_V$  and  $1/GBW$  are posynomials). The slew rate  $SR$  is an inverse posynomial. The phase shift ( $PSH$ ) induced by the dominant pole ( $p1$ ) is usually around 90 degrees, so only the output pole ( $p2$ ), mirror pole ( $p3$ ), and zero ( $z1$ ) will be included in the phase margin calculation. When the phase shift contributed by each pole/zero is less than 25 degrees,  $\tan^{-1}x \approx x$ , the phase shift can be modeled in a posynomial form. Now the circuit performance measures can be effectively constrained with GP constraints. In Table 5.1,  $A_{V\_spec}$ ,  $GBW_{spec}$ ,

**Table 5.1:** Two-stage op-amp models and constraints

Perform.	Models	Constraints
$A_v$	$\frac{g_{m1}}{g_{ds1}+g_{ds3}}$	$\frac{A_{V\_spec}}{A_V} \leq 1$
$GBW$	$\frac{g_{m1}}{2\pi C_c}$	$\frac{GBW_{spec}}{GBW} \leq 1$
$SR$	$\min\left(\frac{I_{tail}}{C_c}, \frac{I_{ds6}}{C_c+C_{out}}\right)$	$\frac{SR_{spec}}{SR} \leq 1$
$p1$	$\frac{g_{m1}}{2\pi A_V C_c}$	
$p2$	$\frac{g_{m6}}{2\pi(C_L+C_1+C_2)}$	
$p3$	$\frac{g_{m4}}{2\pi C_3}$	
$z1$	$\frac{g_{m6}}{2\pi C_c}$	
$PSH$	$\tan^{-1}\left(\frac{GBW}{p1}\right) + \tan^{-1}\left(\frac{GBW}{p2}\right) + \tan^{-1}\left(\frac{GBW}{p3}\right) + \tan^{-1}\left(\frac{GBW}{z1}\right)$	
$PM$	$180^\circ - PSH$	$\frac{PSH}{180^\circ - PM} \leq 1$

$$C_1 = C_{db7} + C_{dg7} + C_{db6} + C_{gs6}, \quad C_2 = C_{db1} + C_{dg1} + C_{db3} + C_{dg3}$$

$$C_3 = C_{gs4} + C_{gs3} + C_{dg3} + C_{dg1} + C_{db1}, \quad C_{out} = C_L + C_{db7} + C_{dg7} + C_{db6} + C_{dg6}$$

$SR_{spec}$ , and  $PM_{spec}$  are specifications for  $A_V$ ,  $GBW$ ,  $SR$ , and  $PM$ , respectively. The same bias constraints used in [Hers 01] are employed in this GP synthesis.

As it has been mentioned before, five design variables ( $L$ ,  $W$ ,  $V_{gs}$ ,  $V_{ds}$ , and  $I_{ds}$ ) are used in our sub-space-based models for  $g_m$ ,  $1/g_m$ , and  $g_{ds}$ . Ideally, design variables in posynomial models should be independent. In fact,  $I_{ds}$  is dependent on other four variables. To ensure  $I_{ds}$  has the right value in the solution, a constraint on  $I_{ds}$  has to be added:

$$I_{ds\_cal} = I_{ds} \tag{5.1}$$

where  $I_{ds\_cal}$  is the drain-to-source current calculated from a monomial model, and  $I_{ds}$  is a design variable.

To achieve an efficient search, the search spaces for each transistor must be constrained according to the performance requirements. This circuit has a symmetrical structure in the input stage ( $M1$ ,  $M2$ ,  $M3$ ,  $M4$ , and  $M5$ ), so only a half part ( $M1$ ,  $M3$ , and  $M5$ ) needs to be considered. Hence, the total number of active transistors is reduced to five. The common-mode input range also exerts a constraint on the searching space. The minimum and maximum common-mode input voltages are:

$$V_{com\_min} = V_{dsat5} + V_{gs1} \quad (a)$$

$$V_{com\_max} = V_{dd} + V_{gd1\_max} - V_{sd3} = V_{dd} + V_{gs1} - V_{dsat1} - V_{sd3} \quad (b) \quad (5.2)$$

where  $V_{dsat1}$  and  $V_{dsat5}$  are drain-to-source voltage at which  $M1$  and  $M5$  saturate respectively, and  $V_{gd1\_max}$  is the maximum gate-to-drain voltage to keep  $M1$  in the saturation region. If we substitute Equation 5.2(a) into Equation 5.2(b), we get

$$V_{com\_max} = V_{dd} + V_{com\_min} - V_{dsat5} - V_{dsat1} - V_{sd3} \quad (5.3)$$

To have  $V_{com\_max} = V_{dd}$ ,  $V_{sd3}$  must satisfy

$$V_{sd3} = V_{com\_min} - V_{dsat5} - V_{dsat1} \quad (5.4)$$

In our defined search spaces (as shown in Table 3.2), the minimum  $V_{dsat}$  is  $\sim 0.11$  V and therefore  $V_{sd3} < (V_{com\_min} - 0.22 \text{ V})$ . For example, if  $V_{com\_min} = 0.7\text{V}$ , then

$V_{sd3} < 0.48V$ . We know that  $V_{sg3} = V_{sd3}$ , so the transistor  $M3$  should work in the moderate inversion region. The search spaces for  $M3$  are therefore sub-spaces mod1, mod4, mod7, and mod10. The source-to-gate voltage in  $M6$  is the same as  $M3$ , so the search spaces for  $M6$  are the same as for  $M3$ . From Equation 5.2(a), we know that

$$V_{dsat5} = V_{com\_min} - V_{gs1} \quad (5.5)$$

and we have  $V_{dsat5\_max} = V_{com\_min} - V_{gs1\_min} = V_{com\_min} - 0.4 V$ . If  $V_{com\_min} = 0.7 V$ , then  $V_{dsat5\_max} = 0.3 V$ . To have  $V_{dsat5} \leq 0.3 V$ ,  $V_{gs5}$  should satisfy  $0.4 V \leq V_{gs5} \leq 0.7 V$ . Therefore the search spaces for  $M5$  are mod1 – mod12, st1, st9, st17, st25, and st33. A transistor working in strong inversion has a larger  $V_{dsat}$ . If a small  $V_{com\_min}$  is wanted, the search spaces for  $M5$  can be restricted to the moderate inversion region (mod1 – mod12).  $M7$  has the same gate-to-source voltage as  $M5$ ; thus  $M5$  and  $M7$  can be searched in the same spaces at the same time.

From Equation 5.2 (a), we can have

$$V_{gs1} = V_{com\_min} - V_{dsat5} \quad (5.6)$$

That means  $0.4 V \leq V_{gs1} \leq (V_{com\_min} - 0.11 V)$ . If  $V_{com\_min} = 0.7 V$ , then  $0.4 V \leq V_{gs1} \leq 0.59 V$ ; therefore the search spaces for  $M1$  are mod1, mod2, mod4, mod5, mod7, mod8, mod10, and mod11. As a result, the total search effort for the two-stage Miller op-amp is  $8 (M1) \times 4 (M3) \times 12 (M5, M7) \times 4 (M6) = 1536$

iterations. A MATLAB based GP solver [Muta 08] is used to solve this optimization problem. It is running on a Xeon workstation with 8 processors at 2.66 GHz. The average computing time for an iteration with one processor is 5 minutes. If we use 8 processors at the same time, it will take 0.67 days to finish the search. This computation effort is acceptable as long as the accuracy of the synthesis is acceptable.

In this two-stage op-amp design, the sum of drain-source voltages along each signal path (from  $V_{dd}$  to ground) equals  $V_{dd}$ , which is called a posynomial equality. As we know from the definition of geometric programming, a posynomial equality is invalid in GP optimization (see section 2.3). Several methods can be used to deal with this posynomial equality. Vanderhaegen, *et al.* [Vand 04] suggested a branch-and-bound algorithm to replace the posynomial equality with a set of posynomial inequalities. GP is solved for each posynomial inequality, and the minimum of all these GPs will be the global optimum. Mandal, *et al.* [Mand 01] introduced an extra procedure to search the DC operating point. Kim [Kim 04] relaxed the posynomial equality to a posynomial inequality. If the relaxed posynomial constraint and the objective can vary with a desired design variable in an opposite monotonic direction, the posynomial equality can be met at the solution. The first two methods need more than one GP; therefore, more computational effort is needed. The third method can solve the problem within the same GP, and it is more computationally

efficient. Therefore, a similar scheme to [Kim 04] is used here to solve our GP power optimization problem. The objective in this problem is:

$$obj = \min(w1(I_{tail} + I_{out}) + w2(1/V_{ds5}) + w3(1/V_{ds7})) \quad (5.7)$$

where  $w1$ ,  $w2$ , and  $w3$  are weights, and they are positive real numbers. Weight  $w1$  usually takes a large value compared to  $w2$  and  $w3$  to reduce the effect of the second and third terms on the power optimization. In this case,  $w1 = 10^6$ , and  $w2 = w3 = 1$ . The same power consumption is observed when the GP is solved from Equation 5.7 and from the objective without the second and third terms. The relaxed posynomial equality in the input stage is

$$V_{ds3} + V_{ds1} + V_{ds5} \leq V_{dd} \quad (5.8)$$

We can see that left side of Equation 5.8 increases as  $V_{ds5}$  increases, but Equation 5.7 decreases as  $V_{ds5}$  increases. Other performance constraints will not be affected by  $V_{ds5}$ . As a result, the equality in Equation 5.8 will be always be satisfied. In the outer signal path, the sum of  $V_{ds}$  is

$$V_{ds6} + V_{ds7} \leq V_{dd} \quad (5.9)$$

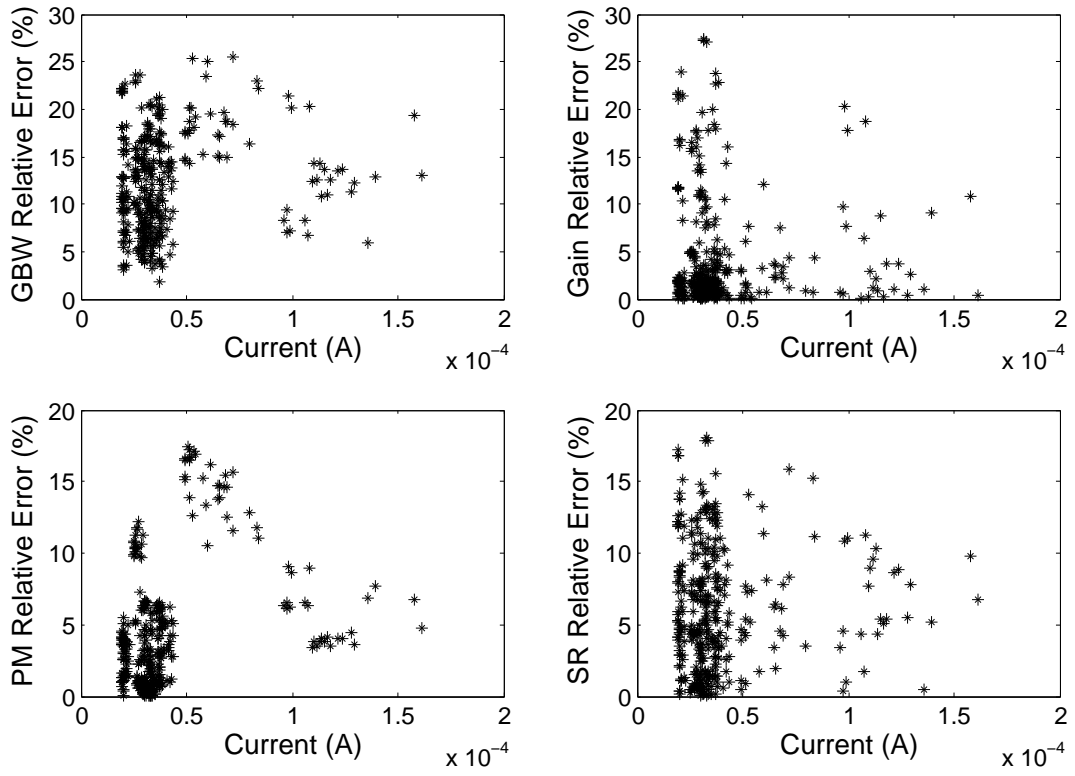
We can also see that left side of Equation 5.9 increases with  $V_{ds7}$ , but Equation 5.7 decreases with  $V_{ds7}$ . Other performance constraints, such as  $A_V$ ,  $SR$ , and  $PM$ , are non-decreasing with  $V_{ds7}$ . Therefore, the equality in Equation 5.9 will always be met.

With constrained search spaces, GP based local power optimization synthesis is performed in each subset of search spaces. In fact, several designs (different transistor sizes and biases) could result in the same minimum static power consumption. In this case, the one with the minimum area overhead will be selected as the solution.

A two-stage op-amp is synthesized with the models and constraints mentioned above. The synthesis gives sizes and biases for all transistors, and the predicted performance of the circuit. The sizes and biases of transistors achieved from the GP synthesis are then input to a HSPICE test bench to verify the performance predicted from the synthesis. To evaluate the sub-space modeling based GP synthesis technique, the design specifications shown in Table 5.2 are applied. Synthesized results from all feasible design spaces are compared to HSPICE simulation results to calculate relative errors of four major performance measures ( $GBW$ ,  $A_V$ ,  $SR$ , and  $PM$ ), which are depicted in Figure 5.3.

It is noticed that relative errors of these four major performance measures are less than 30%. The majority of cases are less than 20%. The large errors in  $GBW$  are caused either by the larger error in bias current or the multi-pole effect. The synthesized bias voltage applied in the HSPICE simulation may generate a very different bias current from the synthesized bias current, as a result, a large error in  $GBW$  can be caused. The error caused by this factor can be reduced by keeping the bias current

in the HSPICE simulation as close to the synthesized bias current as possible. In this GP synthesis problem, the effect of non-dominant poles is ignored because it is assumed to be much larger than  $GBW$ . In HSPICE simulation, this assumption may not be true in some cases, and the multi-pole effect becomes important. The error caused by this factor should be further verified by experimental results. The large errors in  $SR$  are caused by the large error in bias current. Large errors in gain are attributed to a combined effect of two stages; combined gain errors in each stage will be compounded to lead to a large overall error. The large errors in  $PM$  are caused by the large error in non-dominant poles. The power optimal solution is presented in Table 5.3, in which predicted performance results (Predic.) are compared with the equivalent results from HSPICE simulation. The relative errors (RE) between these two results are presented as well, and they are in a very good agreement. Note, however, that these simulations do not include layout effects. Compared with designs using GP in long-channel length technologies [Hers 01, Mand 01], and with a design using GP in a single strong inversion region in  $0.18 \mu\text{m}$  technology [Kim 04], the sub-space-based GP design in an entire design space can give an acceptable accuracy, which can be shown in Table 5.4.



**Figure 5.3:** Relative performance errors in the two-stage Miller op-amp

**Table 5.2:** Op-amp design specifications

$A_V$	PM	GBW	SR	$C_L$
> 40 dB	> 60°	> 10 MHz	> 5.7 V/ $\mu$ s	1 pF

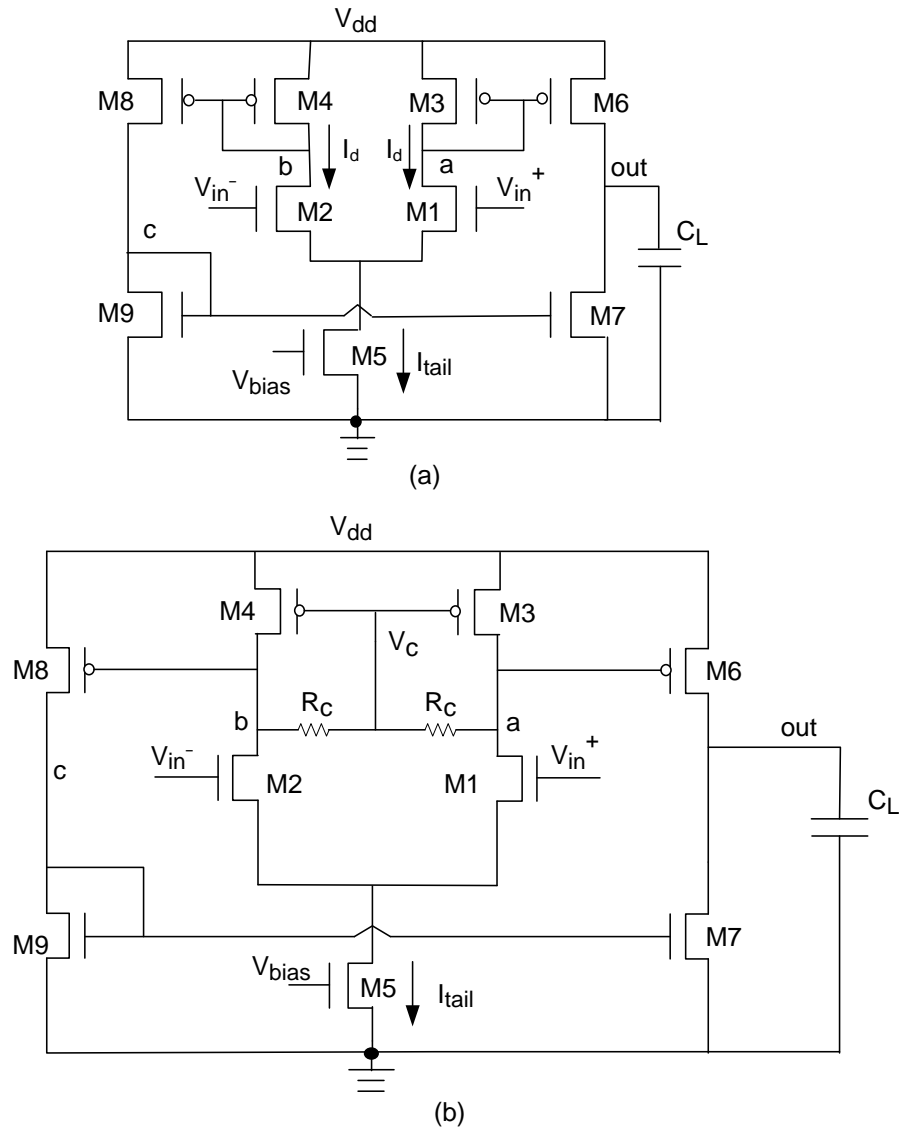
**Table 5.3:** Two-stage Op-amp design verification with HSPICE

Performance	Spec.	Predic.	HSPICE	RE
$A_V$	$\geq 40$ dB	76.7 dB	80 dB	4.1%
GBW	$\geq 10$ MHz	10 MHz	9.04 MHz	10.1%
SR	$\geq 5.7$ V/ $\mu$ s	6.52 V/ $\mu$ s	6.16 V/ $\mu$ s	5.8%
PM	$\geq 70^\circ$	70°	65°	7.7%
$I_{static}$	minimum	15.89 $\mu$ A	16.75 $\mu$ A	5.1%

### 5.1.2 Symmetrical OTA power optimization

A symmetrical OTA design is very attractive in high-speed and low-power applications [Lope 08]. A conventional symmetrical (CS) OTA and a common-mode feedback (CMFB) OTA, as shown in Figure 5.4, will be considered in this study. Model equations and constraints on  $A_V$ ,  $GBW$ ,  $SR$ , and  $PM$  of these two op-amps are listed in Tables 5.5 and 5.6, respectively.

In the CS OTA,  $A_V$ ,  $GBW$ , and  $SR$  are inverse posynomials. The phase shift can be treated in the same way as the two-stage op-amp, and can be approximated as a posynomial. As such, the circuit performances can be effectively constrained using GP constraints. A similar method to the two-stage op-amp design is used in



**Figure 5.4:** Conventional symmetrical OTA (a) and CMFB OTA (b)

**Table 5.4:** GP-based design comparisons

Relative error \ Designs	Long-channel	Long-channel	Strong-inversion	Sub-space
	[Hers 01]	[Mand 01]	<b>0.18<math>\mu\text{m}</math></b> [Kim 04]	
$RE_{AV}$	3.6%	2.6%	4.95%	4.1%
$RE_{GBW}$	2.7%	5.3%	10.6%	10.1%
$RE_{PM}$	3.2%	2.9%	9.7%	7.7%
$RE_{SR}$	2.1%	5.3%	--	5.3%

this GP synthesis to converge the sum of  $V_{ds}$  to  $V_{dd}$ . The objective function in this design is:

$$obj = \min(w1(I_{tail} + I_{out}) + w2(1/V_{ds5}) + w3(1/V_{ds6})) \quad (5.10)$$

where  $w1$ ,  $w2$ , and  $w3$  are weights with positive values. Search spaces for each transistor can be constrained effectively with the same scheme as used in the two-stage op-amp design.

In the CMFB OTA,  $A_V$  and  $GBW$  are inverse posynomials. The phase shift can be approximated by a posynomial. As a result,  $A_V$ ,  $GBW$ , and the phase shift can be constrained with GP constraints. Unlike the CS OTA, the slew rate in the CMFB OTA is independent of the static current in the output path. During the slew, the source-gate voltage at M6 is adaptively adjusted by the common-mode feedback resistor  $R_c$ . The slew current ( $I_{slew\_cal}$ ) drawn from M6 is the current when  $V_{sg6} = V_{sg3} + 1/2(I_{tail}R_c)$ . The slew rate cannot be constrained with a GP constraint

**Table 5.5:** Conventional symmetrical OTA models and constraints

Perform.	Models	Constraints
$A_v$	$\frac{g_{m1}}{g_{m3}} \frac{g_{m6}}{g_{ds6} + g_{ds7}}$	$\frac{A_{v\_spec}}{A_v} \leq 1$
$GBW$	$\frac{g_{m6}}{2\pi C_{out}} \frac{g_{m1}}{g_{m3}}$	$\frac{GBW_{spec}}{GBW} \leq 1$
$SR$	$\frac{2I_{ds6}}{C_{out}}$	$\frac{SR_{spec}}{SR} \leq 1$
$p1$	$\frac{g_{ds6} + g_{ds7}}{2\pi C_{out}}$	
$p2$	$\frac{g_{m3}}{2\pi C_a}$	
$p3$	$\frac{g_{m9}}{2\pi C_c}$	
$PSH$	$\tan^{-1}\left(\frac{GBW}{p1}\right) + \tan^{-1}\left(\frac{GBW}{p2}\right) + \tan^{-1}\left(\frac{GBW}{p3}\right)$	
$PM$	$180^0 - PSH$	$\frac{PSH}{180^0 - PM} \leq 1$

$$C_a = C_{db1} + C_{dg1} + C_{db3} + C_{gs3} + C_{gs6} + C_{gd6}, \quad C_c = C_{db6} + C_{dg6} + C_{db7} + 2C_{gs7} + C_{gd7}$$

$$C_{out} = C_L + C_{db7} + C_{dg7} + C_{db6} + C_{dg6}$$

if the slew current  $I_{slew\_cal}$  is used in the model. To solve GP for CMFB OTA, two new extra variables are introduced; they are slew current ( $I_{slew}$ ) and slewing source-gate voltage of M6 ( $V_{sg6\_slew}$ ).  $I_{slew}$  is a monomial with input variables of  $V_{sg6\_slew}$ ,  $V_{ds6}$ ,  $L_6$ , and  $W_6$ . The slew rate can now be constrained with a GP constraint if  $I_{slew}$  is used in the model. A GP can be solved by adding the following objective and constraint:

$$\begin{aligned}
 obj &= \min(w1(I_{tail} + I_{out}) + w2I_{slew} + w3(1/V_{ds5}) + w4(1/V_{ds6})) \\
 constraint &: \frac{I_{slew\_cal}}{I_{slew}} \leq 1
 \end{aligned} \tag{5.11}$$

**Table 5.6:** CMFB OTA models and constraints

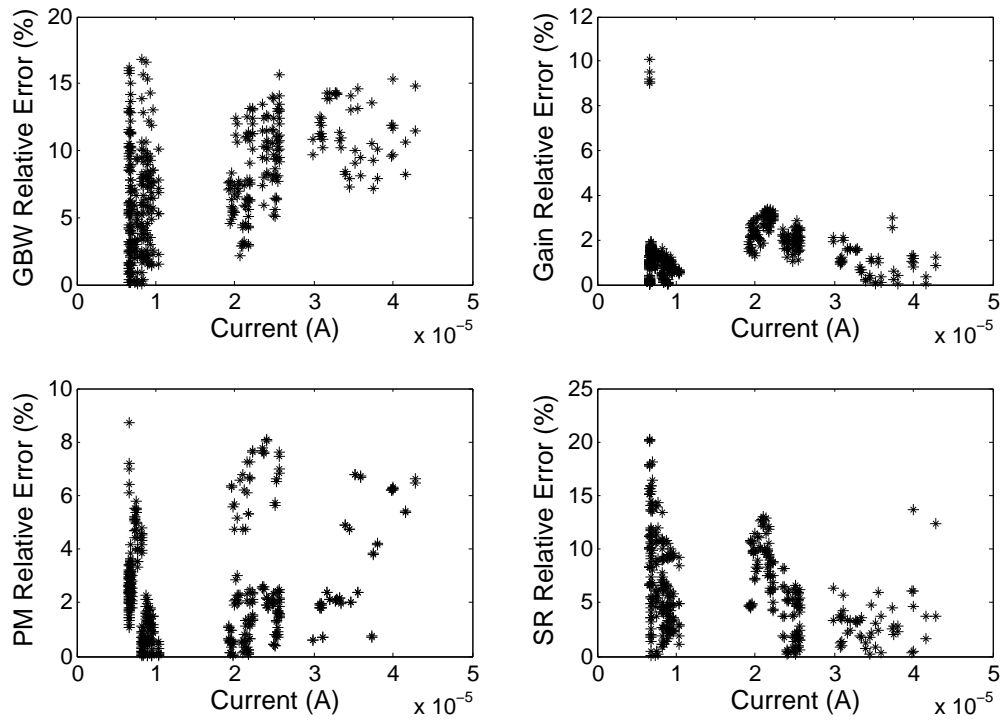
Perform.	Models	Constraints
$A_v$	$\frac{g_{m1}}{g_{ds1}+g_{ds3}+g_{Rc}} \frac{g_{m6}}{g_{ds6}+g_{ds7}}$	$\frac{A_{V\_spec}}{A_v} \leq 1$
$GBW$	$\frac{g_{m1}g_{m6}}{2\pi C_{out}} R_C$	$\frac{GBW_{spec}}{GBW} \leq 1$
$SR$	$\frac{I_{slew}}{C_{out}}, I_{slew} = I_{ds6}$ when $V_{sg6} = V_{sg3} + \frac{1}{2}I_{tail}R_C$	
$p1$	$\frac{g_{ds6}+g_{ds7}}{2\pi C_{out}}$	
$p2$	$\frac{g_{ds1}+g_{ds3}+R_C}{2\pi C_a}$	
$p3$	$\frac{g_{m9}}{2\pi C_c}$	
$PSH$	$\tan^{-1}\left(\frac{GBW}{p1}\right) + \tan^{-1}\left(\frac{GBW}{p2}\right) + \tan^{-1}\left(\frac{GBW}{p3}\right)$	
$PM$	$180^0 - PSH$	$\frac{PSH}{180^0 - PM} \leq 1$

$$C_a = C_{db1} + C_{dg1} + C_{db3} + C_{gs3} + C_{gs6} + C_{gd6}, C_c = C_{db6} + C_{dg6} + C_{db7} + 2C_{gs7} + C_{gd7}$$

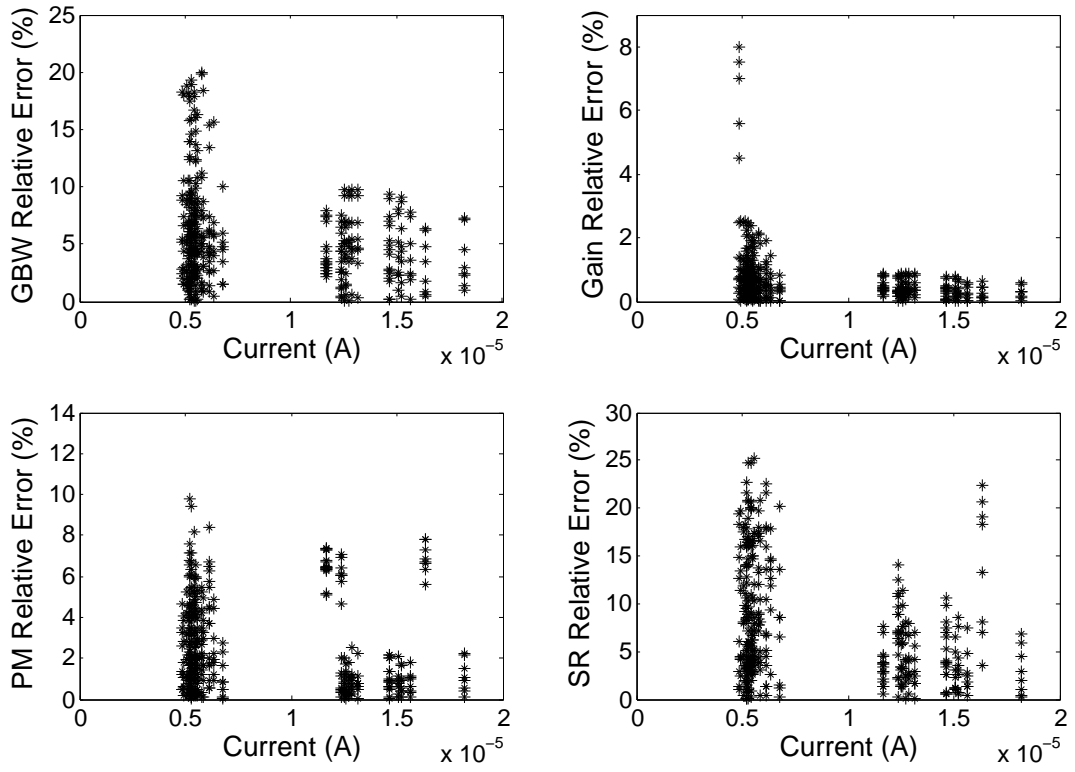
$$C_{out} = C_L + C_{db7} + C_{dg7} + C_{db6} + C_{dg6}$$

where  $w1, w2, w3,$  and  $w4$  are weights with positive values.  $I_{slew}$  is the slew current derived from the slew requirement, and  $I_{slew\_cal}$  is the slew current calculated from  $I_{ds6}$  when  $V_{sg6} = V_{sg3} + \frac{1}{2}(I_{tail}R_c)$ . In this case, we let  $w(2, 3, 4) \ll w1$  to reduce the effect of the 2nd–4th terms on the power optimization. Here a relaxed constraint is used. The objective function is intended to reduce  $I_{slew}$  (the second term in the objective). As long as the power is dominated by the gain-bandwidth product requirement, the equality of the the constraint in Equation 5.11 will always be met. Otherwise, the requirement on the gain-bandwidth should be increased accordingly.

The method used for the CS OTA can also be applied to the CMFB OTA to converge the sum of  $V_{ds}$  to  $V_{dd}$ . The search spaces for each transistor can also be limited with the same scheme as for the two-stage op-amp design. Power optimization designs in CS and CMFB OTAs are synthesized with the models and constraints mentioned above. To evaluate the sub-space modeling based GP synthesis technique, we use the same design requirements as for the two-stage Miller op-amp. Synthesized results from all feasible design spaces are compared to HSPICE simulation results to calculate relative errors of four major performance measures, which are depicted in Figures 5.5 and 5.6. A reasonably good accuracy is observed in these two design cases. Large errors in  $GBW$  and  $SR$  are caused by the large error in the bias current, which can be improved by adjusting  $V_{bias}$ . In each design case, one of the solutions with the minimum power consumption and area overhead is selected as the optimal solution. The performance of the optimized design in each design case is presented in Tables 5.7 and 5.8 respectively. Compared to the performance achieved from the HSPICE simulation, the sub-space modeling based GP can allow an analog circuit to be synthesized with an acceptable accuracy. It is also noted that CMFB OTA is the most power-efficient structure among these three architectures.



**Figure 5.5:** Relative performance errors in the CS OTA



**Figure 5.6:** Relative performance errors in the CMFB OTA

**Table 5.7:** The performance of the optimized design of the conventional symmetrical OTA

Performance	Spec.	Predic.	HSPICE	RE
$A_V$	$\geq 40 \text{ dB}$	40.2 dB	44.7 dB	10.1%
GBW	$\geq 10 \text{ MHz}$	10.2 MHz	10.6 MHz	3.8%
SR	$\geq 5.7 \text{ V}/\mu\text{s}$	5.7 $\text{V}/\mu\text{s}$	6.04 $\text{V}/\mu\text{s}$	5.6%
PM	$\geq 70^\circ$	$75^\circ$	$69^\circ$	8.7%
$I_{static}$	minimum	6.69 $\mu\text{A}$	6.86 $\mu\text{A}$	2.5%

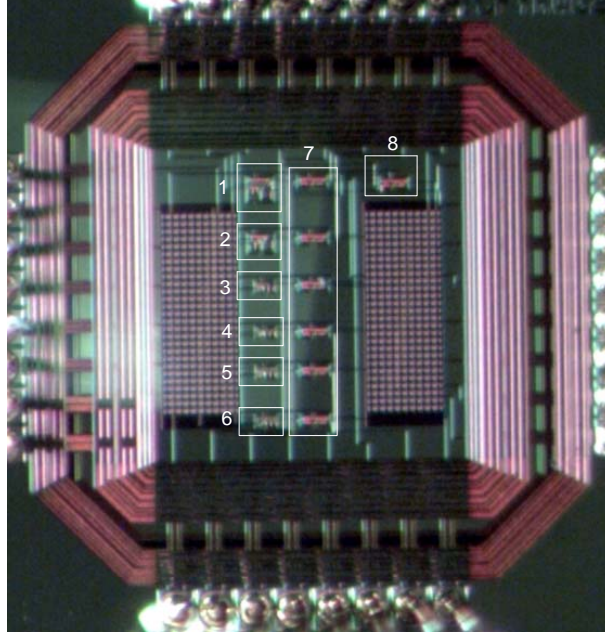
**Table 5.8:** The performance of the optimized design of the CMFB OTA

Performance	Spec.	Predic.	HSPICE	RE
$A_V$	$\geq 40 \text{ dB}$	49.8 dB	46.1 dB	8%
GBW	$\geq 10 \text{ MHz}$	10 MHz	10.4 MHz	3.8%
SR	$\geq 5.7 \text{ V}/\mu\text{s}$	9.21 $\text{V}/\mu\text{s}$	9.57 $\text{V}/\mu\text{s}$	3.8%
PM	$\geq 70^\circ$	$70^\circ$	$69^\circ$	1.4%
$I_{static}$	minimum	4.63 $\mu\text{A}$	5.08 $\mu\text{A}$	8.9%

## 5.2 Experimental verification of synthesized results

The synthesis evaluation with HSPICE does not consider the layout and practical fabrication process factors. To evaluate how well the suggested method can predict the performance of an op-amp in a practical design, some synthesized op-amps were fabricated for testing. The three types of op-amps described above are considered in this study. Two op-amps from each type are synthesized in both moderate and strong inversion regions. All of these six op-amps have the same design specifications as mentioned above. Figure 5.7 is a micro-photograph of the op-amp chip, which is

designed using TSMC 0.18  $\mu\text{m}$  technology. Blocks 1 and 2 in Figure 5.7 are two-



**Figure 5.7:** A micro-photograph of the op-amp chip

stage Miller op-amps, blocks 3 and 4 are CS OTAs, blocks 5 and 6 are CMFB OTAs, and blocks 7 and 8 are output buffers. For the CMFB OTAs, the feedback resistors ( $R_c$ ) are replaced with PMOS transistors which operate in the triode region. Three measurements are performed to extract performance characteristics of these op-amps. They are DC, frequency response, and transient measurements. In DC measurement, each op-amp was biased at synthesized values (bias voltages), and the static power consumption was measured. The results of the static power consumption from the synthesis, HSPICE simulation, and the measurement are presented in Table 5.9. We

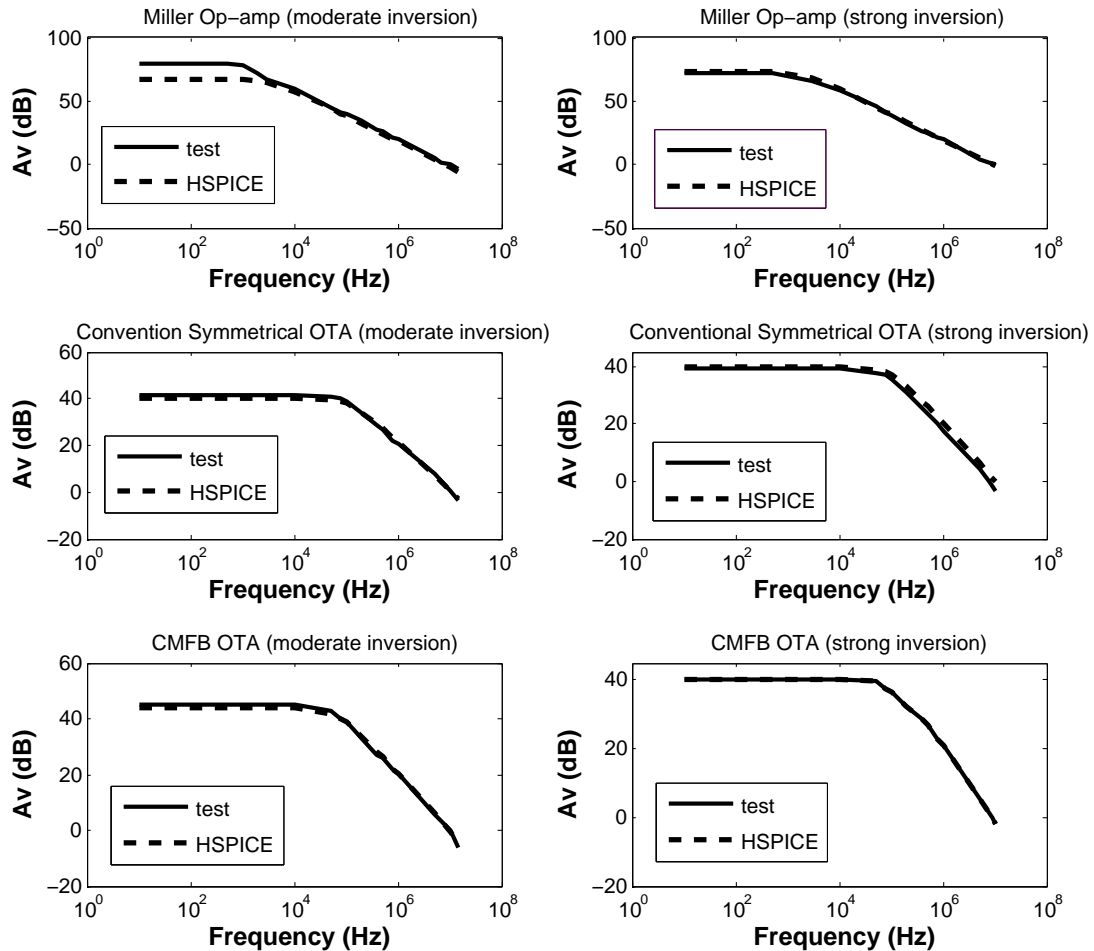
notice that measurement results have a good agreement with predicted results.

**Table 5.9:** Static power measurements comparison

Op-Amp	$P_{\text{static.predict}}$	$P_{\text{static.HSPICE}}$	$P_{\text{static.test}}$
Miller (moderate)	30.74 $\mu W$	29.27 $\mu W$	30.62 $\mu W$
Miller (strong)	39.78 $\mu W$	40.09 $\mu W$	39.96 $\mu W$
CS OTA (moderate)	12.99 $\mu W$	14.85 $\mu W$	14.22 $\mu W$
CS OTA (strong)	26.46 $\mu W$	30.19 $\mu W$	25.68 $\mu W$
CMFB OTA (moderate)	11.29 $\mu W$	12.67 $\mu W$	11.92 $\mu W$
CMFB OTA (strong)	16.74 $\mu W$	20.66 $\mu W$	18.36 $\mu W$

Frequency response measurements were used to measure DC gain, gain-bandwidth product, and phase margin. A sinusoidal signal was applied to the positive input node of each op-amp. A DC signal equal to the DC offset of the sinusoidal signal was applied to the negative node of the op-amp. The frequency of the sinusoidal signal was varied from 10Hz to 15MHz. The frequency response curves (gain versus frequency) of these six op-amps are shown in Figure 5.8, in which HSPICE simulation results are presented for comparison. The final test results of AC performance are depicted in Table 5.10. We notice that synthesized results have a good agreement with test results. We also notice that HSPICE simulation gives unrealistically low values in gain and  $GBW$  in two-stage Miller op-amp when it works in the moderate inversion region. This is because the HSPICE simulation is using the BSIM3 model, and it does not support moderate inversion very well. Unlike CS and CMFB

OTA, the Miller op-amp has two stages. A large overall error will be generated if a significant error occurs in each stage.



**Figure 5.8:** Frequency response of the six studied op-amps

The transient measurement was used to determine the slew rate of an op-amp. A step signal was applied to the positive node of the op-amp, and a DC voltage signal was applied to the negative node of the op-amp. The slew rate results from

**Table 5.10:** AC Performance measurements comparison

<b>Perf.</b>	<b>Miller (moderate)</b>	<b>Miller (strong)</b>	<b>CS-OTA (moderate)</b>	<b>CS-OTA (strong)</b>	<b>CMFB-OTA (moderate)</b>	<b>CMFB-OTA (strong)</b>
$A_{V\_syn}$	78.7 dB	70.7 dB	40 dB	40 dB	43.9 dB	40 dB
$A_{V\_HSPICE}$	66.9 dB	73.2 dB	40.4 dB	40.3 dB	45.4 dB	40.9 dB
$A_{V\_test}$	80 dB	72 dB	41.6 dB	39.3 dB	45.1 dB	40 dB
$GBW_{syn}$	10 MHz	10 MHz	10.4 MHz	10 MHz	10 MHz	10 MHz
$GBW_{HSPICE}$	7.4 MHz	8.9 MHz	11.8 MHz	10.9 MHz	11.1 MHz	11 MHz
$GBW_{test}$	10 MHz	8 MHz	10 MHz	8 MHz	10 MHz	9 MHz
$PM_{syn}$	70°	70°	76.7°	77.1°	70°	70°
$PM_{HSPICE}$	71°	67°	77°	78°	65°	64°
$PM_{test}$	82.6°	70.6°	75.6°	82.1°	75.6°	69.8°

the synthesis, HSPICE simulation, and the measurement are shown in Table 5.11. It is noticed again that the sub-space modeling based GP synthesis can lead to a good reliability in analog circuit design in sub-micron technologies.

**Table 5.11:** Slew rate measurements comparison

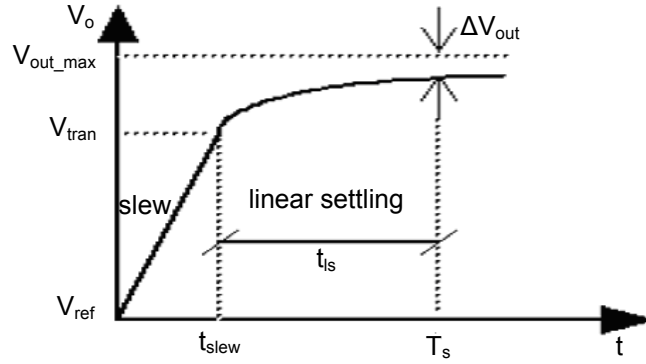
<b>Op-Amp</b>	<b>SR<sub>syn</sub></b>	<b>SR<sub>HSPICE</sub></b>	<b>SR<sub>test</sub></b>
Miller (moderate)	6.5 V/ $\mu$ s	6.9 V/ $\mu$ s	6.4 V/ $\mu$ s
Miller (strong)	10.7 V/ $\mu$ s	13.2 V/ $\mu$ s	11.6 V/ $\mu$ s
CS OTA (moderate)	5.7 V/ $\mu$ s	5.9 V/ $\mu$ s	5.3 V/ $\mu$ s
CS OTA (strong)	11.6 V/ $\mu$ s	11.9 V/ $\mu$ s	10.4 V/ $\mu$ s
CMFB OTA (moderate)	7.3 V/ $\mu$ s	7.9 V/ $\mu$ s	7.7 V/ $\mu$ s
CMFB OTA (strong)	26.7 V/ $\mu$ s	31.2 V/ $\mu$ s	28.6 V/ $\mu$ s

### 5.3 CDS op-amp design

It has been discussed in detail that several different op-amps can be properly synthesized with sub-space modeling based GP method to meet the low power objective. Among the studied architectures, the CMFB OTA can lead to the minimum static power consumption, but a large area is needed to fabricate feedback resistors in a standard CMOS processing technology. An extra pin is needed if a MOS transistor is utilized to implement  $R_c$ . This is not desired in some area stringent applications. The CS OTA is selected as a CDS op-amp to achieve a better trade-off between the power consumption and area overhead.

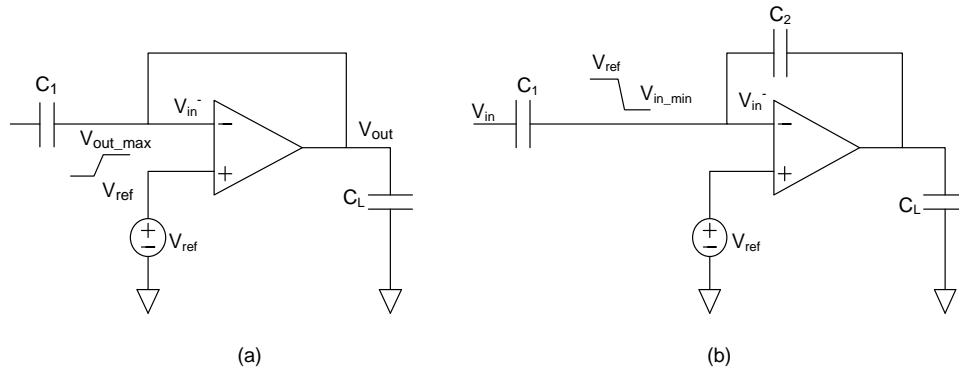
In a practical CMOS image sensor design, the settling time of the CDS op-amp is usually used as a primary design requirement. As a result, it can not be synthesized in the conventional way where  $A_V$ ,  $GBW$ ,  $SR$ , and  $PM$  are known as design requirements. To solve this settling-driven problem, it is better to investigate the settling behavior in the CDS op-amp. The switched capacitor based CDS circuit has two operation modes: reset and data transfer. If the input signal is larger than  $V_{swing}/A_V$ , the CDS op-amp will experience slewing and linear settling, as shown in Figure 5.9. The settling time is the sum of the slew time ( $t_{slew}$ ) and the linear settling time ( $t_{ls}$ ):

$$T_s = t_{slew} + t_{ls} \quad (5.12)$$



**Figure 5.9:** Settling of an op-amp responding to a large step input

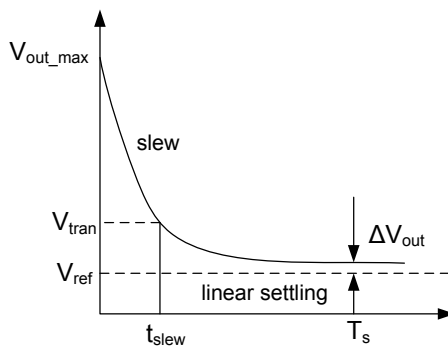
To derive the settling equation we consider the reset and data transfer modes separately. In these two operation modes, the CDS op-amp has different configurations as shown in Figure 5.10. If the worst case is considered, the initial output of the



**Figure 5.10:** CDS op-amp operation modes: (a) reset mode (b) data transfer mode

CDS op-amp is  $V_{out\_max}$ . During the reset, the output node and the negative input node of the CDS op-amp are shorted together. The negative input node will see a

step signal abruptly changing from  $V_{ref}$  to  $V_{out\_max}$ . The output settles from  $V_{out\_max}$  to  $V_{ref}$  as shown in Figure 5.11 and it includes both slew and linear settling.



**Figure 5.11:** CDS op-amp settling during reset

For the linear settling analysis, the CS OTA can be treated as a first order system since the non-dominant pole frequencies are much larger than the GBW. The output signal can be expressed as:

$$V_o = V_{tran} - (V_{tran} - V_{ref})(1 - e^{-t/\tau}) \quad (5.13)$$

where  $V_{tran}$  is the transition point at the output from which the op-amp changes from slew to linear settling or vice versa, and  $\tau$  is the time constant. Equation 5.13 can be simplified as:

$$e^{-t/\tau} = \frac{V_o - V_{ref}}{V_{tran} - V_{ref}} \quad (5.14)$$

During the reset operation, the gain of the CDS op-amp is unity, therefore

$$V_{tran} - V_{ref} = \frac{V_{swing}}{A_V} \quad (5.15)$$

Given a settling accuracy  $\epsilon$ , it is known that

$$\begin{aligned}
 V_o - V_{ref} &= \Delta V_{out} \\
 \epsilon &= \frac{\Delta V_{out}}{V_{ref}} \\
 V_o - V_{ref} &= \epsilon V_{ref}
 \end{aligned} \tag{5.16}$$

The time constant is:

$$\tau = \frac{C_{out}}{G_M} \approx \frac{C_L + C_1}{Sg_{m1}} = \frac{C_L + C_1}{S(g_{m1}/I_{d1})I_{d1}} \tag{5.17}$$

where  $S$  is the ratio between the transistor  $M6$  and  $M3$ ,  $g_{m1}$  is the transconductance of the transistor  $M1$ , and  $I_{d1}$  is the drain current of  $M1$ . By substituting Equations 5.15, 5.16, and 5.17 into Equation 5.14, the linear settling time can be derived as:

$$t_{ls} = \frac{C_L + C_1}{S(g_{m1}/I_{d1})I_{d1}} \ln\left(\frac{1}{\epsilon} \frac{V_{swing}/A_V}{V_{ref}}\right) \tag{5.18}$$

In slew settling,

$$t_{slew} = \frac{\Delta V_{slew}}{I_{slew}/C_{out}} = \frac{V_{out,max} - V_{tran}}{I_{slew}/C_{out}} \tag{5.19}$$

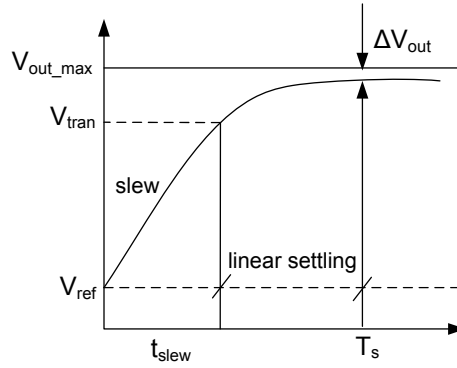
It is known that

$$\begin{aligned}
 V_{out,max} - V_{tran} &= V_{out,max} - (V_{ref} + V_{swing}/A_V) \\
 &= V_{in,step} - V_{swing}/A_V \\
 I_{slew}/C_{out} &= \frac{2SI_{d1}}{C_L + C_1}
 \end{aligned} \tag{5.20}$$

where  $V_{step.in}$  is the amplitude of the input step signal,  $V_{swing}$  is the maximum output swing, and the value of  $V_{swing}/A_V$  indicates the transition point at the input from which the op-amp changes from slew mode to linear settling mode or vice versa. By substituting Equation 5.20 into Equation 5.19, the slew settling time can be derived as:

$$t_{slew} = \frac{C_L + C_1}{2SI_{d1}}(V_{in.step} - V_{swing}/A_V) \quad (5.21)$$

At the end of the reset period, the negative input node of the CDS amplifier has a value of  $V_{ref}$ . When the CDS op-amp enters into data transfer mode by turning on the sampling switch, the negative input node will see a step signal abruptly changing from  $V_{ref}$  to  $V_{in.min}$ . The output of the op-amp will settle from  $V_{ref}$  to  $V_{out.max}$  as shown in Figure 5.12. In linear settling,



**Figure 5.12:** CDS op-amp settling during data transfer

$$V_o = V_{tran} + (V_{out.max} - V_{tran})(1 - e^{-t/\tau})$$

From Equation 5.22:

$$e^{-t/\tau} = \frac{V_{out\_max} - V_o}{V_{out\_max} - V_{tran}} \quad (5.22)$$

It is known that

$$\begin{aligned} V_{out\_max} - V_o &= \epsilon V_{out\_max} \\ V_{out\_max} - V_{tran} &= \beta(V_{swing}/A_V) \end{aligned} \quad (5.23)$$

where  $\beta$  is the gain of the CDS op-amp at the data transfer mode. The time constant at the data transfer mode is:

$$\tau = \frac{C_L + (1 - F)C_2}{FG_M} \quad (5.24)$$

where  $F$  is the feedback factor and it has form:

$$F = \frac{C_2}{C_1 + C_2 + C_{in}} \quad (5.25)$$

If  $C_1 = C_2 \gg C_{in}$ , then  $F = 1/2$ , and the linear settling time can be derived as:

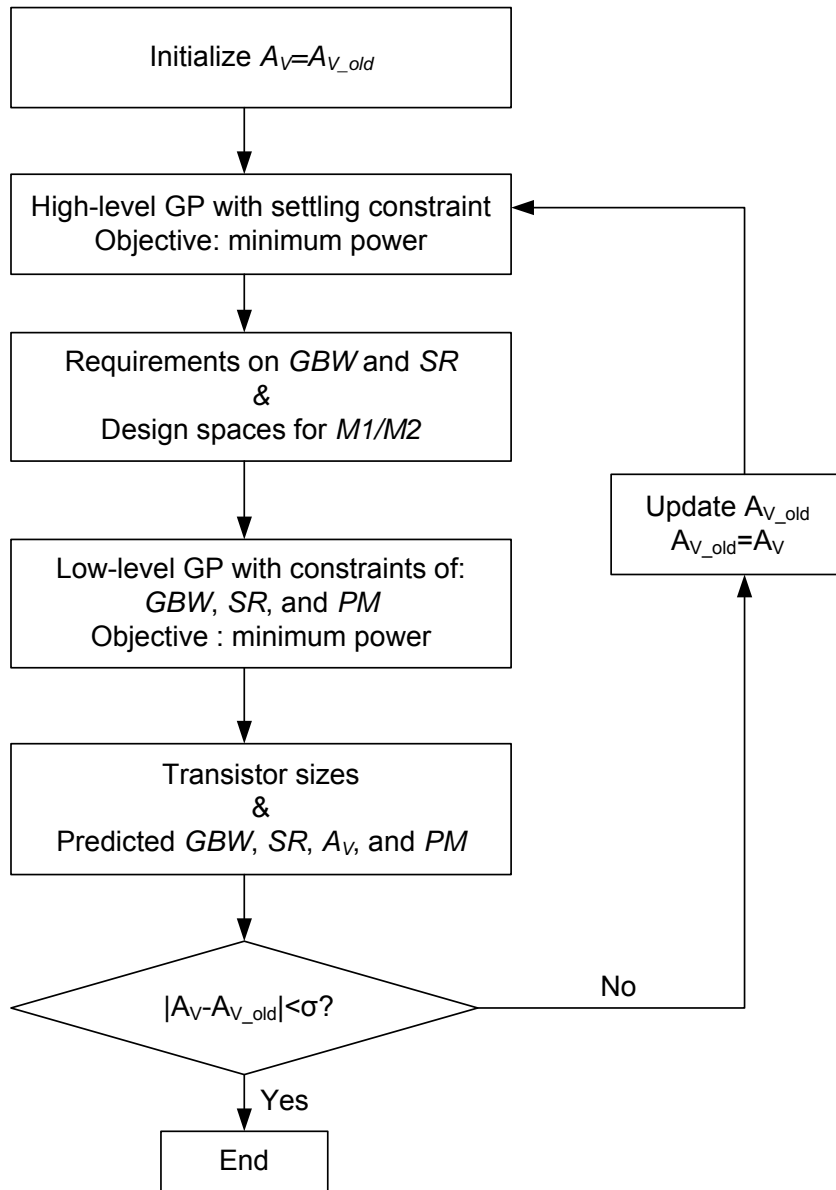
$$t_{ls} = \frac{2C_L + C_2}{S(g_{m1}/I_{d1})I_{d1}} \ln\left(\frac{1}{\epsilon} \frac{V_{swing}/A_V}{V_{out\_max}}\right) \quad (5.26)$$

The slew settling can be derived as:

$$\begin{aligned} t_{slew} &= \frac{\Delta V_{slew}}{I_{slew}/C_{out}} \\ &= \frac{V_{tran} - V_{ref}}{I_{slew}/C_{out}} \\ &= \frac{C_L + C_2}{2SI_{d1}} (|V_{in\_step}| - V_{swing}/A_V) \end{aligned} \quad (5.27)$$

Letting  $C_1 = C_2$ , we notice that slew times in both operation modes are the same. In comparison, we find that the linear settling time in data transfer mode is larger than that in the reset mode. Therefore, we only need to consider settling in the data transfer mode. We notice that the settling time can be modeled in a posynomial form in terms of  $g_m/I_d$ ,  $I_d$ , and  $S$  if the settling accuracy, output swing,  $V_{in.step}$ , and gain of the op-amp are considered constant.

To solve the settling-driven design problem, a new hierarchical geometric programming scheme is proposed, which is shown in Figure 5.13. The high-level GP uses the settling time constraint to solve the power optimization problem. It gives solutions to  $g_{m1}$ ,  $I_{d1}$ , and  $S$ , which can then be used to calculate the gain-bandwidth product and slew rate requirements. Design spaces for input transistors  $M1/M2$  can be derived as well. With GBW and SR requirements, a CDS op-amp with the optimal power consumption can be synthesized with a conventional GP as used for an op-amp design. If the synthesized  $A_V$  from the low-level GP converges to the  $A_V$  in the high-level GP, the synthesized circuit is the solution of the optimal design. Otherwise, the synthesized  $A_V$  will be fed back to the high-level GP, and the process repeats until  $A_V$  converges. Two CDS op-amps are synthesized for settling requirements of 250ns and 500 ns respectively. Both op-amps have a load of 1pF. Synthesized op-amps are evaluated with a HSPICE test bench. Performance results



**Figure 5.13:** A hierarchical GP algorithm for a CDS op-amp

from synthesis and the HSPICE simulation are presented in Table 5.12. Relative errors (RE) between these two results show that they are in a good agreement. In

**Table 5.12:** CDS op-amp design verification with HSPICE

	<b>Performance</b>	<b>Spec.</b>	<b>Predict.</b>	<b>HSPICE</b>	<b>RE</b>
<b>Case1</b>	$T_s$	$\leq 250 \text{ ns}$	250 ns	227 ns	10.1%
	$GBW$	–	15.3 MHz	14.3 MHz	6.9%
	$SR$	–	7.94 V/ $\mu$ s	7.86 V/ $\mu$ s	1.0%
	$PM$	$\geq 70^\circ$	69.99 $^\circ$	71 $^\circ$	1.4%
	$A_V$	$\geq 100$	155	151	2.6%
	$I_{static}$	minimum	9.24 $\mu$ A	9.18 $\mu$ A	
	<b>Case2</b>	<b>Performance</b>	<b>Spec.</b>	<b>Predict.</b>	<b>HSPICE</b>
$T_s$		$\leq 500 \text{ ns}$	500 ns	454 ns	10.1%
$GBW$		–	7.63 MHz	7.483 MHz	2.0%
$SR$		–	4.03 V/ $\mu$ s	4.1 V/ $\mu$ s	1.7%
$PM$		$\geq 70^\circ$	73.1 $^\circ$	73.6 $^\circ$	0.05%
$A_V$		$\geq 100$	200	202	0.99%
$I_{static}$		minimum	4.69 $\mu$ A	4.67 $\mu$ A	

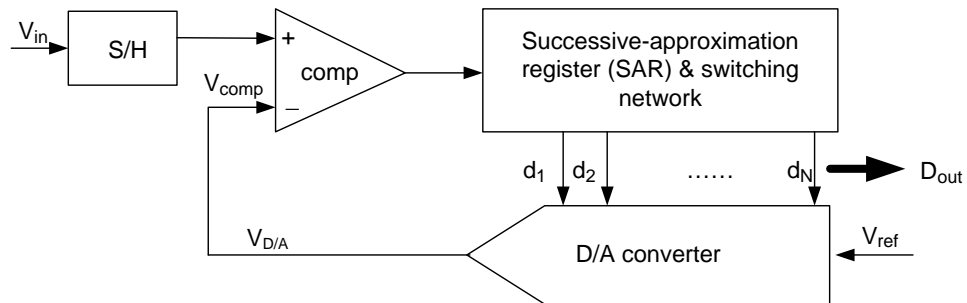
the high-level GP, a compact MOS model [Cunh 98] valid for all inversion regions is used. The solution from the high-level GP is valid for the entire design space. If a conventional symmetrical OTA is used in the circuit, the output swing is equal to  $V_{dd}$ . As a result, only  $A_V$  needs to be converged when settling accuracy,  $V_{in\_step}$ , and output swing are treated as constant.

So far, we have discussed power optimization design in CDS op-amps using GP-based synthesis method. It has been shown that the sub-space modeling based GP

synthesis method can be successfully applied to the power optimization of op-amps in sub-micron technologies. A hierarchical GP technique has also been suggested to solve the speed-driven design issue. In the next chapter, the power optimization of the ADC block is discussed.

## 6 Power optimization in successive-approximation A/D converters

Successive-approximation A/D converters (SAR ADCs) are one of the most popular designs due to their small conversion time and moderate circuit complexity [John 97]. The charge-redistributed SAR ADC can lead to very low power consumption when medium speed and resolution are required [Saue 03, Scot 03, Venu 10, Elza 08, Cho 03]. A typical structure of SAR ADC is shown in Figure 6.1. In a charge-redistributed

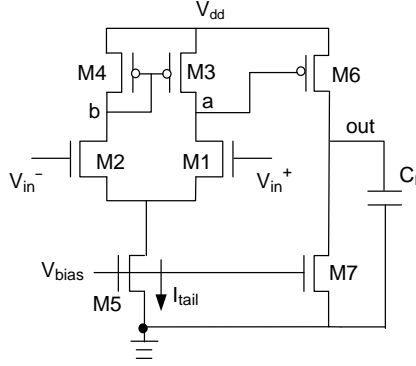


**Figure 6.1:** A typical SAR ADC

SAR ADC, the D/A converter (DAC) consists of a capacitor array with binary-weighted capacitance values. The SAR generates an N-bit digital code to the switching network, which switches the capacitor array in the DAC to generate a voltage at  $V_{comp}$ . The comparator then compares  $V_{comp}$  and  $V_{in}$ . The SAR uses this result to compute its next "guess" for the digital code. In this structure, the power consumption is dominated by two processes: (1) charging the binary-weighted capacitors and (2) the N comparison operations (where N is the number of bits of resolution desired in the resulting conversion). The power consumed in the charging operation is dependent upon  $V_{ref}$ , capacitance value ( $C$ ), and operating speed ( $f$ ). The power consumed in the comparison operation is contributed by the comparator. Generally, values of  $V_{ref}$ ,  $C$ , and  $f$  are set when a design is defined, which leads to a small room for power reduction. Power consumption can be reduced significantly if the comparator is designed properly. For a typical 8-bit ADC, a comparator should have a gain larger than 512, which usually requires a two-stage op-amp. A two-stage open-loop op-amp is selected as our study object because it can be modeled properly in a posynomial form.

## 6.1 Power optimization in a two-stage open-loop op-amp

A two-stage open-loop comparator (as shown in Figure 6.2) is not as simple to synthesize as a miller op-amp [Gao 10] because of the multi-pole effect. Without using



**Figure 6.2:** Two-stage open-loop comparator

a compensation capacitor, a two-stage open-loop comparator has to be treated as a two-pole system. We know that the response of a two-stage open-loop comparator to a step input of  $V_{in}$  is given in [Alle 04]:

$$\begin{aligned}
 v(t) &= A_{v0}V_{in}\left[1 + \frac{p_2 \exp(-tp_1)}{p_1 - p_2}\right] & p_1 \neq p_2 \\
 v(t) &= A_{v0}V_{in}\left[1 - \exp(-tp_1) - tp_1 \exp(-tp_1)\right] & p_1 = p_2
 \end{aligned} \tag{6.1}$$

where  $A_v(0)$  is the DC gain, and  $p_1$  and  $p_2$  are poles at the input and output stages, respectively. If we consider only the worst case where  $V_{in} = V_{in-min}$ , the propagation delay ( $t_p$ ) can be measured at the time when the output reaches half of the maximum

output. At this point, Equation 6.1 can be simplified as:

$$\begin{aligned}\frac{1}{2} &= 1 + \frac{p_2 \exp(-tp_1)}{p_1 - p_2} & p_1 \neq p_2 \\ \frac{1}{2} &= 1 - \exp(-tp_1) - tp_1 \exp(-tp_1) & p_1 = p_2\end{aligned}\quad (6.2)$$

Equation 6.2 shows that the propagation delay of this two-stage open-loop comparator depends on the poles at the input and output stages. The problem of designing a comparator with a specific propagation delay requirement can therefore be converted to a problem of designing a comparator with proper poles at the input and output stages. If we let  $m = p_2/p_1$ , we can derive  $p_1$  and  $p_2$  numerically with given values of  $m$  and  $t_p$ .

Constraints on  $p_1$  and  $p_2$  cannot be applied directly if a geometric programming synthesis is used, because  $p_1$  and  $p_2$  are not in a posynomial form, which can be clearly seen from Equation 6.3:

$$\begin{aligned}p_1 &= \frac{g_{ds1} + g_{ds3}}{C_a} = \frac{g_{ds1} + g_{ds3}}{C_{db1} + C_{dg1} + C_{db3} + C_{dg3} + C_{gs3} + C_{gs6} + A_{V2}C_{gd6}} \\ p_2 &= \frac{g_{ds6} + g_{ds7}}{C_{out}} = \frac{g_{ds6} + g_{ds7}}{C_{db6} + C_{dg6} + C_{db7} + C_{dg7} + C_L}\end{aligned}\quad (6.3)$$

where  $A_{V2}$  is the gain of the output stage.

A new scheme is proposed here to treat the input and output stages separately, in which each can be considered as a single pole circuit. As such, the pole requirement can be converted to a requirement of the  $GBW$  because we know that  $GBW =$

$A_V * pole$  in a single pole circuit. A design with constraints on gain and  $GBW$  can be solved easily with geometric programming. A detailed algorithm suggested for this design is depicted in Figure 6.3. Two major specifications, gain ( $A_V$ ) and propagation delay ( $t_p$ ), are applied to this synthesis, which can be derived from the performance requirement for the ADC. Given a value of  $m$ , the requirement for poles at the input and output stages can be derived. In synthesis of the output stage, a new independent variable  $A_{V2\_spec\_prime}$  (inverse of the DC gain of the output stage) is introduced, which ranges in value from 0.01 to 0.1. The specification of  $GBW$  at the output stage is equal to the product of  $p_2 * A_{V2}$ . Constraints of gain and  $GBW$  have to be relaxed to be used in the geometric programming:

$$\begin{aligned} \frac{1}{A_{V2}} &\leq A_{V2\_spec\_prime} \\ \frac{1}{GBW_2} &\leq \frac{1}{GBW_{2\_spec}} \end{aligned} \quad (6.4)$$

The problem becomes how to ensure the solution converges to

$$\begin{aligned} A_{V2} &= \frac{1}{A_{V2\_spec\_prime}} \\ GBW_2 &= GBW_{2\_spec} \end{aligned} \quad (6.5)$$

If parameters in Equation 6.4 can change in the directions shown in Equation

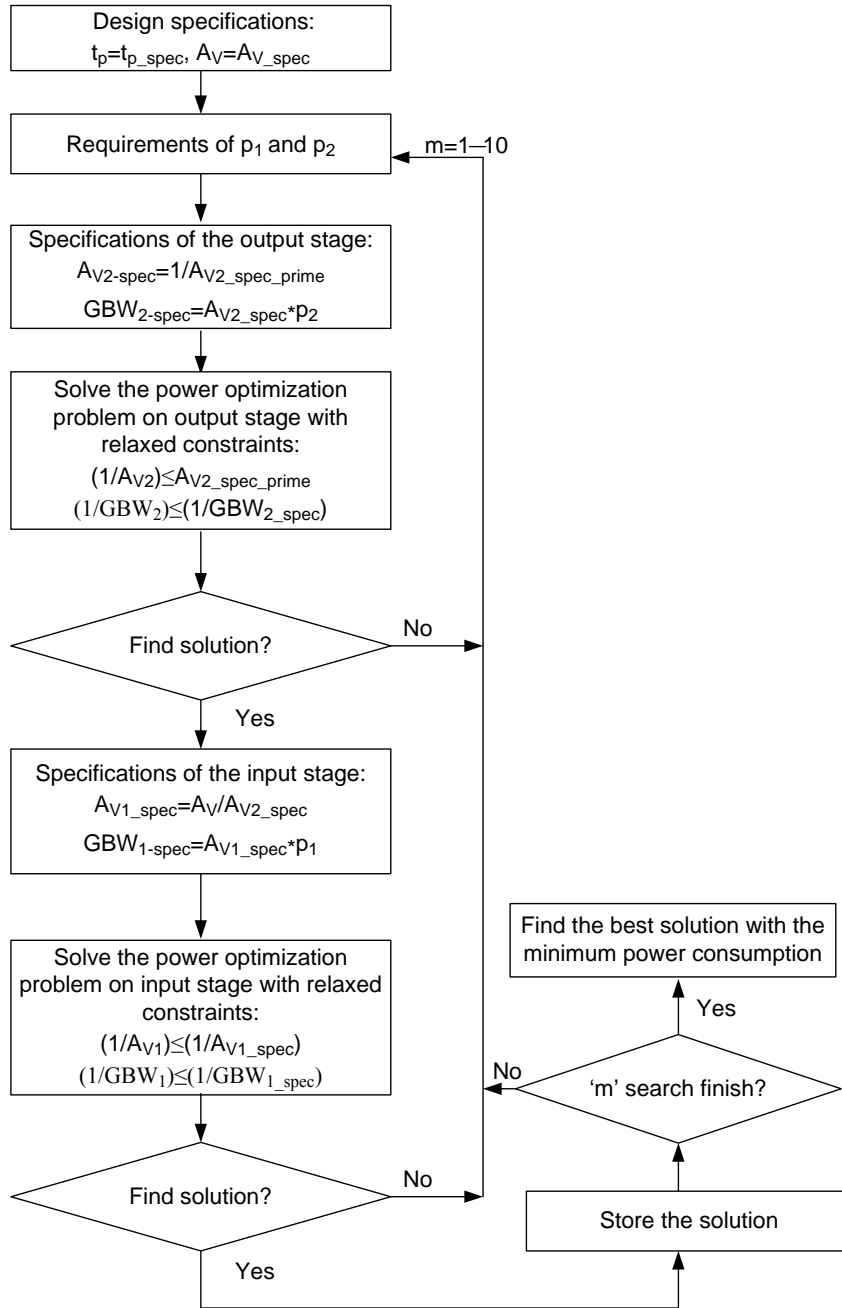


Figure 6.3: Optimization algorithm for a two-stage open-loop comparator

6.6, the solution can converge toward Equation 6.5.

$$\begin{aligned} \uparrow \frac{1}{A_{V2}} &\leq \downarrow A_{V2\_spec\_prime} \\ \uparrow \frac{1}{GBW_2} &\leq = \frac{1}{GBW_{2\_spec}} \quad (constant) \end{aligned} \quad (6.6)$$

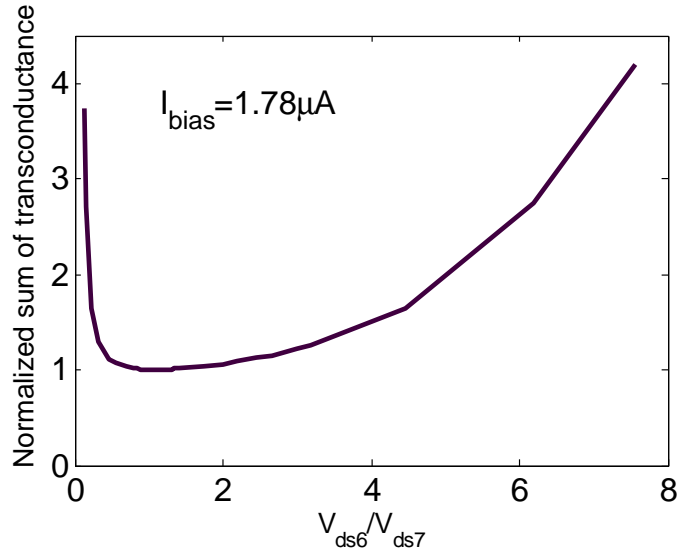
We know that  $GBW$  decreases as  $g_{m6}$  decreases, and the gain of the output stage is:

$$A_{V2} = \frac{g_{m6}}{g_{ds6} + g_{ds7}} \quad (6.7)$$

The sum of transconductance of  $M6$  and  $M7$  has a minimum value when  $V_{ds6} = V_{ds7}$ , and will increase when the bias varies to either side of this point, which can be shown in Figure 6.4. To allow the solution converge to  $GBW_{2\_spec}$  and  $1/A_{V2\_spec\_prime}$ , an objective function should be constructed to force the solution to be searched in the direction of reducing  $GBW_2$ ,  $A_{V2}$ , and  $A_{V2\_spec\_prime}$ . In this case, we construct the objective function as:

$$obj = \min(w_1 I_{bias} + w_2(1/V_{ds7}) + w_3 A_{V2\_spec\_prime}) \quad (6.8)$$

The first term of the objective function will lead to a reduction of  $g_{m6}$ , the second will lead to increasing sum of transconductance, and the third term will result in reducing  $A_{V2\_spec\_prime}$ . As a result, the solution will be searched in the directions shown in Equation 6.6. In sub-space-based geometric programming, a locally optimal solution may not lead to the equality of Equation 6.5 because the boundary of the design



**Figure 6.4:** The normalized sum of transconductance at different bias conditions

space is reached. Different design spaces have to be explored, and only equalized solutions will be used for the input stage synthesis. If the solution can not be found, a new value of  $m$  can be applied.

By knowing  $A_{V2}$  and  $p_1$ , the specifications of the gain and  $GBW$  of the input stage can be derived. The input stage can be synthesized as the similar way to the output stage with relaxed constraints on the gain and  $GBW$ . The objective function is constructed in a way to converge the solution along the direction that reduces  $GBW$  and gain. The transistors,  $M3$  and  $M4$ , in the input stage have the same  $V_{gs}$  as the transistor  $M6$  in the output stage. They will be searched in the same design space as  $M6$ . The transistor  $M5$  in the input stage has the same channel length and

$V_{gs}$  as the transistor  $M7$  in the output stage. It will therefore be searched in the same design space as  $M7$ . As a result, only input transistors ( $M1$  and  $M2$ ) will be searched independently. If an equalized solution can be found, the solution will be stored, otherwise a new value of  $m$  will be applied. If the search for  $m$  has finished, the best solution will be selected from those with the minimum power consumption.

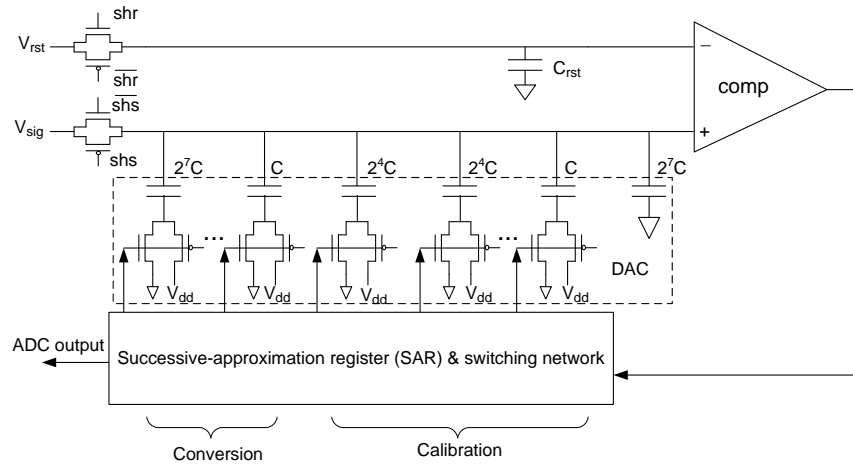
By using this algorithm, a two-stage open-loop comparator is designed with the following requirements which are derived from [Cho 03]:  $A_V = 1000$ ,  $t_p = 20$  ns, and  $C_L = 10$  fF. The design space is limited to the moderate inversion region to achieve low power consumption. A solution is found when  $m=2$ . The synthesized circuit is simulated in a HSPICE test bench, and the propagation delay is measured. The comparison between the synthesis and HSPICE simulation results are shown in Table 6.1. The result reported in [Cho 03] is also listed for comparison. It shows that the proposed power optimization based design is a powerful technique when low power is desired.

**Table 6.1:** Performance measures of the optimized two-stage open-loop comparator

	$I_{\text{bias}}(\mu\text{A})$	$A_V(\text{dB})$	$t_p(\text{ns})$
Synthesis	12.8	61.6	20
HSPICE	12.13	68.9	20.5
[Cho 03]	16	—	—

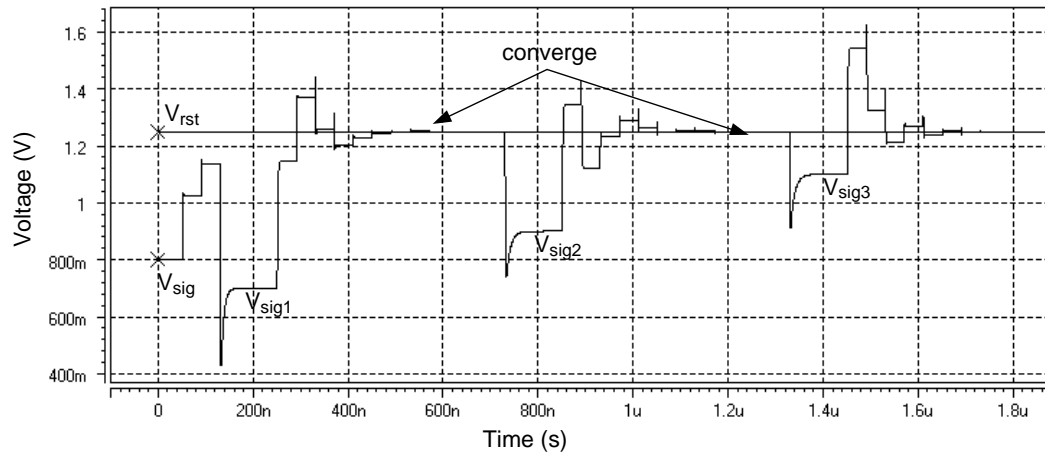
## 6.2 SAR ADC design

A SAR ADC (see Figure 6.5) with the same architecture and the same performance as [Cho 03] was designed, in which the comparator synthesized above is applied. The basic functionality and the linearity of the ADC are evaluated using an HSPICE



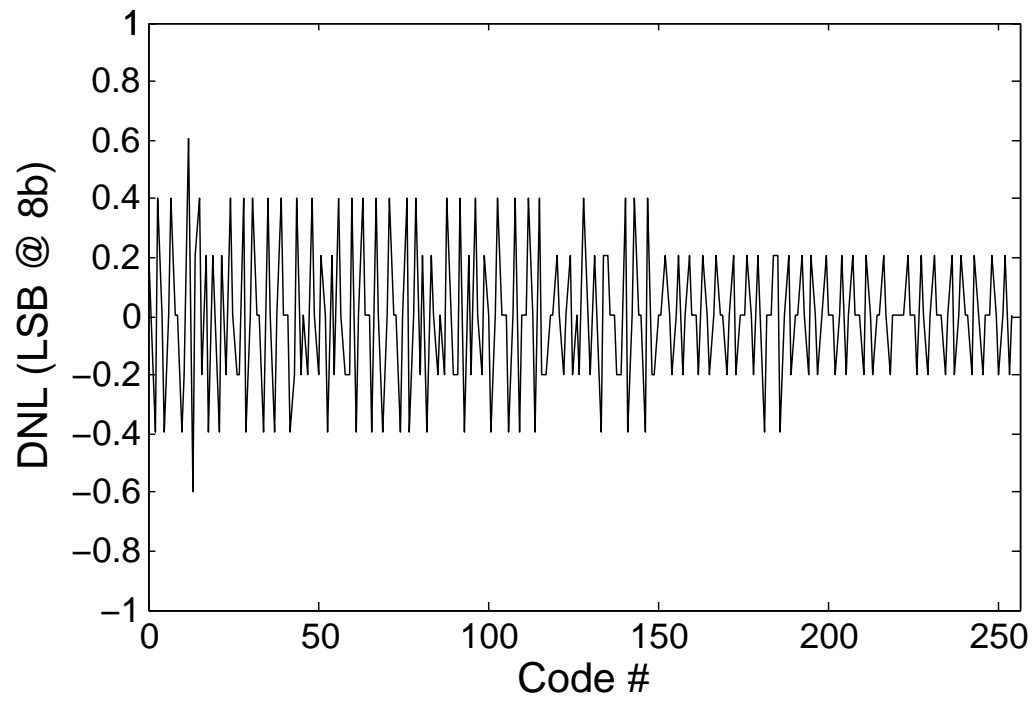
**Figure 6.5:** Low-power 8-b successive approximation ADC

simulation. Arbitrary signal values are applied to the ADC to check if the conversion is performed properly. Figure 6.6 shows that the output of the DAC can correctly converge to the reset value ( $V_{rst}$ ). The linearity of the ADC is checked by determining the differential non-linearity (DNL) and integral non-linearity (INL), which are shown in Figure 6.7 and 6.8, respectively. The DNL is within  $\pm 0.6$  LSB and INL is within  $\pm 0.5$  LSB. No missing codes were observed. By using a geometric-programming-based power optimization synthesis method, a SAR ADC can be efficiently designed

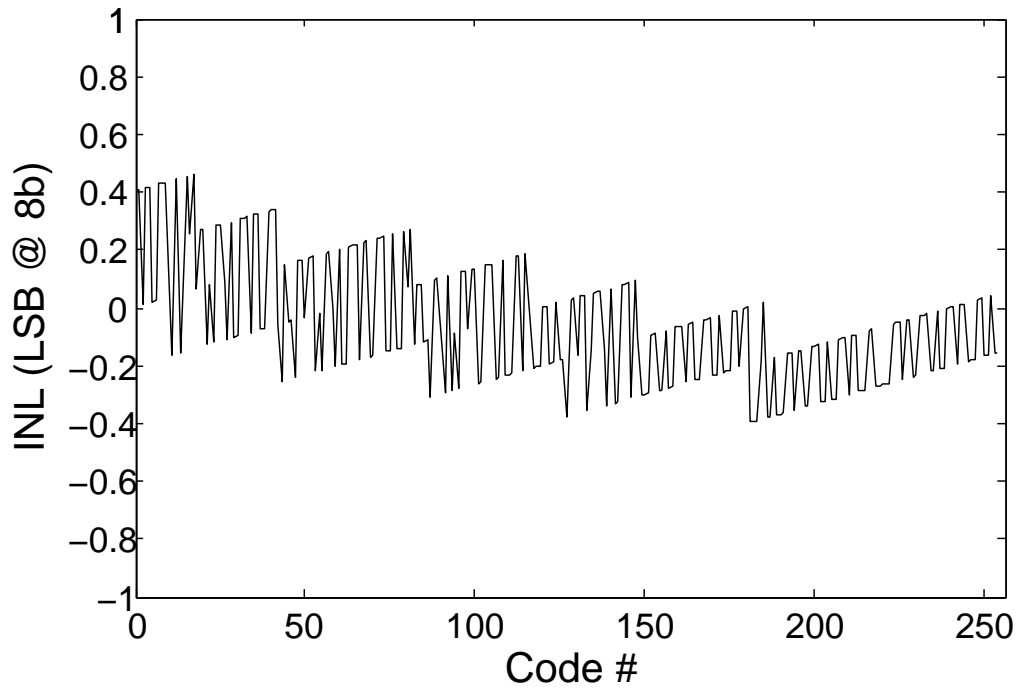


**Figure 6.6:** SAR ADC onversion waveform

to meet the low power consumption requirement.



**Figure 6.7:** Differential non-linearity plot



**Figure 6.8:** Integral non-linearity plot

## 7 System optimization

So far, a power optimization technique using sub-space-based geometric programming methodology to achieve low power design in main analog blocks of CMOS APS image sensors has been discussed. This technique allows designers to find a more power efficient design without changing the existing design architecture. Some new techniques are suggested as well to cope with speed-driven design in op-amps and comparators, which makes GP based analog synthesis more powerful in practical designs. To have a better understanding how this technique can improve the performance in power consumption, a design published in [Cho 03] will be used as an example. This 176 x 144 CMOS image sensor was claimed to only consume 550  $\mu W$  when it operated at 30 frames per second with an internal clock rate of 25.2 MHz. A significant power reduction was achieved in this design by using low-power design methodologies throughout the design process from the system-level to the process-level. At the circuit level, designs for the reduced power supply were addressed, but there was no power optimization technique used for power reduction in analog

blocks. A new design achieved with the GP-based power optimization is compared with this reference design. The comparison between bias currents in the in-pixel source follower, CDS op-amp, and SAR ADC comparator are listed in Table 7.1. A

**Table 7.1:** Bias current comparison between a referenced and an optimized design

	<b>in-pixel source follower</b>	<b>CDS op-amp</b>	<b>ADC comparator</b>	<b>Total</b>
Reference design [Cho 03]	1.25 $\mu A$	30 $\mu A$	16 $\mu A$	266 $\mu A$
GP optimization	264.6 $nA$	28.5 $\mu A$	12.8 $\mu A$	87.9 $\mu A$
% reduction	78.8%	5%	20%	67%

significant improvement of the power consumption in the in-pixel source follower and ADC comparator blocks is noticed, and the power consumption in the CDS op-amp has been reduced as well. It can be concluded that the sub-space modeling based GP is an efficient low power design methodology in analog circuit design in sub-micron technologies.

## **8 Dynamic range enhancement for a laser ranging image sensor with a lowered power supply**

So far, a low-power design methodology based on automatic analog synthesis has been discussed. By using GP analog synthesis, power consumption in analog blocks of CMOS image sensors can be effectively minimized for a required performance using existing circuit architectures. In addition, technological advances and the increased digital functionality on imager chips cause a pressure to reduce supply voltage. Therefore it is also important to design sensors that can offer reasonable performance at these lower voltages. A reduced supply voltage limits the analog voltage swing in CMOS APS pixels, which is undesirable in a sensor intended to operate over a wide range of illuminations. Dual power supplies are adopted by many commercial products to reduce this problem, in which digital blocks are powered with a low value and analog blocks are powered with a high value. This scheme will not

work properly when the processing technology is scaled down to the deep sub-micron level, because devices fabricated at deep sub-micron technologies cannot sustain a large power supply. For example, the analog power supply in 90 nm technology is less than 2.5 V, which is only half of the value in 0.35  $\mu\text{m}$  technology. A more promising solution to this output swing problem is to use dynamic range enhancement techniques, which have a good scalability for different processing technologies. Hence an architecture which provides high DR at a higher supply voltage may produce an acceptable DR at a low supply voltage. Today's commercial pixels are mostly based on a 4T structure, in which pinned photodiodes are used to capture a photo-charge, and a sense node separated from the photodiode converts the photo-charge to a voltage. The small size of these pixels (approaching 1  $\mu\text{m}$ ) leads to a low dynamic range because of the small full-well capacity, so DR enhancement techniques are also beneficial for these structures.

In this chapter, some relevant dynamic range enhancement techniques for CMOS APS image sensors will first be discussed, and then a new predictive-integration design will be proposed to increase the dynamic range of a CMOS image sensor for laser rangefinding. This technique was verified by a fabricated sensor chip, and measurement results are presented.

## 8.1 Dynamic range enhancement technologies

Dynamic range in CMOS image sensors can be enhanced by reducing noise (enhancing the dynamic range toward darker scenes) or expanding incident light saturation level (improving the dynamic range toward brighter scenes). Here, the solutions that extend the DR toward high light intensities will be presented. Many techniques for enhancing DR in CMOS image sensors at the pixel or system levels have been reported recently. A qualitative summary of the existing solutions and their comparisons are presented in [Yadi 99]. Quantitative assessments for different wide-dynamic-range CMOS image sensors are presented in [Spiv 09]. The existing techniques for dynamic range (DR) enhancement in CMOS APS image sensors include logarithmic response techniques [Cham 84, Kava 00, Loos 01, Lai 04, Otim 05], well-adjusting technology [Deck 98], light-to-time conversion [Yang 94, McIl 01, Wang 06, Qi 04, Guo 07], multiple sampling [Yang 99, Yadi 97, Park 05], overflow-charge-storage techniques [Shaf 08, Akah 06], and adaptive integration [Yadi 01, Yadi 03, Acos 04, Yasu 03, Brei 04, Bele 07, Bele 09]. The logarithmic scheme can achieve a very wide DR by converting the photocurrent into voltage in the logarithmic scale. The image quality in this scheme is generally poor due to low signal-to-noise ratio (SNR) [Ricq 95], and long settling time when the illumination changes from high to low intensities [Yadi 99]. The well-adjusting technology can increase the dynamic range of

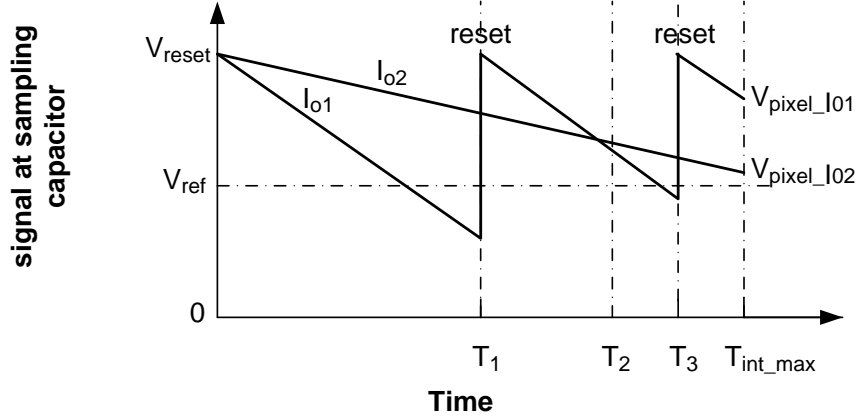
the pixel by dynamically changing the well capacity over the integration time. It can be implemented by simply adjusting the reset signal one or more times during the integration. However, it causes local dips in SNR [Yang 99]. In addition, the sensor response in a well-adjusting scheme is non-linear, which is not efficient in color processing. Light-to-time conversion image sensors convert the light intensity from the voltage domain to the time domain, which are usually represented by the pulse frequency or the pulse width. The dynamic range in light-to-time conversion sensors can be easily enhanced beyond 100 dB even at a power supply voltage of less than 1 V [Wang 06]. This advantage will usually be offset by the small pixel fill-factor due to a large in-pixel readout circuit, which must include a digital counter. Multiple-sampling technology increases intrascene DR by synthesizing multiple images taken at different exposure times. Multiple exposures can result in a longer operating time and motion artifacts compared to a single-exposure scheme. Image sensors implemented with overflow-charge-storage techniques can allow charge generated in a photo-detector to flow into a storage device (either a capacitor [Akah 06] or a pinned diode [Shaf 08]) embedded in a pixel. The dynamic range of the image sensor is therefore enhanced by increasing the maximum charge capacity while keeping the same noise floor. Overflow-charge-storage techniques can lead to low noise with a pinned diode implementation, and small motion blur with a single integration.

Special processing technologies and design efforts are usually required in overflow-charge-storage techniques to reduce the size of the storage capacitor, to reduce FPN, and to improve the response linearity. Adaptive integration can automatically adapt the integration time of a pixel to a local illumination level without affecting the frame operation time. Saturation over a large range of illuminations can be avoided by using this locally adjusted integration time, and the dynamic range is therefore increased. As a specific example of an application requiring a high DR architecture, we consider a CMOS image sensor for laser rangefinding.

A CMOS image sensor for an optical triangulation-based laser rangefinding usually requires a dynamic range larger than 84 dB for accurate peak detection [Blai 04]. With a linear image sensor, only a single range point can be obtained from one frame operation; therefore, a high frame rate is wanted for a laser rangefinding image sensor in order to attain a reasonable measurement time. Among the aforementioned dynamic range enhancement technologies, only the overflow-charge-storage and adaptive integration techniques can meet these requirements for the dynamic range, frame rate, and noise. Since it is subjected to the limitations of the processing technology (only the standard CMOS processing technology is available in this study), only the adaptive integration technique will be considered.

The adaptive integration technique in previously published designs is using mul-

multiple reset method based on overlapping integration scheme, which can be depicted in Figure 8.1. In this technique, the maximum integration time within a frame op-



**Figure 8.1:** Response of a pixel in an adaptive integration scheme.

eration is divided into multiple sub-integration periods whose duration change in a decreasing order, such as  $\frac{T_{int\_max}}{2}$ ,  $\frac{T_{int\_max}}{4}$ ,  $\frac{T_{int\_max}}{8}$ , and etc. At the end of each sub-integration period, the pixel compares its value with a threshold value. If the pixel value is larger than the threshold value, the pixel continues to integrate in the next integration period. Otherwise, the pixel resets itself to avoid saturation, and then the pixel re-starts integrating. The pixel will be checked again at the end of next sub-integration period. The last non-saturation integration period is selected as the optimal integration period of the pixel. By using the adaptive integration, a pixel can select a proper integration time according to the local illumination. The final

pixel value can be calculated using a floating-point representation from the number of resets [Yadi 01]. In Figure 8.1, the pixel resets at  $T_1$  and  $T_3$  for illumination  $I_{o1}$ , and the optimal integration period is from  $T_3$  to  $T_{int\_max}$ .  $I_{o2}$  does not cause any intermediate reset, and the optimal integration period is  $T_{int\_max}$ .

The optimal integration time derived from the aforementioned adaptive integration may not lead to the longest integration time for which the pixel does not saturate. For example, the optimal integration time for illumination  $I_{o1}$  in Figure 8.1 is the minimum integration period. This will lead to a smaller SNR compared to the signal read out using a longer integration time. The signal value used for SNR calculation is the value directly measured at the end of the effective integration time, and it is integration time dependent. The SNR in adaptive integration scheme is [Spiv 09]:

$$SNR = 20\log_{10} \frac{\frac{i_{ph}t_{int}}{C_{int}}}{\sqrt{\sigma_r^2 + \frac{q(i_{ph}+i_{dc})t_{int}}{C_{int}^2} + \sigma_{PRNU}^2 + \sigma_{FPN}^2 + 2\frac{kT}{C_{int}}}} \quad (8.1)$$

where  $i_{ph}$  is the photogenerated current,  $i_{dc}$  is the dark current,  $t_{int}$  is the effective integration time,  $C_{int}$  is the parasitic capacitance of the photodiode,  $\sigma_r$  is the read noise,  $\sigma_{PRNU}$  is the photoresponse nonuniformity, and  $\sigma_{FPN}$  is the fixed pattern noise. When the light intensity is relatively high, the photon-shot noise is dominant.

The SNR becomes:

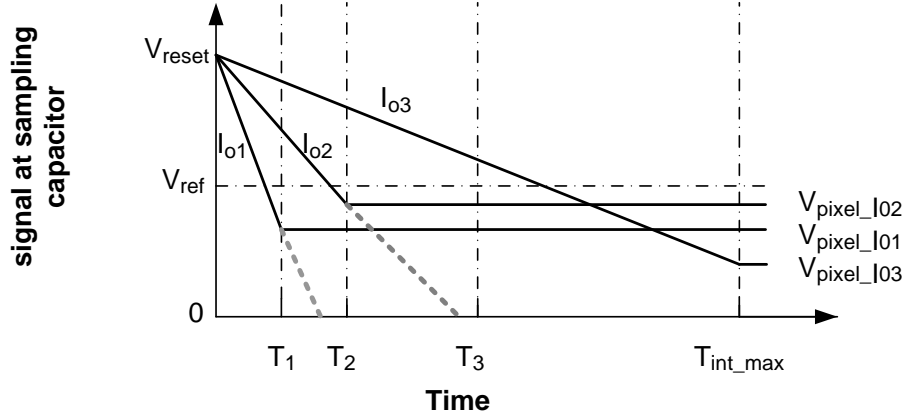
$$SNR \approx 20\log_{10} \sqrt{\frac{i_{ph}t_{int}}{q}} \quad (8.2)$$

It implies that SNR reduces if the integration time decreases.

To improve the SNR using adaptive integration, a new predictive-integration is suggested, in which the longest adaptive integration time can be found without searching all possible sub-integration periods, and only a single reset is needed at the beginning of the frame. In the following section, the principle of the predictive-integration will be explained in detail.

## 8.2 Predictive integration

A pixel of the image sensor in the predictive-integration scheme should automatically determine if it will become saturated within the frame time. As such, an integration time, called predictive-integration time, is adjusted to the local illumination. The predictive integration time is considered to be the longest integration time that does not saturate the pixel. In this scheme, the integration time ( $T_{int\_max}$ ) in a frame operation is divided into multiple sub-integration periods ( $T_n$ ), as shown in Figure 8.2. At the end of each sub-period, the pixel value is compared with a reference value to predict if the pixel will become saturated in the next integration period. We call it the “sat-positive” check. If it is sat-positive, the pixel value at the current period is held by the sampling capacitor, and the pixel stops integrating. The current sub-period is considered to be the predictive-integration time. In Figure 8.2, the pixel



**Figure 8.2:** Response of a pixel to three illuminations for various sub-periods  $T_n$ 's.

is predicted as sat-positive at  $T_1$  and  $T_2$  for illuminations  $I_{o1}$  and  $I_{o2}$ , respectively. The potential saturation before the next sub-period can explicitly be demonstrated by the dashed extension lines of the integration curves. The pixel values  $V_{pixel\_Io1}$  and  $V_{pixel\_Io2}$  are held as the output signal values when a frame operation ends.  $I_{o3}$  does not cause saturation.

By using this predictive-integration, the adaptive integration time for illumination  $I_{o1}$  in Figure 8.1 becomes  $\frac{T_{int\_max}}{2}$ , which is longer than the optimal integration time derived from other adaptive integration schemes, and SNR is improved. In addition, the predictive-integration can allow a better power control, because it can in principle turn off the pixel as the saturation is sensed.

The predictive-integration scheme uses the same principle for DR enhancement

as the multisampling scheme. In a multiple sampling scheme, the maximum and minimum detectable photocurrents can be expressed as

$$\begin{aligned}
 I_{photo\_max} &= Q_{full\_well}/T_{int\_min} \\
 I_{photo\_min} &= Q_{min}/T_{int\_max}
 \end{aligned}
 \tag{8.3}$$

where  $T_{int\_min}$  and  $T_{int\_max}$  are the minimum and maximum integration times, respectively;  $Q_{min}$  is the minimum measurable charge in the photodiode when  $T_{int\_max}$  is applied; and  $Q_{full\_well}$  is the full potential-well charge in the photodiode.  $Q_{min}$  is limited by the read noise. The DR can then be calculated as

$$\begin{aligned}
 DR &= I_{photo\_max}/I_{photo\_min} \\
 &= (Q_{full\_well}/Q_{min})(T_{int\_max}/T_{int\_min}) \\
 &= DR_S(T_{int\_max}/T_{int\_min})
 \end{aligned}
 \tag{8.4}$$

where  $DR_S$  is the DR achieved by a single integration scheme ( $T_{int} = T_{int\_max}$ ). We can see that the DR can be increased by a factor of  $T_{int\_max}/T_{int\_min}$  when multiple integration times are used. In practice,  $T_{int\_max}$  and  $T_{int\_min}$  must be properly selected for speed considerations, which will be discussed in detail later.

To maintain a linear irradiance transfer function, the sub-integration periods and

the reference value should satisfy

$$T_n = kT_{n-1}$$

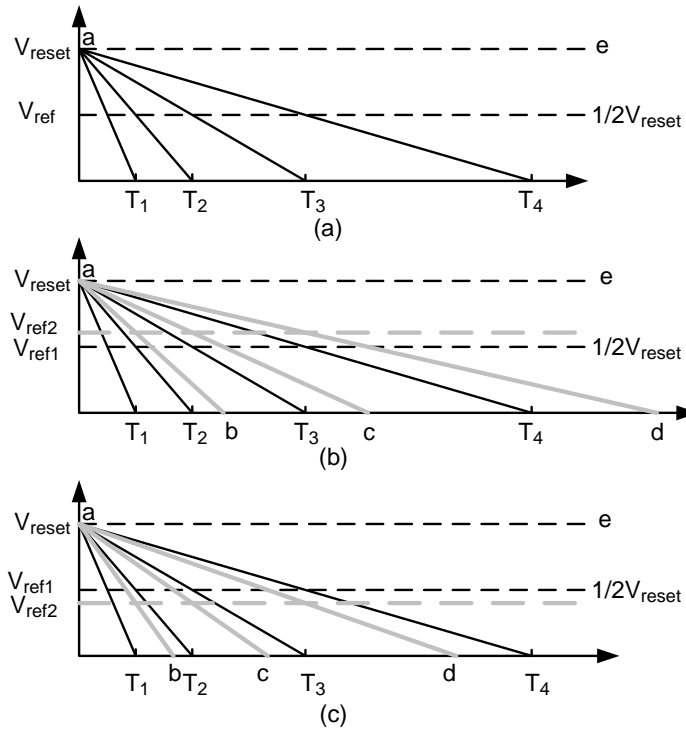
$$V_{ref} \geq V_{reset}/k \quad (8.5)$$

where  $k$  is a constant to be determined by the application. The synthesized output is

$$V_{out\_syn} = V_{out\_pixel}(k^N/k^m) \quad (8.6)$$

where  $V_{out\_pixel}$  is the measured pixel value,  $N$  is the total number of sub-integration periods, and  $m$  is the number of sub-periods that have been searched.

The  $V_{ref} \geq V_{reset}/k$  condition in Equation 8.5 is needed to ensure continuity of the detectable light intensity. Figure 8.3 shows how the reference values can affect the detectable light intensity. Let  $k = 2$  be considered first. If the reference value is half of  $V_{reset}$ , as shown in Figure 8.3(a), the detectable light intensities are defined by the regions of  $a-T_1-T_2$ ,  $a-T_2-T_3$ , and  $a-T_3-T_4$  at times  $T_1$ ,  $T_2$ , and  $T_3$ , respectively. The light intensities falling into  $a-T_4-e$  will be detected at  $T_4$ . As a result, the continuity of the detectable light intensity can be guaranteed. If the reference value is larger than half of  $V_{reset}$ , as shown in Figure 8.3(b), the detectable light intensities are defined by the regions of  $a-T_1-b$ ,  $a-T_2-c$ , and  $a-T_3-d$  at times  $T_1$ ,  $T_2$ , and  $T_3$ , respectively. The light intensities falling into  $a-d-e$  will be detected at  $T_4$ . We can see that the continuity of the detectable light intensity can still be guaranteed due

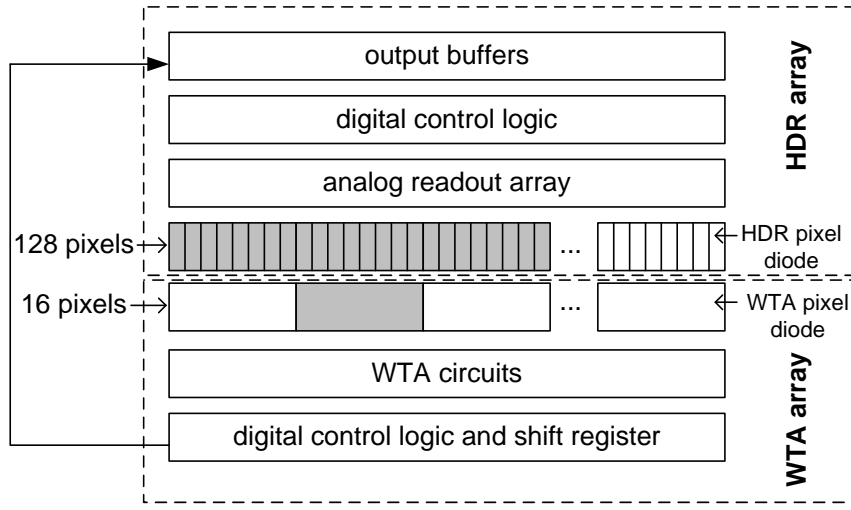


**Figure 8.3:** Effect of the reference value on the detectable light intensity. The letter “a” represents the value of  $V_{reset}$  at  $T = 0$ . Lines  $ab$ ,  $ac$ , and  $ad$  are the integration curves, in which signal value drops to  $V_{ref2}$  at  $T_1$ ,  $T_2$ , and  $T_3$ , respectively.

to the overlap of the regions of  $a-T_1-b$ ,  $a-T_2-c$ , and  $a-T_3-d$ . If the reference value is less than half of  $V_{reset}$ , as shown in Figure 8.3(c), the detectable light intensities are defined by the regions of  $a-T_1-b$ ,  $a-T_2-c$ , and  $a-T_3-d$  at times  $T_1$ ,  $T_2$ , and  $T_3$ , respectively. The light intensities falling into  $a-T_4-e$  will be detected at  $T_4$ . In this case, the continuity of the detectable light intensity *cannot* be guaranteed due to a gap between any two adjacent regions of  $a-T_1-b$ ,  $a-T_2-c$ ,  $a-T_3-d$ , and  $a-T_4-e$ . We can conclude that the reference value should be equal to or greater than half of the reset value. The same analysis can be applied to the case of  $k > 2$ , and  $V_{ref} \geq V_{reset}/k$  can be concluded.

### 8.3 Sensor architecture and operation

The chip architecture is shown in Figure 8.4. It has a similar structure as those suggested by Beraldin *et al.* [Bera 03] and De Nisi *et al.* [De N 05]. It consists of two aligned linear arrays, namely, a high dynamic range (HDR) array and a winner-take-all (WTA) array. These arrays provide independent control of DR enhancement and region of interest (ROI) readout. The HDR and WTA pixel dimensions are  $25 \mu\text{m} \times 250 \mu\text{m}$  and  $200 \mu\text{m} \times 250 \mu\text{m}$ , respectively. The relationship between the HDR and WTA arrays is also shown in Figure 8.4. The winning pixel (shaded) in the WTA array can enable the readout of pixels in the ROI window (shaded) of the HDR



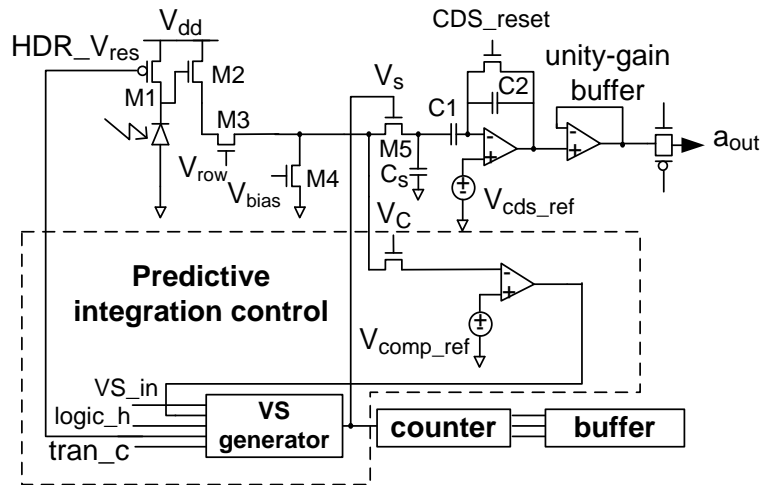
**Figure 8.4:** Laser rangefinding chip architecture

array. In this design, the ROI window size is 24 pixels. Instead of serially reading out all 128 pixels in the HDR array, the centroid of the laser spot can efficiently be extracted from the small ROI window. As a result, the data rate can be increased.

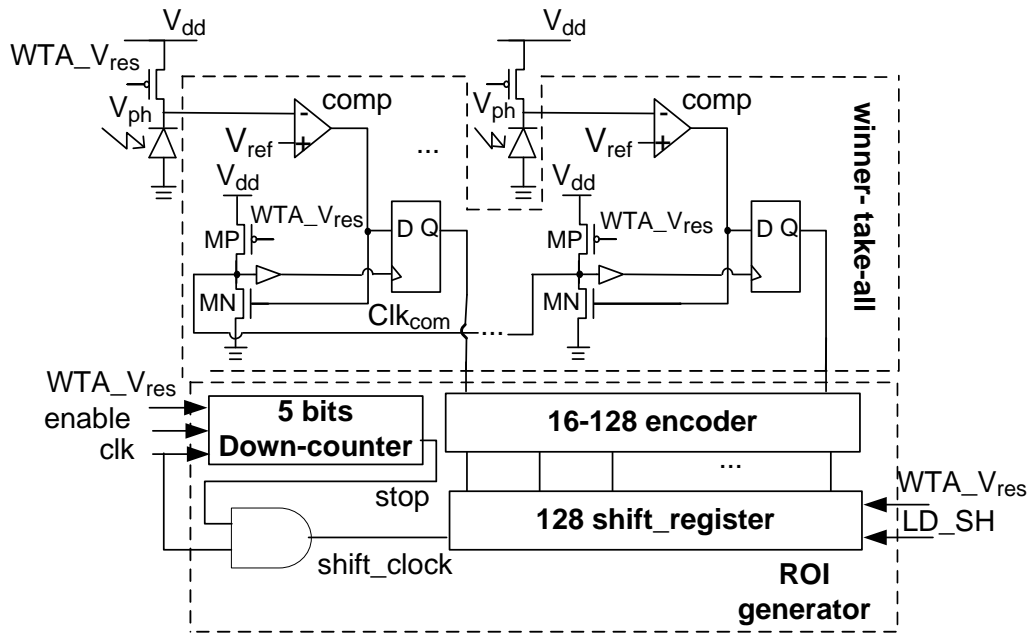
In former designs [Bera 03, De N 05], a passive pixel with a charge amplifier was used in the HDR array. The DR in this design is enhanced by changing the gain of the charge amplifier. Although this implementation can lead to a very high readout speed when using a short integration time, it is disadvantaged by a reduced SNR under low illumination and an increased area cost. We know that the SNR (for both passive and active pixels) under low illumination increases with the integration time at 20 dB/decade [Kavu 06]. If we compare the integration time of the passive pixel design

(12  $\mu\text{s}$ ) to our design (276  $\mu\text{s}$ ), we would expect to see an improvement of  $\sim 27$  dB in the SNR using an APS under low illumination. The extra area caused by the feedback capacitors in the charge amplifier of the passive pixel implementation would be significant. Assuming the use of metal-insulator-metal capacitors with a metal5 and metal6 implementation in the standard 0.18  $\mu\text{m}$  technology (1.29 fF/ $\mu\text{m}^2$ ), an additional area of  $\sim 1.25 \text{ mm}^2$  will be needed for the entire 128 pixel array. By using the APS scheme, we trade off the readout speed and these two performance measures. Another improvement in this design over the former one is power reduction in the WTA array. In the former design, each pixel in the WTA array includes a comparator, two correlated double sampling (CDS) amplifiers, and a cell of the WTA circuit. A bias current is needed in each of these analog components. In our design, each pixel includes a comparator and a cell of the WTA circuit. The cell of the WTA circuit is a digital circuit, which does not draw a static current during the operation. As a result, the power consumption is minimized.

The schematics of the HDR and WTA pixel arrays are shown in Figure 8.5. A typical timing diagram for the HDR and WTA pixel arrays is shown in Figure 8.6.  $HDR\_V_{res}$  and  $WTA\_V_{res}$  are reset signals for the HDR and WTA pixels, respectively.  $VS\_in$  is a control signal to generate a multiple sampling signal  $V_s$ .  $V_c$  is used for signal checking in the comparator of the HDR pixel. Signal  $tran\_c$  enables

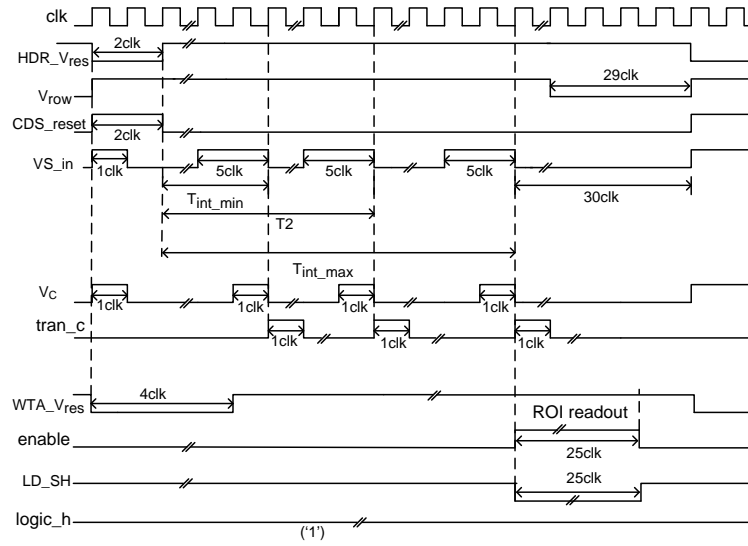


(a)



(b)

Figure 8.5: Schematics for (a) HDR and (b) WTA pixel arrays



**Figure 8.6:** Timing diagram for the HDR and WTA pixel arrays

the sat-positive check.  $V_{row}$  is used for power control to disable the bias current of the source follower during the ROI readout. Signal *enable* generate a signal to enable the pixels in the ROI to be read out. *LD\_SH* is the load/shift signal for the shift register.

The HDR pixel is based on a standard 3-T configuration, comprising a reset transistor (M1), a source follower (M2), and a readout transistor (M3). CDS is implemented for noise reduction. The predictive-integration control circuit allows for sat-positive checking, multiple sampling control, and integration termination when sat-positive is sensed. A 3-bit counter and a 3-bit buffer are also included to count and store the number of checked sub-integration periods. When a HDR pixel is reset,

$V_s$  and  $V_c$  are active. The reset value  $V_{reset}$  is sampled to the sampling capacitor  $C_s$ . At this moment,  $V_{reset} > V_{comp.ref}$ , and the comparator has an output of “0”, which can enable a new sampling at the end of the first sub-integration period,  $V_c$  is ON, and the pixel is compared with the reference again. After the comparison, the transition check signal ( $tran.c$ ) enables a transition check at the comparator. If  $V_{sig} \leq V_{comp.ref}$ , the comparator flips to “1”, and no more sampling pulses can be generated in the next sub-period. The sampling switch is open for the reset of the integration period, and  $C_s$  holds the signal until it is read out at the end of the frame operation. Otherwise, a new sampling pulse is generated, and the pixel can integrate for the next sub-period. This operation is repeated until the sat-positive condition is sensed or the maximum integration period is complete. In this design, the reset and photo signals are sampled within the same frame window, ensuring that the reset noise in these two signals is correlated. As a result, a true CDS can be performed.

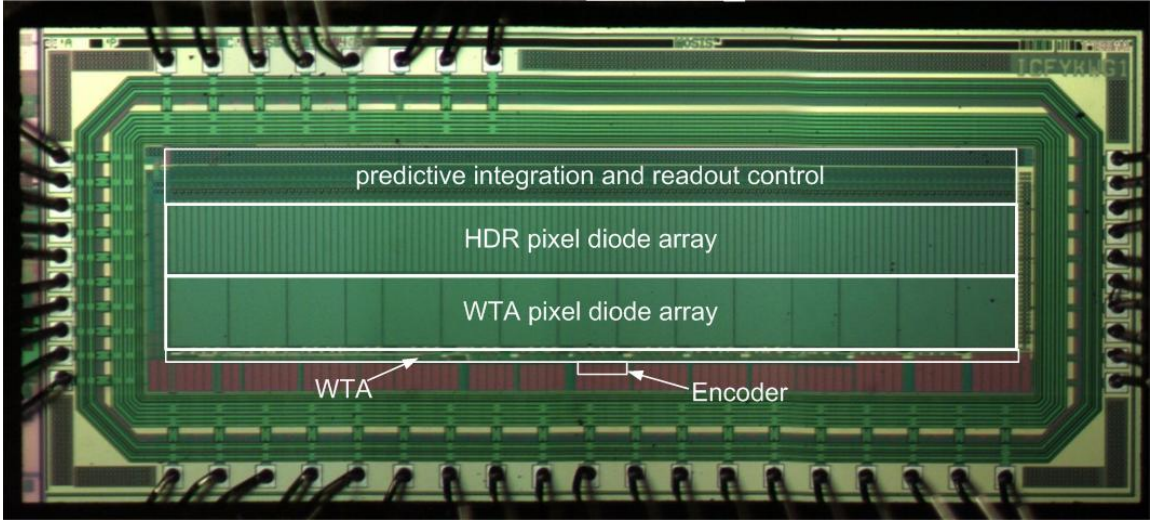
There are two functional blocks in the WTA pixel array: the WTA circuit and the ROI generator. The WTA circuit is implemented to determine which pixel in the WTA array receives the maximum local illumination. The ROI generator is used to generate a readout window for the HDR array. Each element in the WTA circuit consists of a comparator, two control transistors ( $MP$  and  $MN$ ), and a D flip-flop. During the reset, the photodiode in each pixel is charged to  $V_{dd}$ , and the comparator

is “0” at the output. At this moment,  $MN$  is turned off, and the common clock line  $Clk_{comm}$  is charged to “1”. The register in each pixel is initialized to “0”, and the 5-b down counter is initialized to the value of 24 at the same time. During the integration period, the voltage at the photodiode node ( $V_{ph}$ ) decreases due to a photocurrent discharge. The pixel exposed to the maximum local illumination discharges faster than others. When  $V_{ph} \leq V_{ref}$ , the comparator flips to “1”, and the  $Clk_{com}$  line goes to “0” with  $MN$  turned ON. The falling edge of the  $Clk_{com}$  line will trigger the registers to update to the current outputs of the comparators. The first pixel to flip its comparator output will take the control of the  $Clk_{com}$  line. Any late arrival pixel cannot subsequently alter the content of the register. The output from the WTA circuit is passed to the 16-128 encoder in the ROI generator, which will generate a pattern that indicates the starting position of the ROI window. This generated data pattern is then loaded to the shift register with  $LD\_SH$  high. When the *enable* signal is high, the down counter begins to count down from 24, which enables the shift register to choose 24 consecutive HDR pixels in the ROI window.

## 8.4 Chip test and characterization

This prototype chip is fabricated with a standard 0.18- $\mu\text{m}$  one poly-silicon and six metal (1P6M) digital technology. A micrograph of this chip is shown in Figure 8.7.

The performance of this chip has been evaluated from the following aspects.



**Figure 8.7:** Laser rangefinding chip micrograph

*A. Conversion gain measurement*

Conversion gain can be measured by measuring the noise of the pixel, if the photon shot noise is dominant [Beec 96].

$$g = \frac{S^2}{\bar{x}} \quad (8.7)$$

where  $g$  is the conversion gain,  $S^2$  is the variance of signal  $x$ , and  $\bar{x}$  is the mean of the signal. To perform this measurement, the integration time has to be long enough to keep the photon shot noise dominant. The mean of the output signal of pixel  $j$  in

the image sensor can be calculated as:

$$\bar{x}_j = \frac{1}{M} \sum_{i=1}^M x_{j,i} \quad (8.8)$$

where  $M$  is measurement times. The variance of the output signal of pixel  $j$  is:

$$S_j^2 = \frac{1}{M-1} \sum_i^M (x_{j,i} - \bar{x}_j)^2 \quad (8.9)$$

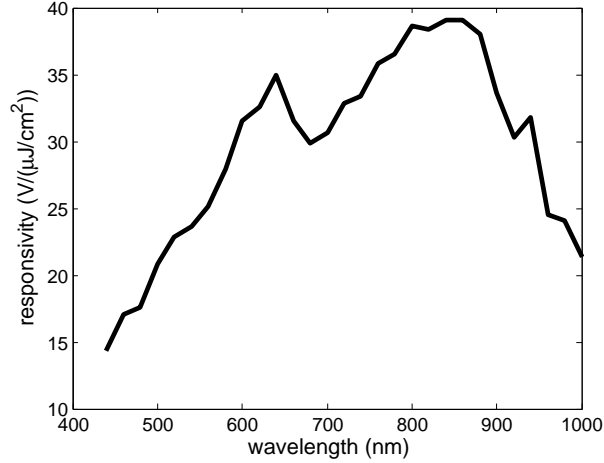
If  $N$  different pixels are considered, the mean conversion gain over  $N$  pixels can be treated as the final conversion gain. In this measurement, 15 pixels were tested, and each pixel was measured 100 times. The incident light has an intensity density of  $1.35 \mu\text{W}/\text{cm}^2$  and a wavelength of 632.8 nm. The measured mean conversion gain is  $0.69 \mu\text{V}/e^-$ . It is noticed that the conversion gain is quite small compared to many reported designs ( $\sim 10 \mu\text{V}/e^-$ ). It is attributed to both the manufacturing process and the large size of the photodiode, which is selected to minimize the speckle noise.

### *B. Spectral response*

An ORIEL monochromator was used for spectral response measurement. Twenty different pixels from HDR array were selected as test objects. For each pixel, we measured the output signal 50 times at a specific wavelength. Wavelength was varied from 440 nm to 1000 nm with a step of 20 nm. The spectral responsivity was first calculated by using the following equation:

$$R(\lambda) = \frac{\Delta V}{I_O} \quad (8.10)$$

where  $\Delta V$  is the output photo signal, and  $I_O$  is the incident light intensity. Figure 8.8 shows the spectral responsivity of the chip.



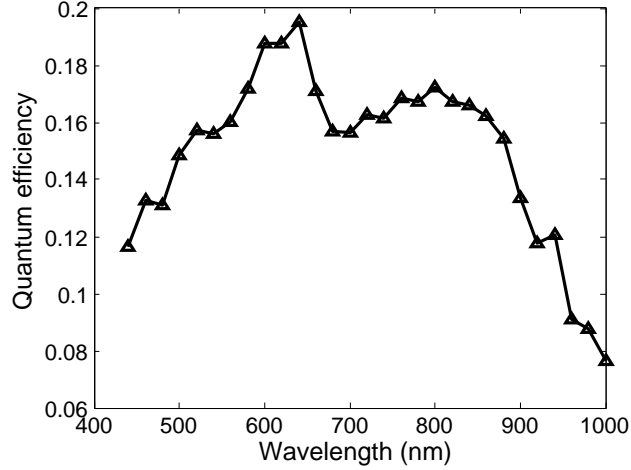
**Figure 8.8:** Measured spectral responsivity

The quantum efficiency of the image sensor can then be calculated from the spectral responsivity:

$$QE(\lambda) = \frac{R(\lambda)E(\lambda)}{Ag} \quad (8.11)$$

where  $E(\lambda)$  is the energy per photon at a specific wavelength,  $A$  is the pixel active area, and  $g$  is the conversion gain. Figure 8.9 shows the spectral quantum efficiency of the chip. The peak response appears around 640 nm. A relatively low quantum efficiency is observed in this chip. This is attributed to the use of a standard digital process technology, which is not optimized for CMOS image sensor performance.

### *C. Noise measurement*



**Figure 8.9:** Measured average spectral response

Read noise and fixed pattern noise (FPN) were measured here to give an information on the noise floor and the uniformity of the image sensor. Read noise of a pixel is measured by calculating the standard deviation of the pixel output under the dark illumination in  $M$  frames.

$$read\_noise_{pixel} = \sqrt{\frac{\sum_{frame=1}^M (x_{pixel} - \bar{x}_{pixel})^2}{M - 1}} \quad (8.12)$$

where  $M$  is the number of frames,  $x_{pixel}$  is output of the pixel,  $\bar{x}_{pixel}$  is the mean value of the output of the pixel. If  $N$  different pixels are measured, the mean read noise over  $N$  pixels is considered to be the final read noise. Ten pixels were selected at random for this evaluation, and each pixel was measured 100 times. The measured read noise is 1.4 mV.

FPN was measured for both the dark and illuminated conditions. Dark FPN is

dominated by the offset mismatches and gain differences in the signal path, and the illuminated FPN is dominated by the shot noise. With no illumination,  $M$  frames of  $N$  random pixels in the array were captured. The temporal noise of each pixel should be removed by averaging output signals over  $M$  frames. FPN under the dark condition is calculated as follows:

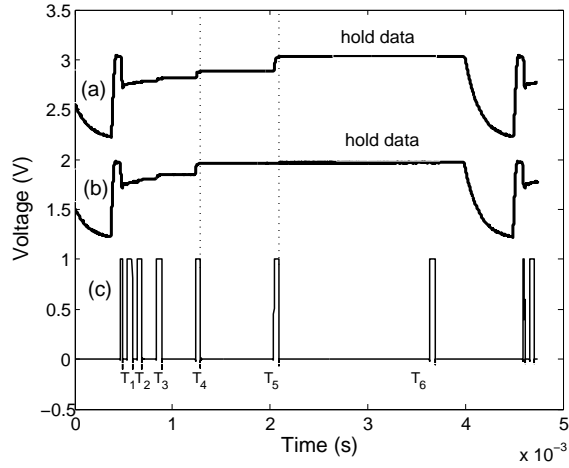
$$FPN = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}} \quad (8.13)$$

where  $x_i$  is the mean output of pixel  $i$ , and  $\bar{x}$  is the mean output of studied pixels. The measured FPN under the dark illumination is 1.2 mV.

The illuminated FPN was measured with the same way as the dark FPN except that an incident light was used. It was calculated as 6.5 mV when the light intensity was  $2.05 \mu\text{W}/\text{cm}^2$  and the wavelength was 632.8 nm.

#### *D. Predictive integration test*

The predictive integration control block should enable a pixel to adjust the integration period to the local illumination. By illuminating the image sensor with different light intensities, we should notice a change in the predictive integration time. Figure 8.10 demonstrates that a HDR pixel stops integration at sub-period 4 with a light intensity of 76.5 lux, and at sub-period 5 when the light intensity is 49.1 lux. After the integration stops, the pixel value will be held until it is read out.

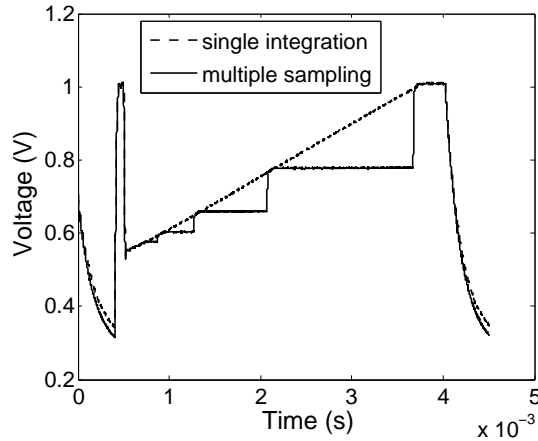


**Figure 8.10:** Predictive integration with multiple sampling. (a) Analog output at a light intensity of 49.1 lux. (b) Analog output at a light intensity of 76.5 lux. (c) Sampling control signal

The charge-injection noise induced by the sampling switch is not a big issue in this multiple-sampling scheme, and only the last switching of the sampling switch will contribute the charge-injection noise. The multiple-sampling should therefore give the same result as the single integration. By using a large sampling capacitor, the charge-injection noise is reduced to an acceptable level. In our design, the sampling capacitor is 1 pF, and the measured charge-injection noise is around 1 mV. Figure 8.11 also shows that multiple-sampling follows the single integration very well.

### *E. DR measurement*

Predictive integration is expected to improve the dynamic range of the image sensor by a factor of  $T_{int\_max}/T_{int\_min}$ . To verify this, a photo-conversion characteristic

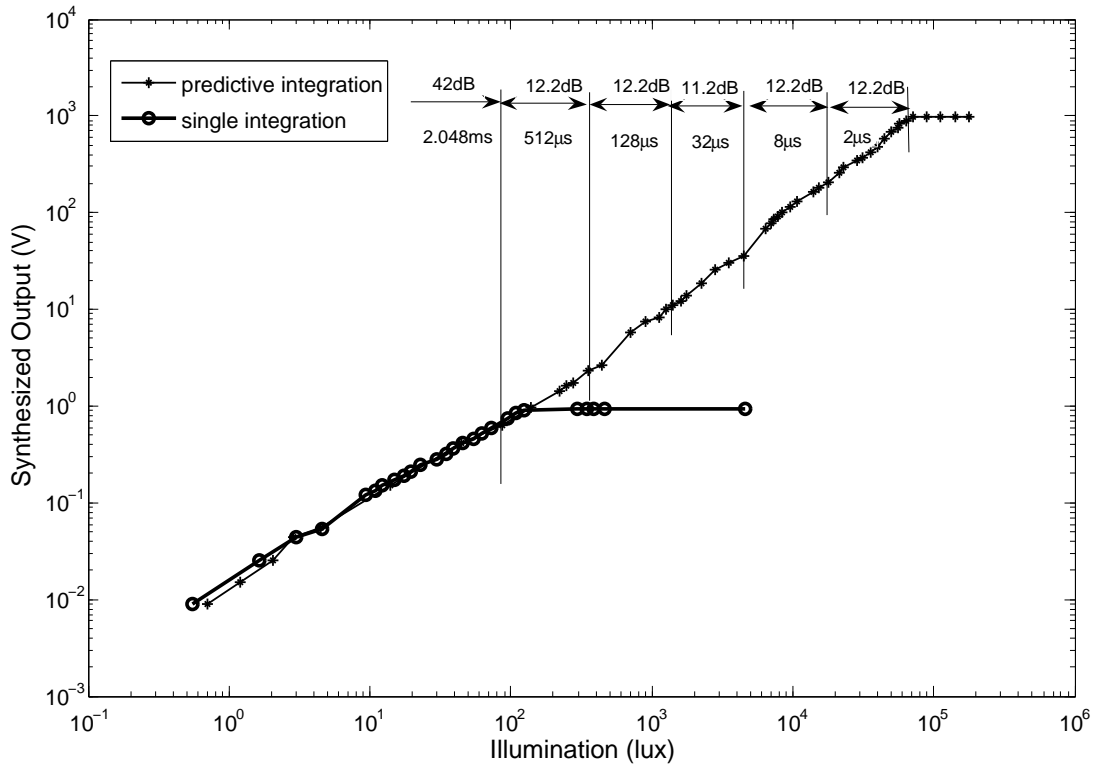


**Figure 8.11:** Multiple sampling and single integration

measurement was performed with 6 sub-integration periods being used. The minimum integration period was  $2 \mu\text{s}$ , and the maximum integration period was 2.048 ms. In the present implementation, each subsequent integration period is four times as long as the previous one. With white light illumination, the dynamic range of this chip can be extended to 102 dB, which is increased by 60 dB compared to a single integration scheme, as shown in Figure 8.12. The output in the predictive scheme is the synthesized output, which can be calculated from Equation 8.6 with  $k = 4$

$$V_{out\_syn} = V_{out\_pixel}(k^N/k^m) = V_{out\_pixel}(4^6/4^m) \quad (8.14)$$

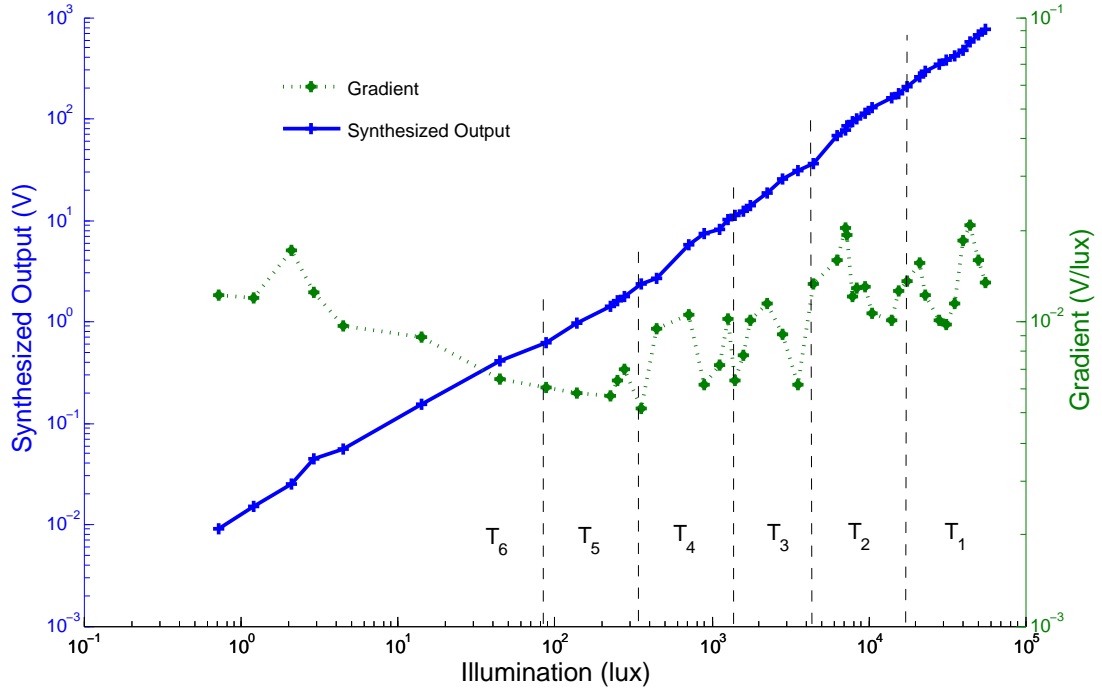
The conversion curve obtained from the predictive integration mode illustrates the conversion characteristics of the HDR pixel along with the incident light intensity, which can be divided into multiple regions defined by different integration times. It



**Figure 8.12:** Photoconversion characteristics for single and predictive integrations.

is noticed that the base dynamic range of this chip is only 42 dB, which is less than the typical value of approximately 50 dB [Saff 07] in a single integration mode. This loss in the low illumination end is due to the relatively low sensitivity of the WTA pixel. The output of the HDR pixel is only enabled when the readout window is activated.

A conversion slope curve from the photo-conversion characteristics is shown in Figure 8.13.  $T_1$ - $T_6$  represent segmented integration times, which are selected by the



**Figure 8.13:** Conversion slope as a function of the incident light intensity.

HDR pixel automatically. It is noticed that the conversion slope under a relatively

low illumination tends to decrease with increasing the light intensity. This is attributed to the non-linearity of the photodiode. After the reset, the photodiode is reverse biased and it works in the accumulation mode with its positive node floated. The potential change in the accumulation mode can be expressed as:

$$C_{PD}(V)\frac{dV}{dt} + I_{ph} + I_d = 0 \quad (8.15)$$

where  $I_d$  is the dark current of the photodiode,  $I_{ph}$  is the photo-current, and  $C_{PD}(V)$  is the junction capacitance of the photodiode, which equals to:

$$C_{PD}(V) = A_{area} \frac{k\epsilon_0}{W_{dep}(V)} \quad (8.16)$$

where  $A_{area}$  is the  $pn$  junction area,  $k$  is the silicon dielectric constant, and  $\epsilon_0$  is the permittivity of empty space,  $W_{dep}(V)$  is the width of the depletion region.  $W_{dep}(V)$  changes as the voltage across the junction changes. The larger the reverse bias voltage, the bigger  $W_{dep}(V)$  becomes.

If the effect of the dark current is ignored in Equation 8.15, we can get:

$$-C_{PD}(V)dV = (I_{ph} + I_d)dt \approx I_{ph}dt = dQ \quad (8.17)$$

The conversion rate can be calculated as:

$$-\frac{dV}{dQ} = \frac{1}{C_{PD}(V)} \quad (8.18)$$

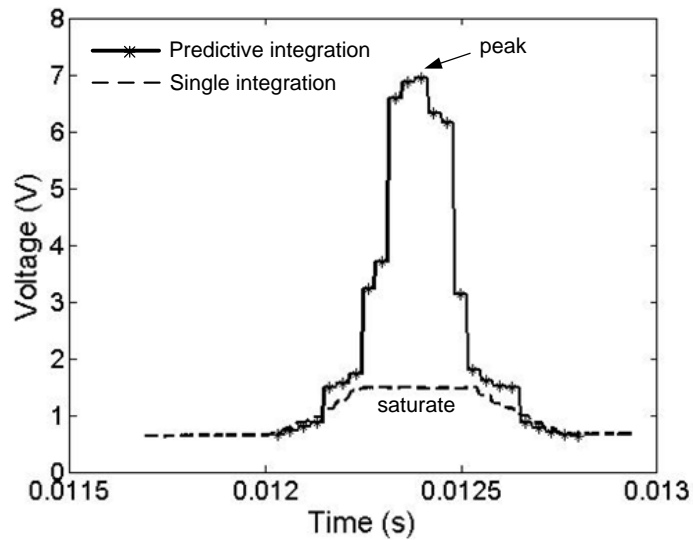
When the light intensity is high, the bias voltage across the photo-diode decreases quickly, which will lead to a larger junction capacitance. As a result, the conversion rate decreases.

When the light intensity is increased beyond the region defined by T5, the conversion behavior changes. The higher light intensity signal has to be held longer on the sampling capacitor before it is readout. The discharge makes it appear that there is more photo-signal, and it increases towards the right, hence the increased gradient. This effect may be enhanced, because M5 (see Figure 8.5) is not optically shielded and leakage (causing additional discharge of  $C_S$ ) is expected at higher illuminations. If a segment fitting is done for each region, slopes (V/lux) for these fitting curves are: 0.0073, 0.0063, 0.0085, 0.0086, 0.0118, 0.0145. It is noticed that the conversion slope increases under a relatively high illumination condition due to an extra photo-induced leakage. This undesired effect can be eliminated by putting a shielding metal on top of other components than the photodiodes.

We note that there appear to be no features on the curve associated with the transitions between integration times, indicating that the predictive integration is working well.

To demonstrate how well the predictive integration scheme can be used for laser spot peak location detection, the array was operated in both single and predictive

integration modes. The same laser spot was used for comparison. The peak location of the laser spot, cannot be measured in the single integration mode due to saturation (Figure 8.14), can easily be localized by the predictive integration scheme.

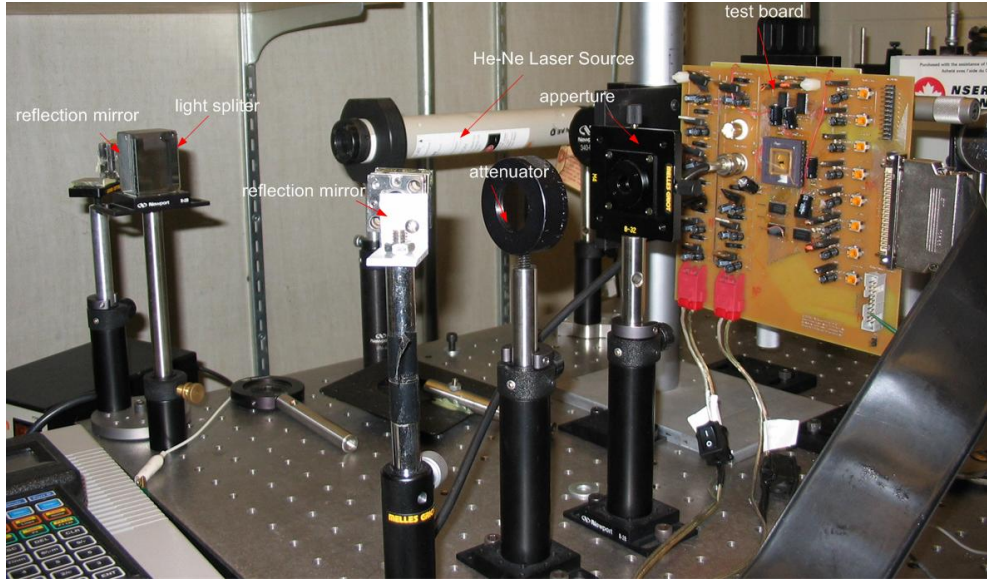


**Figure 8.14:** Laser spots captured with the single and predictive integration schemes.

#### *F. Laser spot location measurement*

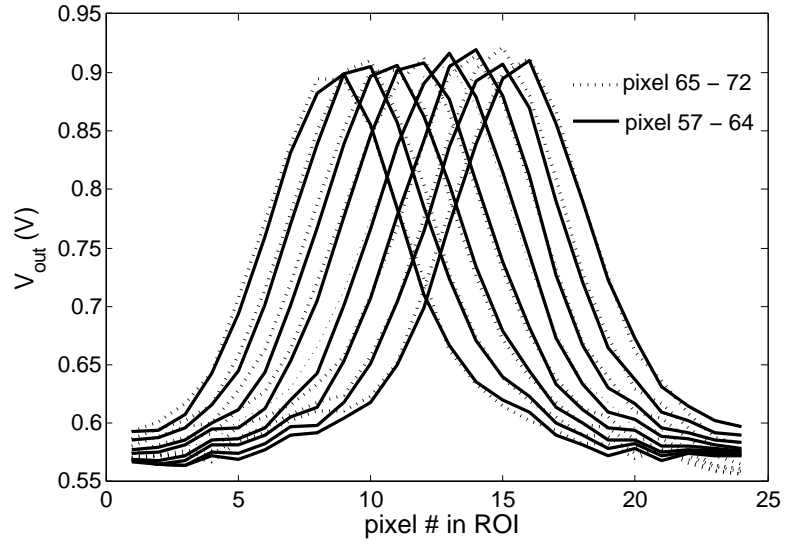
For 3D range data extraction, the centroid of a laser spot should be accurately localized when the laser spot moves along the image sensor. To simulate the movement of the laser spot, the test board was mounted on a micro-translation stage, which can move along the vertical and horizontal directions independently. The minimum step of the translation stage was  $5 \mu\text{m}$ . A He-Ne laser source with a power intensity of

$2.1 \text{ mW}/\text{cm}^2$  was used in this measurement. A  $100 \text{ }\mu\text{m}$  optical aperture was placed just before the image sensor to limit the size of the laser spot. The test setup is demonstrated in Figure 8.15. We first adjusted the sensor vertically to let the laser



**Figure 8.15:** Laser spot centroid detection setup.

spot fall on both HDR and WTA arrays. Next, we moved the sensor horizontally, and made a measurement of the laser spot. The peak of the output signal, representing the centroid of the laser spot, should move across the sensor chip. If the displacement of the micro-stage is  $25 \text{ }\mu\text{m}$ , one pixel displacement in the ROI window should be observed. Measurements were taken at 15 consecutive movements with a step of  $25 \text{ }\mu\text{m}$ . The final result is shown in Figure 8.16. We notice that the peak of the laser spot moves from pixel 9 to pixel 16 in the ROI window when the laser spot



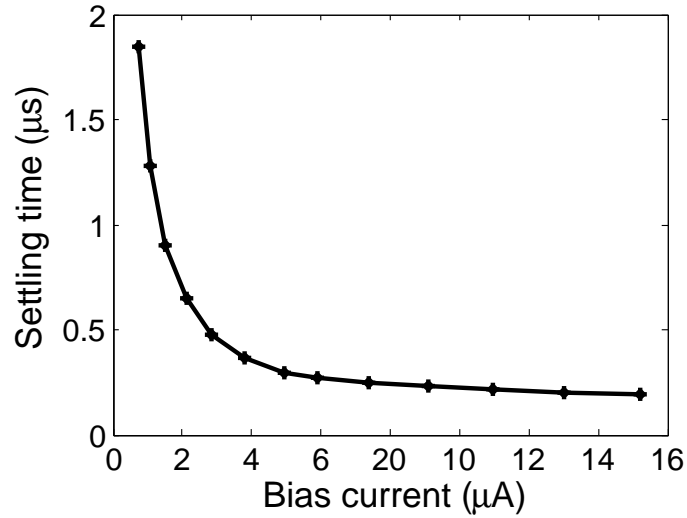
**Figure 8.16:** Laser spots measured at 15 consecutive movements.

moves from pixel 57 to pixel 64 in the image sensor. The peak appears at pixel 9 to pixel 16 of the ROI window again when the movement of the laser spot equals to 8 HDR pixel widths, which is the case for pixel 65–72 in Figure 8.16. This repetition is due to the shift of ROI window over the HDR array. When the centroid of the laser spot moves to an adjacent pixel in the WTA array, the ROI window will have a shift equivalent to 8 HDR pixels. The location of the laser spot can be determined by the winner pixel in the WAT array and the peak location in the ROI window.

#### *G. Data readout rate*

To achieve a high data readout rate, a short frame time must be employed. As such, the minimum integration period ( $T_{int\_min}$ ) within a frame operation must be

reduced in order to keep the same dynamic range. If we look at the timing diagram in Figure 8.6, we notice that the minimum  $T_{int\_min}$  equals to the pulse width of  $V_{S\_in}$ , which is usually selected to guarantee that the sampled signal settles within that specific period. The settling time of the source follower is dominated by the discharging operation, which was measured at different bias currents, as shown in Figure 8.17. The worst case, when the signal drops from  $V_{reset}$  to 0, is considered.



**Figure 8.17:** Settling time at different source-follower bias currents.

We can see that a short settling time requires a large bias current. We also notice that the settling time does not decrease significantly when bias current is larger than  $\sim 6 \mu A$ . For a practical reason,  $6 \mu A$  is selected as the maximum bias to make a better trade-off between the data readout rate and the power consumption.

The  $T_{int\_max}$  is  $1024T_{settling}$ , which is around  $276 \mu\text{s}$ . Assuming an A/D conversion time of  $8 \mu\text{s}$  as in [De N 05], the 3D data readout rate can reach the equivalent of 3500 range-points/s. To achieve a range-map of  $640 \times 480$ , approximately 88 s is needed. While this performance is acceptable in the present application to laser rangefinding, in which the target is static, the technique does not currently permit video-rate operation.

#### *H. Power consumption*

The power consumptions in digital and analog functional blocks of this laser rangefinding chip were measured when the chip operated at the maximum frame rate (3500 frames/s), which is shown in Table 8.1. The chip was powered with a single 1.8 V power supply. It is noticed that a large portion of the power consumption is contributed by the analog output buffers, which are required to drive a large capacitive load. Both digital and analog power will be reduced by slowing down the frame rate. The power consumption can be further reduced if a power optimization design technique is applied, which has been demonstrated in Chapters 4, 5, and 6.

## **8.5 Summary**

A CMOS laser-spot position detection sensor was successfully developed to achieve a high dynamic range using the standard  $0.18 \mu\text{m}$  process when it was powered as

**Table 8.1:** Power consumption with 3500 frames/s at 1.8V power supply

<b>Functional blocks</b>	<b>Current</b>	<b>Power</b>
Source followers	0.768 mA	1.38 mW
CDS amplifiers	0.43 mA	0.77 mW
Comparators (HDR)	1.04 mA	1.87 mW
Comparators (WTA)	0.129 mA	1.23 mW
Output buffers	1.27 mA	2.29 mW
Digital control circuits	0.59 mA	1.06 mW
Total analog power	6.55 mW	
Total digital power	1.06 mW	

low as 1.8 V. The proposed design has achieved an  $\sim 100$ -dB DR by using a new predictive-integration scheme. In this design, the pixel is prevented from saturating by automatically adjusting the integration time to the local illumination level. This predictive-integration scheme is not process-specific, therefore, it should have a good scalability and can be applied to other deep sub-micron processing technologies. A summary of the specifications and performance of the chip is shown in Table 8.2. We notice that the conversion gain in the HDR pixel is quite small in this design due to a large photodiode size. This small conversion gain will not degrade the DR but will lead to a reduced sensitivity.

**Table 8.2:** Specifications and performance of the laser rangefinding chip

<b>Technology</b>	0.18 $\mu\text{m}$ TSMC standard, 1P6M	<b>ROI size</b>	24 pixels
<b>Chip size</b>	4 mm x 1.5 mm	<b>Maximum sub-integration periods</b>	7
<b>Package</b>	84-pin PGA	<b>Dynamic range</b>	102 dB
<b>Array size</b>	1 x 128 $\mu\text{m}$ (HDR) 1 x 16 $\mu\text{m}$ (WTA)	<b>Optimized 3D data readout speed</b>	3500 range-points/s
<b>Pixel size</b>	25 x 250 $\mu\text{m}$ (HDR) 200 x 250 $\mu\text{m}$ (WTA)	<b>Conversion gain (HDR)</b>	0.69 $\mu\text{V}/e^-$
<b>Photodiode</b>	n+/p-substrate	<b>FPN (dark)</b>	1.2 mV
<b>Fill factor</b>	100%	<b>Read noise</b>	1.4 mV

## 9 Conclusion and suggestions for future work

CMOS active pixel image sensors have been widely used in portable devices, such as digital cameras, mobile phones, laptop computers, and bio-sensors. CMOS APS image sensors beat their counterparts, CCD image sensors, in the role of portable applications due to their intrinsically low power consumption, and their compatibility with advanced CMOS processing technologies in achieving integration with other functional blocks. Power consumption is a critical issue in portable applications due to the limitation of the battery lifetime. CMOS APS image sensors have benefited from technology scaling by using scaled power supplies, but power consumption in start-of-the-art imaging systems is expected to increase or, at least, not to decrease greatly, due to increased complexity and resolution. To meet stringent power requirements without harming the performance of the image sensors, special low power design techniques must be employed. Low-power design for digital and analog blocks in CMOS APS image sensors have to be treated independently. Power consumption can be effectively reduced by reducing the power supply in digital cir-

cuits. Many digital EDA tools can be readily used for low power design in digital functional blocks with a good accuracy and reasonable expense. Implementing low power design in analog circuits is not as simple as in digital circuits, because simply reducing the power supply can lead to performance degradation. Analog EDA tools are not as mature as digital ones. They usually have a limited circuit library and are very expensive. This research explores simple, non-expensive, and comprehensive ways to perform the low power design in analog blocks of CMOS APS image sensors. Two major issues related to low power design in analog blocks of CMOS APS image sensors in modern sub-micron processing technologies are studied in this research: they are power minimization for a required performance, and output swing reduction caused by lowered power supplies.

Automatic-synthesis-based analog design was chosen as an efficient method to minimize the power consumption of an analog circuit, in which a power optimal design constrained by other performance measures can be easily found by performing power optimization. After comparing three typical global optimization algorithms, the geometric programming method was selected for aiding low power design in basic analog building blocks of CMOS image sensors due to its computation efficiency and its true global property.

To improve GP-based synthesis accuracy in sub-micron processing technologies,

a new sub-space based max-monomial modeling for electrical parameters of CMOS transistors was proposed. This approach was demonstrated to have a better accuracy for sub-micron technologies than single-space models. TSMC 0.18  $\mu\text{m}$  processing technology was considered in this study. The design space ranging from weak- to strong-inversion region was divided into 52 sub-spaces according to the transistor channel length and gate-source voltage (most important to the model accuracy). The computation inefficiency associated with sub-space based scheme can be improved by using a practical performance requirement. The same modeling method can be applied to other deep sub-micron processing technologies.

By using derived sub-space based max-monomial models, three key components (in-pixel source followers, CDS op-amps, and comparators in a successive-approximation A/D converter) were synthesized using geometric programming to achieve the minimum power consumption. In the synthesis of low power in-pixel source followers, the settling time was modeled in a posynomial form in terms of  $L$ ,  $W$ ,  $I_{ds}$ , and  $C_L$  by using a simulation-based method. With the settling time model, a source follower design was treated as a geometric programming problem. This synthesis method was evaluated by comparing synthesis results with HSPICE simulation results for six random design requirements. It was noticed that the maximum relative error of the settling time was less than 25%, and relative errors in most of cases were less than

10%. In fact, this difference can be reduced by slightly adjusting the bias voltage of the bias transistor.

In the synthesis of low power CDS op-amps, three different types of op-amps were investigated. They are two-stage Miller op-amp, a CS OTA, and a CMFB OTA. Performance measures for each op-amp were properly modeled in posynomial forms in terms of design variables, and power optimal designs for these three op-amps were solved using geometric programming. It was noticed that the CMFB OTA can lead to the minimum power consumption among these three topologies. An efficient scheme in dealing with non-convex constraint inherent in Kirchhoff's voltage law was proposed here as well. By using this scheme, the non-convex constraint, such as posynomial equality, can be relaxed to a posynomial inequality without affecting the result. A fabricated chip with six op-amps (two from each type) was tested for verification. Three typical measurements, DC, frequency response, and transition measurement, were performed to extract performance measures of these op-amps. Test results showed that predictive results from synthesis circuits had a good agreement with measurement results, and it can be concluded that the subspace modeling based GP synthesis can lead to a good reliability in analog circuit design in sub-micron technologies.

A hierarchical geometric programming method was proposed to synthesize a CDS

op-amp driven by the settling time. The high-level GP used the settling time constraint to solve a power optimization problem and gave solutions to other performance measures, such as GBW and SR. With GBW and SR requirements, a CDS op-amp was synthesized with a conventional GP as used for an op-amp design.

In the synthesis of comparators for a SAR ADC, a two-stage open loop op-amp is investigated. A new synthesis methodology is proposed to deal with propagation-delay driven problems, in which input and output stages were treated independently by using geometric programming. A SAR ADC, using a comparator synthesized from the suggested method, was evaluated with HSPICE simulation. The differential non-linearity was within  $\pm 0.6$  LSB, integral non-linearity was within  $\pm 0.5$  LSB, and no missing codes were observed.

GP-based power optimization can lead to a power efficient design without changing the existing design architecture. A comparison between basic analog blocks from a published design and a design from the suggested technique was made, and an improvement in power consumption was predicted. Taking more efforts to derive posynomial models for electrical parameters of MOS transistors and performance measures of an individual analog block is the major drawback in low-power analog design with using sub-space-based GP power optimization. Models for electrical parameters of MOS transistors have to be re-derived when the processing technol-

ogy changes, and new sub-spaces have to be determined according to the change in properties of MOS transistors. Models for a specific analog circuit architecture are independent of the processing technology, and may require a greater effort to model them in posynomial forms. Efforts for transistor modeling are considered as initial setup efforts for a specific processing technology, which are shared by many different designs. Efforts for circuit modeling are considered as overall setup efforts, which are shared by all processing technologies. These disadvantages in sub-space-based GP synthesis can be offset by the efficiency of the synthesis.

A new predictive-integration scheme was proposed in this study to overcome output swing reduction of CMOS image sensors caused by lowered power supplies. This scheme can increase the dynamic range of the pixel to compensate for the loss caused by the output swing reduction. The predictive-integration scheme can prevent the pixel from saturating by automatically adjusting the integration time to the local illumination level. Without searching all possible sub-integration periods and without introducing any extra reset within a frame operation, the control in a high dynamic range pixel in this scheme becomes much simpler compared to other similar schemes. This predictive-integration scheme is not subjected to processing technologies, therefore, it has a good scalability and can be applied to other deep sub-micron processing technologies. A CMOS laser-spot position detection sensor

was successfully developed using this methodology. This design has achieved an  $\sim 100$ -dB DR when it was powered as low as 1.8 V.

## 9.1 Future work

### 9.1.1 Model

The current modeling method for electrical parameters of CMOS transistors is based on a piecewise linear modeling method. Each parameter model in a specific design-space comprises of several tens of terms; therefore, it degrades the advantage of the geometric programming by increasing the computation time. To improve the computation efficiency, a new modeling method should be used to derive a posynomial model with fewer terms. A genetic-algorithm-based modeling method might be a good candidate for deriving posynomial models in the sub-space domain.

Geometric-programming-based power optimization of analog blocks of CMOS APS image sensors has been studied at the circuit level, in which each analog component is synthesized to achieve the minimum power consumption and meet its specific performance requirements. The performance of a system can be obtained by a variety of different combinations of specifications for individual sub-blocks. Actually, a different assignment of performance requirements for each analog block can result in different overall power consumption. Power optimization at the system level

should be conducted in the future to allow designers to find the optimal performance requirement for each analog block.

In order to develop a synthesis tool to facilitate low-power analog design by using GP-based power optimization for other applications, a library consisting of posynomial models for different circuit architectures should be generated in the future.

### **9.1.2 Chip**

As mentioned in Chapter 8, the operating speed of the laser rangefinding image sensor is subjected to the settling time of the in-pixel source follower. In this design, the settling time is limited by the discharge operation of the source follower. If a discharging transistor is added across the sampling capacitor as it was mentioned in Chapter 5, the discharge operation of the source follower can be eliminated; therefore, the operation speed can be increased without increasing bias current. Another problem existing in the current design of the laser rangefinding image sensor is the low base dynamic range caused by the low sensitivity in the large WTA pixels. In the current design, the size of a WTA pixel is 8 times as large as a HDR pixel. If a WTA pixel can be reduced to the same size as the HDR pixel, and a new decoding circuit can be designed, the base dynamic range should be improved accordingly. By adding an on-chip ADC and converting the analog signal to a digital signal right

after the sat-positive is sensed can prevent a signal from being discharged at the sampling capacitor, and therefore, the conversion performance can be improved.

Due to time limitations, the sub-space-based GP power optimization has only been applied to designs in 0.18  $\mu\text{m}$  technology. A study on the feasibility of this technique in other sub-micron technologies should be considered as future work. Except for CDS op-amps, synthesized in-pixel source followers and SAR ADCs were only verified by HSPICE simulations. Verifying other analog blocks with fabricated chips would be useful in helping to design a low-power CMOS image sensor by minimizing the power consumption efficiently.

## Bibliography

- [Acos 04] P. Acosta-Serafini, I. Masaki, and C. Sodini. “A 1/3” VGA linear wide dynamic range CMOS image sensor implementing a predictive multiple sampling algorithm with overlapping integration intervals”. *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 9, pp. 1487–1496, Sep. 2004.
- [Afza 03] A. Afzalian and D. Flandre. “Modeling of the bulk versus SOI CMOS performances for the optimal design of APS circuits in low-power low-voltage applications”. *IEEE Trans. on Electron Devices*, Vol. 50, No. 1, pp. 106–110, Jan. 2003.
- [Agar 06] A. Agarwal, S. Mukhopadhyay, A. Raychowdhury, K. Roy, and C. Kim. “Leakage power analysis and reduction for nanoscale circuits”. *IEEE Micro*, Vol. 26, No. 2, pp. 68–80, Apr. 2006.
- [Agga 07a] V. Aggarwal. *Analog Circuit Optimization using Evolutionary Algorithms and Convex Optimization*. Master’s thesis, Massachusetts Institute of Technology, Cambridge, 2007.
- [Agga 07b] V. Aggarwal and U.-M. O’Reilly. “Simulation-based reusable posynomial models for MOS transistor parameters”. In: *Proc. Design, Automation & Test in Europe*, pp. 69–74, Apr. 2007.
- [Agga 09] B. Aggarwal and M. Gupta. “Low-Voltage cascode current mirror based on bulk-driven MOSFET and FGMOS techniques”. In: *International Conference on Advances in Recent Technologies in Communication and Computing*, pp. 473–477, Oct. 2009.
- [Agui 08] P. Aguirre and F. Silveira. “CMOS op-amp power optimization in all regions of inversion using geometric programming”. In: *Proc. 21<sup>st</sup> Symposium on Integrated Circuits and System Design*, pp. 152–157, Bramado, Brazil, Sep. 2008.

- [Akah 06] N. Akahane, S. Sugawa, S. Adachi, K. Mori, T. Ishiuchi, and K. Mizobuchi. “A sensitivity and linearity improvement of a 100-dB dynamic range CMOS image sensor using a lateral overflow integration capacitor”. *IEEE Journal of Solid-State Circuits*, Vol. 41, pp. 851–858, Apr. 2006.
- [Alle 04] P. E. Allen and D. R. Holberg. *CMOS Analog Circuit Design*. Oxford University Press, New York, Oxford, 2004.
- [Alpa 00] G. Alpaydin, G. Erten, S. Balkir, and G. Dundar. “Multi-level optimization approach to switched capacitor filter synthesis”. *IEE Circuits, Devices and Systems*, Vol. 147, No. 4, pp. 243–249, Aug. 2000.
- [Band 06] A. Bandyopadhyay, J. Lee, R. Robucci, and P. Hasler. “MATIA: a programmable 80  $\mu$ W/frame CMOS block matrix transform imager architecture”. *IEEE Journal of Solid-State Circuits*, Vol. 41, No. 3, pp. 663–672, March 2006.
- [Beec 96] B. P. Beecken and E. R. Fossum. “Determination of the conversion gain and the accuracy of its measurement for detector elements and arrays”. *Applied Optics*, Vol. 35, No. 19, pp. 3471–3477, July 1996.
- [Bele 07] A. Belenky, A. Fish, A. Spivak, and O. Yadid-Pecht. “Global shutter CMOS image sensor with wide dynamic range”. *IEEE Transactions on Circuits and Systems II*, Vol. 54, No. 12, pp. 1032–1036, Dec. 2007.
- [Bele 09] A. Belenky, A. Fish, A. Spivak, and O. Yadid-Pecht. “A snapshot CMOS image sensor with extended dynamic range”. *IEEE Sensor Journal*, Vol. 9, No. 2, pp. 103–111, Feb. 2009.
- [Beni 97] L. Benini and G. De Micheli. *Dynamic power management: design techniques and CAD tools*. Kluwer Academic Publishers, Massachusetts, USA, 1997.
- [Bera 03] J.-A. Beraldin, F. Blais, M. Rioux, J. Domey, L. Gonzo, F. De Nisi, F. Comper, D. Stoppa, M. Gottardi, and A. Simoni. “Optimization position sensors for flying-spot active triangulation systems”. In: *Proc. 4th Int. Conf. 3DIM*, pp. 20–36, Banff, Canada, Oct. 2003.
- [Blai 04] F. Blais. “Review of 20 years of range sensor development”. *J. Electron. Imaging*, Vol. 13, No. 1, pp. 231–243, Jan. 2004.

- [Bous 04] F. Boussaid, A. Bermak, and A. Bouzerdoum. “An ultra-low power operating technique for mega-pixels current-mediated CMOS imagers”. *IEEE Trans. on Consumer Electronics*, Vol. 50, No. 1, pp. 46–53, Feb. 2004.
- [Boyd 04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, United Kingdom, 2004.
- [Boyd 07] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi. “A tutorial on geometric programming”. *Optimization and Engineering*, Vol. 8, No. , pp. 67–127, March 2007.
- [Brei 04] A. Breidenassel, K. Meier, and J. Schemmel. “A flexible scheme for adaptive integration time control”. In: *IEEE Proc. of Sensors*, pp. 280–283, Oct. 2004.
- [Cama 05] E. Camacho-Galeano, C. Galup-Montoro, and M. Schneider. “A 2-nW 1.1-V self-biased current reference in CMOS technology”. *IEEE Transactions on Circuits and Systems II*, Vol. 52, No. 2, pp. 61–65, Feb. 2005.
- [Carv 05] R. Carvajal, J. Ramirez-Angulo, A. Lopez-Martin, A. Torralba, J. Galan, A. Carlosena, and F. Chavero. “The flipped voltage follower: a useful cell for low-voltage low-power circuit design”. *IEEE Transactions on Circuits and Systems I*, Vol. 52, No. 7, pp. 1276–1291, July 2005.
- [Cham 84] S. G. Chamberlain and J. P. Lee. “Silicon imaging arrays with new photoelements, wide dynamic range and free from blooming”. In: *Proc. Customer Integrated Circuits Conf.*, pp. 81–85, Rochester, 1984.
- [Chan 95] A. R. Chandrakasan and R. W. Brodersen. *Low power digital CMOS design*. Kluwer Academic Publishers, Massachusetts, USA, 1995.
- [Chen 06] H. Chen, Y. Hao, Z. Zhu, X. Ma, and Y. Cao. “Investigation of gate-induced drain leakage current in ultra-thin gate oxide LDD nMOS-FET’s”. In: *International Conference on Solid-State and Integrated Circuit Technology*, pp. 1159–1161, Oct. 2006.
- [Cho 03] K. Cho, A. Krymski, and E. Fossum. “A 1.5-V 550- $\mu$ W 176 x 144 autonomous CMOS active pixel image sensor”. *IEEE Transactions on Electron Devices*, Vol. 50, No. 1, pp. 96–105, Jan. 2003.

- [Crus 04] C. Crusius. “High-Speed Digital-to-RF Converter”. Sunnyvale, CA, USA Patent, 5668842, Nov. 2004.
- [Cunh 98] A. Cunha, M. Schneider, and C. Galup-Montoro. “An MOS transistor model for analog circuit design”. *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 10, pp. 1510–1519, Oct. 1998.
- [Daem 03] W. Daems, G. Gielen, and W. Sansen. “Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits”. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 22, No. 5, pp. 517–534, May 2003.
- [De N 05] F. De Nisi, F. Comper, L. Gonzo, M. Gottardi, D. Stoppa, A. Simoni, and B. J.-A. “A CMOS sensor optimized for laser spot-position detection”. *IEEE Sensors Journal*, Vol. 5, No. 6, pp. 1296–1304, Dec. 2005.
- [Deck 98] S. Decker, D. McGrath, K. Brehmer, and C. Sodini. “A 256 x 256 CMOS imaging array with wide dynamic range pixels and column-parallel digital output”. *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 12, pp. 2081–2091, Dec. 1998.
- [Dege 00] Y. Degerli, F. Lavernhe, P. Magnan, and P. J. Farre. “Column readout circuit with global charge amplifier for CMOS APS imagers”. *Electronics Letters*, Vol. 36, No. 17, pp. 1457–1459, Aug. 2000.
- [Dege 03] Y. Degerli, N. Fourches, M. Rouger, and P. Lutz. “Low-power autozeroed high-speed comparator for the readout chain of a CMOS monolithic active pixel sensor based vertex detector”. *IEEE Transactions on Nuclear Science*, Vol. 50, No. 5, pp. 1709–1717, Oct. 2003.
- [Degr 87] M. Degrauwe, O. Nys, E. Dijkstra, J. Rijmenants, S. Bitz, B. Goffart, E. Vittoz, S. Cserveny, C. Meixenberger, G. van der Stappen, and H. Oguey. “IDAC: an interactive design tool for analog CMOS circuits”. *IEEE Journal of Solid-State Circuits*, Vol. 22, No. 6, pp. 1106–1116, Dec. 1987.
- [Dhan 06] N. Dhanwada, A. Doboli, A. Nunez-Aldana, and R. Vemuri. “Hierarchical constraint transformation based on genetic optimization for analog system synthesis”. *J. VLSI Integration*, Vol. 39, pp. 267–290, June 2006.

- [Donn 98] S. Donnay, G. Gielen, and W. Sansen. “High-level power minimization of analog sensor interface architecture”. *Integrated Computer-Aided Engineering*, Vol. 5, pp. 303–314, 1998.
- [Dude 05] P. Dudek and P. Hicks. “A general-purpose processor-per-pixel analog SIMD vision chip”. *IEEE Transactions on Circuits and Systems I*, Vol. 52, No. 1, pp. 13–20, Jan. 2005.
- [Duff 67] R. J. Duffin. *Geometric Programming: theory and application*. Wiley, New York, USA, 1967.
- [Duri 10] D. Durini, A. Spickermann, R. Mahdi, W. Brockherde, H. Vogt, A. Grabmaier, and B. J. Hosticka. “Lateral drift-field photodiode for low noise, high-speed, large photoactive-area CMOS imaging applications”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. In Press, Corrected Proof, No. , pp. 470–475, 2010.
- [Elza 08] M. V. Elzaker, E. V. Tuijl, P. Geraedts, D. Schinkel, E. Klumperink, and B. Nauta. “A  $1.9\mu\text{W}$   $4.4\text{fJ}/\text{conversion-step}$   $10\text{b}$   $1\text{MS}/\text{s}$  charge-redistribution ADC”. In: *Dig. IEEE International Solid-State Circuits Conference*, pp. 244–245, Feb. 2008.
- [Enz 95] C. C. Enz, F. Krummenacher, and E. A. Vittoz. “An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current application”. *Analog Integrated Circuits and Signal Processing Journal on Low-Voltage and Low-Power Design*, Vol. 8, No. 1, pp. 83–114, July 1995.
- [Enz 96] C. Enz and E. Vittoz. “CMOS low-power analog circuit design”. In: *Designing Low Power Digital Systems, Emerging Technologies*, pp. 79–133, Aug. 1996.
- [Fish 05a] A. Fish, A. Belenky, and O. Yadid-Pecht. “Wide dynamic range snapshot APS for ultra low-power applications”. *IEEE Transactions on Circuits and Systems II*, Vol. 52, No. 11, pp. 729–733, Nov. 2005.
- [Fish 05b] A. Fish, S. Hamami, and O. Yadid-Pecht. “Self-powered active pixel sensors for ultra low-power applications”. In: *IEEE International Symposium on Circuits and Systems*, pp. 5310–5313, May 2005.

- [Gao 05] W. Gao and R. Hornsey. “Low power realization in main blocks of CMOS APS image sensor”. In: *SPIE Photonics North*, pp. 261–269, Sep. 2005.
- [Gao 09] W. Gao, M. Liscombe, R. Hornsey, and J.-A. Beraldin. “A 100-dB DR Predictive-Integration CMOS Image Sensor for Laser Rangefinding”. *IEEE Transactions on Electron Devices*, Vol. 56, No. 12, pp. 2987–2994, Dec. 2009.
- [Gao 10] W. Gao and R. Hornsey. “A power optimization method for CMOS op-amps using sub-space based geometric programming”. In: *Proc. Design, Automation, and Test in Europe*, pp. 508–513, Dresden, Germany, March 2010.
- [Giel 00] G. Gielen and R. Rutenbar. “Computer-aided design of analog and mixed-signal integrated circuits”. *Proc. IEEE*, Vol. 88, No. 12, pp. 1825–1854, Dec. 2000.
- [Giel 90] G. Gielen, H. Walscharts, and W. Sansen. “Analog circuit design optimization based on symbolic simulation and simulated annealing”. *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 3, pp. 707–713, June 1990.
- [Gira 06] A. Girardi and S. Bampi. “Power constrained design optimization of analog circuits based on physical gm/ $I_D$  characteristics”. In: *Symposium on Integrated circuits and systems design*, pp. 89–93, Aug. 2006.
- [Grue 10] V. Gruev, Z. Yang, J. Van der Spiegel, and R. Etienne-Cummings. “Current mode image sensor with two transistors per Pixel”. *IEEE Transactions on Circuits and Systems I*, Vol. 57, No. 6, pp. 1154–1165, June 2010.
- [Guo 07] X. Guo, X. Qi, and J. Harris. “A time-to-first-spike CMOS image sensor”. *IEEE Sensors*, Vol. 7, pp. 1165–1175, Aug. 2007.
- [Harj 89] R. Harjani, R. Rutenbar, and L. Carley. “OASYS: a framework for analog circuit synthesis”. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 8, No. 12, pp. 1247–1266, Dec. 1989.
- [Hers 01] M. Hershenson, S. P. Boyd, and T. H. Lee. “Optimal design of a CMOS op-amp via geometric programming”. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 20, No. 1, pp. 1–21, Jan. 2001.

- [Hers 02] M. Hershenson. “Design of pipeline analog-to-digital converters via geometric programming”. In: *IEEE/ACM International Conference on Computer Aided Design*, pp. 317–324, Nov. 2002.
- [Holl 75] J. H. Holland. *Adaptation in natural and artificial systems*. University of Michigan Press, USA, 1975.
- [Inou 02] T. Inoue, H. Nakane, Y. Fukuju, and E. Snchez-Sinencio. “A design of a low-voltage current-mode fully-differential analog CMOS integrator using FG-MOSFETs and its implementation”. *Analog Integrated Circuits and Signal Processing*, Vol. 32, pp. 249–256, Sep. 2002.
- [Jin 03] X. L. Jin and J. Chen. “Analysis of novel low-power quasi-dynamic ratio-less readout scanning shift register for CMOS imagers”. In: *International Conference on ASIC, 2003*, pp. 595–598, Oct. 2003.
- [John 97] D. A. Johns and K. Martin. *Analog Integrated Circuit Design*. John Wiley and Sons, Inc, New York, 1997.
- [Kava 00] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts. “A logarithmic response CMOS image sensor with on-chip calibration”. *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 8, pp. 1146–1152, Aug. 2000.
- [Kavu 06] S. Kavusi, K. Ghosh, and A. El Gamal. “Architectures for high dynamic range, high speed image sensor readout circuits”. In: *Proc. IFIP VLSI-SoC*, pp. 36–41, Nice, France, Oct. 2006.
- [Kim 04] J. Kim, J. Lee, L. Vandenberghe, and C.-k. K. Yang. “Techniques for improving the accuracy of geometric-programming based analog circuit design optimization”. In: *Proc. IEEE/ACM Computer Aided Design (ICCAD)*, pp. 863–870, San Jose, USA, Nov. 2004.
- [Ko 95] U. Ko and P. Balsara. “Short-circuit power driven gate sizing technique for reducing power dissipation”. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 3, No. 3, pp. 450–455, Sep. 1995.
- [Kwon 05] M. Kwon, Y. Chae, and G. Han. “Sub- $\mu$ W switched-capacitor circuits using a class-C inverter”. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Science*, Vol. E88-A, No. 5, pp. 1313–1319, May 2005.

- [Lai 04] L. W. Lai, Y. C. King, and J. Shemmel. “A novel logarithmic response CMOS image sensor with high output voltage swing and in-pixel fixed pattern noise reduction”. *IEEE Sensors Journal*, Vol. 4, pp. 122–126, Feb. 2004.
- [Lee 03] J. Lee, G. Hatcher, L. Vandenberg, and C.-K. K. Yang. “Evaluation of fully-integrated switching regulators for CMOS process technologies”. In: *International Symposium on System-on-Chip*, pp. 155–158, Nov. 2003.
- [Loos 01] M. Loose, K. Meier, and J. Schemmel. “A self-calibrating single-chip CMOS camera with logarithmic response”. *IEEE Journal of Solid-State Circuits*, Vol. 36, No. 4, pp. 586–596, Apr. 2001.
- [Lope 05] A. Lopez-Martin, S. Baswa, J. Ramirez-Angulo, and R. Carvajal. “Low-voltage super class AB CMOS OTA cells with very high slew rate and power efficiency”. *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 5, pp. 1068–1077, May 2005.
- [Lope 08] A. Lopez-Martin, R. G. Carvajal, E. Lopez-Morillo, L. Acosta, T. Snchez-Rodriguez, C. Rubia-Marcos, and J. Ramirez-Angulo. *Analog Circuit Design*, Chap. Low-Voltage Power-Efficient Amplifiers for Emerging Applications, pp. 147–165. Springer Netherlands, 2008.
- [Maci 98] E. Macii, M. Pedram, and F. Somenzi. “High-level power modeling, estimation, and optimization”. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 17, No. 11, pp. 1061–1079, Nov. 1998.
- [Magn 09] A. Magnani and S. P. Boyd. “Convex piecewise-linear fitting”. *Optimization and Engineering*, Vol. 10, No. 1, pp. 1–17, March 2009.
- [Maji 06] V. Majidzadeh and O. Shoaie. “A power optimized design methodology for low-distortion sigma-delta-pipeline ADCs”. In: *16th ACM Great Lakes symposium on VLSI*, pp. 284–289, New York, NY, USA, Apr. 2006.
- [Mand 01] P. Mandal and V. Visvanathan. “CMOS op-amp sizing using a geometric programming formulation”. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 20, No. 1, pp. 22–38, Jan. 2001.

- [Marc 06] A. Marczyk. “Genetic algorithms and evolutionary computation”. <http://www.talkorigins.org/faqs/genalg/genalg.html>, download on May, 2006.
- [McIl 01] L. McIlrath. “A low-power low-noise ultrawide-dynamic-range CMOS imager with pixel-parallel A/D conversion”. *IEEE Journal of Solid-State Circuits*, Vol. 36, pp. 846–853, May 2001.
- [Mend 97] S. Mendis, S. Kemeny, R. Gee, B. Pain, C. Staller, and E. Fossum. “CMOS active pixel image sensors for highly integrated imaging systems”. *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 2, pp. 187–197, Feb. 1997.
- [Muta 08] M. Mutapic, K. Koh, S. Kim, and S. Boyd. “A Matlab Toolbox for Geometric Programming”. <http://standford.edu/boyd/ggplab>, download on March, 2008.
- [Nour 98] Y. Nourani and B. Andresen. “A comparison of simulated annealing cooling strategies”. *Journal of Physics A: Mathematical and General*, Vol. 31, No. 41, pp. 8373–8385, July 1998.
- [Nye 88] W. Nye, D. Riley, A. Sangiovanni-Vincentelli, and A. Tits. “DELIGHT.SPICE: an optimization-based system for the design of integrated circuits”. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 7, No. 4, pp. 501–519, Apr. 1988.
- [Ocho 96] E. Ochotta, R. Rutenbar, and L. Carley. “Synthesis of high-performance analog circuits in ASTRX/OBLX”. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 15, No. 3, pp. 273–294, March 1996.
- [Otim 05] S. O. Otim, C. B., J. D., and S. Collins. “Simplifier fixed pattern noise correction for logarithmic sensors”. In: *Proc. of 2005 IEEE Workshop on Charge-coupled Devices and Advanced Image Sensors*, pp. 80–83, June 2005.
- [Park 05] J.-H. Park, M. Mase, S. Kawahito, M. Sasaki, Y. Wakamori, and Y. Ohta. “A 142dB dynamic range CMOS image sensor with multiple exposure time signals”. In: *Asian Solid-State Circuits Conference*, pp. 85–88, Hsinchu, Nov. 2005.

- [Pedr 02] M. Pedram and M. Rabaey. *Power aware design methodologies*. Kluwer Academic Publishers, Massachusetts, USA, 2002.
- [Pigu 05] C. Piguet. *Low-power CMOS circuits: technology, logic, design, and CAD tools*. CRC Press, USA, 2005.
- [Qi 04] X. Qi, X. Guo, and J. Harris. “A time-to-first spike CMOS imager”. In: *Proc. of 2004 International Symposium on Circuits and Systems*, pp. 824–827, May 2004.
- [Raba 96] J. Rabaey and M. Pedram. *Low power design methodologies*. Kluwer Academic Publishers, Massachusetts, USA, 1996.
- [Ragh 98] A. Raghunathan, N. K. Jha, and S. Dey. *High-level power analysis and optimization*. Kluwer Academic Publishers, Massachusetts, USA, 1998.
- [Rami 04] J. Ramirez-Angulo, R. Gonzalez-Carvajal, A. Lopez-Martin, and A. Torralba. “Some techniques for low-voltage continuous-time analog circuit operation”. In: *IEEE Dallas/CAS Workshop on Implementation of High Performance Circuits*, pp. 87–112, Sep. 2004.
- [Ricq 95] N. Ricquier and B. Dierickx. “Active Pixel CMOS Image Sensor with On-Chip Nonuniformity Correction”. In: *Proc. of 1995 IEEE Workshop on Charge-coupled Devices and Advanced Image Sensors*, pp. 169–173, Dana Point, CA, Apr. 1995.
- [Roy 03] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. “Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits”. *Proc. IEEE*, Vol. 91, No. 2, pp. 305–327, Feb. 2003.
- [Saff 07] F. Saffih and R. Hornsey. “Foveated dynamic range of the pyramidal CMOS image sensors”. *IEEE Trans. Electron Devices*, Vol. 54, No. 12, pp. 3422–3425, Dec. 2007.
- [Sand 02] M. A. T. Sanduleanu and E. A. J. M. Van Tuijl. *Power trade-offs and low power in analog CMOS ICs*. Kluwer Academic Publishers, Massachusetts, USA, 2002.
- [Saue 03] J. Sauerbrey, D. Schmitt-Landsiedel, and R. Thewes. “A 0.5V 1  $\mu$ W successive approximation ADC”. *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 7, pp. 1261–1265, Oct. 2003.

- [Scot 03] M. Scott, B. Boser, and K. Pister. “An ultralow-energy ADC for smart dust”. *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 7, pp. 1123–1129, July 2003.
- [Shaf 08] S. Shafie, S. Kawahito, and S. Itoh. “A dynamic range expansion technique for CMOS image sensors with dual charge storage in a pixel and multiple sampling”. *Sensors*, Vol. 8, pp. 1915–1926, March 2008.
- [Sham 06] A. Shameli and P. Heyclari. “A novel power optimization technique for ultra-Low power RFICs”. In: *International Symposium on Low Power Electronics and Design*, pp. 274–279, Oct. 2006.
- [Shen 07] Y. Shen, S. Kiatsupaibul, Z. Zabinsky, and R. Smith. “An analytically derived cooling schedule for simulated annealing”. *Journal of Global Optimization*, Vol. 38, pp. 333–365, July 2007.
- [Shi 09] C. Shi, M. K. Law, and A. Bermak. “A CMOS image sensor with reconfigurable resolution for energy harvesting applications”. In: *IEEE Sensors*, pp. 197–200, Oct. 2009.
- [Silv 02] F. Silveira and D. Flandre. “(Operational amplifier power optimization for a given total (slewing plus linear) settling time)”. In: *15th Symposium on Integrated Circuits and Systems Design*, pp. 247–253, Jan. 2002.
- [Sohn 03] S.-M. Sohn, M. G. Kim, and S. Kim. “A CMOS image sensor (CIS) with low power motion detection for security camera applications”. In: *IEEE International Conference on Consumer Electronics*, pp. 250–251, June 2003.
- [Spiv 09] A. Spivak, A. Belenky, A. Fish, and O. Yadid-Pecht. “Wide-dynamic-range CMOS image sensors - comparative performance analysis”. *IEEE Transactions on Electron Devices*, Vol. 56, No. 11, pp. 2446–2460, Nov. 2009.
- [Sunt 07] V. Suntharalingam, D. D. Rathman, G. Prigozhin, S. Kissel, and M. Bautz. “Back-illuminated three-dimensionally integrated CMOS image sensors for scientific applications”. In: *Proc. of SPIE on Focal Plane Arrays for Space Telescopes III*, pp. 901–910, Sep. 2007.
- [Taka 05] I. Takayanagi, M. Shirakawa, K. Mitani, M. Sugawara, S. Iversen, J. Moholt, J. Nakamura, and E. Fossum. “A 1.25-inch 60-frames/s 8.3-M-pixel

- digital-output CMOS image sensor”. *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 11, pp. 2305–2314, Nov. 2005.
- [Taka 07] H. Takahashi, T. Noda, T. Matsuda, T. Watanabe, M. Shinohara, T. Endo, S. Takimoto, R. Mishima, S. Nishimura, K. Sakurai, H. Yuzurihara, and S. Inoue. “A 1/2.7-in 2.96 Mpixel CMOS image sensor with double CDS architecture for full high definition”. *IEEE Journal of Solid-State Circuits*, Vol. 42, No. 12, pp. 2960–2967, Dec. 2007.
- [Tang 09] L.-F. Tanguay and M. Sawan. “An ultra-low power ISM-band integer-N frequency synthesizer dedicated to implantable medical microsystems”. *Analog Integrated Circuits and Signal Processing*, Vol. 58, pp. 205–214, 2009.
- [Tann 01] S. Tanner, S. C. Lauxtermann, M. Waeny, M. Willemin, N. Blanc, J. Grupp, R. Dinger, E. Doering, M. Ansorge, P. Seitz, and F. Pellandini. “Low-power digital image sensor for still-picture image acquisition”. In: *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications II*, pp. 358–365, SPIE, May 2001.
- [Utha 03] P. Uthaichana and E. Leelarasmee. “Low power CMOS dynamic latch comparators”. In: *Conference on Convergent Technologies for Asia-Pacific Region*, pp. 605–608, Oct. 2003.
- [Vand 04] J. Vanderhaegen and W. R. Brodersen. “Automated design of operational transconductance amplifier using reversed geometric programming”. In: *Proc. 41<sup>st</sup> IEEE /ACM Design Automation Conference (DAC)*, pp. 133–138, San Diego, USA, June 2004.
- [Venu 10] D. D. Venuto. “Ultra low-power 12 bit SAR ADC for RFID application”. In: *Proc. Design, Automation, and Test in Europe*, pp. 1071–1075, Dresden, Germany, March 2010.
- [Vrat 06] M. Vratonjic, B. R. Zeydel, and V. G. Oklobdzija. “Circuit Sizing and Supply-Voltage Selection for Low-Power Digital Circuit Design”. In: *Power and Timing Modeling, Optimization and Simulation*, pp. 148–156, Sep. 2006.
- [Wang 03] X. Wang and H. Kwok. “A low power real-time video processing imager”. In: *IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, pp. 53–56, Aug. 2003.

- [Wang 06] X. Wang, W. Wong, and R. Hornsey. “A high dynamic range CMOS image sensor with in-pixel light-to-frequency conversion”. *IEEE Transactions on Electron Devices*, Vol. 53, pp. 2988–2992, Dec. 2006.
- [Wang 95] F. Wang and R. Harjani. “Dynamic amplifiers: settling, slewing & power issues”. In: *IEEE International Symposium on Circuits and Systems*, pp. 319–322, Apr. 1995.
- [Wilk 08] G. Wilke, R. Fonseca, C. Mezzomo, and R. Reis. “A novel scheme to reduce short-circuit power in mesh-based clock architectures”. In: *symposium on Integrated circuits and system design*, pp. 117–122, Sep. 2008.
- [Wim 95] D. L. Wim Kruiskamp. “DARWIN: CMOS opamp Synthesis by Means of a Genetic Algorithm”. In: *Design Automation Conference*, pp. 433–438, Dec. 1995.
- [Wulf 05] C. Wulff and C. Ytterdal. “0.8V 1GHz dynamic comparator in digital 90nm CMOS technology”. In: *NORCHIP Conference*, pp. 237–240, Nov. 2005.
- [Xu 02] C. Xu, W. Zhang, and M. Chan. “A  $1.0V_{DD}$  CMOS active pixel image sensor with complementary pixel architecture fabricated with a  $0.25\ \mu\text{m}$  CMOS process”. In: *IEEE International Solid-State Circuits Conference*, pp. 28–385, Aug. 2002.
- [Yadi 01] O. Yadid-Pecht and A. Belenky. “Autoscaling CMOS APS with customized increase of dynamic range”. In: *Proc. IEEE International Solid State Circuits Conference*, pp. 100–101, San Francisco, 2001.
- [Yadi 03] O. Yadid-Pecht and A. Belenky. “In-pixel autoexposure CMOS APS”. *IEEE Journal of Solid State Circuits*, Vol. 38, No. 8, pp. 1425–1428, Aug. 2003.
- [Yadi 97] O. Yadid-Pecht and E. Fossum. “Wide intrascene dynamic range CMOS APS using dual sampling”. *IEEE Transactions on Electron Devices*, Vol. 44, pp. 1721–1723, Oct. 1997.
- [Yadi 99] O. Yadid-Pecht. “Wide dynamic range sensors”. *Optical Engineering*, Vol. 38, No. 10, pp. 1650–1660, Oct. 1999.

- [Yan 00] S. Yan and E. Sanchez-Sinencio. “Low voltage analog circuit design techniques: a tutorial”. *IEICE Trans. Analog Integrated Circuits and Systems*, Vol. E00-A, No. 2, pp. 179–196, Feb. 2000.
- [Yang 94] W. Yang. “A wide-dynamic-range, low-power photosensor array”. In: *Dig. IEEE International Solid-State Circuits Conference*, pp. 230–231, San Francisco, CA , USA, Feb. 1994.
- [Yang 99] D. Yang, A. Gamal, B. Fowler, and H. Tian. “A 640 x 512 CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC”. *IEEE Journal of Solid-State Circuits*, Vol. 34, pp. 1821–1834, Dec. 1999.
- [Yasu 03] T. Yasuda, T. Hamamoto, and K. Aizawa. “Adaptive-integration-time image sensor with real-time reconstruction function”. *IEEE Transactions on Electron Devices*, Vol. 50, No. 1, pp. 111–120, Jan. 2003.
- [Yuan 05] J. Yuan, N. Farhat, and J. VanderSpiegel. “GBOPCAD: a synthesis tool for high-performance gain-boosted opamp design”. *IEEE Transactions on Circuits and Systems I*, Vol. 52, No. 8, pp. 1535–1544, Aug. 2005.
- [Zebu 98] R. S. Zebulum, M. A. Pacheco, and M. Vellasco. “A multi-objective optimization methodology applied to the synthesis of low-power operational amplifiers”. In: *International Conference in Microelectronics and Packaging*, pp. 264–271, 1998.
- [Zhan 03] G. Zhang, T. Yang, S. Gregori, J. Liu, and F. Maloberti. “Ultra-low power motion-triggered image sensor for distributed wireless sensor network”. In: *IEEE Sensors*, pp. 1141–1146, Oct. 2003.
- [Zhon 01] C. Zhongjian, L. Xiaoyong, J. Lijiu, H. Jianzhong, and Y. Songlin. “An improved low power CMOS readout circuit for focal plane array ”. In: *International Conference on ASIC*, pp. 854–857, Aug. 2001.