

ABSTRACT

Title of dissertation: ADAPTATION IN STANDARD CMOS PROCESSES WITH FLOATING GATE STRUCTURES AND TECHNIQUES

Yanyi Liu Wong
Doctor of Philosophy, 2007

Dissertation directed by: Professor Pamela Abshire
Department of Electrical and Computer
Engineering

We apply adaptation into ordinary circuits and systems to achieve high performance, high quality results. Mismatch in manufactured VLSI devices has been the main limiting factor in quality for many analog and mixed-signal designs. Traditional compensation methods are generally costly. A few examples include enlarging the device size, averaging signals, and trimming with laser. By applying floating gate adaptation to standard CMOS circuits, we demonstrate here that we are able to:

- Trim CMOS comparator offset to a precision of 0.7mV.
- Reduce CMOS image sensor fixed-pattern noise power by a factor of 100.
- Achieve 5.8 effective number of bits (ENOB) in a 6-bit flash analog-to-digital converter (ADC) operating at 750MHz.

The adaptive circuits generally exhibit special features in addition to an improved performance. These special features are generally beyond the capabilities of

traditional CMOS design approaches and they open exciting opportunities in novel circuit designs. Specifically, the adaptive comparator has the ability to store an accurate arbitrary offset, the image sensor can be set up to memorize previously captured scenes like a human retina, and the ADC can be configured to adapt to the incoming analog signal distribution and perform an efficient signal conversion that minimizes distortion and maximizes output entropy.

The thesis is organized as follows: chapter 1 outlines a general overview for the scope of this research; chapter 2 provides a background for discussion, chapters 3 to 5 describe in detail the implementation of the floating gate adaptation, prototype design, and experimental result for the comparator, the ADC, and the imager, respectively; finally, chapter 6 summarizes the work.

ADAPTATION IN STANDARD CMOS PROCESSES
WITH FLOATING GATE STRUCTURES AND TECHNIQUES

by

Yanyi Liu Wong

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Pamela Abshire, Chair/Advisor
Professor Martin Peckerar
Professor Neil Goldsman
Professor Timothy Horiuchi
Professor Bongtae Han

© Copyright by
Yanyi Liu Wong
2007

Dedication

In memory of my sister

Alison

(1970–2001)

Acknowledgements

My family provided indispensable material, emotional and spiritual support for my entire life and I am eternally grateful to my parents and my grandparents.

I would like to express my gratitude towards my advisor Dr. Pamela Abshire and her husband Dr. Marc Cohen. I thank them for providing generous support, very insightful research direction, high degree of freedom, and lots of patience and tolerance. I also thank them for their resourcefulness which provided me important opportunities to work with the right people both at APL and at my future job in the industry. A PhD student could not have asked for more.

I particularly thank Dr. Kazuo Nakajima for introducing me to the field of the floating gate logics and his support at an early stage in my graduate career, where my interest in floating gates was first developed. I thank Dr. Neil Goldsman, Dr. Timothy Horiuchi, and Dr. Martin Peckerar for further developing my skills that are essential to my career. I also thank them for taking their valuable time to review my thesis. I thank Dr. Timothy Horiuchi for numerous valuable discussions, including a talk on vision and perspective for a PhD career.

I was very fortunate to work with professionals at APL during my graduate study. I thank Dr. Martin Fraeman, Dr. Mark Martin, Dr. Richard Meitzler and others for giving me an opportunity to be involved in a very special EEPROM project and all technical and non-technical experience and training I received.

I enjoyed working at the IBIS lab, where many interesting, fruitful, and fun discussions took place. I thank the very talented labmates for their company and

assistance, and I wish them best in their career: Eric Chen, Marc Dandin, Alfred Haas, Honghao Ji, Jean-Marie Lauenstein, Nicole Nelson, Somashekar Prakash, David Sander, Peng Xu and Yiming Zhai. I was lucky to meet each and every one of you, especially Honghao, who introduced me to the IBIS lab around 2001.

I thank MOSIS, NSF, JHUAPL for providing chip fabrication services and financial support.

Table of Contents

List of Figures	vii
List of Abbreviations	xii
1 Introduction	1
1.1 The Adaptive Floating Gate Comparator (AFGC)	2
1.2 The Adaptive Floating Gate Quantizer (AFGQ)	3
1.3 The Adaptive Floating Gate Imager (AFGI)	5
2 Background	7
2.1 Overview of Non-volatile Memory Technologies	7
2.2 Floating Gate Structures	9
2.3 Using FG Structures in Modern Scaled CMOS Processes	11
2.4 Hot-Electron Injection	13
2.4.1 Impact Ionization in nFET	14
2.4.2 Channel Hot Carriers in nFET and pFET	16
2.4.3 Impact Ionization in pFET	17
2.5 Fowler-Nordheim Tunneling	18
2.6 Existing Floating Gate Circuits and Techniques	22
2.6.1 Autozeroing and General Floating Gate Amplifiers	22
2.6.2 MITEs and Capacitive Division Applications	26
2.6.3 Arrays of FG Storage for Computation and Trimming	29
3 The Adaptive Floating Gate Comparator (AFGC)	32
3.1 Introduction	32
3.2 Adaptive Floating Gate Comparator	33
3.2.1 The Static Injection Method	35
3.2.2 Dynamic Injection on the Floating Gates	39
3.2.3 Trade-off Between Resolution and Speed	44
3.2.4 Monte Carlo Simulation	47
3.3 Experimental Results	52
3.3.1 Input Offset Distribution among the Chips	57
3.3.2 Dependence of Conversion Accuracy on Clock Voltage	59
3.3.3 Programming Input Offset in the $\pm 1V$ Range	59
3.3.4 The Time Course of Offset Cancellation	61
3.3.5 Robustness of Operation with Temperature and Time	63
3.3.6 Overshoot in Static Injection	64
3.4 Summary	65
4 The Adaptive Floating Gate Quantizer (AFGQ)	68
4.1 Introduction	68
4.2 Background Technologies	70
4.2.1 Floating Gate Structures	70

4.2.2	Adaptive Floating Gate Comparator	72
4.3	The AFGQ	72
4.3.1	Comparator Noise and the 6-bit AFGQ Resolution	73
4.3.2	Designing the AFGQ	74
4.4	Signal Adaptation	78
4.4.1	Non-linear AD Conversion	78
4.4.2	Histogram Equalization	80
4.4.3	Implementing Histogram Equalization with QUs	83
4.5	Measurement and Results	85
4.5.1	Performance vs. Varying Input Frequency	87
4.5.2	Performance vs. Varying Input Amplitude	90
4.5.3	Performance vs. Signal Types	93
4.5.4	Data Retention and Temperature	94
4.5.5	Performance Summary	95
4.6	Summary	96
5	The Adaptive Floating Gate Imager (AFGI)	98
5.1	Introduction	98
5.2	Background	100
5.2.1	Mismatch in the Photodiode	101
5.2.2	Mismatch in Subthreshold MOSFET	102
5.3	Adaptive Floating Gate Pixel	103
5.3.1	Circuit Overview	103
5.3.2	Floating Gate Offset Compensation	105
5.3.3	Adaptation Method	106
5.3.4	Layout	109
5.3.5	Chip Architecture	110
5.4	Experimental Results	111
5.4.1	Temporal Noise	113
5.4.2	Vdd and V_E Dependence	114
5.4.3	Power Consumption	114
5.4.4	Output Voltage Distribution	116
5.4.5	Performing Adaptation	118
5.4.6	Adaptation Performance	119
5.5	Applications and Special Effects	124
5.6	Summary	127
6	Conclusion	129
	Bibliography	132

List of Figures

1.1	A simple 5-transistor CMOS comparator	2
1.2	The block diagram for all types of floating gate comparators	3
1.3	A die photo for the adaptive floating gate quantizer	4
1.4	The 144x144-pixel image sensor captured the image of Jefferson Memorial before (a) and after (b) adaptation.	5
1.5	Afterimage effects for the imager	6
2.1	A “Magic” VLSI layout that shows several floating gate structures (a–d). The circuit with control gate is shown in (e).	10
2.2	A node V_o is capacitively coupled to multiple inputs.	11
2.3	Cross-section of nFETs showing impact ionization hot electron injection (a) and (c), channel hot electron injection (b), and pFET impact ionization hot electron injection (d).	14
2.4	The gate current for injection attains maximum around $10\mu\text{A}$. [44].	18
2.5	Injection in the pFET: current limiting configurations in (a) the autozeroing amplifier and (b) the adaptive floating gate comparator.	18
2.6	The wave function decays exponentially in a barrier of width W	19
2.7	This band diagram illustrates (a) the silicon-oxide-silicon tunneling junction (b) band-banding during Fowler-Nordheim tunneling.	21
2.8	A commonly used tunneling structure is a pFET with source, drain, and bulk connected together.	22
2.9	The AFGA uses tunneling and injection currents to establish an operating point.	23
2.10	An simplified diagram of AutoSOS.	23
2.11	Nauta’s g_m -C transconductor.	24
2.12	An example of the MITE circuit.	26
2.13	The tunable first order low pass filter using MITEs.	27

2.14	This floating gate multiplier operates in saturation region.	29
2.15	(a) The e-pot output is provided by an amplifier operating in a voltage follower configuration. The offset stored in C_f constitutes the voltage shift from the virtual ground (V_{ref}) to the V_{out} , and can be adjusted by the tunneling and injection mechanisms on the left. (b) e-pots are serially linked and addressable to external control. [42]	31
3.1	Adaptive Floating Gate Comparator (AFGC): (a) Circuit diagram of the AFGC with pFET input floating gate differential pair, crosscoupled nFET regenerative elements, and reset switch. Dimensions are specified as width/length, with unit $\lambda = 0.2\mu\text{m}$. The coupling capacitors from inputs $V_{i+,-}$ to floating gates $V_{g+,-}$ are 216fF. (b) During evaluation, bias voltage V_{clk} on the reset switch M5 determines conductance of the regenerative elements and overall comparator gain. Transconductance of the crosscoupled pair and switch is determined using HSPICE simulation of circuits extracted from layout, as described in the text.	35
3.2	Bias voltage V_{clk} controls the conductance of the switch and regenerative crosscoupled pair. Simulated current-voltage relationship for the circuit of Fig.3.1b for V_{clk} from (a)1.4V to (e)3V in 0.4V steps. . .	39
3.3	Circuit noise causes uncertainty in the outcome of the comparison: (a) a typical input offset distribution for one device obtained experimentally, showing measured voltage distribution and empirically fitted error function, and (b) corresponding empirical Gaussian probability density function.	42
3.4	Absolute value of the input offset $ \mu_d - \mu^* $ after calibration according to Eqn.(3.1): (a) as a function of ρ for several values of σ , and (b) as a function of σ for several values of ρ	43
3.5	Increasing V_{clk} increases speed, bandwidth, and noise, decreasing signal-to-noise ratio and increasing total capacity: (a) maximum clock speed, (b) SNR, and (c) channel capacity as a function of the clock voltage, determined by simulation of an extracted AFGC circuit. . . .	45
3.6	Simulation traces depicting one calibration cycle in a series of Monte Carlo simulations, depicting (a) input voltages V_{i+} , V_{i-} ; (b) floating node voltage V_{g+} ; and (c) output voltage V_{o+}	49

3.7	The use of floating gate transistors sacrifices gain at the input, but provides the ability to significantly reduce input offset. Histograms of input offset distribution determined through Monte Carlo simulations for (a) non-FG comparator and (b) AFGC before and (c) after calibration.	50
3.8	Photomicrograph of a single AFGC. The floating gate poly2 is sandwiched between the metal-poly enclosure labeled by (a) and (d), which are the input voltages $V_{i+,-}$, respectively. The floating gates are connected to the tunneling electrode and to the input transistors by metal1 labeled (b) and (c). The tunneling structure is illustrated in Fig.2.8. M1 and M2 are indicated by (i) and (j) respectively; M3 and M4 are indicated by (e), (f) indicates M5, (g) indicates the two output nodes $V_{o+,-}$, (h) shows the tail current mirror and (k) shows the tunneling electrode.	53
3.9	Circuit configuration used for testing the voltage comparator. The comparator output drives an output buffer shown in (a) which generates rail-to-rail output signals V_{out+} and V_{out-} . (b) A cascade of geometrically scaled inverters delivers the signal offchip with minimal delay, and the externally filtered output voltage V_A is interpreted as the probability that the output is logic high.	55
3.10	Conversion accuracy depends on the clock voltage V_{clk} . Accuracy increases and input-referred noise decreases with higher circuit gain as V_{clk} is reduced.	60
3.11	Residual input offset voltages remain small over a wide range of programmed offset voltages.	61
3.12	Input offset decreases exponentially in time. Offset is initially 200mV and decreases in time, with decay time constant decreasing with increasing V_{CM} . The decreasing time constant shows a speeding up of the adaptation near the end. This is because the steady state (50% HI and LOW) is reached well before the injection step becomes infinitesimal.	62
3.13	Measured and fitted injection time constant τ_j v.s. V_{CM}	63
3.14	Input offset for the static injection method with clock voltage held constantly high. Offset first decreases, then “overshoots” the desired point and settles at a nonzero offset voltage.	66
4.1	Each storage capacitor stores the reference voltage $V_{c(n)}$ for the n^{th} comparator.	73

4.2	The AFGQ is a flash ADC with offsets programmed into each quantizing unit (QU).	75
4.3	(a) T/H including equivalent transmission line, pad parasitics and 3.8 pF capacitance load. Transistor widths are labeled near each transistor. 0.4 μm gate lengths are used for all transistors. (b) The 3-stage pipelined comparator occupies 130 μm x 16 μm chip area. (c) The pipelined comparator produces results in 1.5 clock cycles.	76
4.4	The QU includes offset storage and feedback mechanisms for adjusting storage.	77
4.5	A non-linear transfer function T can equalize a non-uniformly distributed input cdf F_1	79
4.6	The non-linear conversion equalizes the output codeword histogram (c) from an arbitrary analog input distribution with nonuniform pdf (a) and corresponding nonlinear cdf (b).	80
4.7	3-bit flash AFGQ: (a) The equal-partition algorithm for histogram equalization is implemented with digital AND gates. (b) Partitions and reference levels for hierarchies 1-3.	82
4.8	The reference levels t_n 's in a 4-bit flash ADC adapt as the input distribution changes from uniform, to Gaussian, to exponential.	84
4.9	Photomicrograph showing both the AFGQ (left) and the on-chip large clock buffer (right).	86
4.10	Performance results for AFGQ: (a), (b) FFT plot computed from 16k digital output samples and (c) SNDR vs. input sine frequency. The code histograms for the results with linearly programmed, arc-sine programmed and autonomously adapted reference level are used to compute non-linearity in (d), (e) and (f), respectively.	89
4.11	(a) The ADC achieves the best performance when the signal has 1V V_{pp} . (b) Conversion with reference levels fit to signal (LPS) outperforms fixed reference levels (LPF). (c) The MSE for conversions with adaptive reference levels scale with signal power for $V_{pp} > 0.3\text{V}$. The dashed line is a quadratic curve fit to GPS results.	91
4.12	Although output cdf deviates slightly from an ideal uniform cdf with residue norm of 0.13, the output cdf for three very different inputs are nearly identical.	93
4.13	The AFGQ maintained an error rate of 2.93×10^{-9} /sample during one month of continuous operation.	95

5.1	AFGP circuit: (a) pixel circuit; (b) injection circuit. Channel current I_2 balances with current source I_3 during injection.	103
5.2	Metal-3 windows expose the photodiodes and part of the floating gate in the square and rectangular windows, respectively.	109
5.3	(a) Pixels are activated by row and column shift registers. (b) Columns lines are organized into 18 groups for rapid readout.	111
5.4	The pixel voltage increases with 1) increasing V_{dd} and 2) decreasing V_E	116
5.5	(a) The pixel voltages tends to be Gaussian distributed. (b) UV-erased curves (C and D) are slightly taller and to the right.	117
5.6	The original bell-shaped pixel voltage distribution is shoved from the left forming a new bell-shaped heap during calibration. In a magnified view (b), the 6σ for the new peak is observed to be roughly 8mV, compared to 110mV for the original distribution.	118
5.7	Pixel voltage distribution changes over time during adaptation.	120
5.8	Pixel voltage statistics change over time during adaptation.	121
5.9	(a) FPN σ reduces significantly after adaptation. (b) Pixel voltage follows the logarithm of intensity linearly.	122
5.10	Images of a slide of the Jefferson Memorial taken (a) before and (b) after calibration.	124
5.11	At $V_{dd}=4.3V$, injection causes pixel voltage and FPN σ increase.	125
5.12	Vignetting correction (a)-(c) and afterimages (d)-(h)	128

List of Abbreviations

AFGA	Autozeroing Floating Gate Amplifier
AS	Adapted to the input Signal
AWG	Arbitrary Waveform Generator
cdf	Cumulative Distribution Function
CDS	correlated double sampling
CMOS	Complementary Metal-Oxide-Semiconductor
DCT	Discrete Cosine Transform
DDS	Double Delta Sampling
DNL	Differential Non-Linearity
DSP	Digital Signal Processing
EEPROM	Electrically Erasable Programmable Read-Only-Memory
ENOB	Effective Number Of Bits
ESD	ElectroStatic Discharge
FFT	Fast Fourier Transform
FG	Floating Gate
FPAA	Field-Programmable Analog Array
FPN	Fixed Pattern Noise
GND	electrical GrouND
GPS	Gaussian Programmed to match Signal
INL	Integrated Non-Linearity
LPF	Linearly Programmed to Full-scale
LPS	Linearly Programmed to match Signal
LSB	Least Significant Bit
MITE	Multiple-Input Translinear Element
MOSCAP	Metal-Oxide-Semiconductor CAPacitor
MOSFET	Metal-Oxide-Semiconductor Field Effect Transistor
MSB	Most Significant Bit
MSE	Mean-Squared Error
NEB	Noise Equivalent Bandwidth
nFET	p-type Field Effect Transistor
OP-Amp	OPerational Amplifier
PCB	PC Board
pFET	p-type Field Effect Transistor
pdf	Probability Density Function
pmf	Probability Mass Function
QU	Quantization Unit
RMS	Root-Mean-Squared
SIPO	Serial-In Parallel-Out
SNDR	Signal to Noise-plus-Distortion Ratio
SNR	Signal to Noise Ratio
THD	Total Harmonic Distortion
UV	Ultra-Violet
VLSI	Very Large Scale Integrated circuit

Chapter 1

Introduction

Using floating gate structures for non-volatile storage began in the early years of metal-oxide-semiconductor (MOS) systems [1]. With the advance of fabrication technology, special processes were developed explicitly for non-volatile storage to achieve higher programming efficiency, higher data rate, better reliability and higher storage density [2]. On the other hand, standard CMOS processes have been developed to achieve different goals such as low cost, low power and high speed. However, the non-volatile storage feature is intrinsic to all CMOS processes including cutting-edge CMOS processes specifically tailored for logic applications [3].

We take advantage of the non-volatile features found in standard CMOS processes and use them as techniques to combat a long standing device problem—offsets caused by intrinsic device mismatch that limit performance in precision circuits. In the following sections, we introduce the techniques and results achieved by using floating gate adaptation in different CMOS circuits: a five-transistor comparator, a full-fledged flash analog-to-digital converter, and a wide dynamic range 144x144 pixels image sensor.

The inverting nature of a common source transistor is utilized in the pFET hot-electron-injection, so that the output signal of the comparator forms a stable negative feedback to the adaptation (Fig.1.2), enabling automatic and accurate adaptation results. The in-circuit, on-line learning feature is very attractive to re-programmable mixed-signal circuits, and we have indeed applied this feature in a fully-functional 6-bit flash analog-to-digital converter (see next section). We successfully showed that we are able to apply adaptation to an ordinary comparator, and turn it into an AFGC. The comparator of choice happens to be a very advanced, high-speed specimen [7] found in IEEE JSSC and the result is a powerful comparator that operates at 1.2GHz with an offset of 199uV, or equivalently, 13 ENOB [8].

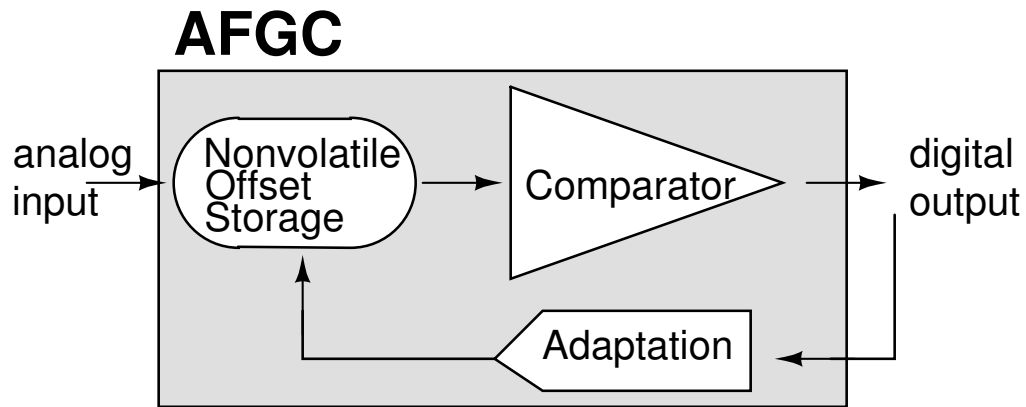


Figure 1.2: The block diagram for all types of floating gate comparators

1.2 The Adaptive Floating Gate Quantizer (AFGQ)

We built a 750MS/s 6-bit flash ADC [9] (Fig.1.3) with 63 high speed AFGCs. The on-line learning feature of the AFGC enables manual and automatic in-circuit

programming of reference levels, completely eliminating both resistor ladders and comparator offset problems. We are able to obtain strictly monotonic output, with integrated non-linearity (INL) and differential non-linearity (DNL) of 0.24 LSB and 0.79 LSB, respectively. Standard FFT based single tone analysis gives 5.7 ENOB and 5.3 ENOB at input frequencies of 200MHz and 387MHz, respectively.

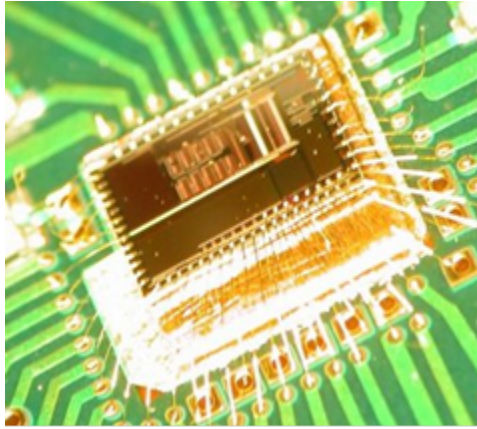


Figure 1.3: A die photo for the adaptive floating gate quantizer

When adaptation is turned on during operation, the ADC learns the input signal distribution and adjusts comparator reference levels such that the ADC converts frequent signal regions with finer detail and greater sensitivity, resulting in an overall lower distortion and higher output entropy. Since we have the ability to accurately trim the reference levels, we have the option of converting the signal non-linearly. By extending adaptation from a DC input to time-varying input signal, the adaptive comparators in the flash ADC directly implements histogram equalization.

1.3 The Adaptive Floating Gate Imager (AFGI)

We applied floating gate pFET injection in the pixels of two wide dynamic range image sensors [10, 11]. The imager operates in the MOSFET subthreshold region, and converts incident light intensity logarithmically, with 100mV per decade intensity. Thus, the imager is extremely sensitive to the mismatch levels (10mV) present in CMOS. Each pixel performs adaptation independently, and the adaptation proceeds in parallel, leading to simple and fast operation. The user simply illuminates the imager uniformly and turns on adaptation for a few seconds and the fixed-pattern noise power will be reduced by one hundred times (Fig.1.4). The adaptation also compensates for distortions in the light path such as the vignetting effects commonly found in optical lenses.

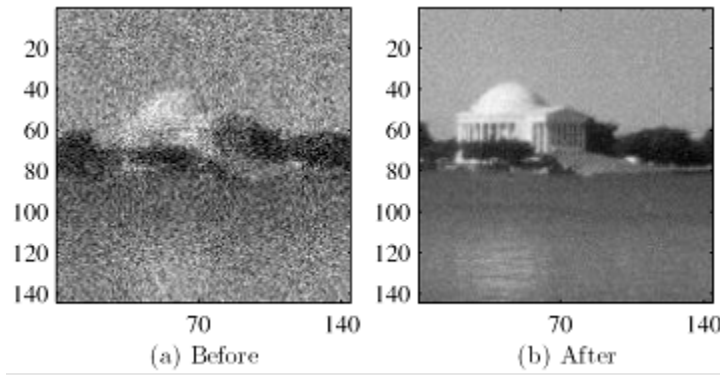


Figure 1.4: The 144x144-pixel image sensor captured the image of Jefferson Memorial before (a) and after (b) adaptation.

If the user directs the imager to a particular pattern during adaptation, the pattern will be imprinted into the floating gate memory and emerges in captured images in negative (Fig.1.5).

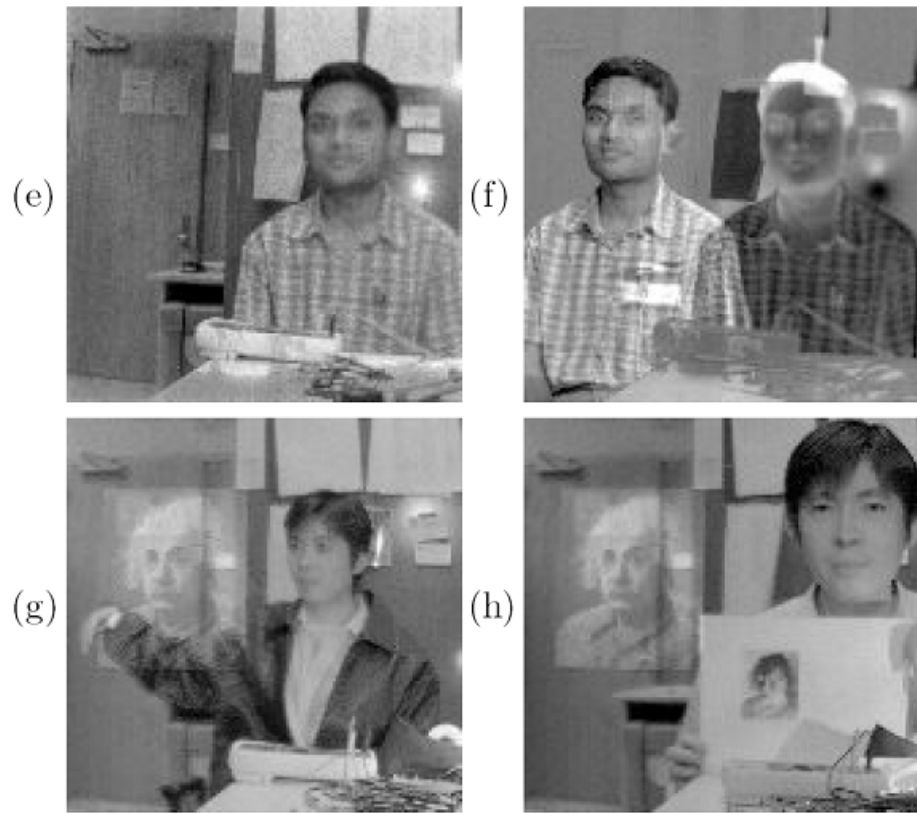


Figure 1.5: Afterimage effects for the imager

Chapter 2

Background

2.1 Overview of Non-volatile Memory Technologies

The memory devices that are expected to retain information in the long term are referred to as “non-volatile memories”. Such devices can usually retain information for more than 10 years without the use of any power source. In addition to floating gate structures [1], alternative non-volatile memory devices exist.

In metal-nitride-oxide-semiconductor (MNOS) [12], silicon-nitride-oxide-silicon (SNOS) [13] and silicon-oxide-nitride-oxide-silicon (SONOS) [14, 15]), electrons and holes are stored in localized traps in the nitride layer. These devices are intrinsically radiation tolerant since the mobilities of electron and hole are similar in nitrides; as electron-hole pairs are generated due to ionizing radiation, they are swept out of the insulator, leaving negligible trapped charge [2]. Both hot-electron and hot-hole injection mechanisms are used to alter trapped charge. Since the charge is localized, a single defect will reduce a portion of the charge, as opposed to the loss of entire charge in the case of a floating gate device.

Ferroelectric random-access memory (FeRAM) uses a layer of ferroelectric material, typically lead zirconate titanate (PZT) [16], as the dielectric layer in a storage capacitor. An applied electric field alters the polarization in the PZT crystal, and the information is stored. During a read operation, an electric field is applied to the

PZT, and a sense amplifier detects the presence of a current pulse. The absence of the current indicates that the polarization is aligned with the field; otherwise, the presence of a current pulse indicates that the polarization was altered, and a subsequent write operation is required to restore the lost information. This destructive read operation is similar to a dynamic random-access memory (DRAM), but the refresh operation is performed only after a read operation as opposed to periodical refresh operations in DRAMs. Difficulties in production include compatibility issues such as the compromised ferroelectric properties during high temperature annealing or deposition [17]. An example of the state of the art is a product reported by Toshiba[®] with high storage densities [18].

Magnetoresistive random access memory (MRAM) uses a magnetized tunnel junction (MTJ) to store information [16]. The MTJ changes resistance for different programmed states. Separate programming row and column metal lines generate required magnetic fields for programming MTJs. When scaled to finer feature sizes, the programming currents will need to increase to keep magnetic flux relatively constant. Since the current scales in the opposite direction, it seems that the device does not scale to smaller sizes very well. However, in 2005 Freescale[®] reported an 8kb array integrated with a 90nm logic CMOS process [19].

Phase-change memory (PCM or PRAM) uses chalcogenide glass as the storage medium [16]. The crystalline and amorphous states of chalcogenide glass have different resistivity. Depending on the heating/cooling cycles, states of the chalcogenide glass and therefore its resistance can be altered. Samsung[®] recently reported a 256Mb PCM device [20]. The programming current is on the order of 1mA. They

report an endurance cycle (a set of full chip programming and erasure procedure) of 10^7 and an access time of 60ns. They mentioned that they are looking for a better material with lower current consumption for phase change.

Although these technologies offer promising features, they are generally not available in standard CMOS processes. Floating gate structures on the other hand, offer direct integration in standard CMOS, with inherent programming mechanisms. These advantages enable the floating gate design techniques that is introduced in this work.

2.2 Floating Gate Structures

A floating gate MOSFET uses an electrically isolated material such as polysilicon to store charge indefinitely. There are no direct electrical connections to this circuit node, so charge on this gate remains trapped for a very long time. Thus floating gate structures provide a nonvolatile storage mechanism, and is widely used to store data in EEPROM [1], to trim current sources [21–23], to autozero amplifiers [24, 25], to store/cancel offset in comparators [8] and ADCs [26], to correct non-uniformity in imagers [27–33], and to store large array of analog parameters [34]. They have also been used in neuromorphic applications [35–37]. Figure 2.1 demonstrates several floating gate layouts. The poly, poly-2, metal-1 and diffusion are shown in red, yellow, blue and brown, respectively. The crosses represent metal contact regions. The floating gate consists of poly in (a), (b) and (d). In (c), the floating gate extends to metal-1 and poly-2 via contacts. A control gate that capac-

itively couples to the floating gate is often useful, and the capacitor can be either poly2-poly1 capacitors (b,c) or a MOSCAP (d).

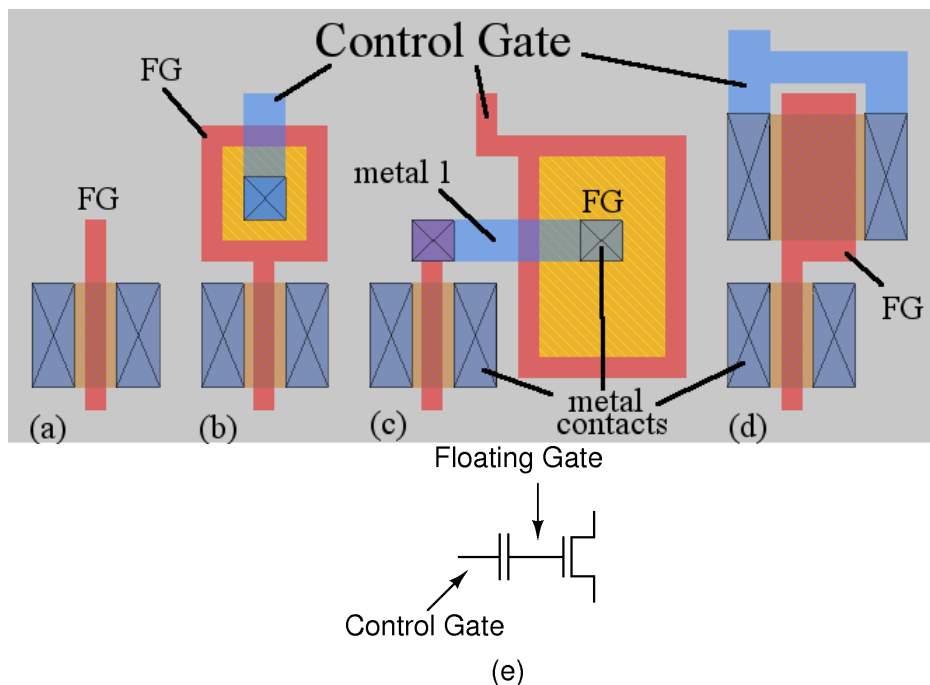


Figure 2.1: A “Magic” VLSI layout that shows several floating gate structures (a–d). The circuit with control gate is shown in (e).

Floating gate structures can be modeled by a capacitive divider (Fig.2.2). Node V_o is capacitively coupled to multiple inputs. Suppose the initial charge on the gate is V_{init} (i.e., $V_o = V_{init}$ when $V_1 = V_2 = V_3 = \dots = V_N = 0$), then by charge redistribution,

$$V_o = V_{init} + \frac{V_1 C_1 + V_2 C_2 + V_3 C_3 + \dots + V_N C_N}{C_{TOT}} \quad (2.1)$$

where $C_{TOT} = C_1 + C_2 + C_3 + \dots + C_N$. Capacitors C_1, C_2, \dots, C_N represent control gate capacitance and parasitic capacitance to the substrate, drain and interconnects.

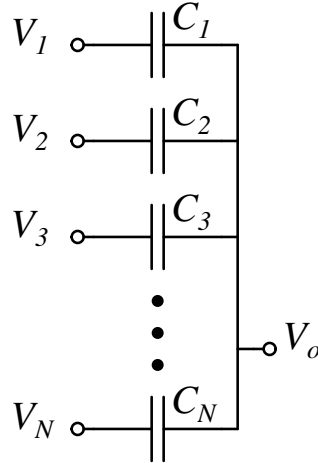


Figure 2.2: A node V_o is capacitively coupled to multiple inputs.

The term “coupling ratio” generally refers to the ratio of control gate capacitance to total capacitance. In the case of multiple control gates, each control gate is associated with its coupling ratio. Designers often use a higher coupling ratio for better control of the floating gate voltage.

2.3 Using FG Structures in Modern Scaled CMOS Processes

Moore’s law predicted doubling the transistor density every fixed amount of period [38], and prior advancements in device fabrication had followed the simple scaling method [39] outlined as follows. For a scale factor α , there are several items that are either divided by α , multiplied by α , or constant:

- $\div\alpha$: Gate length, width, oxide thickness, supply voltage, threshold voltage and capacitance.
- $\times\alpha$: Doping density.

- constant: Electric field (approximately) and power density.

Recently, the scaling trend that followed this 1974 method has seen some obstacles. This is mainly due to two factors:

1. Limit to the level of dopant concentration: subthreshold channel currents prevent further scaling of the threshold voltage and consequently supply voltage, thus to accommodate a high supply voltage the doping density is limited to prevent band-to-band tunneling of the drain-to-substrate junction.
2. Limit to the gate oxide thickness: 1. Direct tunneling across gate oxide occurs when the oxide is too thin. 2. More serious hot-electron injection due to the higher doping densities.

Both of the limitations impose serious challenges for circuit designers attempting to implement floating gate structures in scaled logic CMOS processes:

1. Difficulties in generating high voltages required for erasing and programming floating gates.
2. Charge leakage in floating gates.

Fortunately, many advanced logic processes offer lightly-doped active regions and thick oxide options, mainly for building high voltage MOSFETs for interfacing off-chip I/O signals. These options provide welcoming environments for floating gate structures. Reliability studies [3] have shown 10 years data retention in the floating gate memories implemented in mainstream $0.35\mu\text{m}$, $0.25\mu\text{m}$ and $0.18\mu\text{m}$ logic CMOS processes. For an older $0.5\mu\text{m}$ N-well CMOS process, “the retention

loss is less than $1\mu\text{V}$ at 27°C over 10 years, and less than 1mV at 90°C over 10 years” [40].

2.4 Hot-Electron Injection

Electrons in silicon move like free particles. They possess an average thermal energy of $E_K = 3kT/2$ and average thermal velocity $v_{th} = \sqrt{2E_K/m^*}$, where k is the Boltzmann constant, T the absolute temperature and m^* the effective mass. E_K is also its kinetic energy, and $E_K = E - E_C$ where E_C is the conduction band edge energy and E the total energy. E_K is about 0.04eV at room temperature 300K [41].

When an electric field is applied, electrons gain momentum from the field and experience scattering, resulting from collisions with lattice imperfections, impurities, dopant ions and phonons. In steady state the momentum gained is transferred to the lattice during collision. Thus the momentum gained between collisions contributes to the drift velocity v_d as $q\mathcal{E}\tau_n = m^*v_d$ where q is the electron charge, \mathcal{E} the applied field and τ_n the mean free time between collisions. Expressing drift velocity $v_d = \mu_n\mathcal{E}$ leads to mobility:

$$\mu_n = q\tau_n/m^*.$$

When the electric field is relatively small, the drift velocity v_d is much less than the thermal velocity v_{th} , v_d is proportional to the electric field, and mobility μ_n is roughly constant. However, when v_d is on the order of v_{th} , the excess kinetic energy gained between collisions is effectively absorbed by optical phonons, limiting the speed of electrons as well as v_d . This velocity saturation can be seen as an

effective reduction in the mean free time τ_n and the mobility μ_n . A small fraction of electrons travel longer without scattering than most others. They are accelerated by the field to a high velocity, effectively becoming “hot”.

2.4.1 Impact Ionization in nFET

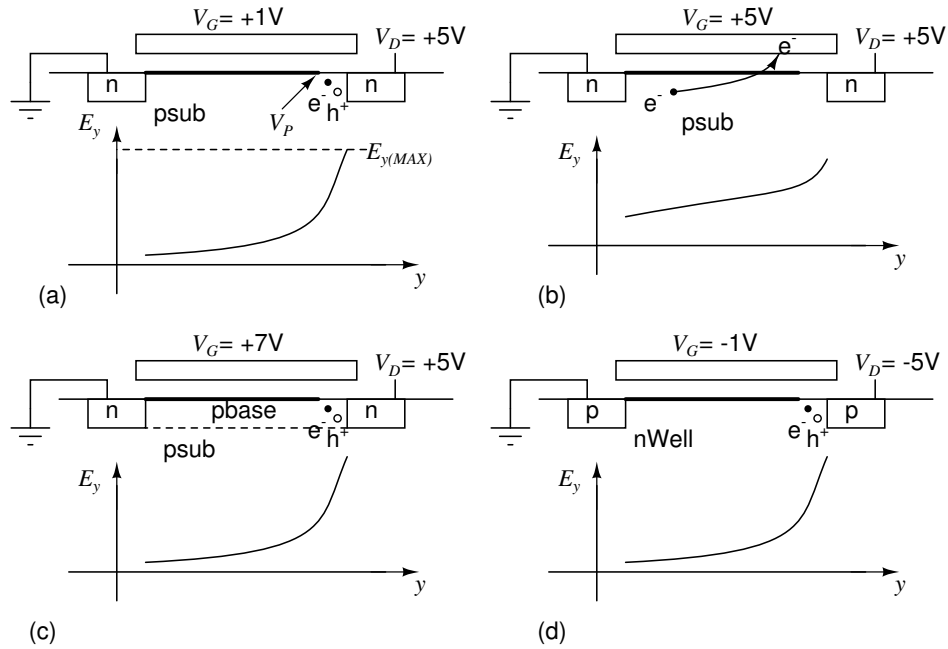


Figure 2.3: Cross-section of nFETs showing impact ionization hot electron injection (a) and (c), channel hot electron injection (b), and pFET impact ionization hot electron injection (d).

When hot electrons acquire more than 1.5eV of energy they can cause impact ionization when they collide with the lattice and produce electron-hole pairs [41]. Impact Ionization generally occurs in the space charge where the electric field is very high. Fig.2.3 (a) shows the cross section of an ordinary nFET biased in saturation, where impact ionization occurs in the high field in the pinch-off space charge. The electric field \mathcal{E}_y is shown below the cross section diagram. The maximum electric

field $\mathcal{E}_{y(MAX)}$ occurs near the drain, and $\mathcal{E}_{y(MAX)} = (V_D - V_{Dsat})/l$, where V_{Dsat} is the drain voltage at which the electrons reach saturation velocity, and $V_{Dsat} < V_P$, the pinch-off voltage that roughly equals to $V_G - V_T$, where V_T is the threshold voltage. $l = \sqrt{x_{ox}x_j\epsilon_s/\epsilon_{ox}}$, where x_{ox} , x_j , ϵ_s and ϵ_{ox} are the oxide thickness, junction depth, permittivities of silicon and oxide, respectively [41]. The impact ionization coefficient α determines the rate of ionization, and is a strong function of the electric field with a constant B : $\alpha \propto \exp(-B/\mathcal{E})$. The generated holes neutralize in the substrate and cause the substrate current. A small portion of the hot holes are attracted by the vertical field and be injected into the oxide. The majority of generated electrons are swept by the lateral field to the drain. A small portion of the generated electrons with energies greater than the oxide barrier of 3.1eV may have the opportunity to travel to the conduction band of the oxide. Most of these will be repelled by the vertical electric field and fall back to the channel. Only those electrons possessing the greatest energy and traveling in the right direction will reach the gate. Thus, gate currents due to impact ionization hot electron injection in nFET is very small. The holes injected to the oxide is essentially immobile, and is generally not considered a viable mechanism for charge manipulation on the floating gate.

Designers in [42] modified the nFET structure by adding higher p-doping in the channel (Fig.2.3 c) to raise the nFET threshold above 6V. The higher p-doping is a layer (pbase) used to make the base of an NPN transistor, and is not available in standard CMOS processes. By raising the gate voltage above the drain voltage, the vertical field now favors attraction of the electrons injected into the oxide conduction

band, and the gate current is much increased.

nFET hot-electron injection is a popular method for programming [2]. Due to the aforementioned low efficiency conditions in an ordinary nFET, special process techniques such as split-gate or source-side-injection have been used [43].

2.4.2 Channel Hot Carriers in nFET and pFET

When the gate voltage is increased in the nFET (Fig.2.3b), V_P will also increase, leading to high electric field in the inversion channel. A small portion of the electrons in the channel do not experience much scattering, and thus possess more energy. Eventually these “lucky” electrons will scatter, and a fraction of those lucky electrons will be bouncing towards the oxide, and enter the oxide conduction band. This does not require impact ionization, and is termed “Channel Hot Electron Injection” (CHEI). The bias condition in Fig.2.3 (b) shows a high gate voltage V_G as well as a high drain voltage V_D . Since the lateral electric field in the inversion channel increases with reducing gate length, a minimum-length nFET gives higher injection. The minimum-length transistor in this bias condition usually consumes exceedingly high current and power.

In a pFET, hot hole can be injected to the oxide, but since mobility of holes in the oxide is very low, this is generally not a useful effect.

2.4.3 Impact Ionization in pFET

Figure 2.3 (d) shows a pFET cross section with impact ionization. In a pFET, the collision of hot holes with the lattice creates hot electron and hole pairs. The ionization coefficient is less than that of nFET, but the vertical electric field is in favor of the gate capturing the electrons. Thus, the result can be a small channel current and higher gate current, which increases the injection efficiency. We have observed pFET injection occurring at 4.8V in a 5V process (see Chap. 5).

The injecting current is mainly exponential to the source-to-drain voltage. An accurate empirical model in [44] suggests that injection produces a current I_{inj} from the floating gate into the channel

$$I_{inj} = \alpha I_s \exp \left[-\frac{\beta}{(V_{gd} + \delta)^2} + \lambda(V_{gd} - V_{gs}) \right] \quad (2.2)$$

where I_s is the source current, V_{gd} and V_{gs} are gate-to-drain and gate-to-source voltages, and α , β , λ , δ are fitting parameters.

Figure 2.4 shows the gate current vs. channel current in a pFET [44]. For all three different source-to-drain bias conditions, the gate current for injection attains maximum around $10\mu A$, when the transistor is in saturation region slightly above threshold. Therefore, it is desirable to bias the pFET injection transistor at the operating current suitable for injection. Figure 2.5 shows the current limiting configurations in the widely-cited autozeroing floating gate amplifier (AFGA) [24, 45] and our adaptive floating gate comparator (AFGC) [6].

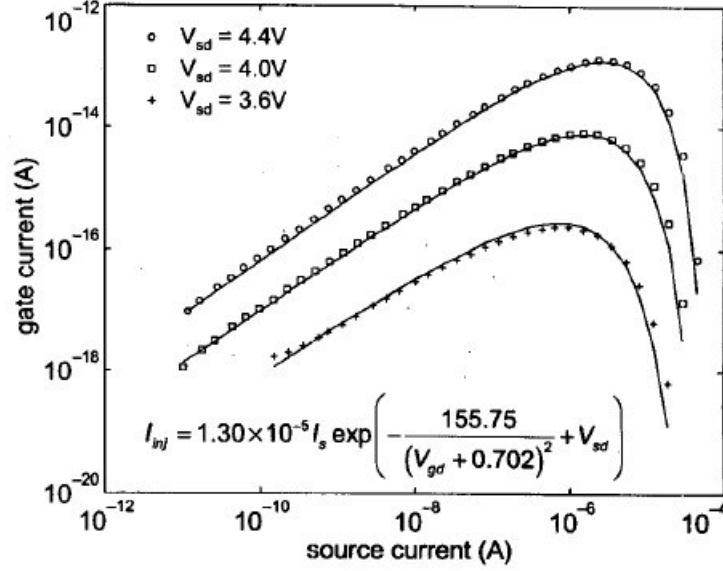


Figure 2.4: The gate current for injection attains maximum around $10\mu\text{A}$. [44].

2.5 Fowler-Nordheim Tunneling

Modern physics suggests that waves possess particle properties as seen in photo-emission effect; and that particles possess wave properties as seen in electron single- and double-slit diffraction patterns. In the latter case, the light/dark “intensity” of the diffraction patterns is in fact the “density” of the electrons. For

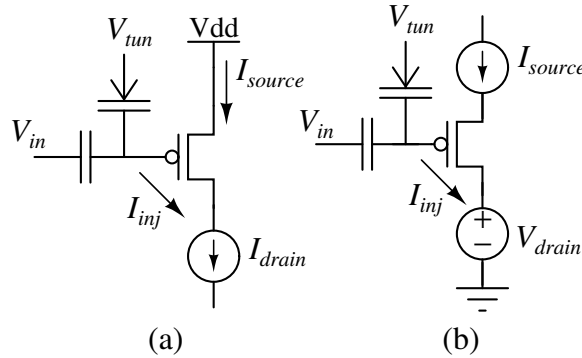


Figure 2.5: Injection in the pFET: current limiting configurations in (a) the autozeroing amplifier and (b) the adaptive floating gate comparator.

lighter area the electron density is low, and for darker area the electron density is high.

The wave function $\Psi(x, t)$ represents an *unmeasurable* probability amplitude for a particle; the complex square $\Psi^*\Psi$ represents a *measurable* probability density.

By using the total energy as a Hamilton control function $E = E_K + U$, where $E_K = mv^2/2 = p^2/2m$ is the kinetic energy and U the potential energy and recognizing $\omega/v = k = p/h$, the wave equation can be re-written as the Schrödinger equation

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{2m}{\hbar^2}(E - U)\Psi = 0$$

where h is the Planck constant, m the mass, $k = 2\pi/\lambda$ the wave number, $\omega = 2\pi f$ the angular frequency, p the linear momentum.

Applying boundary conditions we can solve for the wave function Ψ and the probability density $\Psi^*\Psi$. For $E > U$ the solution is an oscillation (region (a) and (c) in Fig.2.6). In a barrier where $E < U$, Ψ decays exponentially as shown in Fig.2.6(b).

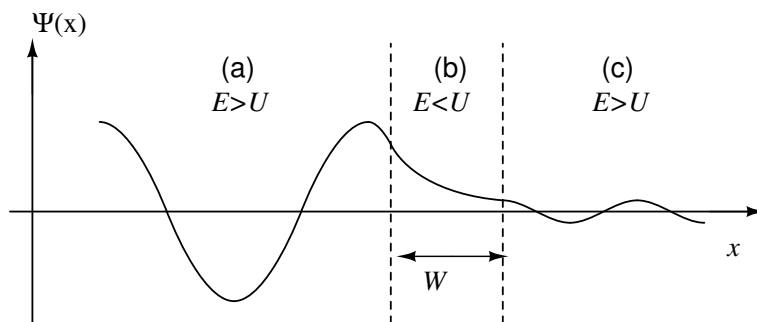


Figure 2.6: The wave function decays exponentially in a barrier of width W .

Thus, if the barrier width W is small enough, we would see significant probabil-

ities that the particle appears on the other side of the barrier. This is the tunneling effect. In silicon and oxide systems, electrons are forbidden in the bandgap of the silicon and the oxide (no allowed state), and only allowed in the conduction band (above energy E_C) and the valence band (below energy E_V). For a tunneling junction made with silicon-oxide-silicon (Fig.2.7 a), the E_C for the oxide is about 3.1eV above the E_C of the silicon, and the oxide forms a barrier. If the width of the barrier W is small enough (a few nanometers), the electrons in the silicon conduction band will have a chance to tunnel through the oxide to the silicon conduction band on the other side. This is generally termed “direct tunneling”. If the width W is big, but a large electric field is applied (Fig.2.7 b) such that the effective width W_T is small enough (a few nanometers), the electrons in the silicon conduction band will have a chance to tunnel through the oxide bandgap to the oxide conduction band and be swept to the silicon conduction band on the other side by the large electric field. This field-assisted tunneling is called “Fowler-Nordheim tunneling” or “field emission”.

Fowler-Nordheim Tunneling has been used extensively in EEPROM applications [2]. Compared to hot-electron injection, this field-assisted electron transport does not use a MOSFET that takes a large amount of channel current. On the other hand, tunneling requires significantly higher voltages compared to injection. Here is a first-order tunneling model [2]:

$$J = \alpha E^2 e^{-E_C/E}$$

where E is the field across the oxide, α the tunneling factor, and E_C the critical

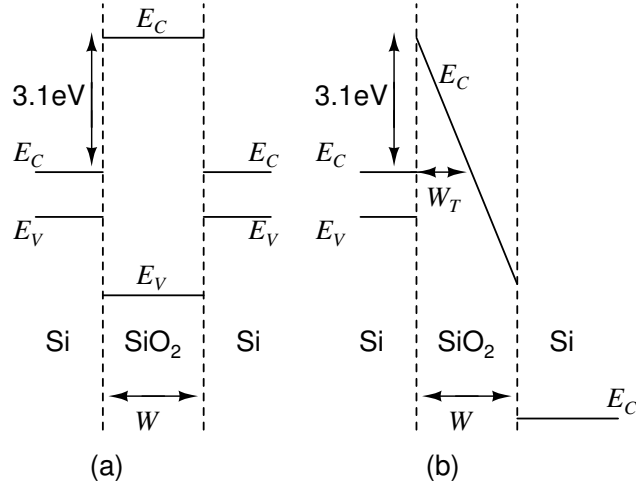


Figure 2.7: This band diagram illustrates (a) the silicon-oxide-silicon tunneling junction (b) band-banding during Fowler-Nordheim tunneling.

field, which is typically 10MV/cm for oxide.

Figure 2.8 shows a tunneling electrode commonly used in this work. The structure is a pFET with source, drain, and bulk connected together. This tunneling voltage V_{TUN} is biased at a very high voltage relative to the gate voltage V_{FG} , and strong inversion is formed under the gate. The high electric field in the thin gate oxide enables Fowler-Nordheim tunneling. The p-n junction of p-substrate to nWell is often lightly doped, and has a high breakdown voltage. In a $.5 \mu\text{m}$ CMOS process available from MOSIS, this breakdown voltage is roughly 17V. Substantial tunneling effect was observed at $V_{TUN} > 13 \text{ V}$ when the V_{FG} is biased near 0 V.

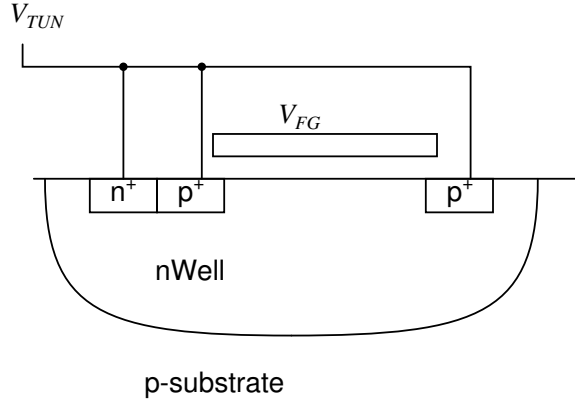


Figure 2.8: A commonly used tunneling structure is a pFET with source, drain, and bulk connected together.

2.6 Existing Floating Gate Circuits and Techniques

2.6.1 Autozeroing and General Floating Gate Amplifiers

An autozeroing floating gate amplifier (AFGA) uses continuous tunneling and injection currents to establish an operating point suitable for a 1-stage inverting amplifier [24]. Because these currents can be extremely small, the time it takes for the circuit to return to the steady state can be extremely long. This feature can be exploited to build extremely low cut-off frequency filter that is not easily done with ordinary RC techniques.

Figure 2.9 shows the schematic of the AFGA. The circuit operates in sub-threshold. The operating current I_d is provided by an nFET. The AFGA attains steady state when $I_{tun} = I_{inj}$. Due to the capacitive feedback provided by C_1 and C_2 , the floating gate is at “virtual ground” that is similar to the inverting input of an ordinary inverting Op-Amp. The AFGA is a band-pass filter with a mid-band gain set by the ratio of the capacitances ($-C_2/C_1$). The DC-blocking capacitor C_1

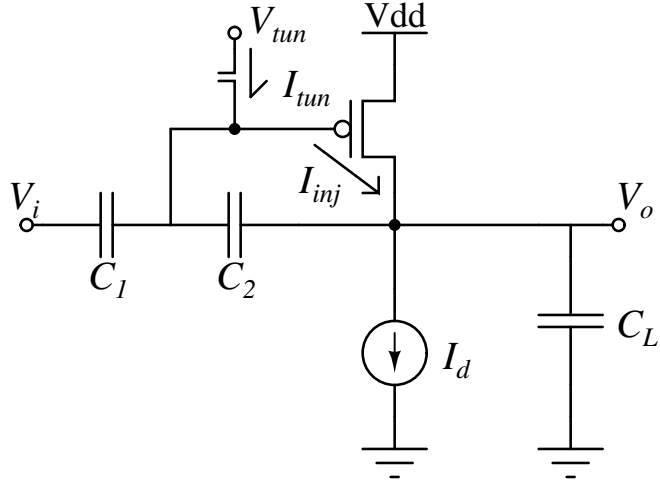


Figure 2.9: The AFGA uses tunneling and injection currents to establish an operating point.

and the oxide currents set a cut-off frequency at the lower end, while the channel current I_d and loading capacitors C_L and C_2 set the cut-off frequency at the upper end.

Capacitive feedback amplifiers such as the AFGA can be purely first-order [46]. If we put three AFGAs in a ring, connected back-to-back, we have a second-order section (SOS), and it is called the AutoSOS [46]. Figure 2.10 gives a simple diagram

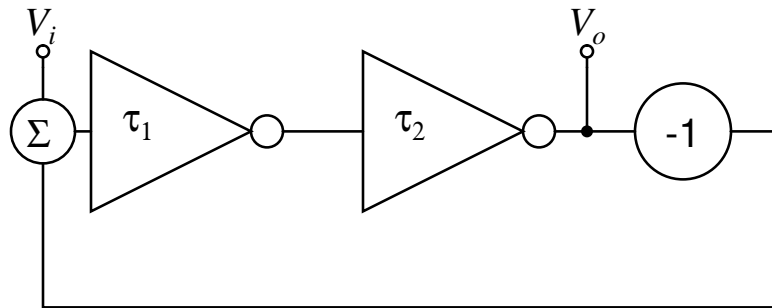


Figure 2.10: An simplified diagram of AutoSOS.

that illustrates the idea. The cut-off frequency of the first two AFGAs are tuned

to very low values f_1 and f_2 by controlling current I_d in Fig.2.9. The third AFGA has a much higher I_d and thus has a higher cut-off frequency comparing to the first two, and it can be seen like a simple -1 multiplier (by matching $C_1 = C_2$). The AutoSOS made use of the fact that the AFGA is a pure 1st order section. It is demonstrated that by adjusting f_1 and f_2 we can tune the circuit to have different quality factor Q and high frequency corner time constant. The sum in front of the first AFGA is simply a capacitor coupled to the floating gate. Offset has been a significant problem due to feedback in an SOS, and compensated with switched-cap method traditionally. However, since AFGA removes the offset, the problem no longer exists.

Floating gate techniques can be also used to tune existing amplifiers or filters, for example, the Nauta's g_m -C filter [47] shown in Fig.2.11. As shown, the filters

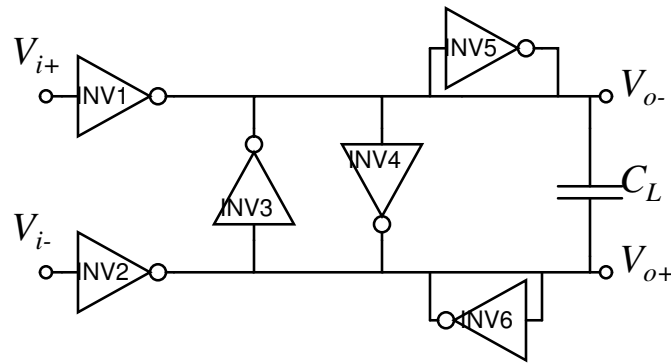


Figure 2.11: Nauta's g_m -C transconductor.

consists of only CMOS inverters. INV1 and INV2 are the main inverters. INV3 and INV4 are connected in a positive feedback method, enhancing the gain. The head-to-tail connected INV5 and INV6 act like resistors. The beauty of the circuit is that it has no internal node. Therefore it is possible to operate at very high

frequency (VHF). The transconductance of these inverters is proportional to their quiescent drain current, which is tunable and is controlled by adjusting the supply rail voltages.

By using floating gate inverters, we have the opportunity to tune the quiescent drain current by biasing the extra inputs coupled to the gate [48]. For example, we could raise and lower the gate voltages of nFET and pFET, respectively, to increase the quiescent drain current. Thus, it provides a way of tuning the filter. However, coupling extra inputs to the gate has an adverse effect that the gain will be lowered, since the signal on the gate is now the average of the input signal and a DC bias voltage. The author interpreted this effect as “an advantage”, because now the input signal can have a wider swing (rail-to-rail) before output is distorted. The other advantage is that we have more controls over the tuning of the filter other than simply changing supply voltage. A drawback negates the benefit of high speed seen in Nauta’s g_m -C filter due to the absence of internal nodes. By introducing floating gate structures, the capacitors coupled to the gate will limit the bandwidth.

Floating gates can also be used to achieve very low voltage operation by effectively raising the gate voltage. Traditionally, a CMOS inverter’s power supply should be at least $V_{thn} + V_{thp}$, the sum of threshold voltages of n- and p-MOSFET. By using the floating gate, it is possible to establish an arbitrary charge on the gate, so that in effect setting V_{thn} and V_{thp} to an advantageous bias for a particular application. This is termed “threshold shifting” in [49]. In [50], designers set up charges on the floating gates by illuminating the chip with UV while applying reverse voltages on the sources of the MOSFETs, i.e., positive voltage to the source

of the nFET and negative voltage to the source of pFET. Note that the substrates are disconnected to prevent the p-substrate-n-well junction from entering forward bias. Afterwards, the substrate and n-well are connected back again, and charges on the capacitors of the floating gate will be established. The author reports several translinear elements with the UV-initialization. An obvious drawback for this configuration is that the sources of the MOSFETs are electrically separated from their wells and need to be connected externally.

2.6.2 MITEs and Capacitive Division Applications

A multiple input translinear element (MITE) uses MOSFETs operating in subthreshold with a gate that is floating and couples to several control gates [51].

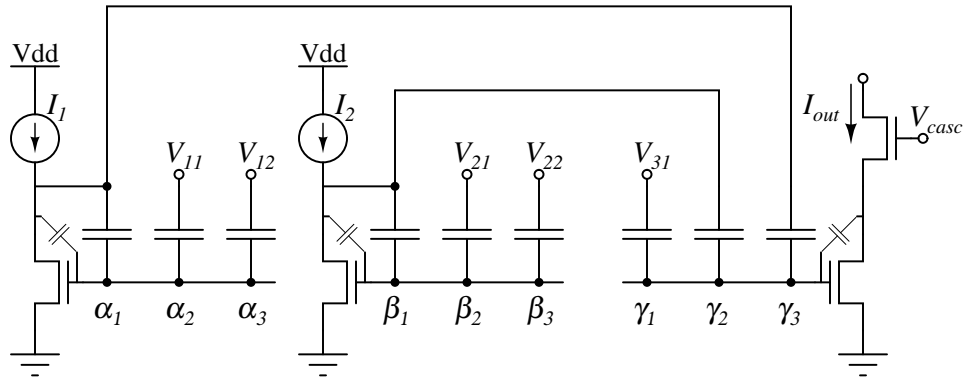


Figure 2.12: An example of the MITE circuit.

Generally, a translinear element has a transconductance that is linear in current. For a MOSFET, the transconductance of a subthreshold transistor

$$g_m = \frac{\partial I}{\partial V_g} = \frac{\kappa}{V_T} \cdot I$$

is proportional to the channel current I , with a factor of $\frac{\kappa}{V_T}$. This gives opportunities

to log-encode an input signal and to perform computation in log-domain. Addition in the log-domain is equivalent to multiplication. Fig.2.12 shows an example. Suppose that α 's, β 's and γ 's are the capacitance coupling ratio on the floating gates for the respective bias voltages V_{11} - V_{31} and that all MOSFETs are in subthreshold, then

$$I_{out} = I_0 \left(\frac{I_1}{I_0} \right)^{\frac{\gamma_3}{\alpha_1}} \left(\frac{I_2}{I_0} \right)^{\frac{\gamma_2}{\beta_1}}$$

where I_0 is a constant. Let V_{11} - V_{31} be grounded, all capacitors are of equal value (i.e., $\alpha_x = \beta_x = \gamma_x = 1/3$), then $I_{out} = I_1 \times I_2 \div I_0$. MITE circuits can compute many other non-linear functions such as the very useful length of a vector $\sqrt{a^2 + b^2}$ and the geometric mean \sqrt{ab} .

A remarkable application with MITEs is the ultra-low power adaptive filter [52]. The system is a tunable first order low pass filter (Fig.2.13) to implement on-line learning of parameters for an unknown target system. The target system is also a first order low pass filter, with unknown gain and time constant. A control

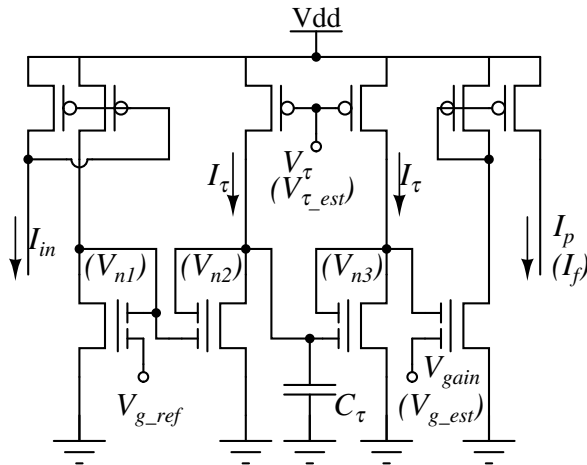


Figure 2.13: The tunable first order low pass filter using MITEs.

circuit based on Lyapunov method [53] controls the gain and the time constant of the adaptive filter by adjusting V_g and V_τ , respectively. The control circuit monitors the outputs of both the adaptive filter and the unknown target and finds the optimal parameters V_g and V_τ to match the target system. Application is mainly in system identification.

Other applications that mainly utilize the capacitive division include:

- The threshold logic neuron-MOS family pioneered by Shibata and Ohmi [54–56], the followers [57–59] with applications [60–63].
- The flash A/D converters and digital multipliers built with threshold logic counters [64–66].
- Analog multipliers that use variable resistor implemented with floating gate MOSFETs in triode region [54, 67]. The linearity is generally poor, as suggested in a survey article for multipliers [68].
- Analog multipliers that use MOSFETs in their saturation region using floating gates [69–71]. They all have the same topology (Fig. 2.14) both with or without the current source I_{tail} at the bottom. It is generally required to fine-tune the bias in order to achieve minimum total harmonic distortion (THD) and non-linearity. With the inherent floating gate addition, it is easy to extend simple multiplication to $(V_1 + V_2) \times (V_3 + V_4)$ by adding more inputs to the floating gate.

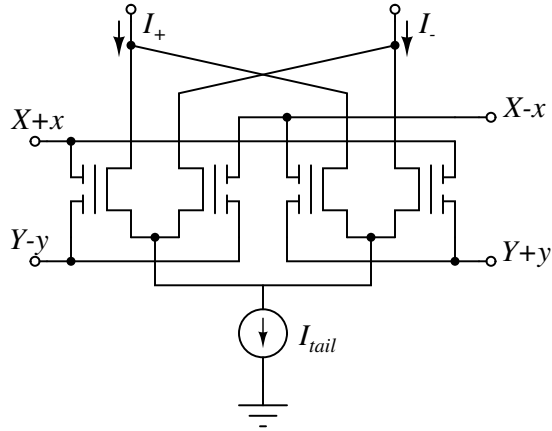


Figure 2.14: This floating gate multiplier operates in saturation region.

2.6.3 Arrays of FG Storage for Computation and Trimming

Since 1967 [1] arrays of floating gate structures in CMOS are selected for massive nonvolatile storage. Modern nonvolatile memories are fabricated in special processes [2] to increase density, efficiency and yield. Here, we focus mainly on arrays of FG structures implemented in standard CMOS that achieve extended functionalities.

Imagers that have pixels with built-in multipliers [34, 40] can perform vector-matrix multiplication easily. Vector-matrix multiplication is useful in computing discrete cosine transform (DCT), discrete sine transform (DST), Hadamard and Haar transformations. Haar transforms can be used for wavelet-based compression similar to JPEG2000. In [34, 40], the transform block parameters are stored in an array of floating gate structures, and are sent to the pixel matrix for vector-multiplication with pixel currents. The result is a compressed image signal on the output, which is then decoded and displayed in a computer. Since the compression is performed in the pixel with transistors operating in subthreshold, the power

consumption is very low compared to traditional digital measures.

A field-programmable-analog-array (FPAA) has been implemented with floating gate structures [72]. A network of central-pattern-generating silicon neurons uses floating gate array to store synaptic weights [73]. A programmable arbitrary waveform generator (AWG) [74] is yet another example of floating gate array parameter storage.

Programmable potentiometers [42] (e-pots, Fig.2.15) are one remarkable example using floating gate storage and trimming. The motivation stems from the limitation of the pin numbers for a VLSI chip, since the pin number scales only to the square root of the area of a (square) chip, and many bias voltages are generally required for mixed-signal circuits. The e-pots are non-volatile, small, tweakable and individually addressable. Each e-pot is monitored and trimmed by hot-electron injection and tunneling. The programmed voltage show a systematic offset error of 19.3mV, but when subtracted, the output voltage show a remarkable $175\mu\text{V}$ standard deviation, and a 2mV deviation for 0.2V power supply fluctuation. Measurements show 20mV drifting in the output voltage for the first 40hr, but is stabilized afterwards.

Precision trimming with floating gates has found many diverse applications: a 14-bit digital-to-analog converters (DACs) [23], analog-to-digital converters (ADCs) [26, 75], current sources [21, 22], high-precision low-drift voltage reference [76] and imagers with non-uniformity corrections [27–31, 33].

In the following chapters, I present methods for enabling local adaptation in floating gate charge transport mechanisms that leads to automatic and accurate

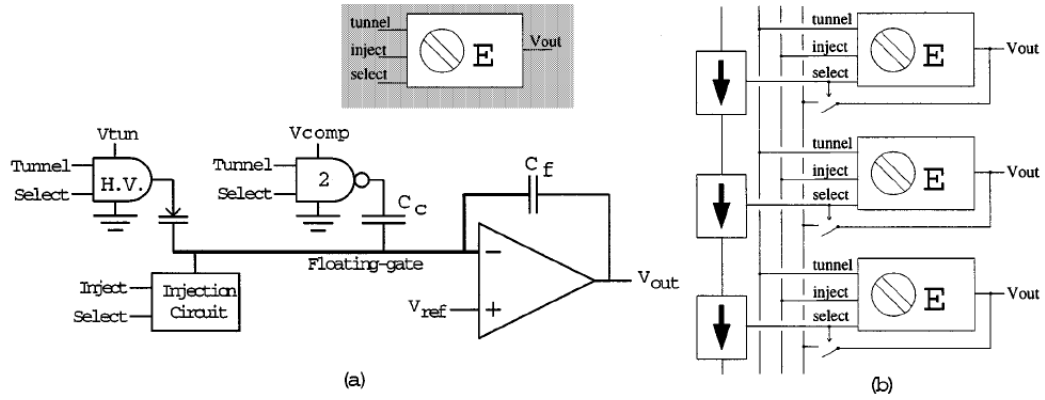


Figure 2.15: (a) The e-pot output is provided by an amplifier operating in a voltage follower configuration. The offset stored in C_f constitutes the voltage shift from the virtual ground (V_{ref}) to the V_{out} , and can be adjusted by the tunneling and injection mechanisms on the left. (b) e-pots are serially linked and addressable to external control. [42]

trimming in applications including comparators (Chapter 3), ADCs (Chapter 4) and imagers (Chapter 5).

Chapter 3

The Adaptive Floating Gate Comparator (AFGC)

3.1 Introduction

Comparators are decision-making circuits that interface between analog and digital signals. Comparators are used in a wide variety of circuit applications, including analog-to-digital converters, memories, dynamic logic, and sense amplifiers. A comparator usually consists of a pre-amplifier stage and a regenerative stage followed by a buffer. Mismatch due to process variation in the pre-amplifier and regenerative stages cause a switch point offset that directly affects resolution. A common and successful approach used to cancel offset is dynamic switching [77], which requires additional circuit components and multiple non-overlapping clocks. We report an adaptation method that requires a single switch and one clock signal to either program or cancel an offset. Since offset is a property of the circuit, it is natural to store it using nonvolatile storage on a floating gate. The ability to program desired nonzero offsets in comparators is a feature that is not readily available using existing offset cancellation techniques but is intrinsic to the voltage comparator we describe here.

We present the design of a comparator that automatically and accurately cancels offset, or depending on the application, can store a predetermined offset [4]. The offset may be cancelled or programmed in either a one-shot or continuous

fashion to calibrate for constant or changing conditions; the offset is retained using nonvolatile local storage, and for many applications it is not necessary to recalibrate dynamically. The calibration mechanism is self-limiting and converges to a stable value without user intervention.

3.2 Adaptive Floating Gate Comparator

The simple five-transistor circuit shown in Figure 3.1 (a) implements the Adaptive Floating Gate Comparator (AFGC), comprising pre-amplification and regenerative stages for the comparison as well as control and local storage for the adaptation. During normal operation (adaptation disabled), floating gate transistors M1 and M2 form the input devices of a differential pair and provide local charge storage. Cross-coupled nFET transistors M3 and M4 form the regenerative elements of the comparator. When the clock signal V_{clk} is “high”, the nFET switch M5 closes and resets the comparator. When V_{clk} is low, switch M5 opens and the evaluation phase begins. The “high” bias voltage on transistor M5 during reset determines the conductance of the regenerative elements and thereby the overall gain and speed of the comparator.

With the power supply V_{dd} set at the nominal operating voltage of 3.3 Volts, there is insufficient electric field between the pFET’s drain and source to produce hot electrons in the channels of M1 and M2. We therefore keep the AFGC’s V_{dd} at 4.5 Volts during normal operation and during adaptation. $V_{dd} = 4.5$ Volts strengthens source-to-drain electric fields thereby increasing the energy of electrons in the chan-

nels of M1 and M2. Adaptation is controlled by the common-mode input voltage V_{CM} : the common source voltage will follow V_{CM} , so raising V_{CM} enables adaptation by increasing the gate-to-drain and source-to-drain electric fields thus attracting hot electrons onto the floating gate, conversely lowering V_{CM} disables adaptation by decreasing the gate-to-drain and source-to-drain electric fields so that hot electrons are no longer attracted to the floating gate. During adaptation, negative charges accumulate on each of the floating gates, lowering their gate-to-drain and source-to-drain voltages and establishing negative feedback between the outputs and the inputs to achieve stable adaptation. While the adaptation mechanism for the AFGC results in reduction in the common mode voltage on the floating nodes, all results reported in this work include any additional error resulting from this shift; thus it does not present a significant limitation to accurate and automatic adaptation.

In the following sections we discuss two methods of injection, a static method and a dynamic method. The static method is simple and serves to illustrate the mechanism of calibration, but its accuracy is limited in practice. The dynamic injection method overcomes the accuracy limitations of the simple static method and provides calibration accuracy under 1mV; however, during dynamic injection the adaptation occurs during the evaluation phase so the output of the comparator is latched. This means that the update direction cannot change during a single cycle, thus accurate calibration must be achieved over many clock cycles. We also discuss the inherent tradeoffs between speed and accuracy, which can be tuned using the clock voltage V_{clk} . We present Monte Carlo simulations and experimental results which demonstrate the efficacy of the calibration using the dynamic injection

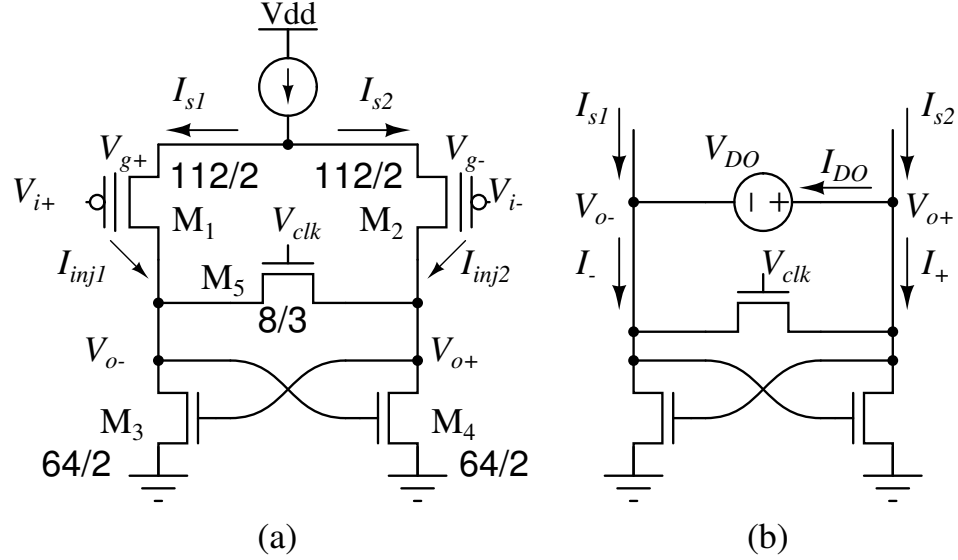


Figure 3.1: Adaptive Floating Gate Comparator (AFGC): (a) Circuit diagram of the AFGC with pFET input floating gate differential pair, crosscoupled nFET regenerative elements, and reset switch. Dimensions are specified as width/length, with unit $\lambda = 0.2\mu\text{m}$. The coupling capacitors from inputs $V_{i+,-}$ to floating gates $V_{g+,-}$ are 216fF. (b) During evaluation, bias voltage V_{clk} on the reset switch M_5 determines conductance of the regenerative elements and overall comparator gain. Transconductance of the crosscoupled pair and switch is determined using HSPICE simulation of circuits extracted from layout, as described in the text.

method.

3.2.1 The Static Injection Method

The static injection method accomplishes adaptation by applying a constant voltage bias to the clock terminal. When the clock is high, the comparator becomes an amplifier whose differential inputs ($V_{i+} - V_{i-}$) and differential outputs ($V_{o+} - V_{o-}$) are related by a finite voltage gain $A_V = \frac{V_{o+} - V_{o-}}{V_{i+} - V_{i-}} = cA_{fg}$. The constant c is the capacitance ratio C_{fg}/C_T , where C_{fg} is the capacitance between nodes V_{i+} and V_{g+} (and between nodes V_{i-} and V_{g-}), and C_T is the total capacitance coupled to the floating node V_{g+} . The voltage gain A_{fg} from floating nodes V_{g+} and V_{g-}

to the differential outputs is greater than the overall voltage gain A_V . The goal of offset cancellation is to balance the differential output ($V_{o+} = V_{o-}$) when the input difference is zero ($V_{i+} = V_{i-}$). Suppose that mismatch causes the outputs to be unbalanced ($V_{o+} > V_{o-}$) when the inputs are equal. When V_{dd} is sufficiently high, injection occurs when the common mode input voltage $V_{CM} = (V_{i+} + V_{i-})/2$ is raised. Since the source-to-drain voltage of M1 is greater than that of M2, the injection current I_{inj1} onto the floating gate of M1 will be greater than the injection current I_{inj2} at M2 and the floating gate potential V_{g+} will decrease faster than V_{g-} . As a result, the differential current I_{s1} will increase with a concomitant decrease in I_{s2} , causing the output voltage V_{o-} to rise and V_{o+} to fall. This feedback cycle will drive the floating gate voltages V_{g+} and V_{g-} to values that compensate for the initial device mismatches.

The input-referred offset after calibration depends on Early voltage, voltage gain and mismatch of both device and injection parameters. The voltage gain A_{fg} is the product of the input transconductance and the equivalent output resistance ($A_{fg} = g_{m12}R_{eq}$), where $R_{eq} = [r_{o12} \parallel r_{o34} \parallel (g_{m5} + g_{m34})^{-1}]$. We define conductance and resistance differentially, e.g., $g_{m12} = \partial(I_2 - I_1)/\partial(V_{g+} - V_{g-})$ and $r_{o12} = \partial(V_{o+} - V_{o-})/\partial(I_1 - I_2)$, where I_1 and I_2 represent the channel currents of M1 and M2, respectively. Note that for the positive feedback pair M3 and M4, the conductance $g_{m34} < 0$. Suppose that there is an initial output offset $V_{o+} - V_{o-} = \Delta V_o > 0$ when $V_{i+} = V_{i-}$ and $V_{g+} = V_{g-}$ (i.e., the inputs are equal and there is no charge on the floating gate). Injection causes V_{g+} to decrease by ΔV_{i+} and V_{g-} to decrease by ΔV_{i-} , so the differential output becomes

$$\Delta V'_o = V'_{o+} - V'_{o-} = (V_{o+} - V_{o-}) + A_{fg}(\Delta V_{i-} - \Delta V_{i+}).$$

The adaptation reaches equilibrium when both sides of the differential pair are decremented equally. Imbalance can result from mismatch of capacitance on the floating gates, mismatch of injection parameters, mismatch of Early voltage, or mismatch in bias conditions between the two sides of the differential pair. In the following development we assume matched capacitances, injection, and Early voltage, and focus on the bias dependence. In this case adaptation is complete when the injection currents reach the same magnitude $I_{inj1} = I_{inj2}$. The time required to achieve equilibrium is a function of initial offset (see Section 3.3.4), so in practice we simply let the system continue injecting for some fixed time. As injection continues, the floating gate voltages and the common source voltage continue to decrease, so the source-to-drain voltages and gate-to-drain voltages decrease and the two injection currents eventually decrease near zero $I_{inj1} = I_{inj2} \rightarrow 0$. Equilibrium is attained when the currents balance and it is not necessary to wait for them to approach zero. Under the operating conditions described, the dominant term in the exponent of (2.2) is a non-linear function $f_1(\cdot)$ of gate-to-drain voltage V_{gd} , so we approximate (2.2) as $I_{inj} = \alpha I_s e^{f_1(V_{gd})}$. Assuming matched injection parameters α and I_s :

$$\alpha I_s e^{f_1[(V_{g+} - \Delta V_{i+}) - V'_{o-}]} = \alpha I_s e^{f_1[(V_{g-} - \Delta V_{i-}) - V'_{o+}]}$$

Therefore $\Delta V'_o = V'_{o+} - V'_{o-} = \Delta V_{i+} - \Delta V_{i-} = \Delta V_o - A_{fg}(\Delta V_{i+} - \Delta V_{i-})$, and $\Delta V_o = (\Delta V_{i+} - \Delta V_{i-})(1 + A_{fg})$, so the input-referred offset after injection is

$$\Delta V'_i = \frac{\Delta V'_o}{A_{fg}} = -\frac{\Delta V_o}{A_{fg}(A_{fg} + 1)}$$

The input-referred offset will be reduced by $(A_{fg} + 1)$ after adaptation. Although

we assumed initial matched gate voltages $V_{g+} = V_{g-}$, this is not required. We can consider an initial gate offset $V_{g+} - V_{g-} = \Delta V_g$ as part of the input offset, which produces an extra term $A_{fg}\Delta V_g$ in the initial output ΔV_o . As adaptation reaches equilibrium, this extra term in the output is eliminated and the net result is the same.

To make the gain as high as possible, we bias V_{clk} so that the conductance $g_{m5} + g_{m34}$ is reduced to a small positive value. Note that $g_{m5} + g_{m34} = \frac{I_+ - I_-}{V_{o+} - V_{o-}}$. We find this operating point from HSPICE simulation of a circuit extracted from layout using the configuration shown in Fig.3.1(b). We set equal currents in the two sides of the differential pair ($I_{s1} = I_{s2}$) by applying equal gate voltages, and introduce an offset voltage source V_{DO} between V_{o+} and V_{o-} . We plot the current difference $-2I_{DO} = I_+ - I_-$ as a function of the voltage difference $V_{o+} - V_{o-}$ in Fig. 3.2. The conductance $g_{m5} + g_{m34}$ depends on V_{clk} and can be found from the slope of the curves at the origin. Negative $g_{m5} + g_{m34}$ causes positive feedback during reset and results in hysteresis in the circuit behavior. Negative $g_{m5} + g_{m34}$ occurs for low clock voltages ($V_{clk} \leq 2.2V$, e.g., traces a,b and c). In order to maximize the gain A_{fg} and avoid hysteresis, V_{clk} is selected so that $g_{m5} + g_{m34}$ is positive ($V_{clk} = 2.6V$ in trace d of Fig. 3.2).

While the method of static injection described above may be used successfully to decrease offsets, its ability to accurately cancel offsets is limited in practice. Both simulation and experiment (see Fig.3.14) demonstrate the phenomenon of “overshoot”— that is, injection does not stop when $V_{o+} = V_{o-}$, resulting in an equilibrium with $V_{o+} < V_{o-}$ or $V_{o+} > V_{o-}$ when injection currents are balanced.

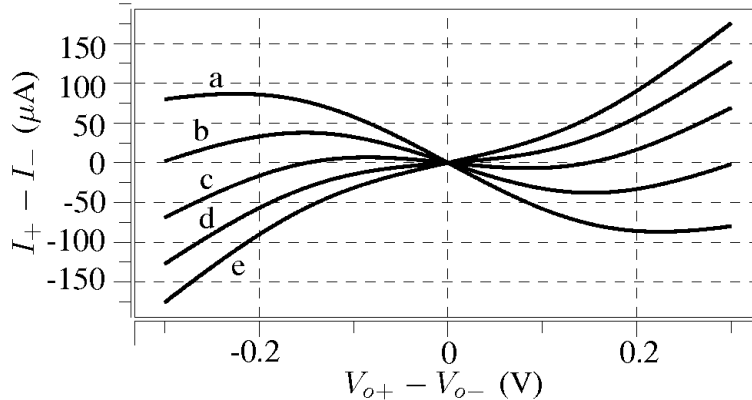


Figure 3.2: Bias voltage V_{clk} controls the conductance of the switch and regenerative crosscoupled pair. Simulated current-voltage relationship for the circuit of Fig.3.1b for V_{clk} from (a)1.4V to (e)3V in 0.4V steps.

This overshoot phenomenon exists because the injection currents become unbalanced during programming. The injection currents are proportional to the channel currents of the pFET differential pair, and these currents are changing in value due to adaptation. Note that the equilibrium does not imply equal channel currents or equal output voltages, since injection depends on both channel current and gate-to-drain voltage. Mismatch in injection, floating node capacitance, or Early voltage will further limit the accuracy of static injection.

3.2.2 Dynamic Injection on the Floating Gates

We describe a dynamic injection technique which overcomes the overshoot problem observed when using static injection. The dynamic technique achieves injection during the evaluation phase when the clock signal V_{clk} is low and the comparator is latched, with adaptation achieved over many evaluation cycles. By injecting with a running clock, we use the outcome of each comparison to correct

offset during the corresponding evaluation cycle. Thus the feedback loop encompasses all mismatch and offset within the circuit, and accurate offset cancellation can be achieved. We bias the common mode input voltage V_{CM} so that the drain-to-channel voltage is insufficient for injection during the reset phase of the clock cycle, but sufficient to produce injection during the evaluation phase when one of the outputs V_{o+} or V_{o-} is close to ground. From a simulation model [44] and our own experimental results, injection begins when drain-to-channel voltage exceeds 3V. For a pFET threshold of 1V, we bias V_{CM} above 2V. During reset both outputs are clamped at approximately the threshold voltage of an nFET $V_{o+} \approx V_{o-} \approx 0.7V$, so we set the desired V_{CM} between 2V and 2.7V. For V_{CM} higher than 2.7V, injection initially occurs during both reset and evaluation, but quickly reduces the common mode voltage of the floating nodes to 2.7V, after which the circuit enters the desired operating range. Suppose that the initial mismatch causes the outputs to be unbalanced $V_{o+} > V_{o-}$ when inputs are equal. When the comparator latches, V_{o-} is pulled to ground, injecting a small charge Q_{inj} on the gate V_{g+} . The charge accumulates on gate V_{g+} for each clock cycle until the gate voltage is low enough that the outcome reverses ($V_{o+} < V_{o-}$). Thereafter, the outcome alternates for each cycle and causes injection on the opposite side of the p-differential pair. Adaptation is controlled by the outcome of the comparison and the offset can be finely tuned.

In practice, any comparator has a limited conversion accuracy that can be defined by the variance of the input-referred noise. Ambiguity exists near the switching point where the outcome is uncertain. This uncertainty is caused by flicker noise and thermal noise generated by the MOSFETs within the circuit, as well as coupled

external noise. The probability that the outcome is correct depends on how far the input is away from the switching point. Empirically we find that this distribution is Gaussian, so we characterize the distribution with the mean and standard deviation obtained from the measured data. Fig.3.3 shows a typical measurement from one AFGC circuit. Fig.3.3(a) plots the measured comparison outcome as a function of the differential input voltage $V_d = V_{i+} - V_{i-}$ with an empirically fitted error function. This outcome is determined by observing the actual outcome through a low pass filter (see Fig.3.9b and detailed description in Sec.IV), and can be transformed to the cumulative distribution function (cdf) of the actual outcome through normalization. Figure 3.3(b) shows the probability density function (pdf) corresponding to the fit with mean $\mu = -25.9\text{mV}$ and standard deviation $\sigma = 1.1\text{mV}$.

Let X be a random variable representing the actual input offset having a nonzero mean μ and variance σ^2 . Then the cdf obtained from Fig.3.3(a) corresponds to $P[X < V_d]$. The goal of adaptation is for μ to approach a desired offset μ_d . Using the dynamic injection method, during each clock cycle μ increases by $\Delta V_1 = C_1^{-1} \int_T I_{inj1} dt \approx Q_{inj1}/C_1$ for $X < \mu_d$, and decreases by $\Delta V_2 \approx Q_{inj2}/C_2$ for $X > \mu_d$. C_1 and C_2 are the total capacitance on the floating gates, and T is the time the clock is low, typically half the clock period for a 50% duty cycle. We express the net shift in μ for one clock cycle as $\Delta\mu = \Delta V_1 P[X < \mu_d] - \Delta V_2 P[X > \mu_d]$. Adaptation finishes when an equilibrium $\Delta\mu = 0$ is reached,

$$\Delta V_1 \Phi\left(\frac{\mu_d - \mu^*}{\sigma}\right) = \Delta V_2 \left[1 - \Phi\left(\frac{\mu_d - \mu^*}{\sigma}\right)\right]$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ is the cdf of a Gaussian random variable with $\mu = 0$

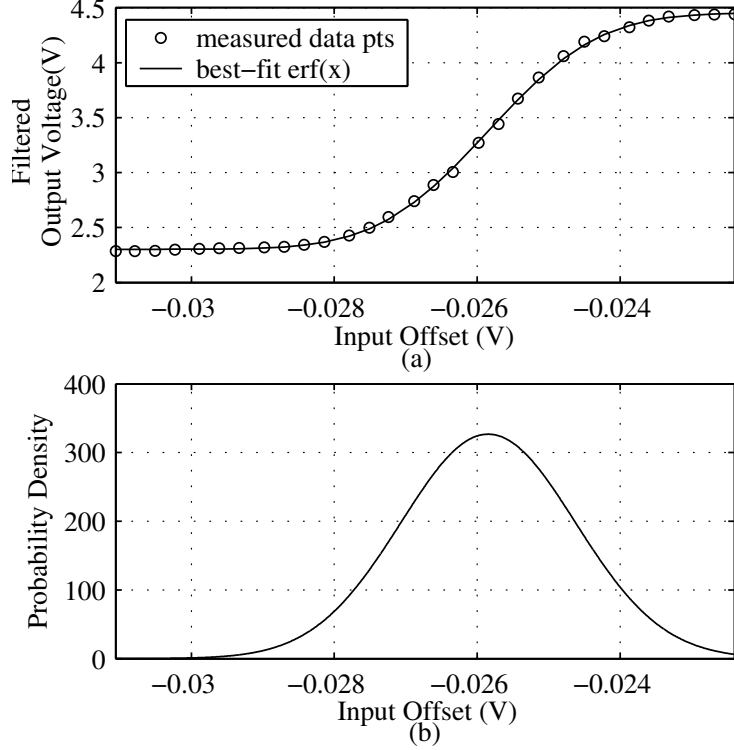


Figure 3.3: Circuit noise causes uncertainty in the outcome of the comparison: (a) a typical input offset distribution for one device obtained experimentally, showing measured voltage distribution and empirically fitted error function, and (b) corresponding empirical Gaussian probability density function.

and $\sigma^2 = 1$ and μ^* the input offset after adaptation. Therefore, we express μ^* as

$$\mu^* = \mu_d - \sigma \Phi^{-1} \left(\frac{\Delta V_2}{\Delta V_1 + \Delta V_2} \right) = \mu_d - \sigma \Phi^{-1} \left(\frac{1 - \rho}{2} \right), \quad (3.1)$$

where $\rho = (\Delta V_1 - \Delta V_2)/(\Delta V_1 + \Delta V_2)$ is the injection mismatch ratio, the normalized difference in voltage change between the two floating nodes due to injection during one clock cycle. Note that this mismatch can result from either mismatch in injection current or from mismatch in floating node capacitance. We can see that the residual input offset is not a function of the device mismatch, but rather a function of both injection mismatch ratio ρ and the standard deviation σ of the input-referred noise.

Figure 3.4 shows the absolute value of the residual input offset $|\mu_d - \mu^*|$: (a) as a

function of ρ for several values of σ , and (b) as a function of σ for several values of ρ . $|\mu_d - \mu^*|$ increases rapidly when injection is extremely unbalanced. For up to 68% injection mismatch ($|\rho| \leq 0.68$), $|\mu_d - \mu^*|$ is bounded by the magnitude of σ . Therefore, even with severely imbalanced injection currents we can obtain accurate calibration. Furthermore, we can improve the accuracy of calibration by increasing the gain A_V and therefore decreasing the input-referred noise $\sigma_{Vi}^2 = \sigma_{Vo}^2/A_V^2$, where σ_{Vi}^2 is the input voltage noise variance and σ_{Vo}^2 is the output voltage noise variance, as long as the output noise does not increase as much as the gain. For matched injection currents, we achieve zero offset ($\mu^* = \mu_d$) regardless of σ .

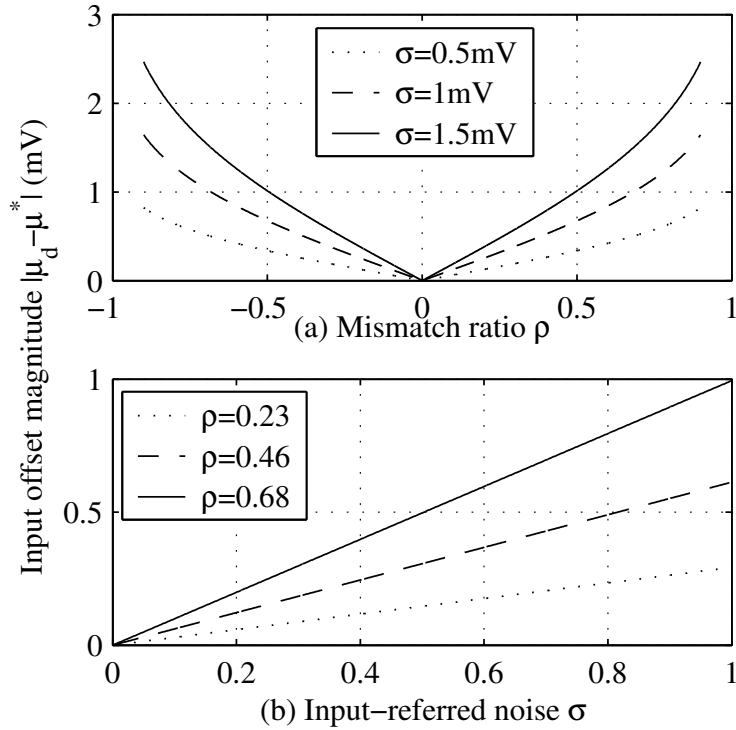


Figure 3.4: Absolute value of the input offset $|\mu_d - \mu^*|$ after calibration according to Eqn.(3.1): (a) as a function of ρ for several values of σ , and (b) as a function of σ for several values of ρ .

During injection, the gate voltages are raised to a high programming common-

mode voltage. During operation, the gate voltages are kept below an operating common-mode voltage limit to prevent injection. This difference in operating conditions during and after adaptation may introduce a small additional offset in the calibrated comparator. This imposes design constraints on the common-mode input voltages used for programming and for normal operation: ideally the common mode voltages should be as close as possible for accurate calibration, but different enough to provide significant injection during programming with negligible injection during normal operation.

3.2.3 Trade-off Between Resolution and Speed

Since offset resulting from device mismatch can be canceled, the resolution of the AFGC is determined by the input-referred noise. For perfectly calibrated devices, the error introduced by this noise will be random and may be reduced by examining the comparator's outcome over many evaluation cycles. When the devices are not perfectly calibrated, the residual offset contributes an additional source of input-referred noise which is deterministic. Under realistic conditions, this deterministic noise is smaller than the random noise (see above). The relative magnitudes of the deterministic and random noise sources are determined by the injection mismatch ratio.

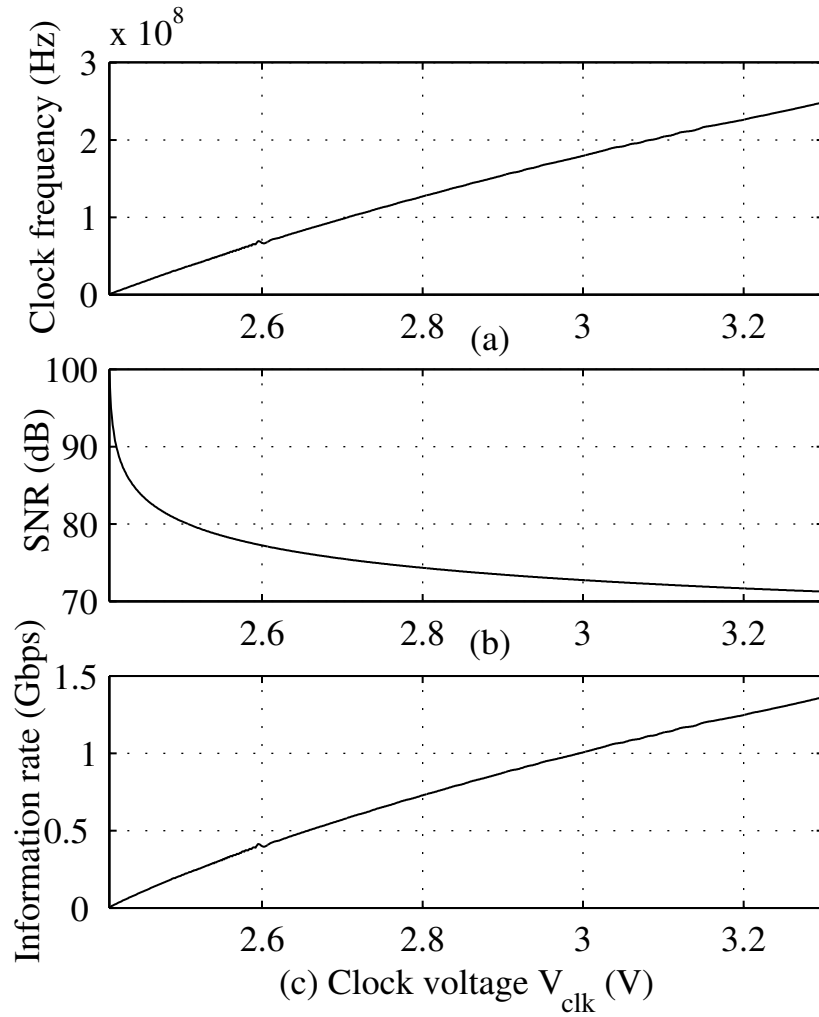


Figure 3.5: Increasing V_{clk} increases speed, bandwidth, and noise, decreasing signal-to-noise ratio and increasing total capacity: (a) maximum clock speed, (b) SNR, and (c) channel capacity as a function of the clock voltage, determined by simulation of an extracted AFGC circuit.

In the remainder of this section, we investigate the inherent trade-off between speed and resolution that occurs for random noise in a single evaluation cycle. For simplicity, we consider only thermal noise. The total mean-squared current noise power across V_{o+} and V_{o-} is $\bar{i}_T^2 = \sum_i \bar{i}_i^2 = 4kT\frac{2}{3}(g_{m1} + g_{m2} + g_{m3} + g_{m4}) \cdot B$ (in units of A^2), where k is Boltzmann's constant, T the temperature in Kelvin, and B the noise equivalent bandwidth (NEB). The input-referred voltage noise power equals $\bar{v}_i^2 = g_{m12}^{-2} \bar{i}_T^2 \approx 8.4 \times 10^{-17} \cdot B$ (in units of V^2), given a tail current of $100\mu A$, a clock voltage of $3.3V$, a room temperature $T = 300K$, the device geometries and the process parameters. Next, we calculate NEB as $B = \frac{\pi}{2} f_{3dB}$ [78] where $f_{3dB} = \frac{g_{m34} + g_{m5}}{2\pi C_L}$. Using capacitance extracted from layout, we find that the RMS input noise is $v_{i,RMS} = \sqrt{\bar{v}_i^2} = 320\mu V$. Under balanced operation with a fixed tail current, g_{m12} and g_{m34} are constant. Since $g_{m34} + g_{m5}$ is a function of V_{clk} , f_{3dB} and $v_{i,RMS}$ also become functions of V_{clk} . As V_{clk} decreases, $g_{m34} + g_{m5}$ and f_{3dB} decrease, and input noise is reduced. Figure 3.5(a) shows the maximum clock frequency as a function of the clock voltage V_{clk} , obtained from the gate delay determined by simulation of an AFGC circuit extracted from layout. The gate delay is taken as the settling time between the clock transition and the convergence of the output voltages to the RMS noise level during reset. The settling time of evaluation is determined by the time required for divergence of the output voltages to within 10% of the power supply and is found to be 1ns. This evaluation time is limited by the tail current and is less than the reset settling time for $V_{clk} \leq 3.3V$. Therefore the reset time dominates the gate delay and the speed depends on V_{clk} . Under the standard assumption that the input signal is a sinusoid with peak-to-peak voltage

equal to the power supply (3.3V), the signal to noise ratio is the ratio between signal power and the input-referred noise power. In Fig.3.5(b) we plot the signal-to-noise ratio ($\text{SNR} = 20 \log_{10} \left(\frac{1.65/\sqrt{2}}{v_{i,RMS}} \right) \text{dB}$) as a function of V_{clk} . Note that the speed increases with V_{clk} , but SNR decreases with V_{clk} . The channel capacity, or maximum number of bits per second for any signal distribution having a peak-to-peak voltage constraint, can be computed from the noise variance and the bandwidth as $C = f \cdot \log_2 \left(1 + \frac{2}{\pi e} \frac{1.65^2}{v_i^2} \right)$ [79]. Figure 3.5(c) plots the channel capacity C as a function of V_{clk} . Whereas increasing V_{clk} increases the operating speed, which tends to increase capacity, it also reduces gain, increases noise and reduces the accuracy of the comparison, which tends to decrease capacity. The net effect is an increase in the capacity as V_{clk} increases. C provides an upper bound of the information transmission rate of an Analog-to-Digital-Converter (ADC) constructed using AFGCs with the conversion outcome determined in a single clock cycle.

3.2.4 Monte Carlo Simulation

In order to verify the performance of the AFGC using dynamic injection, we perform Monte Carlo simulation using HSPICE with the circuit netlist extracted from layout. We use the poly and poly-2 layers to form 216fF capacitors at the input, coupling the input signals V_{i+} and V_{i-} to the floating gates V_{g+} and V_{g-} , respectively. We use the top layer poly-2 as the floating node rather than poly in order to minimize parasitic capacitances to ground. This floating node is connected to the gate of a pFET transistor via metal-1. The gate oxide capacitance is 40fF, so we anticipate

17% reduction in the input voltage swing due to charge sharing. We use a $100\mu\text{A}$ tail current, a 50% duty cycle 340MHz clock, and a “high” clock voltage $V_{clk} = 3.3V$. We augment the extracted netlist using the model from [44] to compute injection current (2.2). We increase the scale factor α by 10^7 to accelerate injection and reduce simulation time. We use the Monte Carlo method to simulate process variation in the following parameters: poly gate length, diffusion width, pFET and nFET threshold voltages, and injection scale factor α mismatch. Each process variation is specified as a Gaussian distribution with a given mean and standard deviation (σ), and values for each transistor are chosen independently. We use $\sigma_L = 0.6\%$ of minimum gate length $0.4\mu\text{m}$, $\sigma_W = 0.012\mu\text{m}$, $\sigma_{V_{thN}} = \sigma_{V_{thP}} = 10\text{mV}$ and $\sigma_\alpha = 20\%$. We approximate device noise by adding 3 parallel sinusoidal current sources across the output nodes V_{o+} and V_{o-} with amplitude $\sqrt{2}\sigma_n/\sqrt{3}$ and frequency 1GHz , $\pi^{-1}\text{GHz}$ and $\pi^{-2}\text{GHz}$, respectively, where σ_n is $0.27\mu\text{A}$. This simple quasi-random model is sufficient for transient analysis because the magnitude of variations on the output nodes matches that expected for random thermal noise. Although the spectral density differs from that expected for random thermal noise, it is of limited importance for transient analysis.

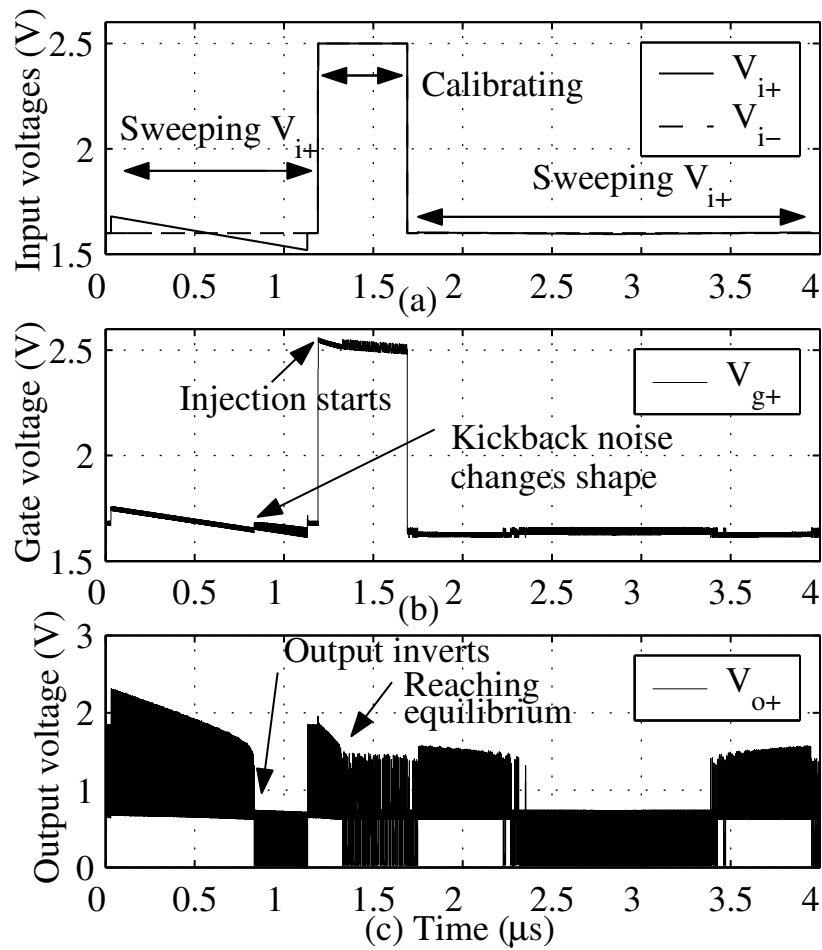


Figure 3.6: Simulation traces depicting one calibration cycle in a series of Monte Carlo simulations, depicting (a) input voltages V_{i+} , V_{i-} ; (b) floating node voltage V_{g+} ; and (c) output voltage V_{o+} .

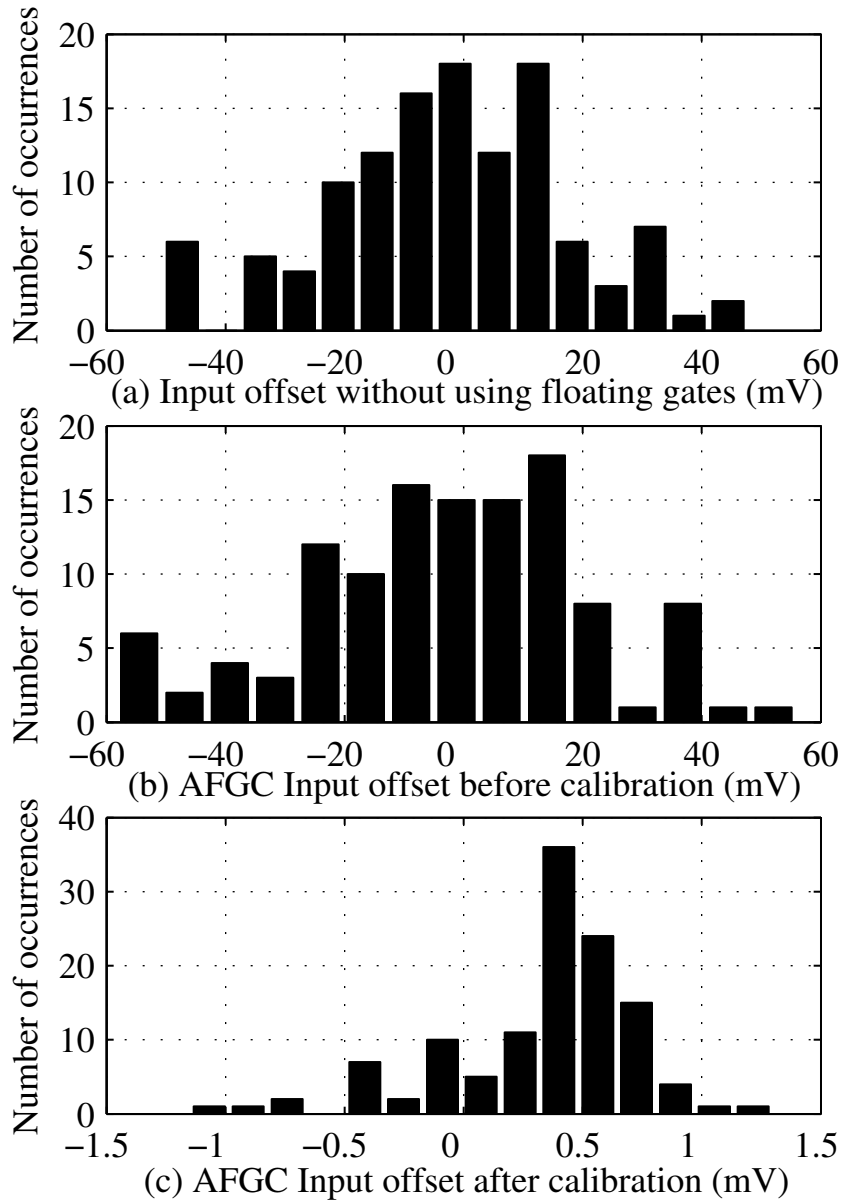


Figure 3.7: The use of floating gate transistors sacrifices gain at the input, but provides the ability to significantly reduce input offset. Histograms of input offset distribution determined through Monte Carlo simulations for (a) non-FG comparator and (b) AFGC before and (c) after calibration.

Figure 3.6 shows results from one trial of the Monte Carlo simulation: Figure 3.6(a) shows the input voltages V_{i+} and V_{i-} , Fig.3.6(b) shows one of the floating gate voltages V_{g+} for clarity, and Fig.3.6(c) shows an output voltage V_{o+} . First, we bias V_{i-} at 1.6V and sweep the positive input V_{i+} to find the point where the output inverts. The input difference at this point is recorded as the input-referred offset before calibration (ΔV_i). The AFGC suffers kickback noise on the floating gate voltages from the switching outputs V_{o+} and V_{o-} , as reflected in the floating gate V_{g+} shown in Fig.3.6(b). Note the correlation between the output states in Fig.3.6(c) and the shape of the kickback noise in Fig.3.6(b). Next, we raise both inputs to 2.5V for $0.5\mu\text{s}$ to enable adaptation. During the calibration phase, the floating gate voltage V_{g+} of Fig.3.6(b) decreases. The output voltage shown in Fig.3.6(c) alternates soon after calibration starts, indicating an equilibrium state. Finally, we sweep V_{i+} from high to low and then back up from low to high and record the differential input voltages at the two points where the output voltage switches. We then take the mean of the two to compensate circuit noise and obtain an estimate for the input referred offset after calibration ($\Delta V'_i$). We perform 120 trials, each with Monte Carlo variables drawn from independent Gaussian distributions, and obtain the input offset distributions shown as histograms with 15 equally-spaced bins in Fig.3.7. Figure 3.7(a) is the simulated input offset distribution of a comparator of identical structure except that floating gate transistors are replaced by normal pFETs. Figure 3.7(b) is the simulated input offset distribution of the AFGC before calibration. Figure 3.7(c) is the simulated input offset distribution of the AFGC after calibration. The standard deviation of ΔV_i ($\sigma_{\Delta V_i}$) in (a) is 20.4mV, in (b) is

23mV, and $\sigma_{\Delta V_i}$ in (c) is $413\mu\text{V}$. The mean of ΔV_i ($\mu_{\Delta V_i}$) in (a) is -2.5mV, in (b) is -2.9mV, and $\mu_{\Delta V_i}$ in (c) is $332\mu\text{V}$. Note that there is a small positive mean in Fig. 3.7 (c) for the offsets after adaptation. It is likely that this offset results from residual injection during the sweeping of the input differential voltage, which is magnified in this simulation by a factor of 10^7 and is further enhanced by biasing the common-mode voltage V_{CM} at 1.6V. This relatively high common mode input voltage does not completely eliminate injection on the floating nodes and was chosen to reduce the applied common-mode voltage change between adaptation and evaluation.

By using floating gate transistors at the input, we sacrifice gain due to capacitive sharing, resulting in larger input deviation $\sigma_{\Delta V_i}$. However, the floating gate transistors allow us to effectively reduce the input offset through adaptation, and under these simulation conditions we achieve a reduction of 55.7 (34.9dB) in offset variance (a factor of 49.4 (33.9dB) relative to the non-FG comparator).

3.3 Experimental Results

The AFGC described in Section 3.2 has been fabricated in a commercially available $0.35\mu\text{m}$ CMOS technology with 2 poly layers and 3 metal layers. The layout has been implemented using scalable submicron rules [80], and one AFGC occupies an area of $52\mu\text{m} \times 38.6\mu\text{m}$ ($65\mu\text{m} \times 38.6\mu\text{m}$ with the tunneling node) with $\lambda = 0.2\mu\text{m}$. A photomicrograph of the fabricated circuit is shown in Figure 3.8.

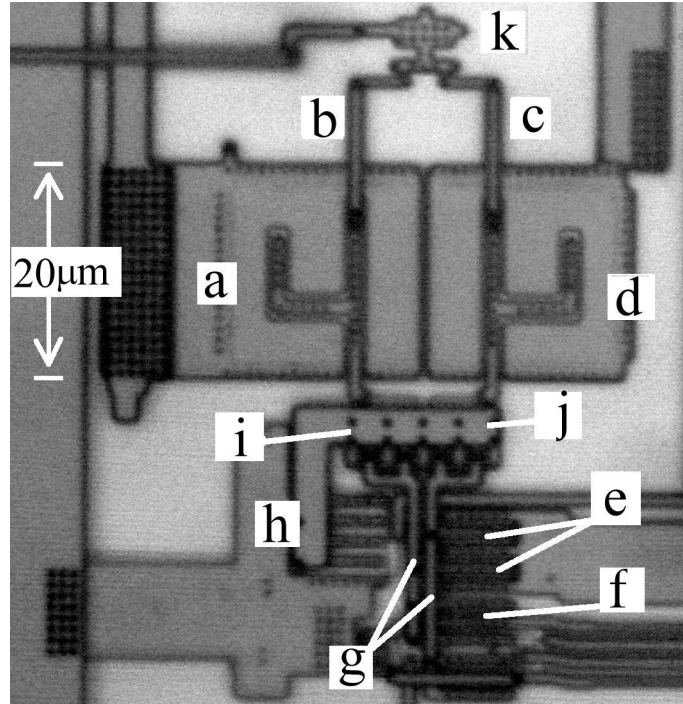


Figure 3.8: Photomicrograph of a single AFGC. The floating gate poly2 is sandwiched between the metal-poly enclosure labeled by (a) and (d), which are the input voltages $V_{i+,-}$, respectively. The floating gates are connected to the tunneling electrode and to the input transistors by metal1 labeled (b) and (c). The tunneling structure is illustrated in Fig.2.8. M1 and M2 are indicated by (i) and (j) respectively; M3 and M4 are indicated by (e), (f) indicates M5, (g) indicates the two output nodes $V_{o+,-}$, (h) shows the tail current mirror and (k) shows the tunneling electrode.

The circuit configuration used for testing the comparator is shown in Fig.3.9. We supply the comparator with V_{CM} at the negative input V_{i-} and a differential voltage V_d between the differential inputs. The comparator depicted in Fig.3.1 drives the output buffer of Fig.3.9(a) to generate rail-to-rail signals $0V \rightarrow 3.3V$ on V_{out+} and V_{out-} . A cascade of geometrically scaled inverters [81] in Fig.3.9(b) deliver the signals to external pads with minimum delay. During reset V_{clk} is set “high” and both outputs of the comparator are high. During evaluation V_{clk} is set low and the outputs are determined by the comparison. We measure a low pass filtered version V_A of the digital output voltage A, as shown in Fig.3.9. We interpret this voltage to determine the probability that the output is logic high. We use a Keithley 236 to supply V_d in $100\mu V$ increments.

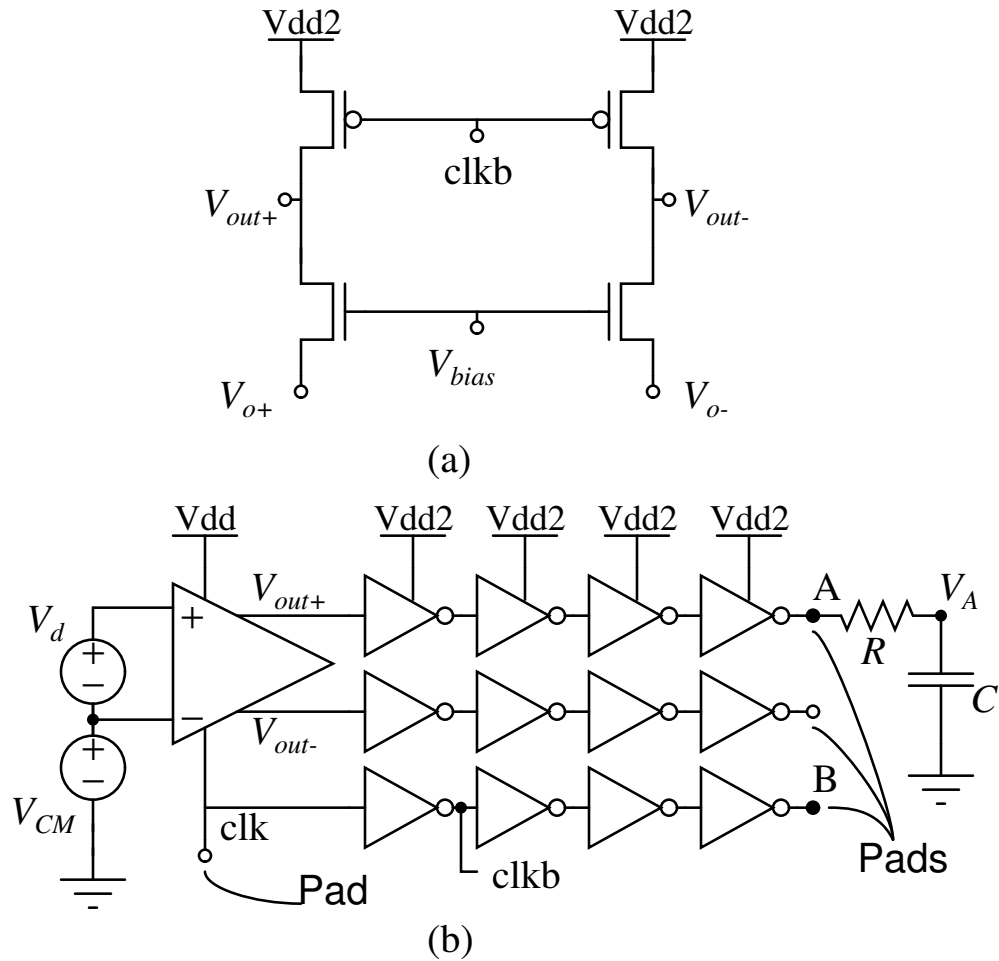


Figure 3.9: Circuit configuration used for testing the voltage comparator. The comparator output drives an output buffer shown in (a) which generates rail-to-rail output signals V_{out+} and V_{out-} . (b) A cascade of geometrically scaled inverters delivers the signal offchip with minimal delay, and the externally filtered output voltage V_A is interpreted as the probability that the output is logic high.

For simplicity, we operate the clock at 100kHz, and choose the time constant of the low pass filter to be $\tau = 2\pi RC = 0.01\text{s}$, so that the clock frequency is much larger than τ^{-1} , which is much larger than the measurement sampling frequency. Therefore, the output of the low pass filter V_A approaches the mean value m of the outcome. As before, let X be the random variable representing the actual input offset, and suppose that the outcome is low ($D_0 = 0$) when the differential input signal V_d is less than X , and high ($D_1 = 1$) when V_d is greater than X . Then, m is equivalent to the cdf $p_1 = P[X < V_d]$ since $m = \sum p_i D_i = p_0 \cdot 0 + p_1 \cdot 1$, where $p_0 = P[X > V_d]$. In practice, we measure V_A as a function of V_d (Fig.3.3a), then translate the filtered output voltage into probability by shifting and scaling the voltage V_A so that it ranges from 0 to 1. We interpret the scaled reading as the Gaussian cdf, and extract μ and σ from the data using a minimum squared-error curve-fitting procedure.

We measure a 5ns propagation delay from the clock edge at node B to the output change at node A in Fig.3.9(b) which corresponds to a sampling frequency of 100MHz. Comparators with sampling frequencies $\approx 1.3\text{GHz}$ have been reported in the same feature size [7]. The AFGC is current-starved with a relatively small tail-current, so it transitions slowly during evaluation. In future work we expect to increase the speed of the floating gate comparator by modifying the latch structure and output buffer.

3.3.1 Input Offset Distribution among the Chips

We measured the offset for AFGC circuits on twelve different chips under three experimental conditions: as received from the foundry before any adaptation (“raw”), after 20 hours of UV irradiation, and after adaptation. $V_{dd} = 4.5V$ for the AFGC, and $V_{dd2} = 3.3V$ for the output buffers for all experimental conditions. $V_{CM} = 1.6V$, except during adaptation when $V_{CM} = 2.5V$ (or higher).

Table 3.1 lists the mean and standard deviation of input offset voltage measured under the three experimental conditions described above for AFGCs from 12 different chips. For “raw” chips, the input offset has mean 45.35mV and standard deviation of 73mV. After 20 hours of UV-irradiation, the mean offset is reduced to 22.02mV with a standard deviation of 6.37mV. This suggests that a significant amount of random initial charge exists on the floating gate when the chip is fabricated and that UV irradiation allows this charge to dissipate. In effect this initial charge constitutes an additional nondeterministic offset which is added to the AFGC during fabrication by implementing floating gates using two polysilicon layers. Rodriguez-Villegas and Barnes report a layout technique to minimize charge trapped on floating nodes [82], but we did not take advantage of this technique in the AFGC structure reported here. We then enable adaptation of the residual offset by briefly raising V_{CM} on the pFET input differential pair, then return to normal operation by reducing V_{CM} . After adaptation, the mean offset is $-109\mu V$ with a standard deviation $\sigma_o = 379\mu V$. The maximum observed residual offset after adaptation is $728\mu V$. Whereas UV irradiation allows charge imbalances to dissipate, the

Table 3.1: Input Offset Statistics

	mean	std
Raw	45.35mV	73mV
UV-irradiated	22.02mV	6.37mV
Programmed	-109 μ V	379 μ V

adaptation technique compensates for offsets due to device mismatch. We achieve a factor of 2 reduction in the input offset mean and an order of magnitude reduction in the standard deviation of the mean after the 12 raw AFGC chips were UV irradiated for 20 hours. We achieve a further two orders of magnitude reduction in input offset mean and one order of magnitude reduction in standard deviation of the mean after adaptation of the 12 UV-irradiated AFGC chips.

According to Eqn.(3.1), residual offset after adaptation is a function of input-referred noise and injection mismatch. We infer the injection mismatch ratio (ρ) from the measured input-referred noise (σ_n) and residual offset. The injection is performed with $V_{clk} = 3.3V$, which results in input-referred noise $\sigma_n \approx 1.025mV$. We calculate the injection mismatch ratio ρ according to Eqn.(3.1), and find that ρ has a mean value of 8.3% with standard deviation 28% and a maximum observed value of 52.2%. Therefore, the pFET injection currents exhibit significant variation among transistors, however this variation does not prevent adaptation from achieving a residual input offset less than the standard deviation of input-referred noise.

3.3.2 Dependence of Conversion Accuracy on Clock Voltage

Figure 3.10 confirms that for lower V_{clk} , a finer resolution comparison can be made on the input signals. The voltage gain in the pFET differential pair is increased by lowering V_{clk} , so the input offset can be adjusted with higher resolution at the cost of longer time required for reset and for overall adaptation. If the clock voltage is too low ($< 2.1V$), reset will be incomplete, resulting in hysteresis in the comparison outcome and adaptation. We can avoid this problem by keeping the clock voltage above a level defined by the nFET threshold ($0.7V$). Experimentally we find that $742\mu V$ is a lower-limit for the input-referred noise σ_n . This exceeds the standard deviation of the input-referred offset (post calibration) $\sigma_n > \sigma_o$, which confirms that the input-referred noise dominates comparator resolution. Assuming that the input signal has a $3.3V$ peak-to-peak swing, $\sigma_n = 742\mu V$ translates into $63.9dB$ SNR or equivalently, 10 effective bits in a single comparison. By averaging over several calibration cycles we can reduce the contribution of circuit noise to reach the limiting resolution provided by the offset calibration procedure ($\sigma_o = 379\mu V$), which translates into $78.8dB$ SNR or 13 effective bits. In this experiment, we used a bias current of $40\mu A$, V_{CM} of $1.2V$, V_{dd} of $4V$, and V_{dd2} of $3.3V$, corresponding to AFGC power consumption of $160\mu W$.

3.3.3 Programming Input Offset in the $\pm 1V$ Range

The AFGC can automatically cancel input offset, as shown above, or program a desired offset over a wide range of input values. This feature leads to compact and

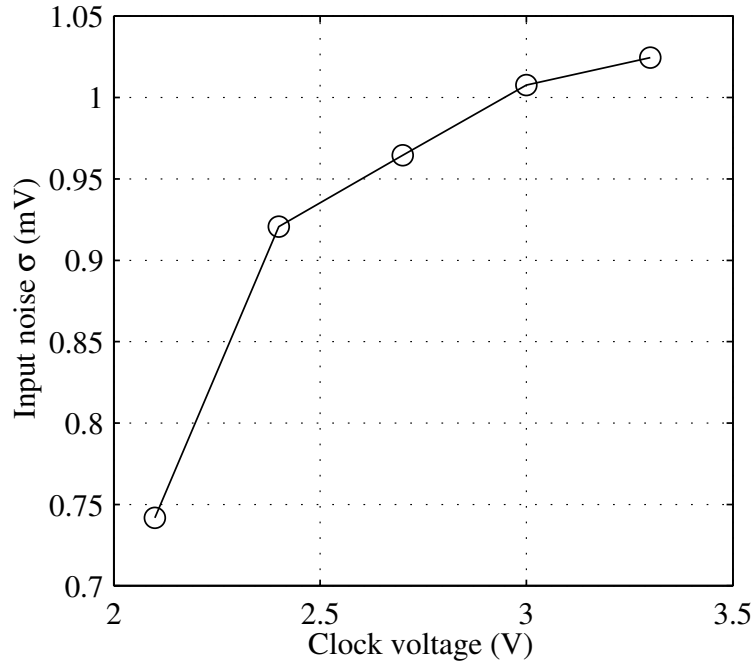


Figure 3.10: Conversion accuracy depends on the clock voltage V_{clk} . Accuracy increases and input-referred noise decreases with higher circuit gain as V_{clk} is reduced.

versatile implementations of flash data converters. Fig.3.11 shows the residual input offset voltages after programming different offsets ranging from -1V to +1V. These experiments used $V_{CM} = 2.5V$ for adaptation. The residual input offset voltage is defined to be the programmed input offset minus the measured input offset. The solid trace shows the input offsets measured at $V_{CM} = 1.9V$, and the dashed trace shows the input offsets when measured at $V_{CM} = 1.6V$. From the figure we can see that larger shifts of V_{CM} from injection conditions result in larger offset errors during operation. This is caused by Early voltage mismatches on the pFET differential pair and channel length modulation on the pFET that sets the bias current for the p-type differential pair.

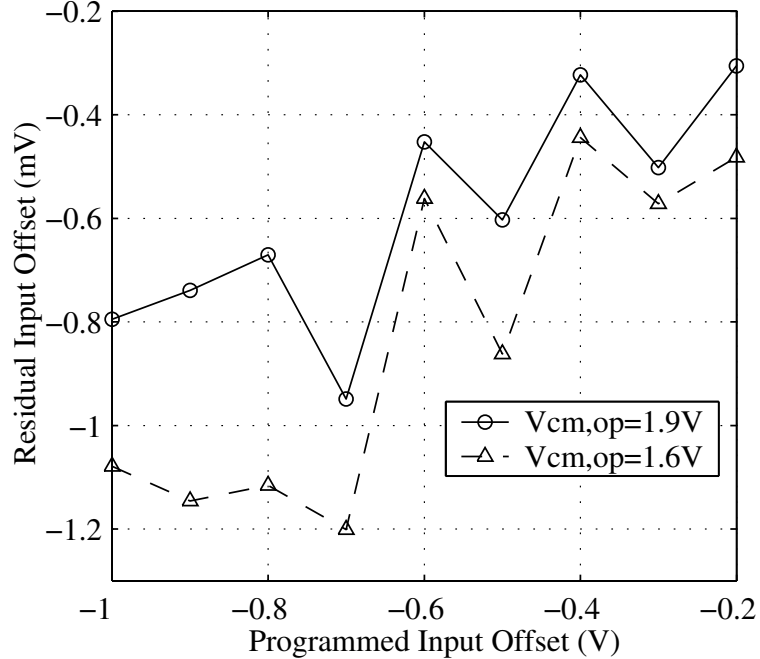


Figure 3.11: Residual input offset voltages remain small over a wide range of programmed offset voltages.

3.3.4 The Time Course of Offset Cancellation

Figure 3.12 shows the time course of offset cancellation for AFGC circuits on four chips. The four traces show the absolute value of the input offset voltage under different input common mode voltages during adaptation. We first program a 200mV input offset on the gate, and then pulse V_{CM} to an appropriate injection voltage (between 3V and 3.3V) for 10ms (1000 clock cycles) with $V_d = 0V$ and a clock frequency of 100kHz. We measure input-referred offset voltage with $V_{CM} = 1.9V$ between each pulse. For higher programming V_{CM} , the residual input offset converges faster. For lower programming V_{CM} , convergence is slower. The time course is roughly exponential, as predicted by the injection model. It is important to note that after the residual offset converges it remains constant below σ_n , for all

values of V_{CM} .

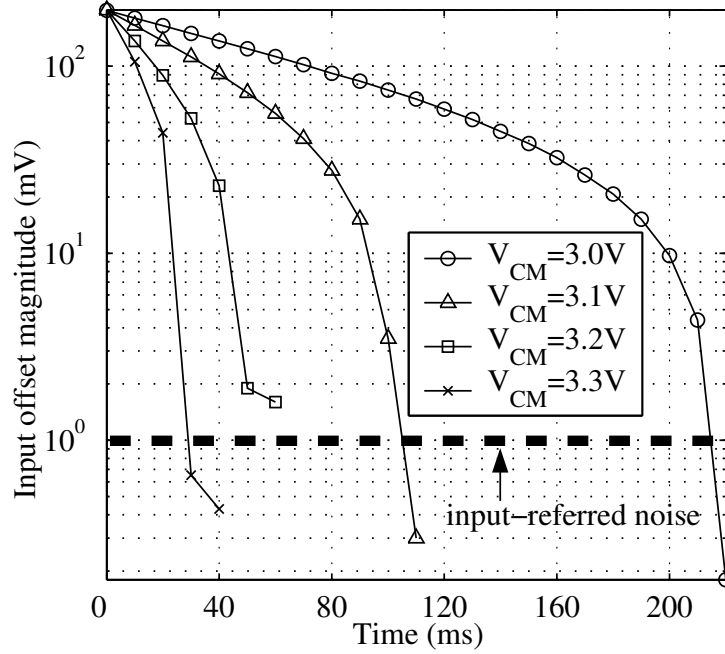


Figure 3.12: Input offset decreases exponentially in time. Offset is initially 200mV and decreases in time, with decay time constant decreasing with increasing V_{CM} . The decreasing time constant shows a speeding up of the adaptation near the end. This is because the steady state (50% HI and LOW) is reached well before the injection step becomes infinitesimal.

We assume that the input offset decays as $\mu(t) = \mu_0 e^{-t/\tau_j}$, from an initial value $\mu_0 = 200\text{mV}$ with injection time constant τ_j . From the experimental data $\mu(t)$ we can estimate τ_j . The estimates are depicted as ‘+’s in Fig.3.13 as a function of V_{CM} . The injection time constant ranges from 100ms for $V_{CM} = 3V$ to 18ms for $V_{CM} = 3.3V$. The injection time constant decreases exponentially with V_{CM} , i.e. $\tau_j(V_{CM}) = \tau_{j0} e^{-k(V_{CM} - V_{thj})}$, which is consistent with our simple model of injection. The time constant τ_j is inversely proportional to the injection current, which in turn is roughly exponential in gate-to-drain voltage. The gate-to-drain voltage scales with V_{CM} , and V_{thj} is the “injection threshold voltage”. We fit the data in Fig.3.13 with

a dashed line corresponding to $\tau_{j0} = 1\text{s}$, $k = 6.43$ and $V_{thj} = 2.64\text{V}$. Experimentally we find injection starting around $V_{CM} = 2.5\text{V}$ with adaptation occurring within seconds, confirming the accuracy of this simple exponential approximation.

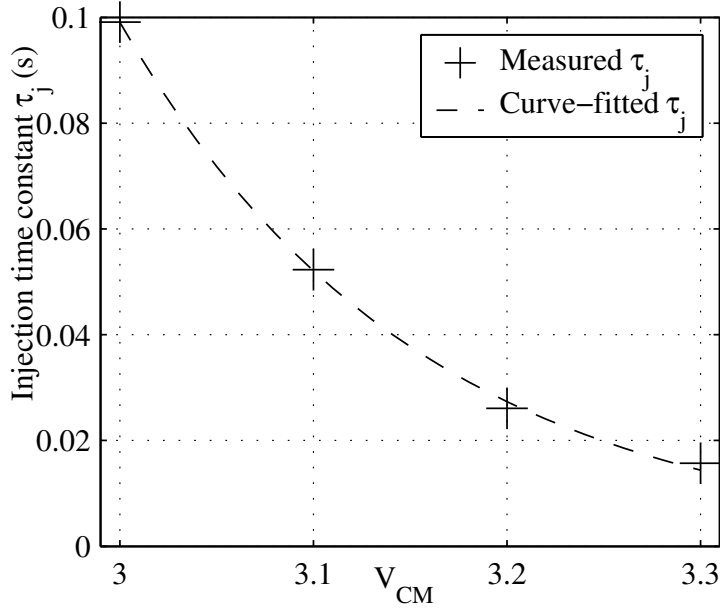


Figure 3.13: Measured and fitted injection time constant τ_j v.s. V_{CM} .

3.3.5 Robustness of Operation with Temperature and Time

Temperature and retention time studies of the AFGC are addressed in detail in [5]. Input offset is sensitive to temperature fluctuations. We measured the residual offset over a range of 20°C after allowing the chip to equilibrate for 2 minutes. The coefficient of variation in input offset with temperature change for one device was $+15\mu\text{V}/^\circ\text{C}$. Adaptation can be accomplished at any desired operating temperature by simply raising V_{CM} as described above, then returning V_{CM} to a suitable voltage for normal operation.

Relaxation of charge stored on the floating nodes after adaptation may cause drift of the input-referred offset over time. We have confirmed experimentally that the AFGC accurately retains offset for more than a month. We programmed initial offsets of 0V and 100mV, and periodically measured the residual input offset. Between each measurement, the chips were removed from the test fixture and stored on conductive foam. We performed the measurements using standard ESD protection without further precautions. For the chip programmed with 0V, the offset drifted down by $691\mu\text{V}$ in the first 3 days, then stayed around $-750\mu\text{V}$ through the end of the experiment. For the chip programmed with 100mV, the error stayed around $-450\mu\text{V}$ throughout the experiment. The initial drift is likely to be due to relaxation of the charge stored on the floating nodes or to further injection. Injection may continue to occur even when the inputs are biased at $V_{CM} = 1.6\text{V}$ used for evaluation, as V_{dd} remains at 4.5V (see simulation results and description in Section 3.2.4).

3.3.6 Overshoot in Static Injection

The results described in previous sections have been obtained using the dynamic injection method, with a running clock and update direction controlled by the outcome of each comparison during evaluation.

If instead of using a running clock, we supply a constant DC voltage at the clock terminal, the comparator becomes an amplifier with gain determined by V_{clk} and update direction controlled by the voltages at the outputs V_{o+} and V_{o-} . In this way, we observe the accuracy and time course of adaptation performed using the

static injection method. Figure 3.14 shows the time course of input-referred offset with a DC voltage of 3V applied to the clock terminal. The setup is similar to the previous experiment (injection time course), but with a higher V_{dd} (5.3V) and a longer V_{CM} pulse width (100ms). As before, we program an initial input offset of 0.2V, and record the input offset once every 100ms during injection. As depicted in Fig.3.14, the input offset voltage does not stop when the offset reaches 0V at time 0.4s, but rather continues to drop. This overshoot phenomenon exists because the currents in the two sides of the differential pair become unbalanced during programming, since the programming changes the gate voltages of the differential pair. The injection current is proportional to the channel current of the pFET transistor, and in order for correction to occur one injection current must be larger than the other. Injection continues until the two injection currents are equal, though the input offset is not.

3.4 Summary

We have described a novel floating gate comparator that can automatically and accurately cancel its input offset or allow programming of a specified offset. The AFGC uses pFET hot-electron injection in a negative feedback loop during calibration and programs a nonvolatile corrective charge on the floating gate. Residual input offset converges to the product of input-referred noise level and the inverse error function of the injection mismatch ratio; thus the residual offset is less than the input-referred noise standard deviation for an injection mismatch as large as

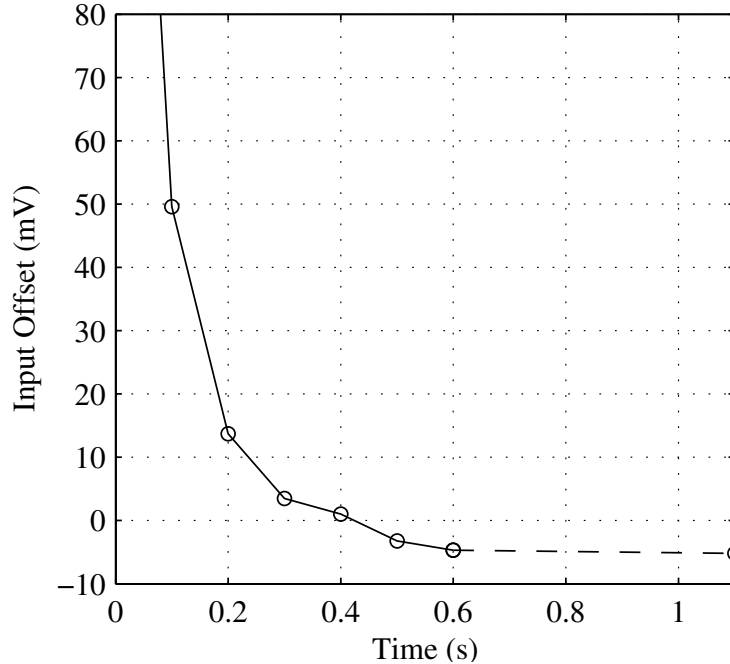


Figure 3.14: Input offset for the static injection method with clock voltage held constantly high. Offset first decreases, then “overshoots” the desired point and settles at a nonzero offset voltage.

68%. Experiments show that adaptation consistently reduces residual offset to a fraction of input-referred noise for all observed values of injection mismatch. We experimentally demonstrate more than two orders of magnitude reduction in offset voltage: the mean offset is reduced by a factor of 416 relative to fabricated chips directly from the foundry and by a factor of 202 relative to UV-irradiated chips. The adaptation mechanism encompasses the entire comparator circuit and therefore the residual offset is independent of device mismatch. Experimental results confirm theoretical predictions for mismatch, injection and adaptation speed. In the presence of observed 8.3% injection mismatch, the AFGC robustly converges to within $728\mu\text{V}$ of the desired input offset (mean offset $-109\mu\text{V}$, standard deviation $379\mu\text{V}$). Offset cancellation is achieved within milliseconds and the AFGC itself consumes \approx

300 μ W.

In addition to canceling offset, the AFGC can accurately store an arbitrary input offset, a feature not readily available in other offset cancellation schemes. While the ability to program offsets is particularly amenable to compact implementations of flash data converters, the AFGC may be used in any data converter in which offset cancellation or programming is desired and two conditions can be satisfied: the desired differential input can be presented across the input terminals (i.e., the input terminals can be shorted for offset cancellation), and the common mode input voltage can be raised to enable programming. Direct external access to the terminals of each comparator is not necessary. The input common mode shift required to enable adaptation can be accomplished using an auxiliary capacitor coupled into the floating nodes.

Chapter 4

The Adaptive Floating Gate Quantizer (AFGQ)

4.1 Introduction

The performance of flash analog-to-digital (AD) conversion is limited by sampling rate and precision, typically determined by the bandwidth and component variations intrinsic to a given technology. In many practical applications, the performance is further limited by disparity between the AD converter (ADC) characteristics and the signal being quantized. This occurs when the conversion range is not equal to the signal range, the amplitude distribution of the signal is not uniform, or the signal characteristics vary with time. In this paper we introduce the adaptive floating gate quantizer (AFGQ), an ADC architecture that stores reference levels using nonvolatile analog memory with a built-in adaptive programming mechanism. The key contributions are novel methods for precise calibration of reference levels in a flash ADC, for programming arbitrary AD mappings, and for autonomous adaptation of ADC characteristics to track a nonstationary signal. The 6-bit flash ADC prototype achieves 37.2 dB SNDR and 48.6 dB SFDR for low input frequencies ($f_{signal} \leq 24$ MHz, $f_{sample} = 750$ MHz) and 36.1 dB SNDR and 45.3 dB SFDR for input frequencies at Nyquist rate ($f_{signal} = 387$ MHz, $f_{sample} = 750$ MHz).

Comparator offset caused by component variations limits conversion accuracy and sets the maximum achievable SNR. Previously reported strategies for combating

offset include dynamic switching [77, 83, 84], averaging [7, 85–87] and background calibration [88–92]. In this work, we investigate the use of an adaptive floating gate comparator (AFGC [5, 6, 8, 93]) to set reference levels. The AFGC is able to accurately cancel offset or program an arbitrary offset automatically. Adaptation compensates for intrinsic device mismatches and achieves an offset error of less than $469 \mu\text{V}$ [6]. While designers usually take advantage of the full input dynamic range to maximize SNR, here we are able to program precise reference levels into each AFGC and achieve high precision conversion at a full-scale input range matched to the signal of interest.

We demonstrate a proof-of-concept 6-bit AFGQ using nonvolatile floating gate storage for reference levels and on-line histogram equalization to adjust reference levels to match signal statistics. Hasler et al. reported flash ADCs that use programmable potentiometers (e-pots [42]) to define reference levels manually for each individual comparator [26, 75]. The AFGQ is programmed by presenting the n^{th} reference voltage at the differential input terminals V_i^+ and V_i^- and issuing a programming pulse to the n^{th} comparator. The user repeats the procedure to program all 63 reference levels. The AFGQ implements an embedded adaptation algorithm for autonomously setting the reference level of each AFGC, so that there is no need for individual programming of each reference level. In a typical experiment, the user simply turns on “autonomous training mode” for a few seconds, after which an equalized output code histogram is observed.

The on-chip, on-line histogram equalization algorithm stems from the autonomous self-adaptive characteristic of the AFGC, and extends the adaptation

to match time-varying input signals. Alternative techniques for histogram equalization include automatic gain control (AGC) and companding. Such techniques introduce nonlinear gain into the signal path to achieve partial equalization of signal amplitudes. However, neither method tracks signal statistics in real-time; AGC is susceptible to outliers and creates difficulty in reconstructing true signal values, and companding uses a static nonlinear gain (i.e., A-law or μ -law [94]) which is matched to a specific signal such as speech. The equalized conversion is generally non-linear, so the analog values must be recovered with a nonlinear digital-to-analog (DA) mapping.

This paper is organized as follows: section II gives a short overview of floating gate technology and briefly describes the AFGC; section III describes the AFGQ design; section IV provides the framework, theory and implementation of the autonomous reference level learning for non-linear AD conversion and histogram equalization; section V presents the measurement setup and detailed experimental results using a wide variety of input signal distributions; finally, section VI summarizes the work and compares performance with other state-of-the-art 6-b ADCs.

4.2 Background Technologies

4.2.1 Floating Gate Structures

A floating gate MOSFET uses an electrically isolated material such as polysilicon to store charge indefinitely. There are no direct electrical connections to this circuit node, so charge on this gate remains trapped for a very long time. Thus

floating gate structures provide a nonvolatile analog storage mechanism, and are widely used to store data in EEPROMs [1], to trim current sources [21–23], to autozero amplifiers [25, 45], to store/cancel offset in comparators [6] and ADCs [26], to correct non-uniformity in imagers [10, 28, 30, 33], and to densely store large arrays of analog parameters [34].

Impact-ionized hot-electron injection [45] and Fowler-Nordheim tunneling [95] are used for adding and removing charge on the floating gate, respectively. High channel electric field near the drain and high vertical gate-to-drain electric field are easily achieved in pFETs, causing high hot-carrier generation rate and high gate collection rate at the same bias condition. Fowler-Nordheim tunneling requires high electric field across the oxide, and the tunneling current is a strong function of the applied electric field between the floating gate and a programming node. In the AFGQ on-chip charge pumps generate all high voltages required for injection and tunneling across gate oxide.

Charge retention in floating gate structures relates to the thickness and quality of the oxide. It has been shown [3] that 10 years of retention is achievable for transistors with the 70 Å oxide thickness available in 0.35 μm , 0.25 μm and 0.18 μm standard logic CMOS processes. Floating gate charge storage is reliable and accurate, and has been commercialized in low-drift, high-precision voltage reference devices [76] and system-on-chip flash memory [96].

4.2.2 Adaptive Floating Gate Comparator

An AFGC has two elements in addition to an ordinary clocked comparator: an offset storage device and feedback to adjust the charge storage. During adaptation, negative feedback leads to a steady state wherein the comparator is operating at its trip point, i.e., the output probabilities of HI and LO are close to 50%. Thus, the DC differential voltage on the input terminals at this steady state becomes the programmed comparator offset. The adjustment of the stored charge is carried out in small amounts for each clock cycle, and the steady state is reached after many clock cycles. We have previously shown that the residual offset error Δ after adaptation is equal to

$$\Delta = \sigma_X \Phi^{-1} \left(\frac{1 - \rho}{2} \right)$$

where σ_X is the input-referred temporal noise std, Φ^{-1} is the inverse Gaussian cdf, and $\rho \in [-1, 1]$ is the inherent mismatch in the programming mechanism [6]. We can thus improve the AFGC accuracy by achieving better comparator precision (lower σ_X) and better programming matching ($\rho \approx 0$). Note that transistor mismatch is irrelevant in the final accuracy of an AFGC. In section 4.4.3, we extend adaptation to time-varying signals, which forms the basis of the autonomous histogram equalization algorithm.

4.3 The AFGQ

Figure 4.1 shows the basic concept of using floating gate storage capacitors to implement reference levels in the flash ADC. Differential input signals V_i^+ and V_i^-

are sampled by track-and-hold (T/H) circuits prior to being fed to the comparators. The negative sampled signal V_{is}^- is connected to all negative inputs of the comparators, and the positive sampled signal V_{is}^+ is connected to the control gates of all 63 storage capacitors. Each storage capacitor stores a unique charge between the control gate and the floating gate, which connects to the positive input terminal of the comparator. In Fig.4.1, the 63rd capacitor stores a voltage $V_{c(63)}$ between the control gate V_{is}^+ and the floating gate $V_{fg(63)}^+$ for the 63rd comparator. The voltage stored across the n^{th} capacitor $V_{c(n)}$ is approximately equal to the reference level t_n for the n^{th} comparator.

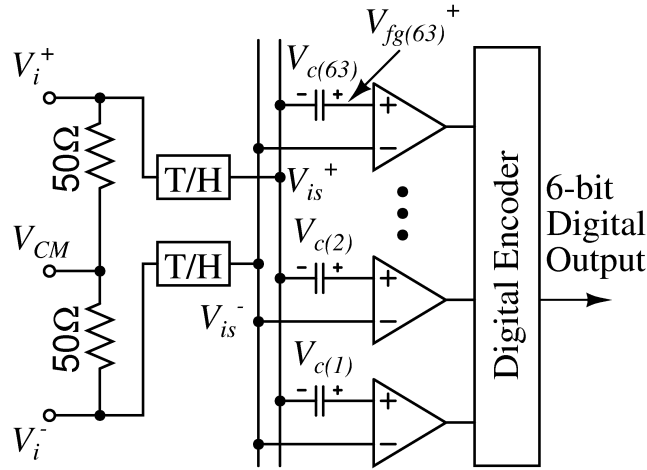


Figure 4.1: Each storage capacitor stores the reference voltage $V_{c(n)}$ for the n^{th} comparator.

4.3.1 Comparator Noise and the 6-bit AFGQ Resolution

The input offset for a comparator is modeled with a random variable X representing stochastic and deterministic effects. The variance σ_X^2 represents stochastic temporal noise resulting from thermal and flicker noise, and the mean $\mu_X = E[X]$

represents deterministic offset due to fabrication imperfection and transistor mismatch. We find empirically that the distribution of X is Gaussian. We have previously demonstrated that we can accurately trim μ_X such that the offset error $\Delta = \mu_X - \mu'_X$ is small, where μ'_X denotes the desired offset [8].

The signal-to-noise ratio (SNR) for an ADC can be expressed as $\text{SNR} = 10 \cdot \log \frac{\sigma_S^2}{\sigma_Q^2 + \sigma_C^2}$ where σ_S^2 , σ_Q^2 and σ_C^2 are signal power, quantization noise power, and conversion noise power, respectively. Neglecting σ_C^2 we obtain the direct relationship between SNR and effective bits (N_{eff}): $\text{SNR} = 6.02N_{\text{eff}} + 1.76$ assuming a sinusoidal input signal.

For flash ADCs both the offset and temporal noise for each comparator contribute to σ_C^2 , and for simplicity we consider them the only contributions to σ_C^2 . Thus $\sigma_C^2 = \sigma_X^2 + \sigma_\Delta^2$, where σ_Δ^2 is the variance of the offset error Δ under the assumptions that the offset errors are identically distributed and that the input signal visited the quantization range for all comparators. For an input sine waveform with peak-to-peak voltage (V_{pp}) of 1 V, $\sigma_\Delta = 469\mu\text{V}$, and $\sigma_X = 1\text{ mV}$ we obtain an SNR of 37.6 dB, significantly higher than the comparable SNR of 28.8 dB obtained for $\sigma_\Delta = 11.9\text{ mV}$ with the non-floating gate version of the same comparator. σ_Δ and σ_X were experimentally measured and reported in [8].

4.3.2 Designing the AFGQ

The AFGQ employs a flash architecture (Fig.4.2). The analog input is first sampled by a track-and-hold circuit (T/H), and then quantized by 63 3-stage pipelined

comparators inside the quantizing units (QU). Resulting thermometer codes are converted to quasi-gray codes by a simple NOR-based ROM decoder.

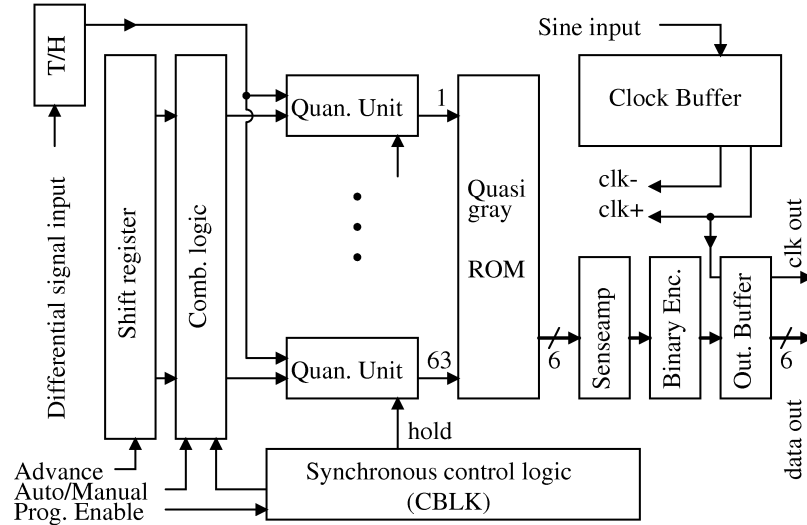


Figure 4.2: The AFGQ is a flash ADC with offsets programmed into each quantizing unit (QU).

The designs for the T/H (Fig.4.3a) and the comparator (b) follow [7]. The T/H delivers a $1V_{p-p}$ sine with 46 dB signal to noise-plus-distortion ratio (SNDR) to the QU array load of 3.8 pF at Nyquist. The differential T/H consists of two copies of the circuit shown in Fig.4.3(a), one for each differential input signal, which is terminated with a 50Ω on-chip resistor connected to a 0.5 V DC common mode voltage V_{CM} . The pipelined comparator produces results in 1.5 clock cycles (Fig.4.3c). The bias currents for stages 1, 2 and 3 are $140 \mu\text{A}$, $270 \mu\text{A}$ and $400 \mu\text{A}$, respectively. The T/H block and QU array consume a total of 31 mA and 51 mA static current, respectively. The AFGQ was fabricated in a 3.3 V, $0.35 \mu\text{m}$ 2-poly CMOS technology.

The QU (Fig.4.4) is an extended version of the AFGC, which uses the comparison outcome for adjusting its input offset in small increments, and accomplishes

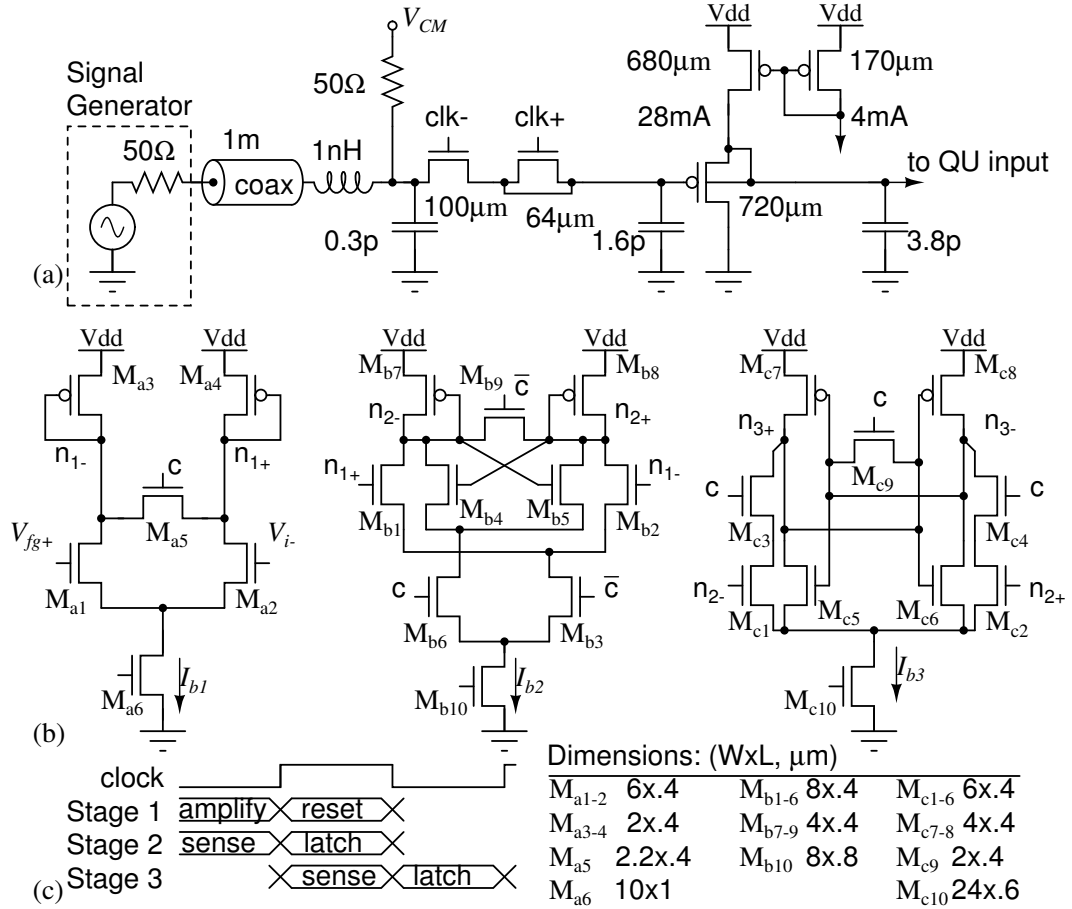


Figure 4.3: (a) T/H including equivalent transmission line, pad parasitics and 3.8 pF capacitance load. Transistor widths are labeled near each transistor. 0.4 μm gate lengths are used for all transistors. (b) The 3-stage pipelined comparator occupies 130 μm x 16 μm chip area. (c) The pipelined comparator produces results in 1.5 clock cycles.

offset adaptation over many adjustment cycles. Each QU consists of one comparator (CP1), positive and negative charge pumps and control logic. The offset V_c is stored on capacitor C_1 , which couples the floating gate V_{fg} to the multiplexer MX1. V_{fg} is connected to the positive input of CP1, and the negative input is connected to V_i^- via dummy devices used for symmetry. The programming enable signal (PE) determines the operational mode; during conversion (PE=0), MX1 passes the positive input signal V_i^+ to the comparator via C_1 . The comparator passes its result to the

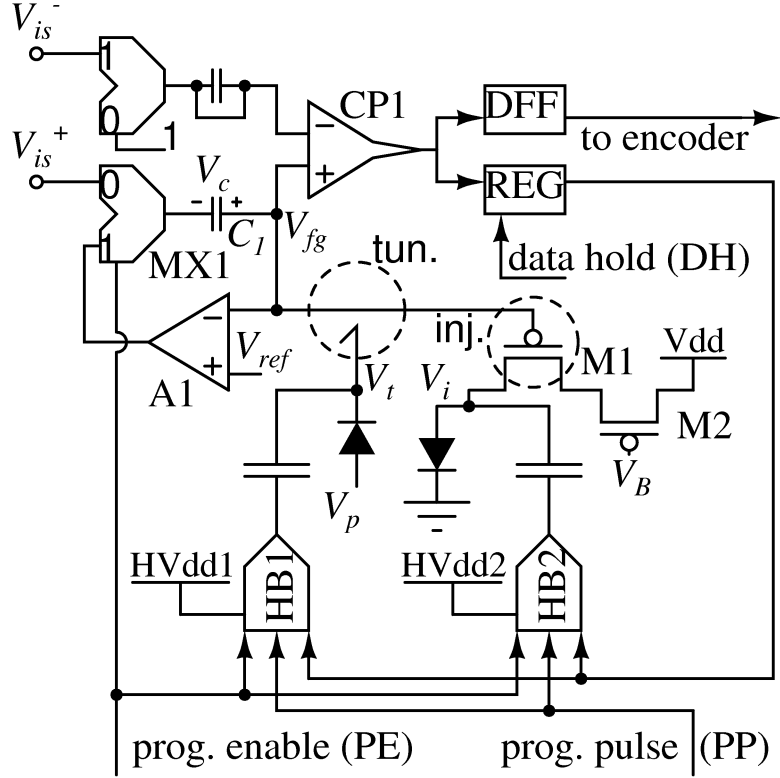


Figure 4.4: The QU includes offset storage and feedback mechanisms for adjusting storage.

encoder. During programming ($PE=1$), MX1 forms a feedback loop for OP-amp A1, configured as a voltage follower. Regardless of the offset V_c present on C_1 , A1 sets V_{fg} close to an externally supplied reference voltage V_{ref} so that programming charges can be applied in controlled increments. Data hold (DH) signal is asserted one clock cycle ahead of PE, and the most recent comparison outcome before entering programming is written to a register (REG). An outcome of LO means that offset V_c is too low, so tunneling is briefly activated to raise V_c . Conversely, an outcome of HI means that V_c is too high, so hot-electron injection is briefly activated to reduce V_c . The QU quickly reaches a steady state wherein V_c is close to the desired offset. The tunneling and injection mechanisms are activated with a programming pulse

(PP), which is asserted when the feedback loop formed by A1 and MX1 reaches a steady state.

The dashed circles mark the tunneling and injection sites. The tunneling site is the gate oxide of a pFET with its source and drain shorted to its nWell (V_t). The injection site is the gate oxide of a pFET (M1), whose drain V_i is pulsed below GND to induce a high electric field in the channel. M2 sets the channel current for M1. High voltage buffers HB1 and HB2 generate the short voltage pulses required during programming. V_t is driven by HB1 to roughly 8 V for tunneling and V_i is driven by HB2 to -2 V for injection. Bias voltages and currents are adjusted to induce ± 1 V/sec rate of change on V_c via tunneling or injection. For a 1 V V_{pp} signal range, the maximum required change in V_c is 1 V, so programming time is set to 1.2 s for each QU to ensure that programming is complete. Each QU occupies $550 \mu\text{m} \times 16 \mu\text{m}$.

4.4 Signal Adaptation

4.4.1 Non-linear AD Conversion

Figure 4.5 depicts 6-bit non-linear AD conversions for (a) sine and (b) Gaussian-distributed signals. A continuous-time sine signal spends more time near the maximum and minimum values than the middle, and a Gaussian-distributed signal spends more time near the middle value. Let $F_1(v) = P[X < v]$ denote the cumulative distribution function (cdf) for input random variable X , $T(n) : n \rightarrow v$ the DA mapping function for 64 discrete n values, and $F_2(n)$ the cdf for the output

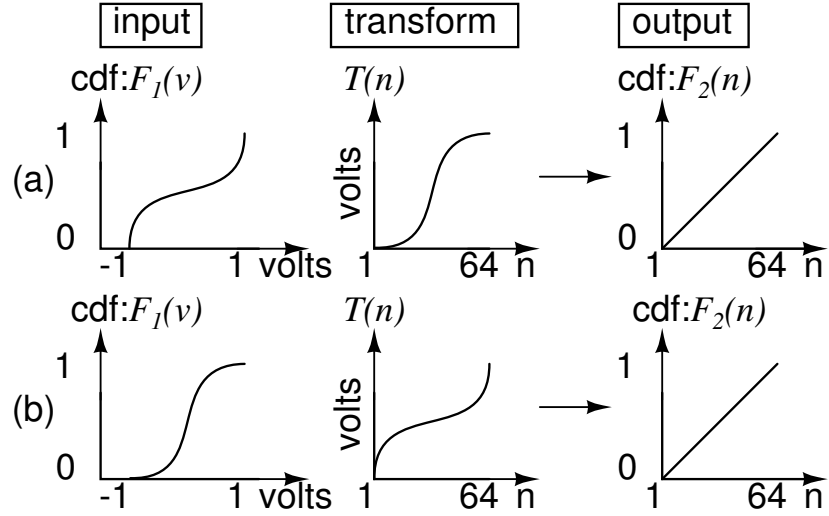


Figure 4.5: A non-linear transfer function T can equalize a non-uniformly distributed input cdf F_1 .

distribution. Note that $F_1(v)$ takes on a continuous value v (input voltage) and $F_2(n)$ takes on a discrete value n (output codeword). By substituting v with $T(n)$ in $F_1(v)$ we obtain $F_2(n)$, i.e., $F_1(T(n)) = F_2(n)$. By taking the inverse of F_1 on both sides, we can easily find the DA mapping function

$$T(n) = F_1^{-1}(F_2(n)) \quad (4.1)$$

Therefore, the input and output cdf together determine the DA mapping necessary to recover the analog input signal. In the case of histogram equalization, the desired probability mass function (pmf) is uniform and the desired output cdf $F_2(n)$ is linear as shown in Fig.4.5. According to (4.1), $T(\cdot)$ is simply the inverse of $F_1(\cdot)$.

In an N -bit flash ADC, there are $2^N - 1$ comparator reference levels t_n and 2^N DA mapping values $T(n)$, respectively. Once $T(n)$ is obtained using (4.1), it is adequate to assume that $t_n = (T(n) + T(n + 1))/2$, for $n=1$ to $2^N - 1$ [97]. The ADC produces output code n if the input X falls between t_{n-1} and t_n , for n from 1

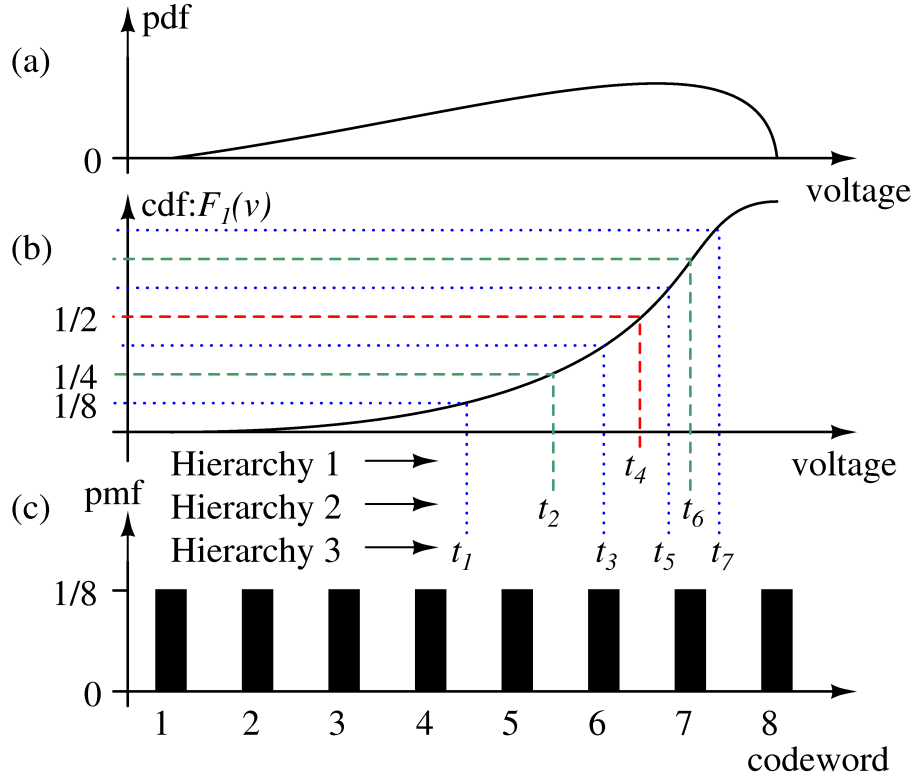


Figure 4.6: The non-linear conversion equalizes the output codeword histogram (c) from an arbitrary analog input distribution with nonuniform pdf (a) and corresponding nonlinear cdf (b).

to 2^N , where $t_0 \equiv -\infty$ and $t_{2^N} \equiv +\infty$.

4.4.2 Histogram Equalization

Figure 4.6 demonstrates histogram equalization in a 3-bit flash converter. Suppose that the input signal has the probability distribution function (pdf) shown in (a), the cdf $F_1(v)$ in (b), and undergoes non-linear AD conversion with the t_n 's marked on the x-axis of (b). The resulting output probability mass function (pmf) is uniform as shown in (c). Codeword 1 occurs with probability $F_1(t_1)$, codeword 2 with $F_1(t_2) - F_1(t_1)$, and so on. Clearly histogram equalization implies that

$F_1(t_n) - F_1(t_{n-1}) = 1/8$, for $n = 1, 2, \dots, 8$, where $t_0 \equiv -\infty$ and $t_8 \equiv +\infty$.

For implementation the algorithm has been structured according to equal-partitioning for reasons discussed in the next section. Level t_4 is the threshold for the middle comparator, which is responsible for the most significant bit (MSB) in the digital output, and we classify it as partition hierarchy 1. Level t_4 partitions the set of input signals into equally probable halves. At the next partition hierarchy, t_2 and t_6 divide the remaining partitions in halves, and so on.

For an N -bit flash ADC, with an input cumulative distribution function (cdf) $F_1(v)$, we achieve output code histogram equalization if we assign the t_n values such that

$$F_1(t_n) - F_1(t_{n-1}) = 1/2^N, \quad (4.2)$$

for $n = 1, 2, \dots, 2^N$, where $t_0 \equiv -\infty$ and $t_{2^N} \equiv +\infty$. For sub-partitions (i.e., hierarchy 2 and higher) we determine suitable reference levels using an equal-partitioning algorithm based on the conditional cdf $F_1(t_n|A) = P[(X < t_n) \cap A]/P[A]$. Summing (4.2) over n , we have a single partition at hierarchy 1: $\sum_{n=1}^{2^{N-1}} (F_1(t_n) - F_1(t_{n-1})) = F_1(t_{2^{N-1}}) = 1/2$. Similarly, we have two partitions at hierarchy level 2: $F_1(t_{2^{N-2}}|X < t_{2^{N-1}}) = 1/2$ and $F_1(t_{2^{N-2}+2^{N-1}}|X > t_{2^{N-1}}) = 1/2$. In general, for hierarchy level l , the partitions resulting from the previous levels $1, 2, \dots, l-1$ are again partitioned into halves: $F_1(t_{2^{N-l}+(n-1)2^{N-l+1}}|t_{(n-1)2^{N-l+1}} < X < t_{n2^{N-l+1}}) = 1/2$, for $n = 1, 2, 3, \dots, 2^{l-1}$. In this way, an update direction for each reference level t_n is determined by monitoring the corresponding conditional cdf, determined from the appropriate subset of all reference levels. If t_n is too high, the conditional cdf is

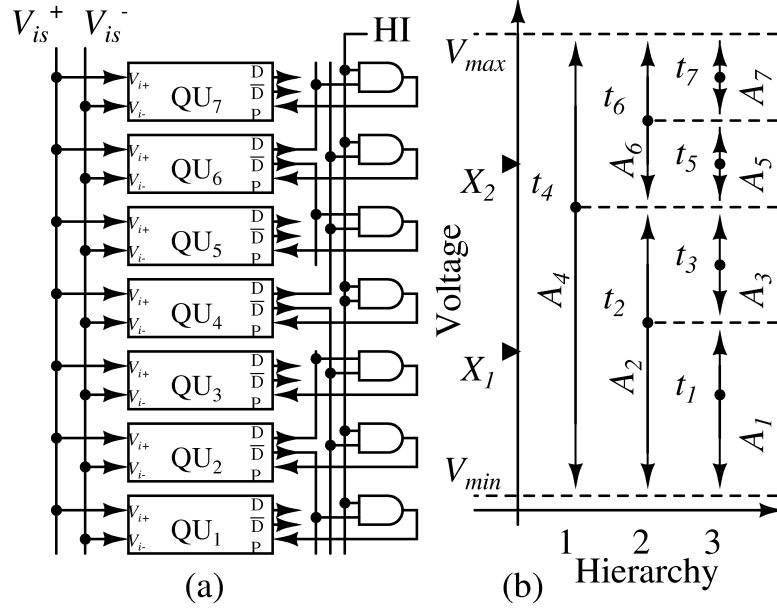


Figure 4.7: 3-bit flash AFGQ: (a) The equal-partition algorithm for histogram equalization is implemented with digital AND gates. (b) Partitions and reference levels for hierarchies 1-3.

greater than $1/2$, and if t_n is too low, the conditional cdf is less than $1/2$. Therefore the update law for the reference level t_n is specified by:

$$\Delta t_n = \alpha_n \text{Sign} \left(F_1(t_n | A_n) - \frac{1}{2} \right) \quad (4.3)$$

where A_n is the signal partition corresponding to reference level t_n and α_n is the magnitude of the update. Examples of partitions A_n are illustrated in Fig.4.7 (b) for the case of $N = 3$: at hierarchy 1, A_4 covers the entire input signal range. If more samples are observed above than below reference level t_4 , then t_4 will increase over time. The same considerations are repeated for hierarchies 2 and 3 in order to determine the sub-partitions corresponding to each t_n . The QU introduced in Sec.4.3.2 implements (4.3) for a given partition A_n , with α_n equal to the matched programming increments and decrements for tunneling and injection.

We simulated this algorithm for a 4-bit flash ADC; the 15 reference levels and output code entropy are plotted against sample number in Fig.4.8 (a) and (b), respectively. The entropy $H = -\sum p \log_2 p$ is a good indication of the flatness of the output histogram. For an equalized histogram, the entropy is equal to the number of bits [98]. The 15 reference levels were initialized to random values, simulating typical initial floating gate voltages after manufacturing. As the t_n 's gradually converge to their steady state positions, the entropy steadily rises to 4 bits. The t_n 's quickly track the input, which changes from uniform to Gaussian distribution at the 128000th sample and subsequently to exponential distribution at the 192000th sample. The uniform signal is distributed between +1 V and -1 V, the Gaussian signal has zero mean with $\sigma = 0.33$ V, and the exponential signal has mean 0.4 V and offset -1 V. We set the update increment (α_i) to 37.5 μ V per sample for partition hierarchy 1. Since hierarchy 2 is updated half as frequently as hierarchy 1, the adjustment increment for these levels is set to 75 μ V. Hierarchies 3 and 4 are set to 0.15 mV and 0.3 mV, respectively. For higher adjustments we observed faster convergence but coarser t_n values; for lower adjustments we observed slower convergence but finer values.

4.4.3 Implementing Histogram Equalization with QUs

We first describe how the static adaptation mechanism is adapted for time-varying signals, then we describe how equalization is achieved. From section 4.3.2, each QU is capable of adapting its reference level t_n to a DC value supplied at the

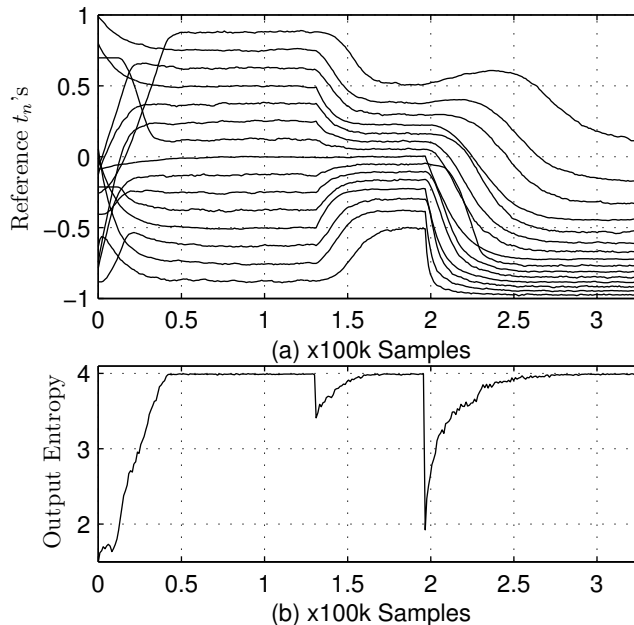


Figure 4.8: The reference levels t_n 's in a 4-bit flash ADC adapt as the input distribution changes from uniform, to Gaussian, to exponential.

input $V_i = V_i^+ - V_i^-$ by adjusting its internal offset V_c . From the law of large numbers, it is easy to show that when the input signal is a random variable X , the reference level t_n will be adapted to the mean of X (i.e., $t \rightarrow E[X]$), under the assumptions that X is stationary, the adaptation is carried out many times, and the increment Δt for each adaptation is small.

This intrinsic behavior of the QU directly implements the equal-partition algorithm for histogram equalization described in Sec.4.4.2. For the first hierarchy, updates to the reference level for the middle QU ensure that $F_1(t) = F_1(E[X]) = 1/2$. For subsequent hierarchies, the updates for the reference levels of the corresponding QUs are conditioned according to the partition in which a particular sample falls. Fig.4.7 depicts an example for a 3-bit AFGQ: an array of 7 QUs is shown, and the terminal P is used to enable or disable adaptation. P is computed locally within the

array based on the outcomes of nearby comparisons. Each QU performs adaptation only if the input signal falls within its corresponding partition.

Figure 4.7 (b) illustrates the partitions and reference levels for the 3-bit AFGQ. The reference levels for $QU_{1,2,\dots,7}$ are $t_{1,2,\dots,7}$, respectively, and $t_1 < t_2 < \dots < t_7$. At hierarchy 1, the partition A_4 covers the entire signal range and the corresponding reference level is that of the middle QU, t_4 . At hierarchy 2, there are two partitions, A_2 and A_6 : one below t_4 , with corresponding reference level t_2 ; and one above t_4 , with corresponding reference level t_6 . At hierarchy 3, there are 4 partitions A_1 , A_3 , A_5 and A_7 with partition boundaries defined by t_2 , t_4 and t_6 , and corresponding reference levels t_1 , t_3 , t_5 and t_7 . For an example in which the input value X_1 is less than t_2 , the sample falls within partitions A_1 , A_2 and A_4 . Thus, QU_4 and QU_2 output LO (i.e., D is LO and \bar{D} is HI). Following the signal paths in (a), we see that $QU_{1,2,4}$ perform adjustments and $QU_{3,5,6,7}$ are inactive. For another example in which the input value X_2 is between t_4 and t_6 , the sample falls within partitions A_4 , A_5 and A_6 . Thus, QU_4 is HI and QU_6 is LO, and we see that $QU_{4,5,6}$ perform adjustments and $QU_{1,2,3,7}$ are inactive. This simple circuit implementation realizes the equal-partition algorithm outlined in Sec.4.4.2.

4.5 Measurement and Results

The QFN package containing the 3 mm x 4.5 mm chip (Fig.4.9) was attached to a 4-layer PCB with two SMA connectors for the differential input and one SMA connector for the clock input. The PCB was attached to a thermoelectric plate, a

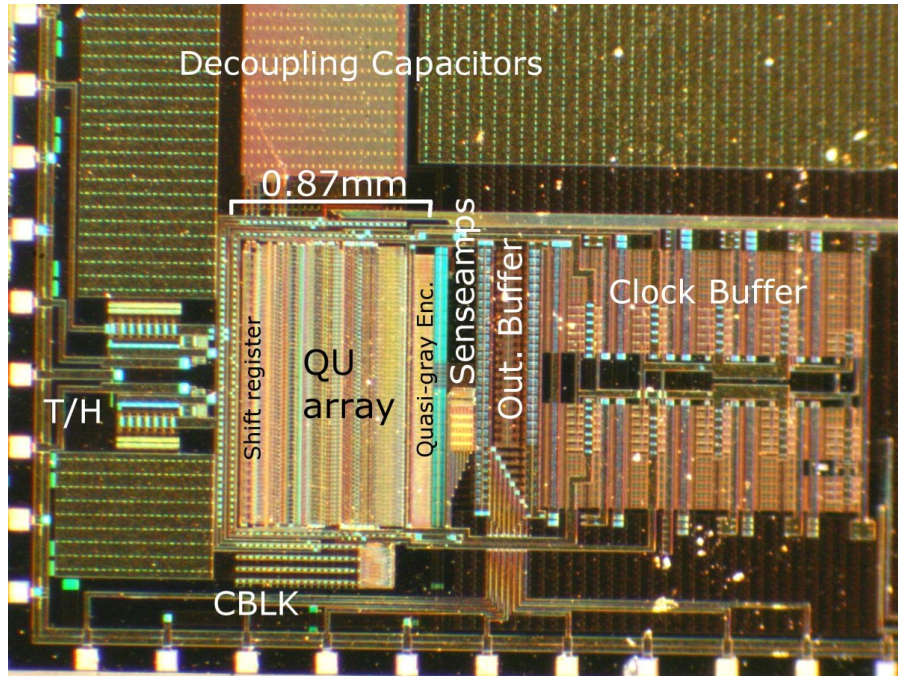


Figure 4.9: Photomicrograph showing both the AFGQ (left) and the on-chip large clock buffer (right).

heat sink and a fan. The chip surface temperature was maintained at 23°C during testing. An RF signal generator was used to drive a phase splitter, which generated and supplied the differential input signals to the PCB. A second RF signal generator supplied the clock signal as well as a 10MHz reference signal in order to synchronize the first signal generator to allow standard DSP-based coherent single tone analysis [99]. We used an 8-bit arbitrary waveform generator (AWG) to supply differential DC references to the AFGQ during manual programming, and a differential time-varying signal to the AFGQ during characterization. The AWG shared the same clock signal as the AFGQ via a power splitter, and the data from the AWG was synchronized with the sampling clock. We were able to supply arbitrary data at exactly half the sampling rate by repeating each data point twice. A logic ana-

lyzer (LA) captured the digital output from the AFGQ synchronously. A PC with data acquisition cards provided analog and digital interfaces to the PCB. A software interface controlled floating gate tunneling and injection increments, selected mode of operation, transferred data to the AWG and from the LA, and executed sequential manual programming and automatic adaptation. The software analyzed and decoded the captured data, performed Fast Fourier Transformation (FFT) for sine wave inputs, and re-aligned captured data with programmed data for arbitrary waveforms before calculating performance.

During adaptation, we used both the AWG and the signal generator to provide signals of known distribution for calibration. On-chip inverters occupy 43% of the active area, and generate non-overlapping digital clock signals from the sine wave clock input. Typical total power consumption is 1.1 W at 3.3 V supply and a sample rate of 750 MS/s, of which 60% is used in the clock buffer. The AFGQ is able to perform programming and signal conversion up to a sampling rate of 750 MS/s and 800 MS/s, respectively. We were unable to confirm operation above 800 MS/s due to LA bandwidth limitation.

4.5.1 Performance vs. Varying Input Frequency

Before capturing data, we programmed the QUs with matched injection and tunneling rates of 1.2 V/sec. We programmed the t_n 's manually for both linear and arc-sine values between 0.5V to -0.5 V using the AWG, then used a 1.9 dBm sine wave as input. We also performed adaptation using a 2 dBm sine wave at

each frequency. We decoded the data using Eqn.(4.1) and analyzed 16384 captured samples in the frequency domain using FFT. The sampling rate was 750 MS/s. Two plots of the spectrum for the linearly programmed AFGQ are shown in Fig.4.10(a), (b). The SNDR is the ratio of signal power at the fundamental frequency to the sum of all other power excluding DC. The SNDR is plotted versus input signal frequency in Fig.4.10(c) for the three cases. The sine power is adjusted to obtain full scale in the digital codes. As the frequency of the input sine wave increased from 24 MHz to 387 MHz, we observed an increase of 11.8 dB in HD2.

Code histograms are obtained from the digital codes by counting code occurrences. We determined code histograms, DNL and INL for the three cases at 750 MS/s near Nyquist rate. The histogram for the linearly programmed result gives DNL and INL of less than 0.27 LSB (Fig.4.10d) when normalized using an ideal sine histogram. The flat histograms for the arc-sine and adaptation results confirm histogram equalization. The histogram for the arc-sine programmed result (plotted as crosses) gives peak DNL and INL of 1.93 LSB and 1.38 LSB, respectively (Fig.4.10e). The DNL and INL near the center is small. For adaptation, the histogram gives a peak DNL of 0.83 LSB and INL of 2.22 LSB (Fig.4.10f).

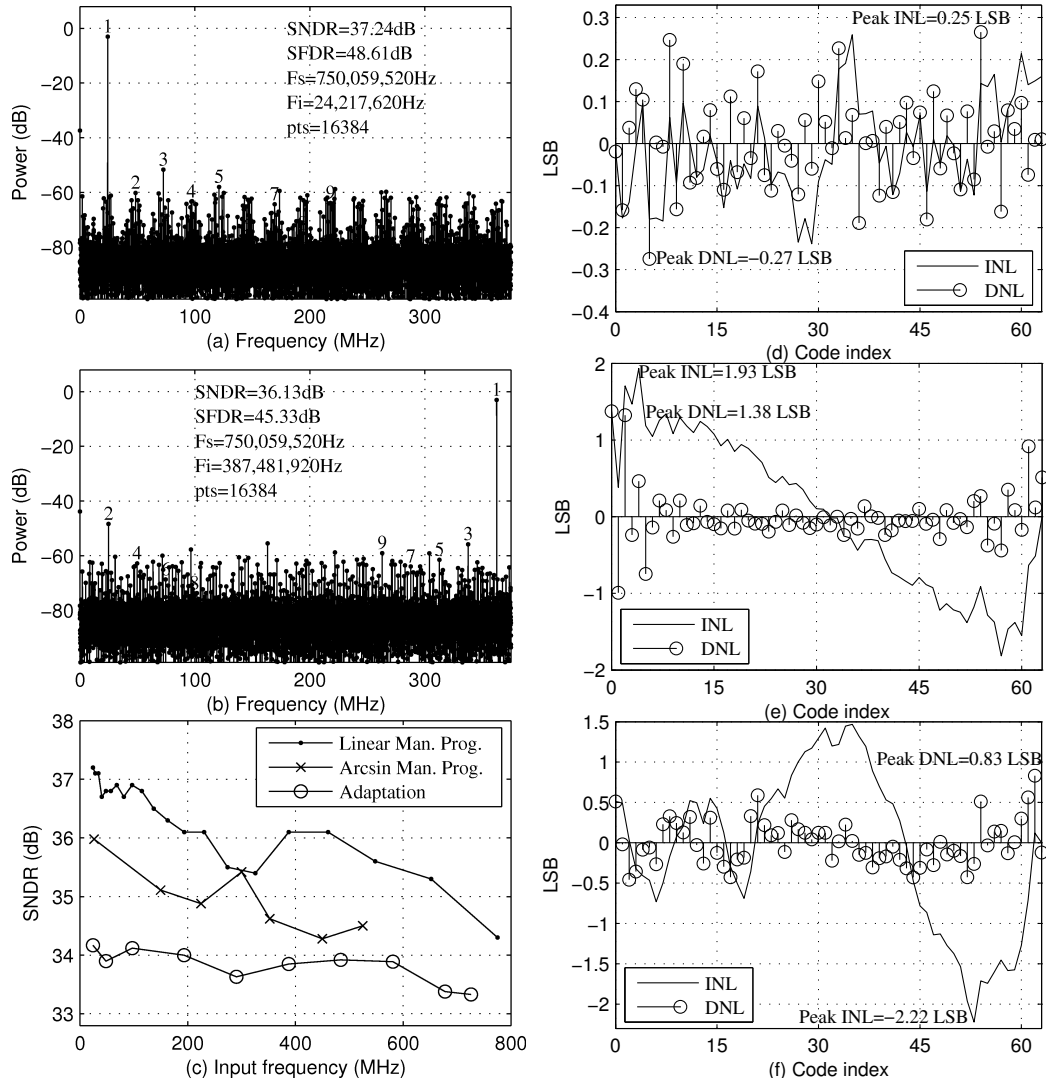


Figure 4.10: Performance results for AFGQ: (a), (b) FFT plot computed from 16k digital output samples and (c) SNDR vs. input sine frequency. The code histograms for the results with linearly programmed, arc-sine programmed and autonomously adapted reference level are used to compute non-linearity in (d), (e) and (f), respectively.

4.5.2 Performance vs. Varying Input Amplitude

An RF signal generator was used to supply the AFGQ with sine waves having V_{pp} ranging from 50 mV to 2 V at Nyquist rate with a sampling frequency of 750 MHz. Before capturing data, the reference levels were matched to the input signal by manually programming the t_n 's linearly with $t_1 = -V_{pp}/2$ and $t_{63} = +V_{pp}/2$. The resulting SNDR is shown as a function of V_{pp} in Fig.4.11(a). The SNDR reaches a maximum value for a $V_{pp} \approx 1$ V. Below 1 V the distortion caused by comparator residual offset and temporal noise dominates performance.

We emulate sensor data by supplying a 32768-point Gaussian random waveform with zero mean and a standard deviation $\sigma_S = V_{pp}/6$ at a data rate of 87.5 MS/s with a sampling frequency of 700 MHz. We varied V_{pp} from 44 mV to 1 V and analyzed performance. For each input amplitude, we captured the data with the reference levels (a) linearly programmed to full-scale (LPF, $t_1 = -0.5$ V and $t_{63} = +0.5$ V), (b) linearly programmed to match the signal (LPS, $t_1 = -V_{pp}/2$ and $t_{63} = +V_{pp}/2$), (c) Gaussian programmed to match the signal (GPS), and (d) adapted to the input signal (AS). For (a)–(c), we used the corresponding calculated $T(n)$ to decode digital data. For (d), we drove the AFGQ with a ramp signal after adaptation to obtain $F_2(n)$ in Eqn.(4.1). Since $F_1(v)$ is linear (a ramp signal), we were able to calculate $T(n)$ using Eqn.(4.1). Note that this procedure removes the transfer function distortion that causes large INL as shown in Fig.4.10(f).

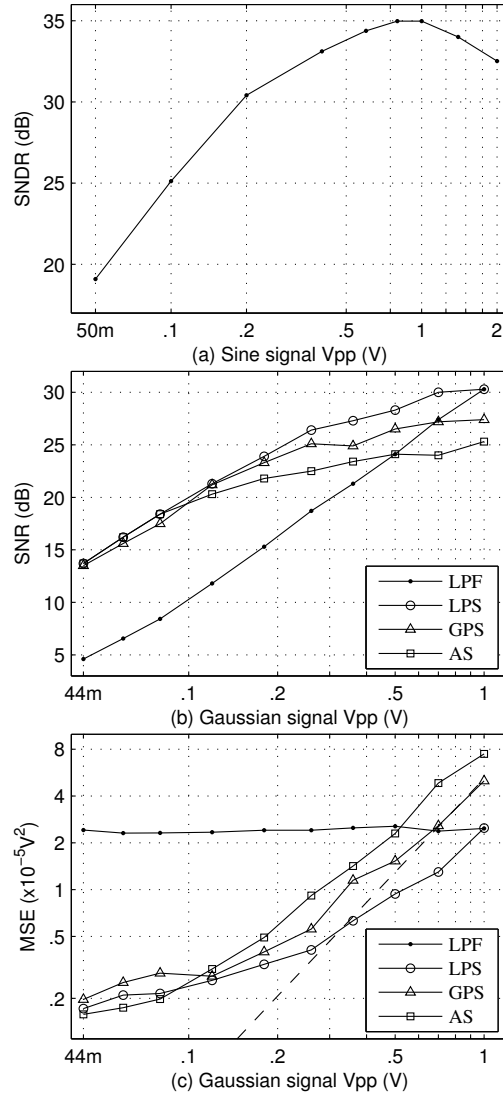


Figure 4.11: (a) The ADC achieves the best performance when the signal has 1V V_{pp} . (b) Conversion with reference levels fit to signal (LPS) outperforms fixed reference levels (LPF). (c) The MSE for conversions with adaptive reference levels scale with signal power for $V_{pp} > 0.3V$. The dashed line is a quadratic curve fit to GPS results.

We computed the MSE as $\sigma_N^2 = E[|x - \hat{x}|^2]$, where x represents the programmed data points and \hat{x} represents the decoded captured data points. The mean was taken over all samples. We express the signal-to-noise ratio (SNR) as $10 \log_{10}(\sigma_S^2/\sigma_N^2)$. As seen in Fig.4.11(b), the linear-fit result (LPS) gives 2 dB more SNR than the Gaussian-fit (GPS), which gives 2 dB more than adaptation (AS) at higher signal amplitudes. All of the above mappings outperform the LPF result at lower signal amplitudes. Fig.4.11(c) shows the MSE for all cases. The MSE for $V_{pp} > 0.3$ V is dominated by quantization noise, and is proportional to signal power (V_{pp}^2). The dashed line is a quadratic fit to the GPS curve. For $V_{pp} < 0.3$ V the deviation from the dashed line is more pronounced, as distortion caused by comparator residual offset and temporal noise sets a lower limit for the MSE.

We anticipate that the performance of the nonlinearly programmed and autonomously programmed reference levels would improve with error correction. Nonlinearly distributed reference levels are more likely to disrupt monotonicity in the comparators than linearly distributed reference levels. Error correction has not been implemented in this prototype, and we believe that such errors are the primary reason for the reduced performance in the nonlinear programmed reference levels in Fig.4.10 and Fig.4.11. This leads to the possibility of using gray-code error detection for restoring monotonicity of the reference levels through adaptation.

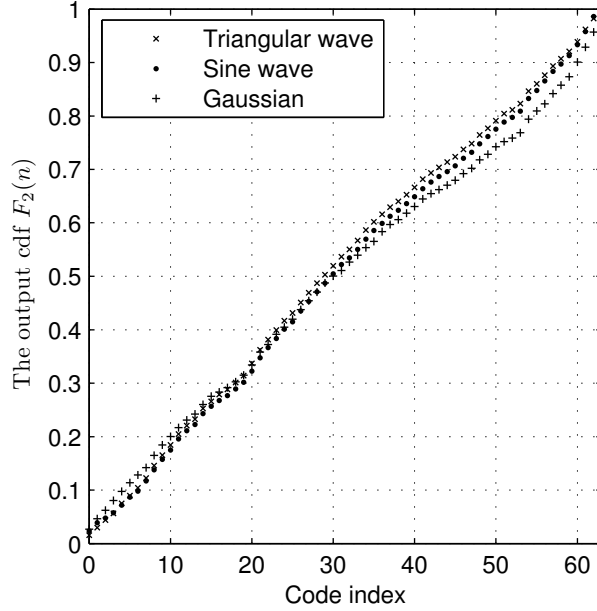


Figure 4.12: Although output cdf deviates slightly from an ideal uniform cdf with residue norm of 0.13, the output cdf for three very different inputs are nearly identical.

4.5.3 Performance vs. Signal Types

Next, we perform adaptation with 3 different signals at input: a) 1.54MHz, 1V peak-to-peak triangular wave, b) 3dBm 387MHz sine wave and c) a Gaussian random signal with zero mean, $\sigma = 166\text{mV}$ at a data rate of 375MS/s. a) and c) are generated using the AWG. We normalize the histogram to obtain pmf and subsequently cdf. The resulting cdf are plotted in Fig.4.12. The slight deviation from ideal cdf can be attributed to offset in the op-amp, mismatch of the injection and tunneling currents, and mismatch in charge pumps in each AFGC. The calculated entropies are 5.96, 5.95, and 5.90 bits, and the maximum DNL are 0.62, 0.83 and 1.75 LSB for a), b) and c), respectively. For comparison, we programmed the ADC manually with uniformly distributed t_n 's and observed entropy of 5.99, 5.73, and

5.34 bits for a), b) and c), respectively. Adaptation to periodic signals as in a) and b) requires careful selection of the input frequency such that sufficiently distinct values are sampled for adaptation. We have chosen an input frequency such that 256 distinct values in each period of the signal are used for adaptation.

4.5.4 Data Retention and Temperature

We monitored continuous operation for one month, and observed sporadic bit errors with an error rate of 2.93×10^{-9} /sample with no sign of amplitude drift, offset drift or SNDR degradation. We monitored bit error rate by sampling a low frequency 0.97 MHz full scale sine wave at 750 MS/s. From a captured sequence of 2^{20} samples, we counted each instance of two or more LSB changes in consecutive code transitions as a single error. The AFGQ was operated continuously, and each measurement was repeated approximately every 40 seconds. Figure 4.13 plots cumulative error samples vs. captured samples. The error curve follows a straight line with a slope of 2.93×10^{-9} /sample. We observed at most 2 errors in every captured sequence.

We programmed a chip at 20°C and measured SNDR at $f_{signal} = 387$ MHz, $f_{sample} = 750$ MHz. The SNDR dropped from 36.47 dB to 36.02 dB when the temperature was raised to 30.6°C. The SNDR rises to 36.49 dB when we cooled the chip to 21.5°C.

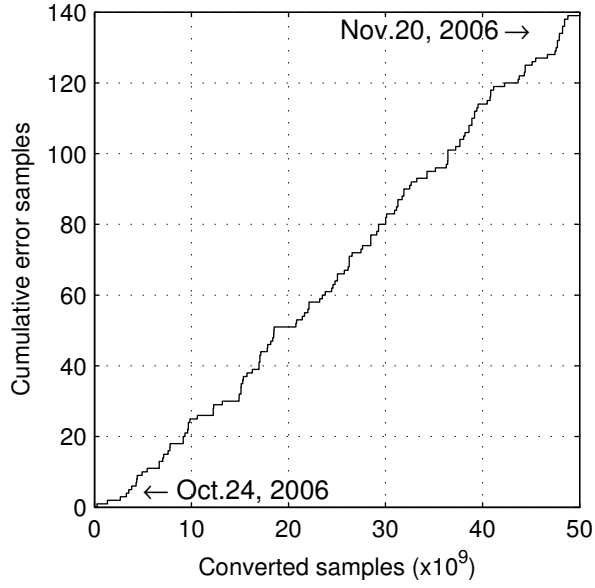


Figure 4.13: The AFGQ maintained an error rate of 2.93×10^{-9} /sample during one month of continuous operation.

4.5.5 Performance Summary

Table 4.1 summarizes performance for the AFGQ prototype. Sampling speed is limited by a critical path in the control logic that has a fanout of 63 and drives all QUs during adaptation, and thus adaptation is not feasible at frequencies above 750 MS/s. Future improvements will include using the foundry native design rules as opposed to MOSIS scalable CMOS design rules [80], reducing clock buffering power consumption with a more efficient clocking scheme, optimizing logic blocks, reducing QU input capacitance, adopting a more robust thermometer encoder, employing digital correction, and migrating to a smaller feature size process. Table 4.2 compares the AFGQ performance with that of other state-of-the-art 6-bit ADCs.

Table 4.1: Performance Summary

Process	0.35 μm 2P4M CMOS
Sampling Rate	750 MS/s
Input Range	Programmable
INL / DNL	< 0.27 LSB
SFDR	48.6 dB at 24 MHz
SNDR	37.2 dB at 24 MHz 36.1 dB at 387 MHz
Bit error rate	2.93×10^{-9}
Chip/Active Area	13.5 mm ² /2.9 mm ²
Core/Total Power	0.4 W / 1.1 W
Supply Voltage	3.3 V

4.6 Summary

A flash ADC architecture was demonstrated using nonvolatile storage of reference levels and on-chip adaptation of reference levels for histogram equalization. The AFGQ realizes 36.1 dB SNDR at Nyquist rate and 37.2 dB SNDR at lower input frequency sampled at 750 MS/s in standard 0.35 μm technology. The design methodology for adaptation is independent of operating speed and can be used to improve the performance of CMOS comparators and mixed-signal circuits in general. Arbitrary comparator reference levels are conveniently programmed on-chip with good data retention. For input signals near Nyquist rate, lower distortion was observed when performing AD conversion with reference levels that match the

Table 4.2: 6-bit ADC Comparison

Ref.	Year	Architecture	Technology	Vdd	Input V_{pp}	Sample Rate	Best reported SNDR at f_{in}
[100]	2004	Interleaved SAR	90 nm	1 V	-	0.6 GS/s	34 dB at 30 MHz 31 dB at 329 MHz
[101]	2005	Interleaved flash	0.18 μm	1.8 V	1 V	2 GS/s	36 dB at 4 MHz 30 dB at 921 MHz
[84]	2007	Interleaved pipelined	0.18 μm	1.8 V	0.4 V	0.8 GS/s	33.7 dB at 100 MHz 31.5 dB at 400 MHz
[102]	2005	Flash	0.13 μm	1.5 V	-	1.2 GS/s	35.8 dB at 51 MHz 32.8 dB at 700 MHz
[103]	2002	Flash	0.18 μm	1.95 V	-	1.6 GS/s	36 dB at 263 MHz 31.9 dB at 660 MHz
[104]	2003	Flash	0.25 μm	1.8 V	-	1.3 GS/s	33.2 dB at 133 kHz 32 dB at 500 MHz
[87]	2001	Flash	0.35 μm	3.3 V	1.5 V	0.9 GS/s	35.7 dB at 30 MHz 32.7 dB at 450 MHz
[7]	2001	Flash	0.35 μm	3.3 V	1.6 V	1.0 GS/s	36 dB at 100 MHz 34.8 dB at 630 MHz
This work	2007	Flash	0.35 μm	3.3 V	50mV-2V	0.75 GS/s	37.2 dB at 24 MHz 36.1 dB at 387 MHz

signal amplitude. Autonomous learning of signal amplitude statistics is directly implemented using the self-adapting characteristics of the AFGC, and used to determine reference levels that achieve equalized output code probabilities. This approach eliminates the necessity for trimming and calibration after fabrication, and also provides the capability to autonomously optimize the ADC characteristics for time-varying input signals.

Chapter 5

The Adaptive Floating Gate Imager (AFGI)

5.1 Introduction

Image sensors are transducers that convert optical images into electrical signals. Fabrication process variations cause circuit mismatch that creates unwanted artifacts in the image and compromises the maximum dynamic range of an imager. Mismatch between identical transistors in CMOS VLSI occurs both randomly and deterministically; sources of deterministic mismatch include “edge”, “striation”, and “gradient” effects [105]. The deterministic variations result in non-temporal spatial noise across the array of pixels, known as fixed pattern noise (FPN). Both deterministic and random variations impose severe limitations on the dynamic range and picture quality of CMOS imagers.

A common approach used in active pixel imagers is to cancel offset with multiple sampling techniques such as correlated double sampling (CDS) [106] or double delta sampling (DDS) [107]. Such techniques usually produce satisfactory results for integrating-type imagers that perform the relatively simple job of directly transducing an optical scene. However, CDS is difficult to implement in current-mode continuous-time imagers that offer wide dynamic range or for smart sensors that perform sophisticated computation on the image plane such as motion detection, edge enhancement, or feature extraction [108–110]. Massively parallel high dynamic

range image plane computation is most compactly implemented by current-mode continuous-time image sensors. This paper reports an offset cancellation technique that is compatible with high density image plane computation, and is expected to improve the accuracy and dynamic range of such image plane processors.

Since FPN is a static characteristic of each pixel comprising an imager, it seems natural to reduce it by using nonvolatile analog storage of a fixed charge on a floating gate in each pixel. Floating gate techniques have long been used for adaptation and calibration purposes. They have been used to correct non-uniformity in imagers [27–31, 33]. In [30, 33] a comparator and dedicated programming logic were used to detect and control correction. In [33], a randomly selected pixel readout was compared to the previously selected pixel in order to determine the local update direction. This method achieved both FPN non-uniformity correction and intensity histogram equalization for a large number of iterations.

We describe a new five-transistor pixel circuit that eliminates the need for any additional supporting circuitry for automatic adaptation by exploiting local feedback inside the signal path, and thus enabling parallel adaptation of each pixel to a desired common voltage given arbitrary incident light patterns. The adaptation extends naturally beyond FPN cancellation to correct optical distortions in intensity by modeling non-uniformities in the incident light intensity as a form of offset. We experimentally confirm the ability to reduce FPN variance to that of the temporal noise. Since the local feedback mechanism is engaged for all pixels in parallel by simply raising the power supply voltage, adaptation is fast and accurate. A brief report of a similar image sensor was presented at the IEEE International Symposium

Table 5.1: Adaptive Floating Gate Imagers

Ref	Sensor	Array Size	Sensitivity	Calibration Method	FPN Reduction
[32]	Photodiode	26x1	50mV/decade	AFGA	Resolution: 4.4 bit to 8.5 bit
[27]	Photodiode	4	-	Tun.,In-pixel,Parallel	-
[28]	Photodiode	8x9	-	Tun.,In-pixel,Parallel	Δ Pulse width: 2.5ms to 0.5ms
[29]	Floating bulk	-	1.2V/decade	Tun.,Ext-pixel, -	-
[30]	Floating bulk	128x1	2V/decade	Tun.,Ext-pixel,Sequential	Vout FPN: 0.8V to 0.1V at 1W/m ²
[31, 33]	Vertical PNP	64x64	-	Inj.,Ext-pixel,Sequential	$\frac{\sigma I_{out}}{I_{out}}$: 70% to 10%
[10]	Vertical PNP	128x128	0.08V/decade	Inj.,In-pixel,Parallel	σV_{out} : 24mV to 4.8mV at 0.6W/m ²
This Work	Photodiode	144x144	0.11V/decade	Inj.,In-pixel, Parallel	σV_{out} :16.2mV to 1.37mV at 0.61W/m ²

on Circuits and Systems (ISCAS) in 2005, and appears in [10]. Since then, the circuit and sensor have been modified significantly to achieve superior performance and lower power consumption. The new design has been fabricated and characterized, and we report these results here.

The remainder of this paper is organized as follows: in section 5.2 we present our design approach and discuss background material. In section 5.3 we describe the design of the adaptive floating gate pixel (AFGP) and its adaptation method. In section 5.4 we describe experimental results from fabricated chips and their interpretation. In section 5.5 we illustrate novel applications of the new imager. Finally, section 5.6 summarizes the work.

5.2 Background

The adaptive imager presented here was inspired by the floating gate imager previously developed by Cohen and Cauwenberghs [31, 33]. In that imager, some components of the adaptation mechanism were implemented off-chip, and pixels were updated one at a time over many iterations. In the novel architecture reported

here, the adaptation mechanism is in-pixel and all pixels are updated in parallel. In both circuits, the update is stored in a floating gate current mirror within each pixel. Two earlier architectures reported offset cancellation using nonvolatile floating gate storage, with adaptation accomplished using tunneling mechanisms. However, these designs used non-conventional photosensors (in [27, 28], with binary output; in [29, 30], with a floating bulk for a pFET) with large pixel sizes and did not produce high density image sensor arrays. An earlier imager uses an autozeroing floating gate amplifier (AFGA) in the signal path for filtering out DC signal (including FPN) with a widely tunable time constant [32]. Table 5.1 shows a brief comparison of reported floating gate image sensors with offset cancellation.

5.2.1 Mismatch in the Photodiode

Incident photons deliver energy to electrons in a semiconductor, causing electrons to be excited into the conduction band from the valence band and leaving behind empty states, or holes, in the valence band. Photo-generated electron-hole pairs produce intrinsic photocurrent in a depletion region where the built-in electric field serves to separate and collect the carriers. Here we analyze the mismatch in current density. Photocurrent density across an illuminated depletion region is

$$J = -q\Phi_0 \left(1 - \frac{e^{-\alpha W}}{1 + \alpha L} \right) - J_s$$

where Φ_0 is the flux of photons per unit area, α is the optical absorption coefficient, W is the depth of the depletion region, L is the minority carrier diffusion length and J_s is the dark current [111]. W depends on reverse bias voltage and W , α , L

and J_s depend on doping concentrations.

The AFGP uses an n^+ active- p^- substrate photodiode. The photo current is $I_p = AJ$, where the area A is sensitive to the fabrication process.

Parameters that depend on doping concentration and geometry are susceptible to mismatch induced during fabrication. By grouping terms, we obtain $I_p = \kappa \cdot \Phi_0 - J_s A \approx \kappa \cdot \Phi_0$ where $\kappa = -qA \left(1 - \frac{e^{-\alpha W}}{1 + \alpha L}\right)$. The photocurrent I_p is approximately proportional to the photon flux Φ_0 with a poorly-controlled gain κ that varies from diode to diode. This relationship is valid when the photocurrent is much larger than the dark current $J_s A$. In the remaining discussion, we explicitly model mismatch as

$$I_p = \alpha_p C_0 \Phi_0, \quad (5.1)$$

where C_0 is the nominal value for gain κ and is assumed to be the same for all pixels. α_p is the gain mismatch among individual photodiodes, with a mean value of 1.

5.2.2 Mismatch in Subthreshold MOSFET

Channel current for a MOSFET operating in subthreshold is an exponential function of terminal voltages:

$$I_D = \frac{W}{L} I_0 e^{\frac{V_G}{nU_T}} \left(e^{-\frac{V_S}{U_T}} - e^{-\frac{V_D}{U_T}} \right) \quad (5.2)$$

where V_G , V_D and V_S are gate, drain and source voltages, respectively, U_T is the thermal voltage $U_T = kT/q$, $\frac{W}{L}$ is the width to length ratio, I_0 the characteristic current, and n the slope factor. For $V_S = 0$ and $V_D > 4U_T \approx 100\text{mV}$, (5.2) can be approximated by $I_D \approx \frac{W}{L} I_0 e^{\frac{V_G}{nU_T}}$.

Fig.5.1(a) shows the circuit implementation of the AFGP. The AFGP directly transduces photocurrent as a continuous-time waveform. This is appropriate for subsequent integration with high density current-mode computation for massively parallel image plane processing. Photodiode D_1 is exposed to incident light and produces a photocurrent I_p at its emitter. This photocurrent is translated into voltage logarithmically by diode connected pFET transistor M_1 . In contrast to integrating voltage mode pixels commonly used in CMOS imaging [113, 114] that exhibit a linear relationship between incident light intensity and pixel output voltage, the AFGP produces a continuous-time output voltage that is logarithmic with the incident light intensity. Neglecting parasitic capacitances on M_2 , we see that M_1 and M_2 form a “floating current mirror” such that $I_2 = I_p * f_1(V_{C_1})$, where V_{C_1} is the voltage drop across capacitor C_1 ($\sim 49.7\text{fF}$) between V_A and the floating node V_B , and f_1 is exponential in V_{C_1} . In addition to the capacitor C_1 , the floating node is capacitively coupled to a globally connected node V_E , through a much smaller capacitance C_2 ($\sim 1.5\text{fF}$). The global node V_E provides an external control to the floating node and is especially useful for compensating the common mode shifts of the floating node voltages induced by injection. The mirrored current is then translated into voltage V_D by a current conveyor composed of transistors M_3 and M_4 and current I_c , where I_c is a strong bias current that increases the driving strength for fast column readout. The current conveyors for a row of pixels can be turned on and off by controlling M_5 with a bias voltage common to transistors in the row. During row activation the column voltage settles to V_D . Off-chip A/D converters convert the analog column voltage V_D into digital form for acquisition by

a PC or microcontroller.

5.3.2 Floating Gate Offset Compensation

We analyze the AFGP circuit to find the charge q that should be stored on the floating gate to compensate for mismatch between pixels. We define a constant $I' = \frac{W}{L}I_0$ and variables $V_a = V_{dd} - V_A$, $V_b = V_{dd} - V_B$ for convenience. The channel currents for M_1 , M_2 and M_3 are

$$I_p = I' \exp [(V_a + \Delta V_1)/n_p U_T] \quad (5.3)$$

$$I_2 = I' \exp [(V_b + \Delta V_2)/n_p U_T] \quad (5.4)$$

$$I_2 = I' \exp [(V_D + \Delta V_3)/n_n U_T] \quad (5.5)$$

respectively, neglecting body effect and Early effect and assuming that M_4 and M_5 are biased such that M_3 is effectively diode-connected. The differences in individual transistor geometry and characteristic current I_0 for both p- and n-type MOSFETs are incorporated into mismatch quantities ΔV_1 , ΔV_2 and ΔV_3 ; I' is a mismatch-free quantity that is consistent among all transistors. n_p and n_n represent the slope factors for p- and n-type MOSFETs. Next, we express the floating gate voltage that results from charge-sharing as

$$V_b = \lambda_1 V_a + V_0 \quad (5.6)$$

where λ_1 is the ratio of C_1 to the total capacitance C_T on the floating node, and V_0 is the voltage from the charge q stored on the floating gate, $V_0 = q/C_T$.¹

¹We assume that all parasitic capacitances are connected to fixed voltages.

We relate the pixel output voltage V_D to the photocurrent and mismatch terms by combining (5.3), (5.4) and (5.5). Simplifying, we obtain:

$$V_D = n_n U_T \ln \left(\frac{I_p}{I'} \right) + \frac{n_n}{n_p} \left(V_b - V_a + \Delta V_2 - \Delta V_1 - \frac{n_p}{n_n} \Delta V_3 \right)$$

We further relate the pixel voltage V_D to the floating gate charge and photon flux by using (5.6) and substituting (5.1) to obtain $V_D = n_n U_T \ln \left(\frac{C_0 \Phi_0}{I'} \right) + n_n U_T \ln \alpha_p + \frac{n_n}{n_p} \left\{ V_0 - [\bar{\lambda}_1 n_p U_T \ln \left(\frac{C_0 \Phi_0}{I'} \right) + \bar{\lambda}_1 n_p U_T \ln \alpha_p - \bar{\lambda}_1 \Delta V_1] + \Delta V_2 - \Delta V_1 - \frac{n_p}{n_n} \Delta V_3 \right\}$, where $\bar{\lambda}_1 = 1 - \lambda_1$.

By setting $V_0 = \lambda_1 \Delta V_1 - \Delta V_2 + \frac{n_p}{n_n} \Delta V_3 - \lambda_1 n_p U_T \ln \alpha_p$, we obtain

$$V_D = \lambda_1 n_n U_T \ln \left(\frac{C_0 \Phi_0}{I'} \right) \quad (5.7)$$

The pixel voltage V_D is logarithmic in the photon flux Φ_0 . Offset contributions from D_1 , M_1 , M_2 and M_3 are entirely eliminated. Uniform V_D across the entire pixel array can be achieved for uniform temperature distribution across the chip, and conversely temperature nonuniformities will contribute apparent FPN. Prior work has shown that floating gate calibration is best when performed at the desired operating temperature [6], as was the case for the data shown in Sec. IV.

5.3.3 Adaptation Method

Each pixel automatically adapts to cancel its unique offset value by exploiting the negative feedback property of pFET hot-electron injection with a bias current on

the drain. During adaptation, we use incident light of uniform intensity to illuminate the imager, but I_p , V_A , V_B , I_2 and V_D still differ from pixel to pixel. Each pixel adapts by injecting appropriate charge onto its floating gate so that all pixel output voltages approach a desired constant voltage V_D^* .

Fig.5.1(b) shows the mechanism for self-regulated pFET hot electron injection used in the AFGP. The drain of the floating gate transistor M_2 is connected to a current source I_3 implemented by M_3 with a gate bias voltage V_D^* . I_3 is also a source of inter-pixel mismatch. Recall that I_2 is produced by the path from D_1 , M_1 , C_1 to M_2 , and that the mismatch of these transistors and the voltage stored on C_1 are responsible for inter-pixel variations in I_2 . We apply a large enough bias V_D^* to M_3 such that for every pixel $I_2 < I_{3(sat)}$, where $I_{3(sat)}$ denotes the channel current of M_3 in the saturation region. Thus for every pixel M_3 enters the triode region, and the source-to-drain voltage V_{SD} on M_2 is approximately equal to the power supply Vdd. A normal operating Vdd is chosen such that the lateral electric field across the channel E_L is insufficient for hot electron injection.

During adaptation, we increase the power supply voltage Vdd to enable injection. As electrons are injected onto the floating gate, the floating gate voltage decreases at a rate proportional to the injection current and inversely proportional to the total capacitance on the floating node. The decreased gate voltage increases I_2 , pulling V_3 higher. Eventually M_3 enters the saturation region when I_2 approaches $I_{3(sat)}$, and V_3 begins to rise rapidly. The rapid decrease in V_{SD} on M_2 turns off injection. This intrinsic feedback loop leads to self current calibration for each pixel independently. The calibration time must be set to the maximum time required for

any pixel, and pixels with less mismatch will automatically turn off injection early.

The transition from operation to adaptation is simple. In Fig.5.1(a), M_3 forms part of a current conveyor during normal operation, and in Fig.5.1(b) M_3 is the current source that provides I_3 during adaptation. For global adaptation, all conveyor switches M_5 are turned off, and all columns are driven with global voltage V_D^* , where V_D^* is set to the maximum voltage among all measured pixels to ensure that initially all pixels have $I_2 < I_{3(sat)}$. With the conveyor off, the AFGP enters adaptation mode as in Fig.5.1(b). Next, Vdd is raised and hot electron injection proceeds until a steady state is reached where $I_2 = I_{3(sat)}$ for all pixels. Because this current calibration loop encompasses the entire pixel, we compensate the offsets due to D_1 , M_1 , M_2 and M_3 simultaneously in each pixel. It is easy to extend the calibration loop to include incident photon flux Φ_0 mismatch. By doing so, we compensate any intensity distortion in the optical path that can be modeled by a mismatch gain coefficient for Φ_0 . We demonstrate this by performing vignetting correction in Sec. V.

Exceeding the recommended power supply shortens lifetime and is a method for accelerating aging in common endurance testing [2]. The technology used to fabricate this circuit has a nominal power supply voltage of 5V. We observed sufficient injection for this application at power supplies as low as 5.3V, the nominal adaptation power supply voltage used in all experiments reported here. The increased power supply voltage is applied only for a short period of time during adaptation with very small resulting current, and is expected to have minimal effect on overall life expectancy.

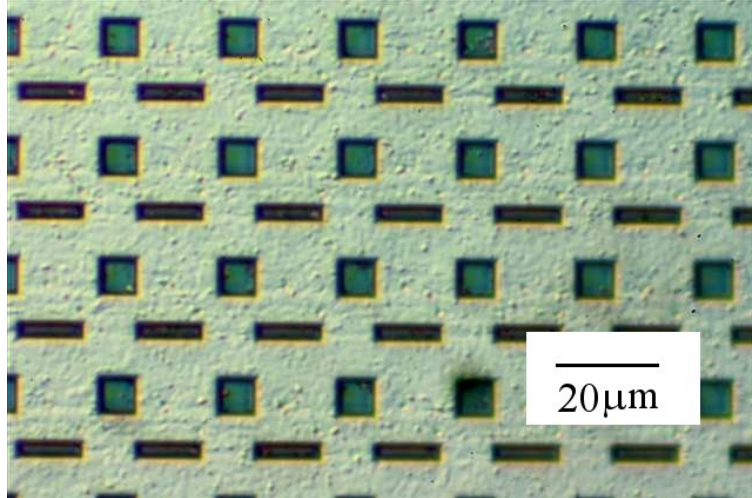


Figure 5.2: Metal-3 windows expose the photodiodes and part of the floating gate in the square and rectangular windows, respectively.

5.3.4 Layout

The pixel pitch is $18\mu\text{m} \times 18\mu\text{m}$. The fill factor is 13.4%. The chip was fabricated in a commercially available 2-poly, 3-metal $0.5\mu\text{m}$ nWell CMOS technology. The design including pad frame occupies $3\text{mm} \times 3\text{mm}$ silicon area. The sensor array is covered by a metal-3 shield with openings for each phototransistor and floating gate. Figure 5.2 is a microphotograph showing the top metal-3 layer and its openings. The square openings expose the photodiodes and the rectangular openings expose part of the floating gate material for UV erasure. The floating node consists of the top plate of a poly-poly2 capacitor, the poly gate of a pFET, and the metal that connects them. This arrangement minimizes stray capacitances to ground. On top of the poly2 there is a layer of metal-1 which serves as the global node V_E .

5.3.5 Chip Architecture

Figure 5.3(a) shows the rows and columns of the imager with connections to supporting circuitry. We use serial-in-parallel-out (SIPO) shift registers for activating rows and columns. The user activates a row by first flushing the row SIPO with logic LO (i.e., setting row data $D_R=LO$ and pulsing row clock CLK_R many times) and then selects the first row (i.e., by driving D_R with logic HI and pulsing CLK_R once). The row selection is propagated to the next row by setting $D_R=LO$ and pulsing CLK_R again. The row clock is pulsed repeatedly (144 times) to scan through all rows. The row SIPO has 144 digital outputs and connects directly to the pixel row input.

Each column line is connected to a set of transistors M_7 , M_8 and M_9 . During column activation, M_7 provides the column current I_c . The switches M_8 and M_9 are turned on, passing the column voltage to a readout pin V_{COLa} .

The 144 column lines are organized into 18 groups of 8 lines (Fig.5.3b). The 8 column lines in each group shares one readout pin V_{COLx} , and a total of 18 pins (V_{COLa} to V_{COLr}) are measured simultaneously. Each line in a group is selected by the 8-bit column SIPO (COLSIPO) sequentially. Figure 5.3(b) shows only M_9 and omits M_7 and M_8 for each row.

During adaptation, all row lines are set to LO, and all bits in COLSIPO as well as V_{CG} are set to HI. Switches M_7 and M_8 for all columns are turned on, connecting all column lines to V_{CS} , which can be set to the desired voltage V_D^* for global adaptation.

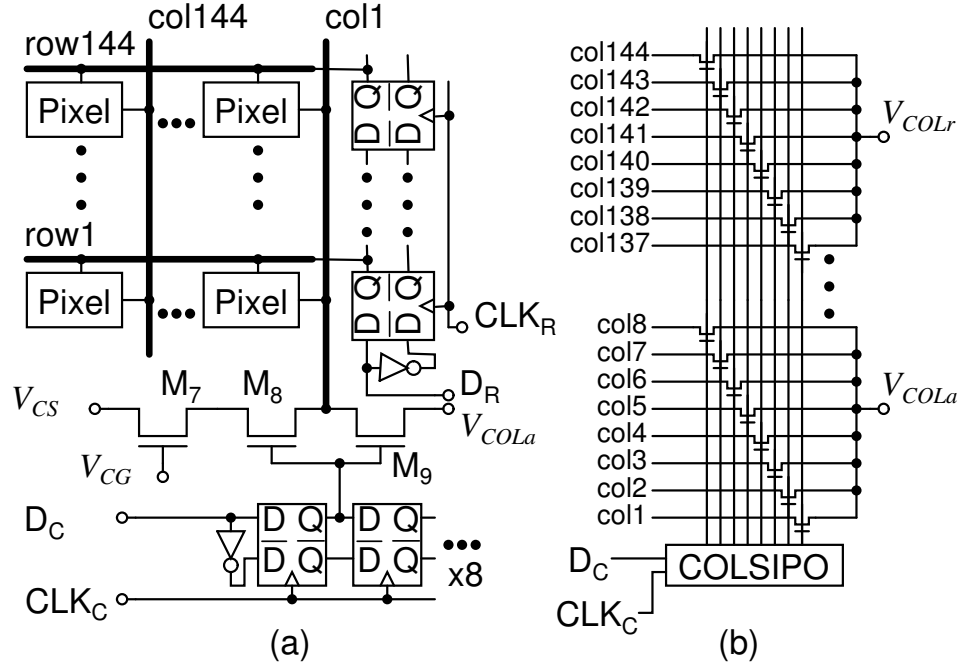


Figure 5.3: (a) Pixels are activated by row and column shift registers. (b) Columns lines are organized into 18 groups for rapid readout.

5.4 Experimental Results

To supply optical inputs, the chip was aligned on an optical rail with either uniform or patterned optical sources. We used a halogen light source and an integrating sphere to supply uniform light intensity directly onto the die surface for calibration and fixed-pattern/temporal noise measurement. We introduce neutral density filters into the optical path in order to obtain measurements at different intensities. We denote the unfiltered source intensity as $L \approx 61\text{W/m}^2$ at the chip surface. The imager chip is mounted on a 4-layer PC board that is shielded electrically and optically. The column line voltage is buffered by discrete surface mount precision op-amps (AD8574, Analog Devices, Inc.). They drive the 16-bit A/D converter in a personal computer based data acquisition card (MCC PCI-DAS6031),

with 2V range, $31\mu\text{V}$ resolution, $24\mu\text{V}$ root-mean-square (RMS) noise, and an absolute accuracy of $790\mu\text{V}$, which includes the effects of noise, offset, quantization and temperature drift. Before any measurements, The ADC was calibrated so that sampling error was minimized.

In the current implementation, the frame rate is limited by the settling time required for small currents driving capacitive loads. 18 columns of the imager can be sampled simultaneously, and it takes 8 acquisition cycles to capture an entire row. After row activation, the column outputs are allowed to settle before samples are acquired. The delay depends strongly on the channel currents I_p and I_2 . We observe artifactual column patterns if the delay is too short. The subjective criteria used to determine minimum delay for each intensity level is to gradually reduce the delay from a large value, until the image captured yields more than 3mV pixel voltage deviation. We were able to use a delay of 0.1ms for intensities $10^{-2}L$ and above, which theoretically gives a frame rate of $(8 \times 144 \times 10^{-4})^{-1} = 8.7$ frames/s (fps), provided the time required for A/D conversion and row/column activation is negligible. However, due to software issues, we are not able to reduce the time required for column and row activation. We were able to capture an image at 0.93fps at intensities $10^{-2}L$ and above, while capturing an image at intensity $10^{-6}L$ takes 40 minutes.

The inter-pixel voltage differences on nodes V_3 during column activation cause slight voltage changes on the photodiode nodes V_A via capacitance coupling, and a delay is required for the pixel circuit to reach steady state. Since the weakest current is I_p , this delay depends on incident light intensity and can be very long. Possible

methods for reducing response time include precharging column groups in advance (pipelining), adopting a feedback pixel architecture as in [115] for virtually grounding V_A , and using a microcontroller for data acquisition rather than a software-controlled data acquisition card. On-chip buffer amplifiers would also improve response time, but would introduce additional mismatch into the readout path.

To determine the proper subthreshold voltage and current range for operation, we measure the I_D - V_{GS} curve of an isolated nFET having the size of M3. I_D is roughly exponential with V_{GS} for gate voltages ranging from 0.3V to 0.8V, with current values of 1.02pA and 81.64nA, respectively.

5.4.1 Temporal Noise

Temporal noise arises from fundamental shot noise in the photocurrent and bias currents as well as thermal noise and $1/f$ noise associated with the transistors and photodiode, and is expected to limit the precision of pixel adaptation. Thus during operation the temporal noise will vary amongst pixels due to differing photocurrent, bias current, and transistor parameters such as threshold. We examined the temporal noise for each pixel by acquiring 288 consecutive images under identical conditions: $V_{dd} = 4.3V$, $I_c = 2.2\mu A$, $V_E = 2.6V$ and illumination $10^{-2}L$. We computed standard deviation (σ) of pixel voltage for each pixel using all 288 samples. Among all 144x144 pixels, the minimum observed standard deviation was $\sigma_{min} = 0.75mV$ and the maximum observed standard deviation was $\sigma_{max} = 1.29mV$, and the average $\sigma_{avg} = 0.90mV$. σ denotes the total RMS temporal noise [99] of the

system including the image sensor, the op-amp and the A/D converter, and sets an upper bound on the accuracy of the sampled output of the image sensor at any instant in time. The above statistics were obtained with a calibrated chip. For an uncalibrated chip, the temporal noise was slightly higher with minimum, maximum and mean standard deviations of 1.28mV, 1.80mV and 1.50mV, respectively.

5.4.2 Vdd and V_E Dependence

The pixel voltage is affected by power supply Vdd and electrode voltage V_E ; this dependency should be taken into account during operation. The channel current I_2 increases as the global voltage V_E decreases, resulting in a higher pixel voltage V_D . V_D also increases for larger values of Vdd. Figure 5.4 shows the mean pixel voltage $\overline{V_D}$ as a function of $V_{ED} = V_E - \text{Vdd}$ for Vdd of 4.3V and 3.3V. Both curves are straight lines, with a standard deviation of residue from least-square fits of 0.72mV and 0.38mV, respectively. From the fits we find that $\overline{V_D}$ changes with V_E with slope $\frac{\Delta\overline{V_D}}{\Delta V_E} = -69\text{mV/V}$, and $\overline{V_D}$ changes with Vdd with slope $\frac{\Delta\overline{V_D}}{\Delta V_{dd}} = 24\text{mV/V}$. An isolated test structure without the floating gate (i.e., V_B is connected to V_A) exhibits slope $\frac{\Delta V_D}{\Delta V_{dd}} = 5\text{mV/V}$. Thus, much of the dependency on Vdd results from charge sharing through parasitic capacitances on the floating gate.

5.4.3 Power Consumption

We estimate power consumption by analyzing the currents in each pixel and comparing to experimental results. When the pixels are not being read out, all row

Table 5.2: Power Consumption

	Power (μW) @ $10^{-1}L$		Power (mW) @ L	
V_E (V)	Idle	Read	Idle	Read
3.3	116	261	1.099	1.274
2.3	116	290	1.099	1.419
1.3	116	416	1.099	2.112
0.3	116	990	1.099	4.719

control lines are held at GND, turning off M5. V_D is held at GND, and V_3 increases towards Vdd. The only current flowing is the photocurrent I_p through M₁ and D₁. Thus, the idle power consumption depends only on the illumination intensity and junction leakage. During operation, additional currents I_2 and I_c are turned on. I_2 depends on the floating node voltage V_B and increases as V_E decreases. I_c is set to approximately $2.2\mu\text{A}$, so the power contribution from column currents are $18 \times 2.2\mu\text{A} \times 3.3\text{V} = 130.7\mu\text{W}$ during a read operation. Table 5.2 lists measured power consumption under varying conditions for V_E and incident illumination, for Vdd=3.3V. The idle power consumption during projection of a slide (Fig.5.10) and of indoor scene (Fig.5.12) is less than $10\mu\text{W}$. The equivalent illumination intensities for (Fig.5.10) and (Fig.5.12) are 2.2×10^{-2} and 10^{-4} , respectively. At these intensities, power consumption during operation is about $140\mu\text{W}$, dominated by the column currents.

The global voltage V_E provides an opportunity for adjusting the trade-off between power consumption and readout speed, especially for high incident light in-

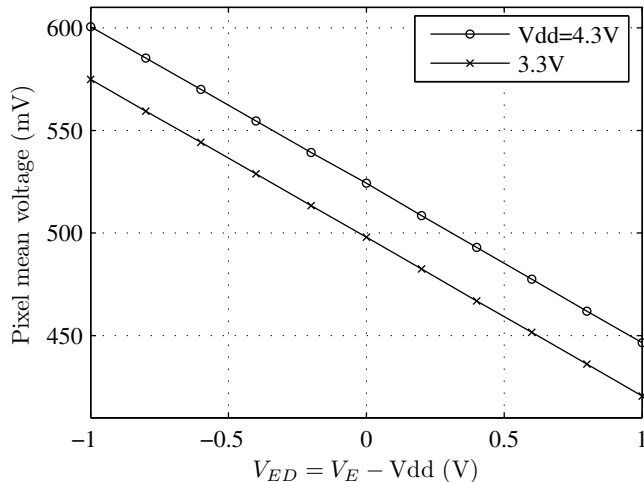


Figure 5.4: The pixel voltage increases with 1) increasing V_{dd} and 2) decreasing V_E .

tensities. For lower V_E the readout is faster but consumes more power. However, V_E should be limited so that M_2 remains in the subthreshold region for the maximum possible incident light after programming.

5.4.4 Output Voltage Distribution

We compute statistical characteristics for the pixel voltages from their empirical distribution. Figure 5.5 shows histograms of the output voltage V_D for all pixels for a bin size of 0.5mV. In Figs. 5.6 and 5.7, the bin size is 0.1mV and 1mV, respectively. For a large number of identical pixels (20.7k pixels) with the same incident intensity (spatially uniform illumination), the histogram approximates the probability density function (pdf) if we consider the output voltage V_D as a random variable. Therefore, we quantify the FPN noise power according to the variance σ^2 and standard deviation σ obtained from the pdf of V_D . In Fig.5.5(a), we measured V_D for each pixel of a UV-erased chip under $10^{-1}L$ intensity and plot the histogram

with dots. The solid trace is a Gaussian fit using least squared error curve-fitting. As expected, the FPN is approximately Gaussian. The σ for the Gaussian fit is 16.675mV, and that of the empirical distribution of V_D is 16.638mV.

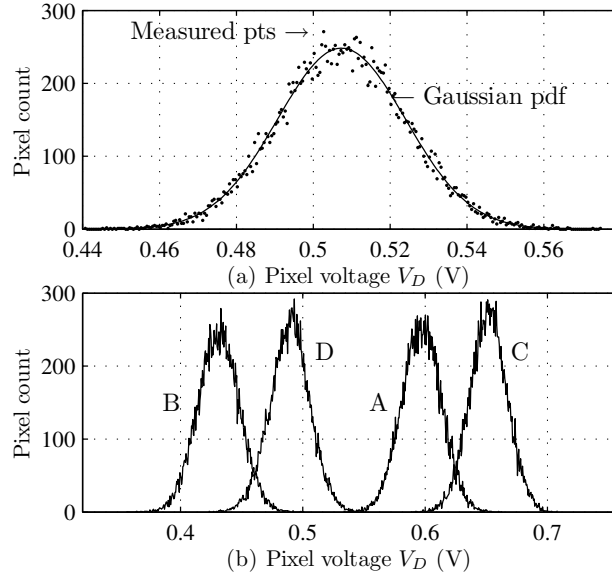


Figure 5.5: (a) The pixel voltages tends to be Gaussian distributed. (b) UV-erased curves (C and D) are slightly taller and to the right.

Before programming, residual charge on the floating gates was reduced through UV erasure. The chip was exposed to UV illumination in a standard EPROM eraser for at least 20 hours. In Fig.5.5(b), we plot the results of four V_D measurements: A) before UV erasure with $10^{-1}L$ illumination, B) before UV with $10^{-3}L$, C) after UV with $10^{-1}L$ and D) after UV with $10^{-3}L$. The UV-erased chip has consistently higher pixel voltages for all intensity levels, as indicated by the rightward shifts from B to D and from A to C. The FPN is slightly lower for the UV-erased chips, as indicated by the taller, narrower distributions C and D as compared to A and B. The FPN σ for A and C are 16.2mV and 15.3mV, respectively, and the mean $\overline{V_D}$

shift from A to C is 54.6mV. Thus, illuminating with UV results in more negative charge on the floating gate, and reduces the variation of pre-existing charge on the floating gates.

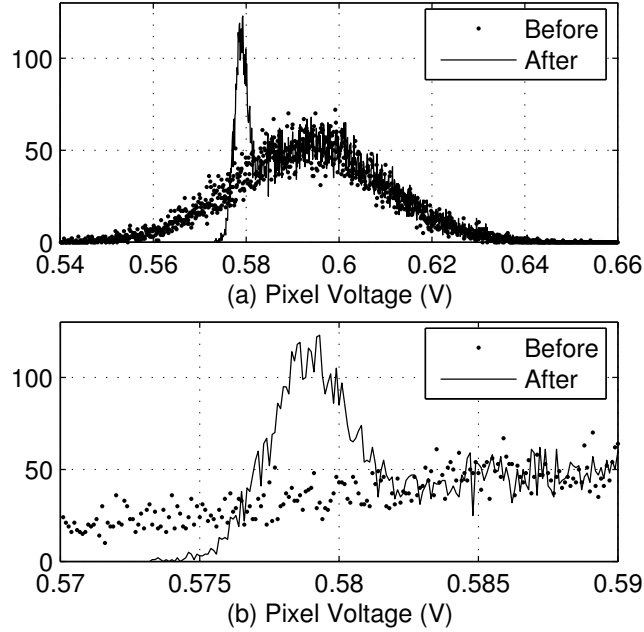


Figure 5.6: The original bell-shaped pixel voltage distribution is shoved from the left forming a new bell-shaped heap during calibration. In a magnified view (b), the 6σ for the new peak is observed to be roughly 8mV, compared to 110mV for the original distribution.

5.4.5 Performing Adaptation

Figure 5.6 is a demonstration of partial adaptation. The sensor is illuminated with an intensity of $10^{-2}L$ and pixel voltages are read out with $V_{dd}=4.3V$ and $V_E=2.4V$. The minimum and maximum pixel output voltages are 527mV and 654mV, respectively. The pixel voltage distribution is plotted as dotted curves. Next, we enable adaptation by setting all column voltages to $V_D^* = 615mV$ and then raising V_{dd} to 5.3V for 2 minutes, while keeping V_{ED} constant. Finally we

measure pixel output voltages again for the same readout conditions and obtain the distribution which is plotted as solid curves. Comparing the dotted and solid curves, we see that the original curve shifts to the right after adaptation. The pixels with initially lower readout voltages were adapted through hot-carrier injection onto the floating gates, resulting in new readout voltages that are closer to the programmed voltage V_D^* . The pixels with initially higher readout voltages were not adapted and remain at their initial voltage. This partial injection results in a new bell-shaped distribution with a peak around 578mV (b), 37mV below the programmed V_D^* . This voltage-shifting effect of adaptation is a combination of the Vdd dependency and injection that both tend to increase I_2 . Therefore, in order to achieve complete adaptation in all pixels, we find empirically that it is adequate to set V_D^* 60mV above the maximum initial pixel value during adaptation. We follow this rule of thumb in all following experiments.

5.4.6 Adaptation Performance

We report observed FPN improvements over adaptation time, across multiple illumination intensities, and at different values of power supply Vdd and global voltage V_E , as well as performance retention over time. A UV-erased chip is exposed to spatially uniform illumination, then alternately adapted with Vdd=5.3V and $V_D^* = 0.7V$ for one second and measured with Vdd=4.3V in order to observe the progression of adaptation. Figure 5.7 plots the measured distribution with time, for the first 18 seconds. The pixel voltages gradually shift to the right, forming

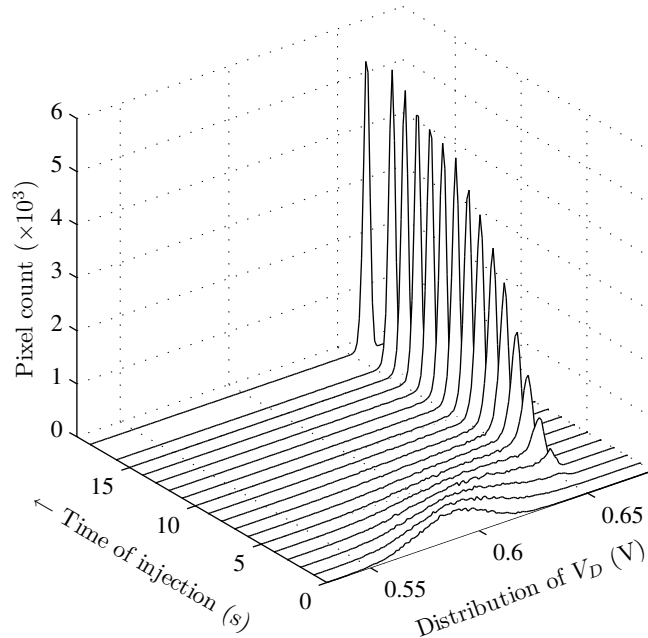


Figure 5.7: Pixel voltage distribution changes over time during adaptation.

a new bell-shaped curve. Pixels with initially higher voltages have higher channel current and higher injection rate and thus move faster than those with initially lower voltages. This is better illustrated in Fig.5.8, which plots the FPN σ , minimum, mean, and maximum voltages with time. As shown, the mean settles as early as 10s, but the minimum voltage does not settle until about 30s. Afterwards, the statistics stay roughly the same, but the FPN σ continues decreasing. At 30s $\sigma = 1.35\text{mV}$; at 9 minutes $\sigma = 1.16\text{mV}$. The FPN σ has been reduced to roughly the same level as the RMS temporal noise. Not surprisingly, the speed of injection is accelerated for higher channel currents I_2 , which can be globally adjusted with V_E . The FPN σ for a chip with an initially minimum pixel voltage of 629mV, approximately 100mV higher than that shown in Fig.5.7 reduces from 14.20mV to 1.44mV in only 4 seconds.

We then perform calibration on an un-programmed and un-erased chip and

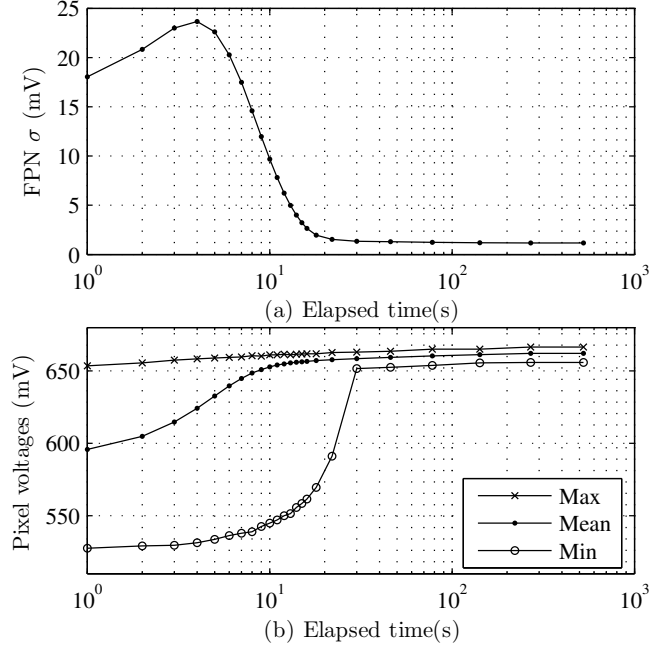


Figure 5.8: Pixel voltage statistics change over time during adaptation.

measure performance over a range of illumination intensities. Figure 5.9(a) plots the FPN σ vs. intensities from $10^{-6}L$ to L . The black bars show the FPN σ before calibration. The white bars show the FPN σ after performing calibration with intensity L . The gray bars show the FPN σ after calibration at intensity $10^{-2}L$. FPN σ is minimum when the imager is operated at an illumination intensity equal to that at which the calibration was performed, as shown in both cases. As the intensity deviates from the calibration intensity, FPN σ increases. In a hypothetical image that contains intensities ranging from $10^{-6}L$ to L , Fig.5.9(a) indicates that calibration at intensity $10^{-2}L$ would result in better performance. FPN σ reduces from 14.31mV to 1.07mV at intensity L , and from 16.20mV to 1.37mV at intensity $10^{-2}L$, which corresponds to an FPN power reduction of 178x and 140x, respectively. For intensities from $10^{-5}L$ to L , the $10^{-2}L$ calibration gives at least 34x FPN power

reduction.

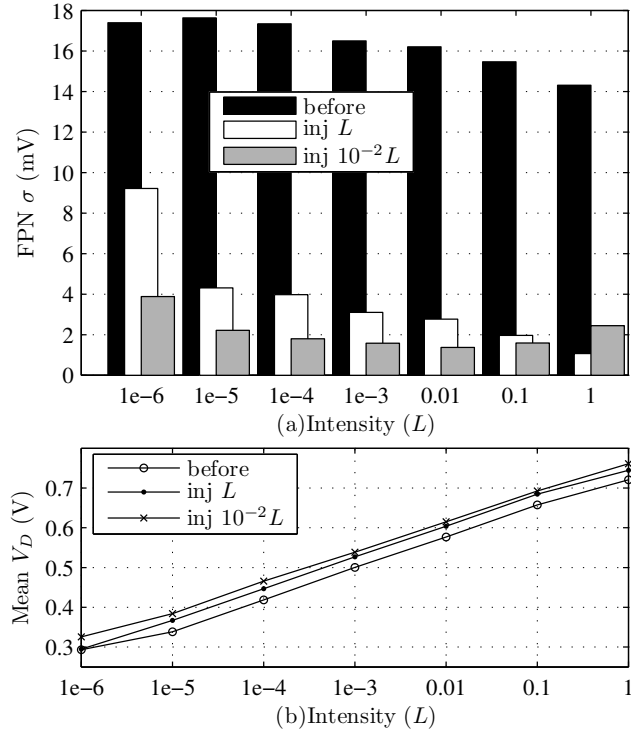


Figure 5.9: (a) FPN σ reduces significantly after adaptation. (b) Pixel voltage follows the logarithm of intensity linearly.

V_{ED} was $-2.5V$ before injection, $-2V$ during calibration and measurement at intensity L , and $-1.7V$ during calibration and measurement at intensity $10^{-2}L$. We increased V_E before a new adaptation cycle to compensate for the increase in pixel voltages caused by prior adaptation. Figure 5.9(b) shows the mean pixel voltage as a function of intensity for these three cases. At intensity L the nFET is starting to leave subthreshold operation, and at intensity $10^{-6}L$ the current level is comparable to that of junction leakage. Between intensities $10^{-5}L$ to $10^{-1}L$ the pixel voltage is a logarithmic function of the illumination intensity, with $79mV/decade$ L . It is worth noting that for a non-FG test pixel (V_A connected to V_B , with all

other circuit elements identical) the response is 112mV/decade L . This difference reflects the reduction in gain from capacitive division with capacitances other than C_1 (Fig.5.1), including C_2 and parasitic capacitances to the floating node.

For best performance, the operating Vdd should be close to the calibration Vdd. When the imager was operated with Vdd = 3.3V rather than Vdd = 4.3V, after being calibrated at Vdd = 5.3V, we observed about 0.5mV increase in the FPN σ for a 3.3V Vdd over all intensities.

A standard 35mm camera lens was used to focus 35mm slides positioned 24 inches from the lens. They were illuminated from the back. A test slide containing a triangle with sharp angles was used to manually focus the lens. Figure 5.10 shows images of a slide of the Jefferson Memorial taken (a) before and (b) after calibration at uniform illumination of $10^{-2}L$. All pictures shown in this paper have been normalized: Pixels having a voltage less than or equal to $\mu - 3\sigma$ are shown in black, and those greater than or equal to $\mu + 3\sigma$ are shown in white, where μ and σ are mean and standard deviation of all pixels voltages, respectively. From the voltage response we calculate that the illumination intensity for the slide shown in Fig.5.10 is approximately equal to $2.2 \times 10^{-2}L$.

Finally, we investigated the retention of programming after calibration. A chip was calibrated at intensity $10^{-2}L$ and then read out at Vdd = 4.3V continuously. Figure 5.11 shows its performance over time. The pixel voltages increased gradually after 4 hours of operation, and the FPN σ increased after 24 hours. Although the magnitude of hot-carrier injection current is low, its effects are cumulative and result in significant changes at Vdd=4.3V. At the reduced power supply Vdd=3.3V,

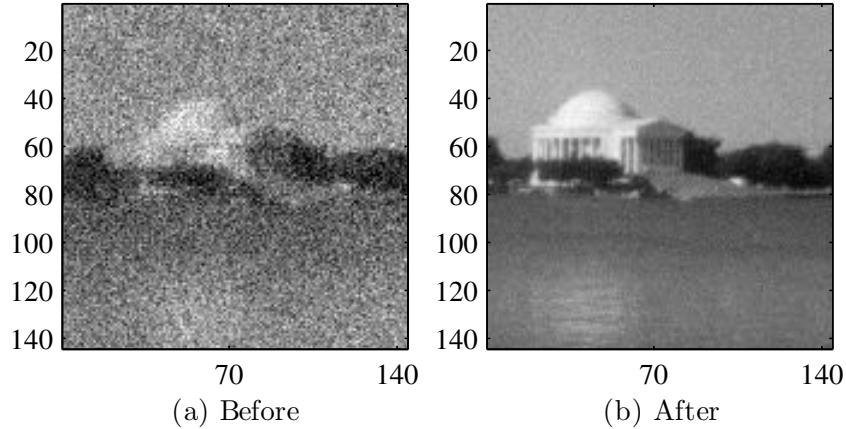


Figure 5.10: Images of a slide of the Jefferson Memorial taken (a) before and (b) after calibration.

the magnitude was insufficient to cause noticeable change in the statistics during a monitoring time of 5 days.

5.5 Applications and Special Effects

We have demonstrated the ability to effectively reduce fixed pattern noise within each pixel of a current mode imager. This adaptation mechanism can be applied to correct any distortion of the illumination intensity in the optical path. A common problem for lenses is vignetting, which causes unintended darkening of the image corners [116]. We do not observe significant vignetting with the lenses in our experimental setup. To illustrate the technique, we created a similar effect by positioning a point source near the back of a slide such that the center is brighter than the edges. We used a chip calibrated at a uniform illumination of $10^{-2}L$ to capture an image of such a back-illuminated slide. The slide shows a building in Oldtown San Diego, having wide light-colored walls (Fig.5.12a). We removed

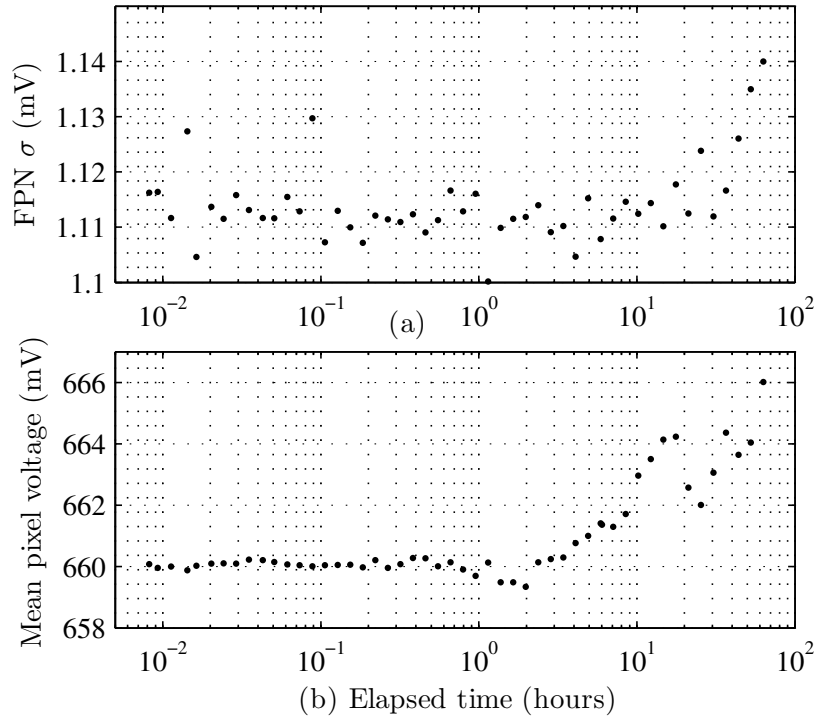


Figure 5.11: At $V_{dd}=4.3V$, injection causes pixel voltage and FPN σ increase.

the slide and performed calibration with the point source-illuminated screen with $V_{dd} = 5.3V$ for 10 seconds, then replaced the slide and captured the image shown in Fig.5.12(b). The calibration imprinted a pattern in the floating gates to compensate for the bright center and dark corners. If we then use the calibrated sensor to capture a natural scene, the center looks darker than the edges, as shown in Fig.5.12(c). This image was taken in our lab under ordinary fluorescent lighting, at an illumination level equivalent to approximately $10^{-4}L$.

Human vision exhibits temporal adaptation. If one stares at an object for a period of time, the features of the object itself and surrounding objects fade. If we now look at other views we find the residue of the previous view, opposite in color, commonly called an “afterimage” [117]. In the following experiments,

we created an afterimage with our image sensor by performing adaptation with a scene that is spatially non-uniform, and then used the scene-adapted sensor to capture another image. A volunteer posed in front of the imager in our lab, and we performed adaptation with $V_{dd} = 5.3V$ for 10 seconds. Figure 5.12 (d) shows an image acquired with $10^{-2}L$ uniform illumination after this adaptation. We recovered the scene by inverting the acquired image in software (e). Before acquiring the next image (f), we asked the volunteer to move slightly to his right, remove one flyer from the wall behind and turn on a lamp on the table. Since the scene has been captured with a scene-adapted sensor, only changes in the scene appeared in the image. For example, the relocation of the volunteer results in two figures, one that is negative at the original location, and another one that is positive at the new location. The removal of the flyer results in a dark residue on the cabinet, and the additional lighting results in a visible test bench, wires and a multimeter, as well as reflections on the cabinet. Other stationary objects in the original scene such as the top three flyers disappear.

Finally, we created an illusion akin to double exposure of film negatives in Fig.5.12 (g) by first performing adaptation using a scene constructed as the negative of a portrait of Albert Einstein printed on white paper. We then invited another volunteer to enter the scene and captured his image with the “ghost” of Einstein. The paper in his hands in (h) is the target paper used to create the afterimage.

5.6 Summary

We have described a novel adapting floating gate continuous-current pixel for high quality imaging that has the ability to automatically remove fixed pattern noise (FPN) simultaneously from all pixels. We have described theory and method for adapting the voltage on the floating gate of a pFET that leads to accurate calibration. The mechanism that is used to adapt out FPN is hot electron injection inside each pixel. Injection stops when two currents inside the pixel balance each other through a negative feedback loop. In addition to canceling offset, each pixel can be used to accurately set up an arbitrary input offset for various applications ranging from vignetting compensation to creating afterimages. This feature is not readily available in other FPN removal schemes. During adaptation, an external voltage is applied globally to all pixels and the imager is uniformly illuminated. We have experimentally demonstrated that FPN can be reduced by a factor of 178x. The pixel output voltage is logarithmically related to the photon flux providing a large dynamic range exceeding 5 orders of magnitude. Each pixel measures $18\mu\text{m}$ on a side and has a fill factor of 13.4%. The chip consumes $140\ \mu\text{W}$.

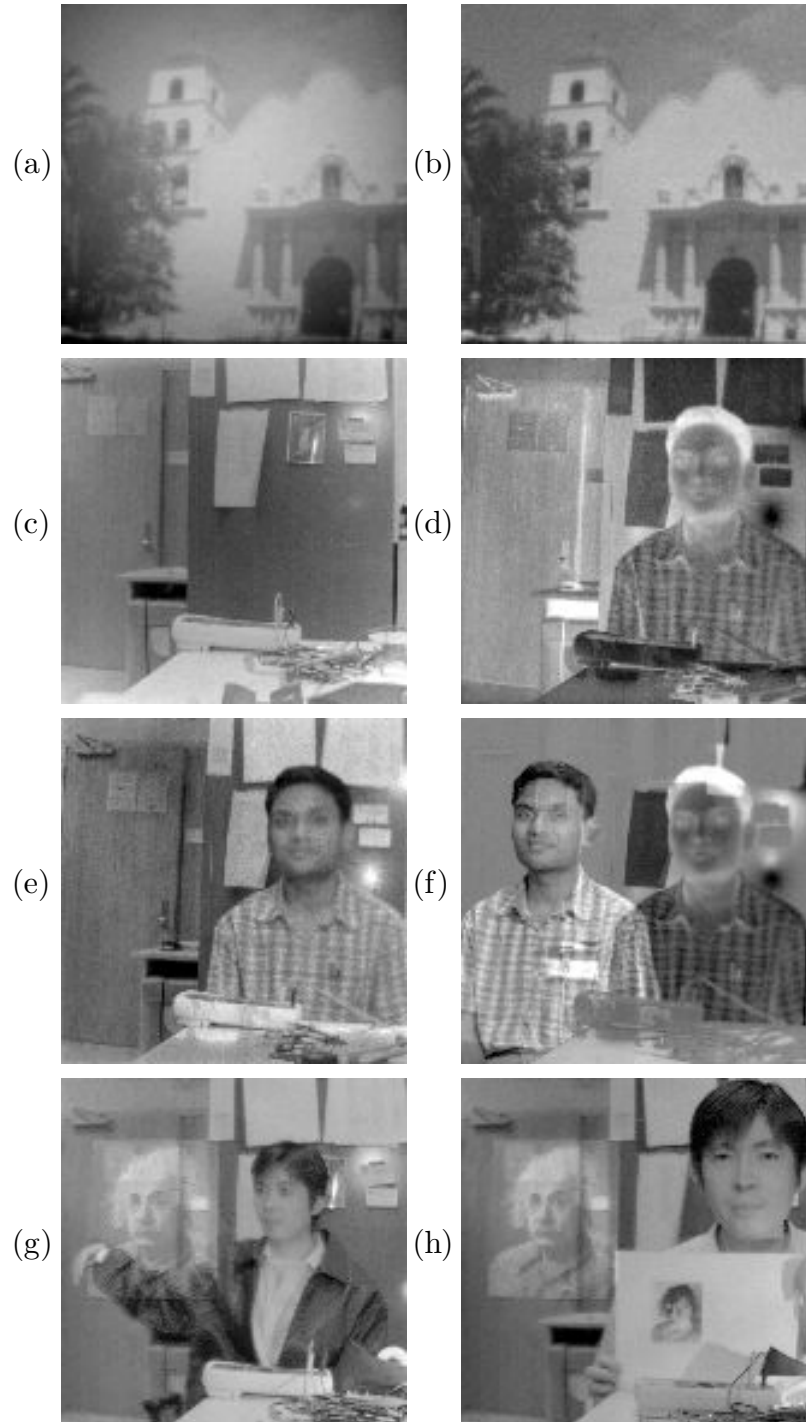


Figure 5.12: Vignetting correction (a)-(c) and afterimages (d)-(h)

Chapter 6

Conclusion

CMOS VLSI device technology has enjoyed continuous and steady development for the past several decades with major capital support from the industry. With no other technologies mature enough to be a replacement as of this writing, CMOS VLSI is expected to continue the evolution into newer generations with even cheaper, faster and lower power consumption.

Surprisingly, with continuous multi-billion-dollar investments into the CMOS industry, the basic physics in the metal-oxide-semiconductor operations and even basic manufacturing has not changed much over the past decades. Same can be said to the basic circuit techniques that were invented to help achieve higher gain, higher operating speed, and wider operating temperature and voltage range. These circuit techniques still apply to today's complex VLSI designs. However, as we are beginning to push the limit set by the Moore's Law, we anticipate modifications and even deviations to the traditional CMOS technologies in the near future.

Circuit designers are usually told by device engineers what they should or allowed to do with a given process. They exercise their creativities and imaginations in the given limit. However, circuit designers should be encouraged to explore why they were given such limits, and what physical implications can occur if they deviate from such rules and limits. By doing so, circuit designers gradually discover and

exploit hidden benefits in a given process.

The p-FET in a standard CMOS process has substantial gate currents in a certain bias condition that easily occurs in normal operating range, as seen earlier. The phenomenon seen by a device engineer as a reliability concern is exploited by floating gate pioneers to build new and creative circuits and techniques. Since the earliest publications, we have seen many examples of floating gate applications in standard CMOS. Recently the floating gate techniques are mature enough to see the first signs of commercialization. Some examples include non-volatile EEPROMs to be sold as intellectual property blocks (IP) and embedded by the customers into their standard logic CMOS products (Virage Logic[®], Impinj[®]), and the Floating Gate Array (FGA[™]) that generate low-drift, high precision voltage references (Intersil[®]). During the commercialization process, the team including device engineers and circuit designers will need to find trade-offs that gives best possible reliability and performance.

In this work, we attack the long standing problem in devices with floating gate techniques. Offsets caused by intrinsic device mismatch limits performance in precision circuits. We successfully applied floating gate techniques in comparators, ADCs and image sensors and demonstrated significant performance gains. In all reported examples, we applied both implicit and explicit feedbacks to control mismatch adaptation, achieving automatic and accurate results. In addition, we achieved special features that are otherwise difficult or impossible thanks to the non-volatile storage characteristics of the floating gate structures.

We have confirmed the excellent charge retention characteristics reported in

the literature with our fabricated chips. Improvements in manufacturing lead to higher quality gate oxide, essential to reliable floating gate implementations. The promising future for a more powerful CMOS is evident in the continuing advancements in manufacturing techniques as well as in circuit techniques such as the new and useful utilization of floating gate structures.

Bibliography

- [1] D. Kahng and S. M. Sze, "A floating gate and its application to memory devices," *Bell Syst. Tech. J.*, vol. 46, p. 1283, 1967.
- [2] W. D. Brown and J. E. Brewer, *Nonvolatile Semiconductor Memory Technology : A Comprehensive Guide to Understanding and Using NVSM Devices*. Piscataway, NJ: Wiley-IEEE Press, 1997.
- [3] Y. Ma, T. Gilliland, B. Wang, R. Paulsen, A. Pesavento, C.-H. Wang, H. Nguyen, T. Humes, and C. Diorio, "Reliability of pFET EEPROM with 70Å tunnel oxide manufactured in generic logic CMOS processes," *IEEE Trans. Device and Materials Reliability*, vol. 4, no. 3, pp. 353–358, Sept. 2004.
- [4] Y. Wong, P. Abshire, and M. Cohen, "Floating gate comparator with automatic offset manipulation capability," in *Proc. IEEE ISCAS*, vol. 1, May 2004, pp. I-529–532.
- [5] Y. L. Wong, M. H. Cohen, and P. A. Abshire, "Differential hot electron injection in an adaptive floating gate comparator," *Analog Integrated Circuits and Signal Processing*, vol. 43, no. 3, pp. 281–296, June 2005.
- [6] —, "A floating-gate comparator with automatic offset adaptation for 10-bit data conversion," *IEEE TCAS I*, vol. 52, no. 7, pp. 1316–1326, July 2005.
- [7] M. Choi and A. Abidi, "A 6-b 1.3-Gsample/s A/D converter in 0.35- μm CMOS," *IEEE JSSC*, vol. 36, no. 12, pp. 1847–1858, December 2001.
- [8] Y. Wong, M. Cohen, and P. Abshire, "A 1.2GHz adaptive floating gate comparator with 13-bit resolution," in *Proc. IEEE ISCAS*, May 2005, pp. 6146–6149.
- [9] —, "On-line histogram equalization for flash ADC," in *Proc. IEEE ISCAS*, May 2007, to appear.
- [10] —, "A 128x128 floating gate imager with self-adapting fixed pattern noise reduction," in *Proc. IEEE ISCAS*, May 2005, pp. 5314–5317.
- [11] —, "A 144x144 image sensor with self-adapting mismatch reduction," *IEEE TCAS I*, 2007, to appear.
- [12] J. Sewell, F., H. Wegener, and E. Lewis, "The variable threshold FET: Theory and experiment," in *IEEE ISSCC Tech. Dig.*, vol. XII, Feb 1969, pp. 182–183.
- [13] T. Hagiwara, Y. Yatsuda, R. Kondo, S. Minami, T. Aoto, and Y. Itoh, "A 16 kbit electrically erasable PROM using n-channel Si-gate MNOS technology," *IEEE JSSC*, vol. 15, no. 3, pp. 346–353, June 1980.

- [14] P. C. Chen, "Threshold-alterable Si-gate MOS devices," *IEEE J. Electron Devices*, vol. ED-24, p. 584, 1977.
- [15] E. Suzuki, K. Ishii, Y. Hayashi, and H. Hiraishi, "A low-voltage alterable EEPROM with metal-oxide-nitride-oxide-semiconductor MONOS structures," *IEEE J. Electron Devices*, vol. ED-30, pp. 122–128, February 1983.
- [16] E. Prinz, "The zen of nonvolatile memories," in *IEEE 43rd Design Automation Conf.*, 2006, pp. 815–820.
- [17] H. Toyoshima, S. Kobayashi, J. Yamada, T. Miwa, H. Koike, H. Takeuchi, H. Mori, N. Kasai, Y. Maejima, N. Tanabe, T. Tatsumi, and H. Hada, "Feram device and circuit technologies fully compatible with advanced cmos," in *IEEE Conf. Custom IC*, 2001, pp. 171–178.
- [18] K. Hoya, S. Takashima, D. and Shiratake, R. Ogiwara, T. Miyakawa, H. Shiga, S. Doumae, S. Ohtsuki, Y. Kumura, S. Shuto, T. Ozaki, K. Yamakawa, I. Kunishima, A. Nitayama, and S. Fujii, "A 64Mb chain FeRAM with Quad-BL architecture and 200MB/s burst mode," in *IEEE Int. Solid-State Circuits Conf.*, Feb 2006, pp. 459–466.
- [19] J. Slaughter, R. Dave, M. Durlam, G. Kerszykowski, K. Smith, K. Nagel, B. Feil, J. Calder, M. De Herrera, B. Garni, and S. Tehrani, "High speed toggle MRAM with MgO-based tunnel junctions," in *IEEE IDEM*, Dec 2005, pp. 873–876.
- [20] S. Kang, W. Y. Cho, B.-H. Cho, K.-J. Lee, C.-S. Lee, H.-R. Oh, B.-G. Choi, Q. Wang, H.-J. Kim, M.-H. Park, Y. H. Ro, S. Kim, C.-D. Ha, K.-S. Kim, Y.-R. Kim, D.-E. Kim, C.-K. Kwak, H.-G. Byun, G. Jeong, H. Jeong, K. Kim, and Y. Shin, "A 0.1- μm 1.8-V 256-Mb phase-change random access memory (PRAM) with 66-MHz synchronous burst-read operation," *IEEE JSSC*, vol. 42, no. 1, pp. 210–218, Jan 2007.
- [21] S. Shah and S. Collins, "A temperature independent trimmable current source," in *Proc. IEEE ISCAS*, vol. 1, May 2002, pp. 713–716.
- [22] S. Jackson, J. Killens, and B. Blalock, "A programmable current mirror for analog trimming using single poly floating-gate devices in standard CMOS technology," *IEEE TCAS II*, vol. 48, no. 1, pp. 100–102, January 2001.
- [23] J. Hyde, T. Humes, C. Diorio, M. Thomas, and M. Figueroa, "A 300-MS/s 14-bit digital-to-analog converter in logic CMOS," *IEEE JSSC*, vol. 38, no. 5, pp. 734–740, May 2003.
- [24] P. Hasler, B. Minch, and C. Diorio, "An autozeroing floating-gate amplifier," *IEEE TCAS II*, vol. 48, no. 1, pp. 74–82, January 2001.

- [25] T. Constandinou, J. Georgiou, and C. Toumazou, "An auto-input-offset removing floating gate pseudo-differential transistor," in *Proc. IEEE IS-CAS*, vol. 1, May 2003, pp. 169–172.
- [26] P. Brady and P. Hasler, "Offset compensation in flash ADCs using floating-gate circuits," in *Proc. IEEE ISCAS*, May 2005.
- [27] F. Devos, M. Zhang, Y. Ni, and J.-F. Pone, "Trimming CMOS smart imager with tunnel-effect nonvolatile analogue memory," *IEE Electron. Lett.*, vol. 29, pp. 456–459, September 1993.
- [28] M. Zhang, F. Devos, and J.-F. Pone, "Trimming smart imagers for an image converter with a nonvolatile analog memory," *Sensors and Actuators A: Physical*, vol. 47, pp. 456–459, March 1995.
- [29] A. Aslam, W. Brockherde, B. Hosticka, H. Vogt, and G. Zimmer, "A CMOS image sensor with combined analog nonvolatile storage capability," in *IEEE Int. Elec. Dev. Mtg.*, December 1996, pp. 923–926.
- [30] A. Aslam-Siddiqi, W. Brockherde, M. Schanz, and B. Hosticka, "A 128-pixel CMOS image sensor with integrated analog nonvolatile memory," *IEEE JSSC*, vol. 33, pp. 1497–1501, October 1998.
- [31] M. Cohen and G. Cauwenberghs, "Focal-plane on-line nonuniformity correction using floating-gate adaptation," in *Proc. IEEE ISCAS*, vol. 5, May 2000, pp. 153–156.
- [32] A. Pesavento, T. Horiuchi, C. Diorio, and C. Koch, "Adaptation of current signals with floating-gate circuits," *Analog Integrated Circuits and Signal Processing*, vol. 30, no. 2, pp. 137–147, Feb 2002.
- [33] M. Cohen and G. Cauwenberghs, "Floating-gate adaptation for focal-plane online nonuniformity correction," *IEEE TCAS II*, vol. 48, no. 1, pp. 83–89, January 2001.
- [34] A. Bandyopadhyay, J. Lee, R. Robucci, and P. Hasler, "A 80uW/frame 104x128 CMOS imager front end for JPEG compression," in *Proc. IEEE IS-CAS*, vol. 5, May 2005, pp. 5318–5321.
- [35] C. K. Sin, A. Kramer, V. Hu, R. R. Chu, and P. K. Ko, "EEPROM as an analog storage device with particular applications in neural networks," *IEEE J. Electron Devices*, vol. 39, no. 6, p. 1410V1419, July 1992.
- [36] H. Castro, S. Tam, and M. Holler, "Implementation and performance of an analog non-volatile neural-network," *Analog Integrated Circuits and Signal Processing*, vol. 4, no. 2, p. 97V113, 1993.

- [37] A. F. Murray, S. Churcher, A. Hamilton, A. J. Holmes, G. B. Jackson, H. M. Reekie, and R. J. Woodburn, "Pulse stream VLSI neural networks," *IEEE Micro*, vol. 14, no. 3, pp. 29–39, 1994.
- [38] R. Schaller, "Moore's law: past, present and future," *IEEE Spectrum*, vol. 34, no. 6, pp. 52–59, June 1997.
- [39] R. Dennard, F. Gaensslen, V. Rideout, E. Bassous, and A. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE JSSC*, vol. 9, no. 5, pp. 256–268, October 1997.
- [40] P. Hasler and T. Lande, "MATIA: a programmable $80\mu\text{W}$ /frame CMOS block matrix transform imager architecture," *IEEE JSSC*, vol. 41, no. 3, pp. 663–672, March 2006.
- [41] R. Muller, T. Kamins, and M. Chan, *Device Electronics for Integrated Circuits*. New York, New York: John Wiley & Sons, 2002.
- [42] R. Harrison, J. Bragg, P. Hasler, B. Minch, and S. Deweerth, "A CMOS programmable analog memory-cell array using floating-gate circuits," *IEEE TCAS II*, vol. 48, no. 1, pp. 4–11, January 2001.
- [43] J. Van Houdt, P. Heremans, L. Deferm, G. Groeseneken, and H. Maes, "Analysis of the enhanced hot-electron injection in split-gate transistors useful for EEPROM applications," *IEEE J. Electron Devices*, vol. ED-39, no. 5, pp. 1150–1156, May 1992.
- [44] K. Rahimi, C. Diorio, C. Hernandez, and M. Brockhausen, "A simulation model for floating-gate MOS synapse transistors," in *Proc. IEEE ISCAS*, vol. 2, May 2002, pp. 532–535.
- [45] P. Hasler, B. Minch, D. Diorio, and C. Mead, "An autozeroing amplifier using pFET hot-electron injection," in *Proc. IEEE ISCAS*, vol. 3, May 1996, pp. 325–328.
- [46] P. Hasler, T. Stanford, and B. Minch, "A second-order section built from autozeroing floating-gate amplifiers," *IEEE TCAS II*, vol. 48, no. 1, pp. 116–120, January 2001.
- [47] B. Nauta, "A CMOS transconductance-C filter technique for very high frequencies," *IEEE JSSC*, vol. 27, no. 2, pp. 142–153, February 1992.
- [48] F. Munoz, A. Torralba, R. Carvajal, J. Tombs, and J. Ramirez-Angulo, "Floating-gate-based tunable CMOS low-voltage linear transconductor and its application to HF g_m -C filter design," *IEEE TCAS II*, vol. 48, no. 1, pp. 106–110, January 2001.

- [49] Y. Berg, T. Lande, Ø. Næss, and H. Gundersen, “Ultra-low-voltage floating-gate transconductance amplifiers,” *IEEE TCAS II*, vol. 48, no. 1, pp. 37–44, January 2001.
- [50] Y. Berg, T. S. Lande, and Ø. Næss, “Programming floating-gate circuits with UV-activated conductances,” *IEEE TCAS II*, vol. 48, no. 1, pp. 12–19, January 2001.
- [51] B. Minch, P. Hasler, and C. Diorio, “Multiple-input translinear element networks,” *IEEE TCAS II*, vol. 48, no. 1, pp. 20–28, January 2001.
- [52] P. Abshire, E. Wong, Y. Zhai, and M. Cohen, “Adaptive log domain filters using floating gate transistors,” in *Proc. IEEE ISCAS*, vol. 1, May 2004, pp. I–29–32.
- [53] K. Narendra and A. Annaswamy, *Stable Adaptive Systems*. New Jersey: Prentice-Hal, 1989.
- [54] T. Shibata and T. Ohmi, “A functional MOS transistor featuring gate-level weighted sum and threshold operations,” *IEEE J. Electron Devices*, vol. 39, pp. 1444–1455, July 1992.
- [55] —, “Neuron MOS binary-logic integrated circuits—part I: Design fundamentals and soft-hardware-logic circuit implementation,” *IEEE J. Electron Devices*, vol. 40, no. 3, pp. 570–576, March 1993.
- [56] —, “Neuron MOS binary-logic integrated circuits—part II: Simplifying techniques of circuit configuration and their practical applications,” *IEEE J. Electron Devices*, vol. 40, no. 5, pp. 974–979, May 1993.
- [57] P. Celinski, S. Al-Sarawi, and D. Abbott, “Delay analysis of neuron-MOS and capacitive threshold-logic,” in *The 7th IEEE International Conference on Electronics, Circuits and Systems, 2000. ICECS 2000.*, vol. 2, December 2000, pp. 932–935.
- [58] P. Celinski, J. Lopez, S. Al-Sarawi, and D. Abbott, “Complementary neu-GaAs structure,” *IEE Electron. Lett.*, vol. 36, no. 5, pp. 424–425, March 2000.
- [59] A. Rantala, S. Franssila, K. Kaski, J. Lampinen, M. Aberg, and P. Kuiv-alainen, “High-precision neuron MOSFET structures,” *IEE Electron. Lett.*, vol. 35, no. 2, pp. 155–157, January 2001.
- [60] N. Yu, T. Shibata, and T. Ohmi, “A real-time center-of-mass tracker circuit implemented by neuron MOS technology,” *IEEE TCAS II*, vol. 45, no. 4, pp. 495–503, April 1998.
- [61] A. Nakada, T. Shibata, M. Konda, T. Morimoto, and T. Ohmi, “A fully parallel vector-quantization processor for real-time motion-picture compression,” *IEEE JSSC*, vol. 34, no. 6, pp. 822–830, June 1999.

- [62] K. Kotani, T. Shibata, M. Imai, and T. Ohmi, "Clock-controlled neuron-MOS logic gates," *IEEE TCAS II*, vol. 45, no. 4, pp. 518–522, April 1998.
- [63] W. Tuttlebee, "Software-defined radio: facets of a developing technology," *IEEE Personal Communications*, vol. 6, no. 2, pp. 38–44, April 1999.
- [64] D.-C. Kang, C.-I. Kim, H.-K. Park, S.-B. Cho, and J.-H. Lee, "Design of an 8-bit neuron MOSFET A/D converter using subranging method," in *6th International Conference on VLSI and CAD, 1999. ICVC '99.*, October 1999, pp. 533–536.
- [65] Y. Leblebici, H. Ozdemir, A. Kepkep, and U. Cilingiroglu, "A compact high-speed (31,5) parallel counter circuit based on capacitive threshold-logic gates," *IEEE JSSC*, vol. 31, no. 8, pp. 1177–1183, August 1996.
- [66] K. Hirose and H. Yasuura, "A comparison of parallel multipliers with neuron MOS and CMOS technologies," in *IEEE Asia Pacific Conference on Circuits and Systems, 1996.*, November 1996, pp. 488–491.
- [67] J. Ramirez-Angulo, S. Choi, and G. Gonzalez-Altamirano, "Low-voltage circuits building blocks using multiple-input floating-gate transistors," *IEEE TCAS I*, vol. 42, no. 11, pp. 971–974, November 1995.
- [68] G. Han and E. Sanchez-Sinencio, "CMOS transconductance multipliers: a tutorial," *IEEE TCAS II*, vol. 45, no. 12, pp. 1550–1563, December 1998.
- [69] Z. Hong and H. Melchior, "Four-quadrant CMOS analog multiplier," *IEE Electron. Lett.*, vol. 20, pp. 1015–1016, November 1984.
- [70] H. R. Mehrvarz and C. Y. Kwok, "A large-input-dynamic-range multi-input floating gate MOS four-quadrant analog multiplier," in *Proc. IEEE Int. Solid-State Conf.*, February 1995, pp. 60–61.
- [71] J. F. Schoeman and T. H. Joubert, "Four-quadrant analog CMOS multiplier using capacitively coupled dual gate transistor," *IEE Electron. Lett.*, vol. 32, pp. 405–408, February 1996.
- [72] J. D. Gray, C. M. Twigg, D. N. Abramson, and P. Hasler, "Characteristics and programming of floating-gate pFET switches in an FPAA crossbar network," in *Proc. IEEE ISCAS*, May 2005.
- [73] F. Tenore, R. Vogelstein, R. Etienne-Cummings, G. Cauwenberghs, and P. Hasler, "A floating-gate programmable array of silicon neurons for central pattern generating networks," in *Proc. IEEE ISCAS*, 2006, pp. 3157–3160.
- [74] R. Chawla, C. M. Twigg, and P. Hasler, "An analog modulator/demodulator using a programmable arbitrary waveform generator," in *Proc. IEEE ISCAS*, May 2005.

- [75] V. Krishnan, C. Duffy, D. Anderson, and P. Hasler, "Optimal quantization employing programmable flash analog to digital converters," in *Proc. 38th Asilomar Conf. Sig. Syst. Comp.*, vol. 1, 2004, pp. 816–819.
- [76] B. Ahuja, H. Vu, C. Laber, and W. Owen, "A very high precision 500-nA CMOS floating-gate analog voltage reference," *IEEE JSSC*, vol. 40, no. 12, pp. 2364–2372, December 2005.
- [77] B. Razavi and B. Wooley, "Design techniques for high-speed, high-resolution comparators," *IEEE JSSC*, vol. 27, no. 12, pp. 1916–1926, December 1992.
- [78] R. J. Baker, *CMOS Mixed-Signal Circuit Design*. Piscataway, NJ: Wiley–IEEE Press, 2003.
- [79] P. Furth and A. G. Andreou, "A design framework for low power analog filter banks," *IEEE TCAS I*, vol. 42, no. 11, pp. 966–971, November 1995.
- [80] *MOSIS Scalable CMOS (SCMOS) Design Rules rev. 8.00*, MOSIS, Oct. 2004, <http://www.mosis.org/Technical/Designrules/scmos/scmos-main.html>.
- [81] R. Jaeger, "Comments on 'An optimized output stage for mos integrated circuits'," *IEEE JSSC*, vol. SC-10, no. 1, pp. 185–186, June 1975.
- [82] E. Rodriguez-Villegas and H. Barnes, "Solution to trapped charge in FGMOS transistors," *Electronics Letters*, vol. 39, no. 19, pp. 1416–1417, September 2003.
- [83] G. Al-Rawi, "A new offset measurement and cancellation technique for dynamic latches," in *Proc. IEEE ISCAS*, vol. 1, May 2002, pp. 149–152.
- [84] D.-L. Shen and T.-C. Lee, "A 6-bit 800-MS/s pipelined A/D converter with open-loop amplifiers," *IEEE JSSC*, vol. 42, no. 2, pp. 258–268, Feb 2007.
- [85] P. Figueiredo and J. Vital, "Analysis of the averaging technique in flash ADCs," in *Proc. IEEE ISCAS*, vol. 1, May 2003, pp. 849–852.
- [86] O. Carnu and A. Leuciuc, "Optimal offset averaging for flash and folding A/D converters," in *Proc. IEEE ISCAS*, vol. 1, May 2004, pp. 133–136.
- [87] G. Geelen, "A 6b 1.1 Gsample/s CMOS A/D converter," in *IEEE Int. Solid-State Circuits Conf.*, May 2001, pp. 128–129.
- [88] U.-K. Moon and B.-S. Song, "Background digital calibration techniques for pipelined ADCs," *IEEE TCAS II*, vol. 33, no. 12, pp. 102–109, February 1997.
- [89] D. Fu, K. Dyer, S. Lewis, and P. Hurst, "A digital background calibration technique for time-interleaved analog-to-digital converters," *IEEE JSSC*, vol. 33, no. 12, pp. 1904–1911, December 1998.

- [90] S. Jamal, D. Fu, P. Hurst, and S. Lewis, "A 10b 120MSample/s time-interleaved analog-to-digital converter with digital background calibration," in *IEEE Int. Solid-State Circuits Conf.*, vol. 1, February 2002, pp. 172–457.
- [91] X. Wang, P. Hurst, and S. Lewis, "A 12-bit 20-Msample/s pipelined analog-to-digital converter with nested digital background calibration," *IEEE JSSC*, vol. 39, no. 11, pp. 1799–1808, November 2004.
- [92] C.-C. Huang and J.-T. Wu, "A statistical background calibration technique for flash analog-to-digital converters," in *Proc. IEEE ISCAS*, vol. 1, May 2004, pp. 125–128.
- [93] Y. Wong, M. Cohen, and P. Abshire, "A 1.2GHz comparator with adaptable offset in 0.35 μ m CMOS," *IEEE TCAS I*, 2007, in review.
- [94] *Federal Standard 1037C: Glossary of Telecommunications Terms*, The Institute for Telecommunication Sciences, Oct 2006, <http://www.its.bldrdoc.gov/fs-1037/fs-1037c.htm>.
- [95] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO₂," *J. App. Phys.*, vol. 40, no. 1, pp. 278–283, 1969.
- [96] L. Larcher, P. Pavan, and A. Maurelli, "Flash memories for SoC: An overview on system constraints and technology issues," in *5th Int. Workshop on System-on-Chip for Real-Time App.*, 2005, pp. 73–77.
- [97] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Info. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct 1998.
- [98] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, New York: John Wiley & Sons, 1991.
- [99] M. Burns and G. W. Roberts, *An introduction to mixed-signal IC test and measurement*. New York, New York: Oxford University Press, 2001.
- [100] D. Draxelmayr, "A 6b 600MHz 10mW ADC array in digital 90nm CMOS," in *IEEE Int. Solid-State Circuits Conf.*, February 2004, pp. 264–265.
- [101] X. Jiang and M.-C. Chang, "A 1-GHz signal bandwidth 6-bit CMOS ADC with power-efficient averaging," *IEEE JSSC*, vol. 40, no. 2, pp. 532–535, Feb 2005.
- [102] C. Sandner, M. Clara, A. Santner, T. Hartig, and F. Kuttner, "A 6-bit 1.2-GS/s low-power flash-ADC in 0.13- μ m digital CMOS," *IEEE JSSC*, vol. 40, no. 7, pp. 1499–1505, July 2005.
- [103] P. Scholtens and M. Vertregt, "A 6-b 1.6-Gsample/s flash ADC in 0.18- μ m CMOS using averaging termination," *IEEE JSSC*, vol. 37, no. 12, pp. 1599–1609, December 2002.

- [104] K. Uyttenhove and M. Steyaert, "A 1.8-V 6-bit 1.3-GHz flash ADC in 0.25- μ m CMOS," *IEEE JSSC*, vol. 38, no. 7, pp. 1115–1122, July 2003.
- [105] A. Pavasovic, A. Andreou, and C. Westgate, "Characterization of subthreshold MOS mismatch in transistors for VLSI systems," *Journal of VLSI Signal Processing*, vol. 8, no. 1, pp. 75–85, July 1994.
- [106] O. Yadid-Pecht and R. Etienne-Cummings, *CMOS imagers: from phototransduction to image processing*. Boston, MA: Kluwer, 2004.
- [107] R. H. Nixon, S. E. Kemeny, C. O. Staller, and E. R. Fossum, "256 \times 256 CMOS Active Pixel Sensor Camera-on-a-Chip," *IEEE JSSC*, vol. 31, pp. 2046–2050, December 1996.
- [108] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Z. Zhang, "A foveated silicon retina for two-dimensional tracking," *IEEE TCAS II*, vol. 47, no. 6, pp. 504–517, June 2000.
- [109] A. Pesavento and C. Koch, "A CMOS imager with focal-plane computation for feature detection," in *Proc. IEEE ISCAS*, vol. 3, May 2001, pp. 624–627.
- [110] P. Dudek and P. Hicks, "A general-purpose processor-per-pixel analog SIMD vision chip," *IEEE TCAS I*, vol. 52, no. 1, pp. 13–20, January 2005.
- [111] M. A. Karim, *Electro-optical Devices and Systems*. Boston, MA: PWS-Kent Pub. Co., 1990.
- [112] E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operations," *IEEE JSSC*, vol. 12, no. 3, pp. 224–231, June 1977.
- [113] E. R. Fossum, "Active pixel sensors: are CCDs dinosaurs?" in *Proc. SPIE Vol. 1900, p. 2-14, Charge-Coupled Devices and Solid State Optical Sensors III, Morley M. Blouke; Ed.*, July 1993, pp. 2–14.
- [114] S. Kleinfelder, S. Lim, X. Liu, and A. El Gamal, "A 10000 frames/s CMOS digital pixel sensor," *IEEE JSSC*, vol. 36, pp. 2049–2059, December 2001.
- [115] T. Delbruck and C. Mead, "Adaptive photoreceptor with wide dynamic range," in *Proc. IEEE ISCAS*, vol. 4, May 1994, pp. 339–342.
- [116] P. B. Catrysse, X. Liu, and A. El Gamal, "QE reduction due to pixel vignetting in CMOS image sensors," in *Proc. SPIE Vol. 3965, p. 420-430, Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications, Morley M. Blouke; Nitin Sampat; George M. Williams; Thomas Yeh; Eds.*, May 2000, pp. 420–430.
- [117] Y. A. Petrov and A. V. Popple, "Effects of negative afterimages in visual illusions," *Optical Society of America Journal A*, vol. 19, pp. 1107–1111, June 2002.