Variability Characterization of Imaging Readout Integrated Circuits in Deeply Scaled CMOS

Amy Whitcombe
Borivoje Nikolic, Ed.

Electrical Engineering and Computer Sciences
University of California at Berkeley

December 1, 2016

Variability Characterization of Imaging Readout Integrated Circuits in Deeply Scaled CMOS
by Amy Whitcombe

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

_____

Borivoje Nikolic
Research Advisor

_____

Date

* * * * * *

_____

Elad Alon
Second Reader

_____

Date

Contents

Abstract

As CMOS technologies scale to deep submicron technology nodes, improvements in transistor density and cost due to scaling inevitably come with degraded analog performance and increased device variability. Variation presents a major challenge to reliably building large arrays of mixed-signal circuits in nanometer scale CMOS processes. To determine the feasibility of creating highly sensitive readout integrated circuits (ROICs) for imaging detector arrays in 28 nm bulk CMOS, this work involves developing a chip to characterize performance variability in areas specifically relevant to image sensors: leakage current and random telegraph signaling (RTS) noise. High leakage current in crucial per-pixel reset switch devices can lead to pixel failure, while low-frequency RTS noise can cause pixels to blink erroneously. In imager arrays with tens of thousands of pixels, devices operating at extreme 3σ variation levels may result in dead or defective pixels. To improve the likelihood of measuring this behavior experimentally and obtain statistically meaningful results, the test chip includes nearly 7,000 devices under test (DUTs) for each type of device characterized. NMOS and PMOS devices with multiple threshold voltage flavors and dimensions of interest to digital and analog chip designs are included. The characterization platform consists of 4,608 pairs of test device cells and measurement circuit cells arrayed using a typical ROIC framework. By applying well-tested techniques in ROIC design to the variability measurement space, device measurements can be performed mostly in the digital domain, and a statistically significant subset of RTS noise data can be collected quickly. Because high-performance imagers are often cooled to cryogenic temperatures to reduce noise, the test chip was characterized from room temperature down to 78 K. Initial measurements indicate that while leakage current reduces significantly with temperature, RTS noise is still a concern down to cryogenic temperatures.

Acknowledgements

Chapter 1

Introduction

CMOS technology scaling driven by Moore's Law has enabled the development of increasingly complex yet inexpensive integrated circuits within the past few decades. In deep submicron technology nodes, however, the benefits of scaling have been limited by high design and fabrication costs, power density concerns, and analog device performance degradation. In advanced semiconductor technologies where minimum sized transistors may contain fewer than hundreds, or even tens, of electrons within a channel, device performance has become especially susceptible to random variation introduced during fabrication. Coupled with increasing device densities on chip, this performance variation can considerably heighten the probability of failure for large mixed-signal systems, such as readout integrated circuits (ROICs) for imaging detectors. Accurate technology variability statistics are therefore necessary to prevent wasteful overdesign and to realistically evaluate the feasibility of using the newest semiconductor technologies to construct mixed-signal systems. Moreover, large-scale device characterization studies can help provide information about the fundamental processes underlying technology variability.

For infrared imaging detectors, deep submicron technologies have increased transistor density to the point that analog to digital conversion may be implemented within a single pixel. In turn, this can enhance detector sensitivity and allow image processing algorithms to be implemented directly on-chip. [1–3] Moving to smaller technology nodes can also enable infrared (IR) detection at shorter wavelengths that are useful for low-light sensing applications. Nevertheless, the failure of a handful of devices in ROICs with upwards of tens of thousands of pixels can render some pixels useless and cause unwanted degradation in image quality. This is especially a cause for concern in high-performance detector applications, in which imagers are typically cooled to cryogenic temperatures (78 K) to reduce noise.

While building imagers in nanoscale technologies may enable per-pixel image processing and enhance robustness to noise by converting information directly to the digital domain, the performance degradation and enhanced variability associated with shrinking device size presents a growing concern. As a result, this research focuses on characterizing performance variation in a 28 nm bulk CMOS technology to determine the feasibility of implementing precision ROICs in shrinking technology nodes.

1.1 Scope of research

While detector readout integrated circuits (ROICs) have been demonstrated in technologies down to 65 nm CMOS [2], they have yet to be demonstrated in 28 nm. Moving to smaller process nodes can improve detector sensitivity and allow signal processing functions to be performed directly within the sensor; however, heightened performance variation also sacrifices reliability. To evaluate

the feasibility of implementing a highly sensitive detector in deep submicron processes, this research focuses on the development and characterization of a test chip in 28 nm bulk CMOS to characterize statistical variation in two major areas of concern to image sensors. Specifically, this work focuses on quantifying variation in leakage current — transistor drain current that flows even when the device's gate to source voltage $|V_{GS}|$ is 0 V — and random telegraph signaling (RTS) noise — discrete jumps in a device's drain current with time as the gate oxide traps and releases carriers, shifting the threshold voltage (V_{TH}) of a device.

1.2 Prior work

Comprehensive threshold voltage, random telegraph noise, and leakage current variability studies have previously been published for 90 nm and 45 nm CMOS, but only limited variability information is available for technology nodes at 28 nm CMOS and below. [4,5] Moreover, prior experimental work has focused on either simplified test structures such as ring oscillators designed to capture general performance metrics [5] or addressable device arrays with analog switches that require four-point Kelvin sensing techniques to accurately characterize both current and voltage. These approaches require long design times and cumbersome testing setups with separate bench supplies for each device terminal. [6–8] To address this issue, some prior work [4, 9, 10] has demonstrated custom technology characterization chips with fully digital interfaces, including a complete capacitance vs. voltage (C-V) and current vs. voltage (I-V) characterization chip with an integrated digital to analog converter (DAC) and analog to digital converter (ADC). Additional efforts have also integrated on-chip ADCs for highly sensitive leakage current measurements, and leveraged ring oscillators to characterize the dominant variability sources (within die, die to die, and wafer to wafer) in a 45 nm process. [5] This work uses an array of integrated ADCs to achieve high measurement throughput for characterizing RTS noise with long time constants.

Prior technology variation studies have largely focused on understanding digital performance reliability at room temperature. However, high-performance imaging ROICs are often cooled to cryogenic temperatures to reduce noise, so the effect of temperature on device performance must also be considered. Earlier research has confirmed that reducing temperature can enhance some aspects of transistor performance, resulting in higher carrier mobility and sharper subthreshold swing which boosts the ratio of ON-state to OFF-state current in a transistor. [11,12] Maintaining circuit functionality at cryogenic temperatures and below, however, can be challenging due to the risk of freeze-out effects that can alter transistor operation, so prior efforts have used careful design techniques to build an analog to digital converter that can operate from 4.2 K to 300 K. [13] From a technology characterization perspective, some previous experimental results in 90 nm CMOS have shown that transistor threshold mismatch actually increases with temperature. [14] However, no similar technology performance or variability studies have been performed for 28 nm CMOS.

This research aims to build on prior work by investigating new methods of variability characterization, applying well-tested techniques in ROIC design to the technology characterization space. It also examines device performance as a function of temperature to provide a clearer understanding of how variability may impact cryogenic temperature applications, and to provide insight into the physical sources of RTS noise.

1.3 Thesis organization

This report first discusses background in CMOS imager design and technology-dependent performance limitations in Chapter 2. Subsequently, the test chip design is described in Chapter 3, which provides a high-level overview of the design and then discusses in detail major blocks within the measurement system. The measurement procedure and preliminary test chip results are detailed in Chapter 4. Finally, Chapter 5 describes main conclusions from this work and future directions.

Chapter 2

Technology Limitations in CMOS Imaging Circuits

Advances in semiconductor manufacturing capabilities enabling higher transistor densities have helped increase the capability and reduce the cost of CMOS imagers within the past few decades. Fabrication improvements have allowed imaging technology to become ubiquitous in a myriad of sensing applications ranging from portable electronics to infrared detector arrays for medical and defense applications, and the field has expanded into a multi-billion dollar industry. [15–17]

Future application areas, such as gigapixel photography, could also benefit from higher levels of technology scaling. For instance, the AWARE-2 imager developed at Duke University uses an array of 98 microcameras with dedicated lenses to build a 1 gigapixel imager. However, this system expends 430 W while an image is taken and occupies a $0.75 \times 0.75 \times 0.5$ m^3 volume, largely to facilitate heat dissipation and electronic control. [18] An integrated approach with custom electronics could allow more of this processing to be accomplished on-chip in a smaller area, requiring less energy for heat dissipation.

While advances in CMOS technology scaling have helped drive many advances in integrated image sensors, the benefits of technology scaling will not continue as device sizes keep shrinking. Many of the challenges in scaling digital circuits to deep submicron technology nodes — increased leakage current, low-frequency noise, and process variability — are exacerbated in imaging applications due to the large-scale, mixed-signal nature of these detectors. This chapter provides the background needed to help understand these challenges. First, it illustrates the benefit of digital focal plane array technology over conventional analog conversion techniques, and then discusses two major sources of performance degradation at nanoscale nodes: random telegraph noise and leakage current.

2.1 Digital focal plane array technology

In CMOS imagers, a photosensitive element generates current proportional to the amount of light it absorbs at a particular wavelength. As summarized in Fig. 2.1, this current is typically converted to a digital signal that represents the detected image. The implementation of the process shown in Fig. 2.1 can vary greatly depending upon the range of wavelengths to be detected.

Silicon is sensitive to visible light, allowing photodiodes to be integrated directly on chip for conventional imaging applications. As a result, each imager pixel contains a photodiode with only a small amount of supplemental circuitry to maximize the ratio of photosensitive area to total pixel size (fill factor). However, these sensors do not benefit much from technology scaling past the 180 nm technology node. Camera lens diffraction limits restrict minimum pixel sizes to 5μm$\times5\mu$m, so
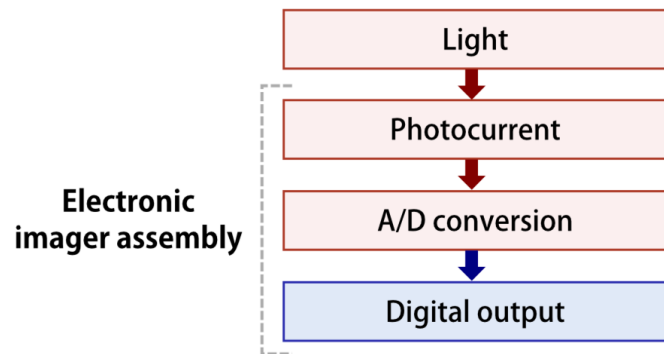
Figure 2.1: General processing flow of an imaging sensor.
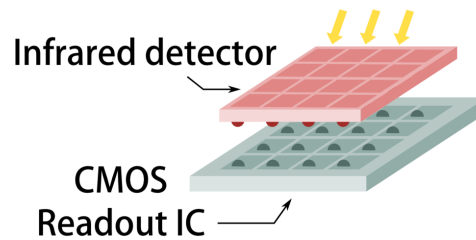
Figure 2.2: Infrared sensors require a separate detector array fabricated from a low-bandgap material, such as HgCdTe, to be bonded on top of a conventional CMOS readout IC.

while scaling can improve pixel fill factor or enable additional functionality within a single pixel, it cannot be used to improve pixel resolution. Moreover, modifications to fabrication processes intended to boost transistor performance at nanoscale technology nodes often degrade photodiode sensitivity. Heightened doping levels needed to mitigate drain-induced barrier lowering effects will reduce diffusion length. The need to use silicides during processing to reduce resistance also reduces the substrate's responsivity to light, and parasitic leakage currents further degrade detector performance and serve as additional noise sources which degrade imager performance. Using smaller devices also results in a larger sensitivity to crosstalk between adjacent photodiodes. [19] Therefore, process scaling does not provide significant improvements for sensors built directly in silicon.

In contrast to visible light detectors, infrared sensors typically require a non-silicon photodetector array to be fabricated and bonded on top of a standard CMOS readout integrated circuit (ROIC), as depicted in Fig. 2.2, because silicon is not sensitive to infrared light. The major barrier to implementing visible light sensors in technologies below 180 nm is the tradeoff between photodiode sensitivity and transistor performance; however, this is not a restriction for infrared detector ROICs because the photodetector is fabricated separately. Because of this, recent trends in IR focal plane array (FPA) technology have moved away from typical analog sensing techniques and towards digital FPA implementations with per-pixel analog to digital converters (ADCs). Fig. 2.3 compares the amount of processing that occurs on-pixel between analog and digital FPAs.

In a conventional analog FPA, photocurrent is integrated onto a capacitor for a fixed amount of time and stored using a sample and hold circuit. Individual pixel voltages are read out using an analog multiplexer and per-pixel source follower buffers, as shown in Fig. 2.4. From there, the array of analog voltage readings may be processed and converted to a digital reading using a separate ADC at the row or column level.

Using this analog approach, the detector storage capacity (measured in number of electrons,
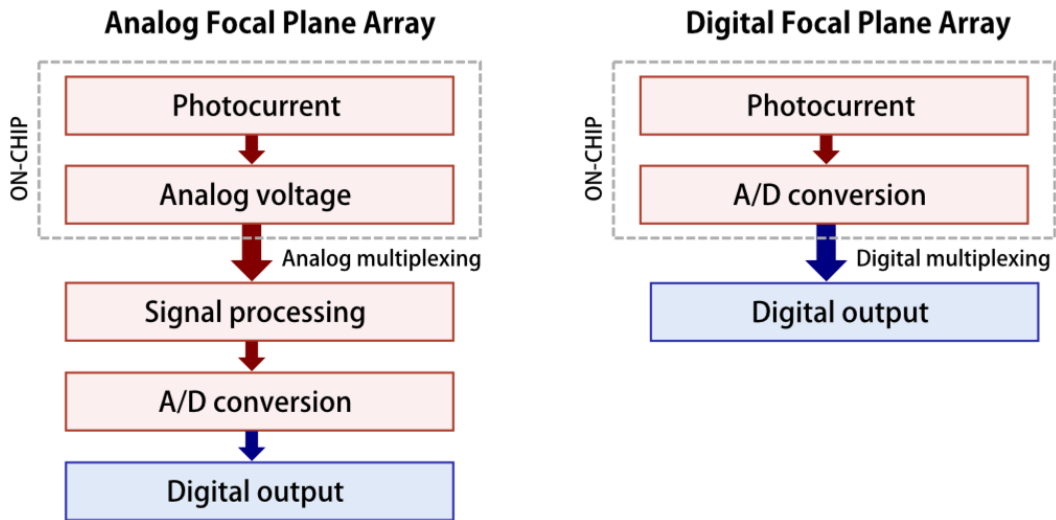
Figure 2.3: Comparison of processing levels in analog and digital FPAs (adapted from [1]).
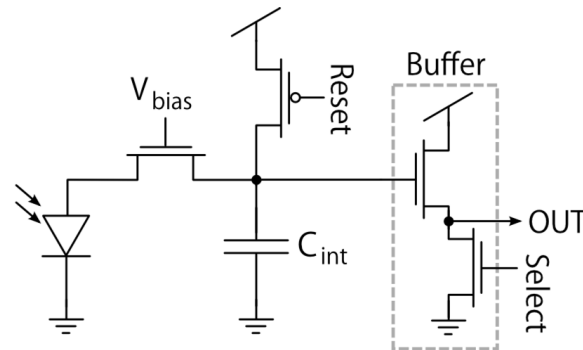
Figure 2.4: Example schematic of an analog focal plane array pixel.
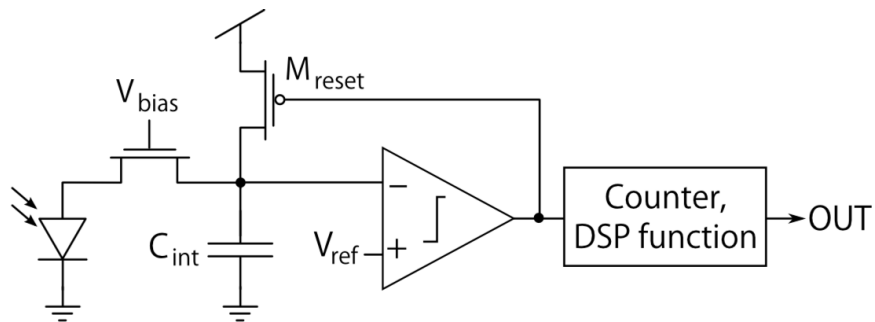
Figure 2.5: Example schematic of an digital focal plane array pixel.

N_{e-}) is set by the maximum voltage drop allowed by a given process technology (V_{max}), the integration capacitance (C_{int}), and charge on an electron (e^-):

$$N_{e^-} = C_{int} V_{max} \frac{6.24 \times 10^{18} e^-}{\text{Coulomb}}$$

Given a maximum of 5V across a 0.8 pF integration capacitance for a detector with 30 μm pitch, the number of photoelectrons is typically limited to <25 million for analog FPAs. [2] This maximum well capacity in turn determines the maximum signal to noise ratio (SNR) of the detector. Assuming a well-designed detector limited only by the shot noise of the photodiode, the signal power will be set by the maximum number of signal photoelectrons, N_{e^-}, while the noise power will be set by $\sqrt{N_{e^-}}$. The corresponding SNR upper bound is therefore $\sqrt{N_{e^-}}$. This does not account for the fact that keeping the pixel readings in the analog domain increases opportunities for noise to be introduced in the imager during voltage transfer.

In contrast, a digital FPA converts analog photocurrent into a digital reading at the pixel level, typically using a current to frequency converter and a digital counter, as shown in Fig. 2.5. Higher counter resolution and longer integration times can increase the maximum number of detectable photoelectrons significantly. Assuming 16 bits of resolution and a 1V drop across a 1 fF parasitic integration capacitance, the DFPA can count a maximum of 230 million photoelectrons — nearly an order of magnitude increase over the capacity of an analog FPA. [2] In addition to this improvement in SNR, the DFPA can be configured to apply digital signal processing (DSP) algorithms directly during readout. Moreover, converting the pixel readings directly to the digital domain significantly reduces the probability of introducing additional noise during processing.

Building DFPA ROICs in finer-resolution process nodes may enable further resolution improvements and support more in-pixel computation in advanced sensors. Recently, work at MIT Lincoln Labs has demonstrated digital ROICs in 65 nm and 90 nm CMOS for long-wavelength IR (LWIR) imaging focal plane arrays [1–3, 20]. This approach has allowed demonstrations of up to 640 × 480 pixel arrays with a 20-30 μm pitch, and up to 21 bits per pixel. Moving to sub-32 nm technologies can enable a 4028 × 4028 array to be built with $\leq 12\mu$m pixel pitch and more than 28 bits per pixel [2]. However, while significant advances can be made by moving to 32 nm and smaller manufacturing processes, deep submicron technologies demonstrate large levels of process variability and degradation in analog performance. Among these concerns are random telegraph signaling (RTS) noise, a low-frequency noise phenomenon that results in discrete changes in MOSFET drain current, and high leakage current levels that may discharge internal nodes over long integration times.

2.2 Random telegraph signaling noise

Random telegraph signaling (RTS) noise is a low-frequency noise phenomenon that results in discrete jumps in the drain current of a MOSFET. This discrete shift in current occurs when charge carriers are caught and released in "traps" — such as positively charged dangling hydrogen bonds — within the gate oxide or at the interface between the gate oxide and the FET channel, as shown in Fig. 2.6. Fig. 2.7 shows a representative plot of drain current as a function of time due to RTS noise in a device with a single oxide interface trap. When an oxide trap is filled, the total charge within the MOSFET channel will decrease, which raises the effective threshold voltage of the transistor. In turn, this reduces the channel current by a fixed amount. The current returns to its previous level when the electron is released back into the channel. Each trap can be characterized by the mean time to capture $\overline{\tau_c}$, measured as the time between generating an empty trap and filling the same trap, and time to emission $\overline{\tau_e}$, the time that an electron is contained within
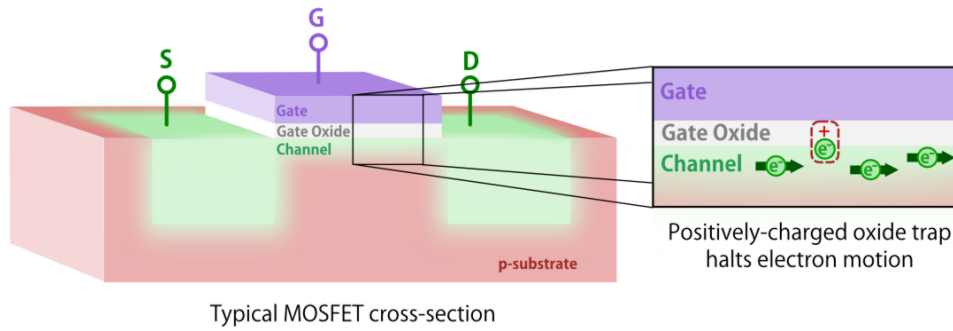
Figure 2.6: Illustration of occupied, positively-charged oxide trap restricting the flow of current between the MOSFET's drain and source.
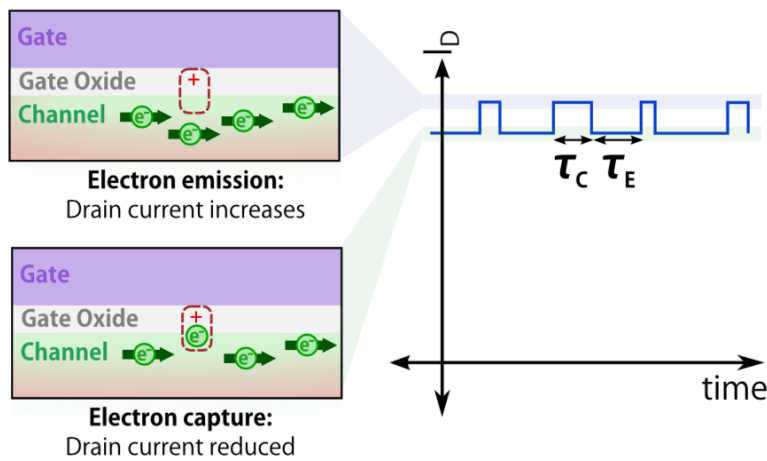
Figure 2.7: Basic illustration of I_D vs. time due to only RTS noise introduced by a single trap.

the gate oxide. Each trap also has a characteristic magnitude ΔI_D, which can be mapped to a corresponding change in threshold voltage.

2.2.1 Impact of RTS noise

As device sizes shrink, fewer carriers exist within a MOSFET channel, and the effect of trapping and de-trapping a single electron in the gate oxide becomes more pronounced. The change in threshold voltage ΔV_{th} introduced by RTS caused by a single carrier trap is roughly given by the following expression [21]:

$$\Delta V_{th} = \frac{e^-}{WLC_{ox}}$$

To first order, RTS noise amplitude is inversely proportional to a transistor's channel area ($L \times W$) and gate capacitance per unit area (C_{ox}). Under conventional scaling laws, this would maintain constant RTS as the oxide thickness scales with device dimensions [22]. However, the suspension of t_{ox} scaling due to gate tunneling leakage concerns has led to an increase in RTS noise in modern manufacturing processes — in the deep submicron regime, RTS noise is predicted to have a larger influence than variations due to random dopant fluctuation (RDF) [21, 23].

RTS noise is beginning to impact many nanoscale CMOS applications. In the digital space, RTS has been attributed to logic delay fluctuations of up to 16% in a 40 nm technology operating

at low supply voltages frequently used to reduce power consumption [21]. In contrast, at the 65 nm technology node, ring oscillator frequency fluctuations due to RTS of only 0.48% were observed, even at low supply voltages. [24] Because digital circuits must perform with adequate timing margins even under worst-case operating conditions, this sharp increase in RTS noise presents a growing concern as technologies shrink further. RTS noise also presents a large challenge for SRAM, where small device sizes are required to achieve high memory density. To accurately retain data, memory cell stability margins must be large enough to contend with random shifts in threshold voltage and drain current resulting from RTS noise. Because stability margins are already low in technologies with shrinking supply voltages and significant process variability, the additional impact of RTS noise presents a significant challenge to developing reliable scaled SRAM. [25, 26]

In conventional CMOS image sensors, RTS noise has been identified as a main contributor to pixel read noise, and can result in perceptible image flickering under low-light conditions. [27, 28] While correlated double sampling (CDS) techniques have been developed to reduce thermal noise by sampling the output voltage twice during a single integration period, RTS noise in the source follower device of an analog pixel will appear as a discrete jump in the pixel's output voltage if trap occupancy states change between CDS samples. Because the probability of trap occupancy is related to the capture and emission time constants of a trap, RTS noise is dependent upon the CDS time period. The impact of RTS noise typically increases when the time between CDS samples is larger than the typical trap time constant. [29] This behavior has been exploited to reduce RTS noise by modifying the CDS sampling intervals in accordance with trap capture and emission time constants. Other RTS mitigation methods propose biasing the devices prior to readout in order to force a known trap occupancy state. [30]

For the digital FPA ROIC shown in Fig. 2.5, RTS noise in the direct injection transistor which fixes the photodiode bias voltage will introduce fluctuations in the voltage drop maintained across the photodiode. In turn, this will alter the photodiode's responsivity in a manner that is indistinguishable from jumps in signal current. Given that trap time constants can be on the order of seconds, this change in current may cause blinking pixels. For imaging applications that require high accuracy, such as astronomy or medical imaging, these effects may be unacceptable.

2.2.2 Temperature dependence

RTS noise is temperature dependent because the addition of thermal energy facilitates the trapping and de-trapping process. Previous studies of RTS noise in high-κ/metal gate transistors has demonstrated that as temperature increases from 20° C to 60° C, the mean time to capture ($\overline{\tau_C}$) and emission ($\overline{\tau_E}$) of RTS traps can reduce by an order of magnitude. While the absolute magnitude of $\overline{\tau_C}$ and $\overline{\tau_E}$ shifted with temperature, the relationship between the two parameters with respect to bias voltage remained largely the same over this temperature range. [26, 31]

A theoretical model for the relationship between $\overline{\tau_C}$, $\overline{\tau_E}$, temperature, and other device parameters is presented in the next section, in Eqn. 2.3. This model predicts that at low temperatures, the ratio between $\overline{\tau_C}$ and $\overline{\tau_E}$ will increase. As a result, oxide traps will remain mostly empty — electrons lack the thermal energy required to enter the trap. The required level of thermal energy will depend on the trap depth, which will vary randomly. Some prior studies have found that RTS noise is evident in leakage current even down to 4.2K. [32] Additional studies have found RTS noise is still significant in amplitude at temperatures down to 100K. [33]

2.2.3 Extracting RTS parameters

As previously described, the major measurable characteristics of RTS noise are the mean capture time $\overline{\tau_c}$, mean emission time $\overline{\tau_e}$, and trap amplitude ΔI_D. Many physical parameters associated with a given trap can be extracted from these values. The ratio between the mean capture and

emission times, known as the mark-space ratio, can be used to extract the probability of trap occupancy α. If $\overline{\tau_e} \gg \overline{\tau_c}$, only a small amount of time is required to fill a trap relative to the amount of time it remains occupied, meaning that α approaches unity. Conversely, if $\overline{\tau_c} \gg \overline{\tau_e}$, α approaches 0 because an electron is emitted from the trap quickly after being captured; α can be expressed as:

$$\alpha = \frac{1}{1 + \frac{\overline{\tau_c}}{\overline{\tau_e}}} \tag{2.1}$$

The mark-space ratio can be written in terms of the trap energy E_T, Fermi level E_F, and trap degeneracy g [22, 23]:

$$\frac{\overline{\tau_c}}{\overline{\tau_e}} = ge^{\frac{E_T - E_F}{k_B T}} \tag{2.2}$$

This can be rewritten in terms of a constant K characteristic of each trap, the charge on an electron q, surface potential Ψ_s, and trap depth relative to the gate oxide thickness $\frac{z}{t_{ox}}$: [22]

$$\ln \frac{\overline{\tau_c}}{\overline{\tau_e}} = K - \frac{q}{k_B T} \left[\left(1 - \frac{z}{t_{ox}}\right) \Psi_S + \frac{z}{t_{ox}} V_G \right] \tag{2.3}$$

The above relationship demonstrates that the relative trap depth can be extracted from the partial derivative of $\ln \frac{\overline{\tau_c}}{\overline{\tau_e}}$ with respect to V_G, provided the dependence of Ψ_S on V_G is small. Moreover, the natural logarithm of the mark space ratio is inversely proportional to temperature, suggesting that traps remain mostly empty as devices are cooled.

While multiple theoretical models exist to predict trap magnitude, few can accurately predict the $\Delta I_D / I_D$ measured from experimental data over a full range of applied bias voltages [22]. As a result, a common empirical metric for trap magnitude is to translate the measured change in drain current to an effective change in threshold voltage ΔV_{TH} using the device transconductance g_m extracted from the IV characteristic of the transistor. Using this technique, $\Delta V_{TH} = \Delta I_D / g_m$.

2.2.4 Measuring RTS noise

The most straightforward means of characterizing RTS noise is to measure the drain current of a MOSFET over time under static biasing conditions (constant V_{DS} and V_{GS}). However, the time required to observe traps buried deep within the gate oxide may be prohibitive for measuring RTS in a statistically significant set of devices; trap time constants as high as 10 minutes have been observed in a 45 nm CMOS process [22]. The relationship between probability of trap occupancy and gate bias voltage can be used to force traps to a known occupancy state prior to conducting measurements. This can be accomplished simply by applying V_{DD} or 0V to the gate of the test device to force a trap to be occupied or empty, respectively. Fig. 2.8 demonstrates the test process required to characterize RTS using this technique.

Once an RTS waveform has been measured, the mean capture and emission time constants for each trap must be determined. Extracting τ_c and τ_e is trivial in the case of a single-trap device, but identifying the characteristics of multi-trap systems can be difficult. One common graphical method of extracting RTS levels is to plot a histogram of current samples, as demonstrated in Fig. 2.9. Histogram peaks can be used to identify distinct current levels caused by RTS even in the presence of sampling noise, though measurement discretization errors can limit the accuracy of peak identification [34]. As the number of traps increases, distinct levels become increasingly difficult to identify. Time-lag plots, in which the current measured at one time sample is plotted against the current measured in a previous sample, provide another means of identifying RTS levels [35, 36].
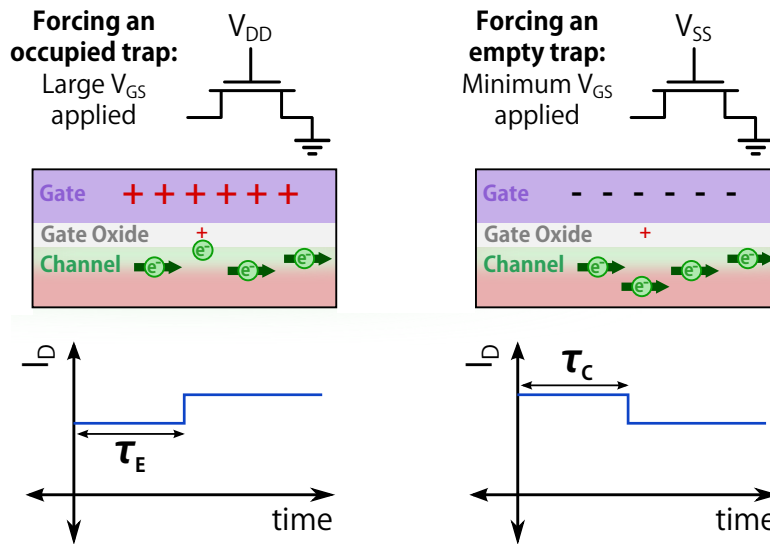
Figure 2.8: By maintaining a constant, large bias across a transistor, oxide traps can be forced to a "full" or "empty" state.

The rightmost graphs in Fig. 2.9 present a histogram of measurement peaks along the diagonal of a time lag plot, which can be used as a more accurate means of identifying distinct RTS levels. In addition to these graphical techniques, multi-trap RTS signals can be analyzed by applying hidden Markov models to RTS measurements. In this case, the underlying RTS states are treated as a signal obscured by measurement noise, and a probabilistic model is developed to determine the most likely sequence of state transitions. [26, 37]

2.3 Leakage current

Simplistic device models assume that the drain to source current of a transistor is zero if the applied gate to source voltage (V_{GS}) falls below the threshold voltage of the device, and that no current flows into the gate of a transistor. Realistically, leakage current can exist between any two MOSFET terminals, even when $V_{GS} = 0$V. One of the most common forms of leakage current is between the drain and source of the transistor, which is caused by the parasitic bipolar junction transistor (BJT) present in a MOSFET. Fig. 2.10 shows an example I/V characteristic of a MOSFET presented on a logarithmic scale. Below the transistor's threshold voltage, leakage current drops exponentially. With supply voltages shrinking as device dimensions shrink, the ratio between ON state and OFF state current also becomes smaller. This is especially problematic for low-power applications operating at reduced supply voltages.

While drain leakage is typically the largest source of leakage current in a transistor, gate leakage can also occur due to the tunneling of charge carriers from the gate to the channel through the gate oxide. Gate leakage is increasing at smaller process nodes where the gate oxide thicknesses scale with device dimensions to maintain constant electric field strength. New high-κ dielectric materials which increase electric field strength while keeping large oxide thicknesses have helped mitigate this problem, but gate leakage may still be a significant concern in analog ROICs designed to retain a fixed voltage on a capacitor for an extended time. This work focuses predominantly on characterizing drain to source leakage effects, but gate leakage parasitics must also be considered when designing the test chip to accurately measure small I_{DS} levels.
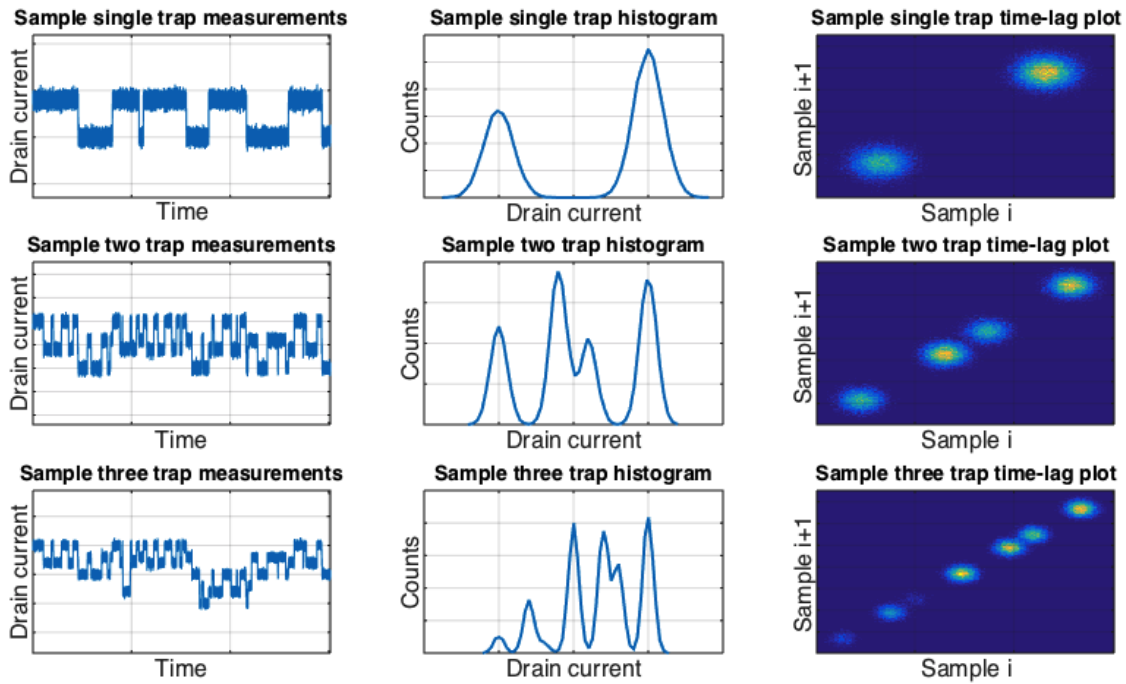
Figure 2.9: The above plots present randomly-generated single-trap (top), two-trap (center), and three-trap (bottom) RTS signals. Two typical methods for identifying RTS amplitude levels include creating a histogram of current measurement levels (center) and creating a time-lag plot (right), in which future drain current readings are plotted against previous readings.
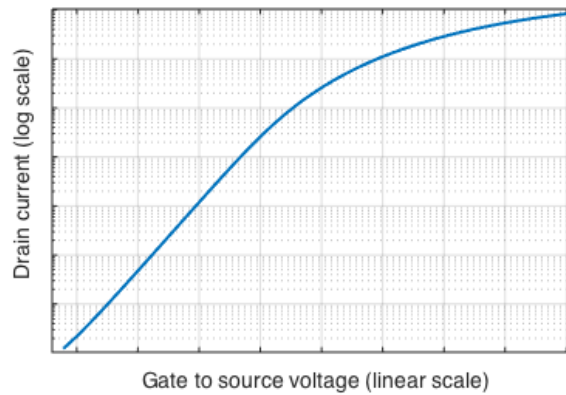
Figure 2.10: Sample MOSFET $I_{DS}vs.V_{GS}$ characteristic presented on a logarithmic scale.

2.3.1 Impact of leakage current

As transistor dimensions scale and supply voltages shrink, leakage current levels are increasing relative to the ON current of a device. In digital designs, smaller I_{ON}/I_{OFF} ratios reduce reliability margins, and the increase in leakage current coupled with larger transistor densities have raised the standby power consumption of digital circuits significantly. In some low-speed applications, this can be the largest source of power consumption in a chip.

For imaging applications, leakage current can degrade imager performance or, in extreme cases, introduce faulty pixels. In analog ROICs, leakage current can cause a sizeable loss in stored charge. Considering only leakage current due to gate oxide tunneling, a capacitor fabricated in a thick-oxide 28 nm process with 10 fF / μm^2 capacitance density and 100 fA / μm^2 of leakage current will lose 10 mV each millisecond. For video applications with integration times in the tens of milliseconds, this results in a significant degradation in signal level. Moreover, any variation in leakage current will cause variations at the the output of the imager.

In digital ROIC applications, the charge stored on the integration capacitor is quickly reset after exceeding a fixed threshold. This avoids the effect of charge loss due to leakage in analog imagers, which must maintain the sampled charge prior to readout. However, most digital ROICs use a single-slope ADC, in which photocurrent is integrated onto a capacitor, and the capacitor voltage is reset upon reaching a target threshold. In this case, leakage current in the reset device which surpasses the desired test current will prevent the capacitor voltage from integrating properly, causing a faulty pixel. For large imaging arrays containing tens or hundreds of thousands of pixels, multiple pixels are statistically likely to exhibit 3 σ variation levels that may cause faulty pixels. Accurately understanding leakage current variability can therefore help predict imager yield.

2.3.2 Temperature dependence

The model for drain leakage current (where μ_{eff} is the carrier mobility, C_{ox} is the oxide capacitance per unit area, m is the subthreshold slope factor — determined by the ratio between the depletion layer capacitance and C_{ox} — k is Boltzmann's constant, T is temperature, q is electron charge, and V_{TH} is the threshold voltage) is:

$$I_{DS,leak} = \mu_{eff}C_{ox}\frac{W}{L}(m-1)\left(\frac{kT}{q}\right)^2 e^{\frac{V_{GS}-V_{TH}}{mkT/q}}\left(1 - e^{-\frac{V_{DS}}{kT/q}}\right) \tag{2.4}$$

As shown in the above expression, drain to source leakage current exists even when $V_{GS}-V_{TH} \leq 0$ provided that a positive drain to source voltage is applied to the device. Moreover, leakage current decreases with temperature, provided $V_{GS} - V_{TH} < 0$. However, the temperature also sets the exponential dependence of leakage current upon device-dependent parameters such as V_{TH} and m. As a result, while reducing temperature may reduce average leakage current levels, it may also increase the variation in leakage current between devices. This is consistent with previous studies regarding device performance at cryogenic temperatures for 90 nm CMOS, in which mismatch levels were found to increase with temperature. [14]

2.3.3 Measuring leakage current

Because leakage currents must be characterized down to the femtoampere range, accurate leakage characterization requires careful test chip design. Some approaches to leakage current measurement use an array based scheme in which independent device terminals can be forced or sensed in the analog domain [6, 38]. However, sensing extremely low current levels requires a highly sensitive current meter and careful experimental setup. An alternate proposed method has used an analog CMOS imager array structure to generate a voltage proportional to leakage current by measuring

current integrated on very small test capacitances. However, accurate leakage characterization using this approach requires precise capacitance measurement and accurate array calibration [39]. In other work, an on-chip ADC has been used to measure leakage current by measuring the time required to integrate current across a fixed reference capacitor [5]. The work in this thesis uses a modified digital FPA ROIC structure to generate digital measurements corresponding to device leakage.

Chapter 3

Test Chip Design

The major goals for this chip are as follows:

1. Accurately characterize the statistical distribution of random telegraph signaling (RTS) noise and leakage current in a 28 nm bulk CMOS process to provide design insight for CMOS readout integrated circuits (ROICs).

2. Demonstrate a technology characterization platform that can streamline the variability characterization process.

To meet the first goal, the test chip contains thousands of test structures designed to measure both leakage current and RTS noise. To meet the second goal, the design utilizes a typical DFPA ROIC structure to convert test current to a digital output. This enables a fast readout process that can leverage well-tested ROIC techniques to perform variability measurements continuously, even as previous measurement results are being read.

The test chip details are discussed as follows: first, Section 3.1 presents a high-level overview of the architecture for the test chip. Section 3.2 subsequently provides an overview of the system and block level performance specifications required to obtain accurate RTS and leakage current measurements. Finally, more detail about the implementation of individual blocks used in the architecture are presented in 3.3. Specifically, Section 3.3.1 details the implementation of the current measurement block, which converts the current of a device under test (DUT) into a digital reading, while Section 3.3.2 discusses the implementation of the block of test devices (DUTs) that generates this test current.

3.1 Architecture overview

The characterization chip mirrors the design of a digital FPA ROIC. It contains a central array of test devices and measurement cells, with peripheral circuitry to address and decode individual test cells. Using a structure similar to a standard ROIC allows the digital interface circuitry to be repurposed from previous work. It also enables test chip readings to be processed in the same manner as a traditional digital image, accelerating the variability characterization process to address goal (2) from the chip design objectives.

Fig. 3.1 shows a high-level block diagram of the major chip components. At the core of the design is a 96×48 array of unit cell pairs dedicated to measuring either RTS noise or leakage current. Individual cell pairs are addressed using digital row and column decoders, and readings from each cell are propagated to the chip-level output using the row and column addresses and a digital multiplexer. A simple current mirror generates bias currents for calibrating individual pixel measurements.
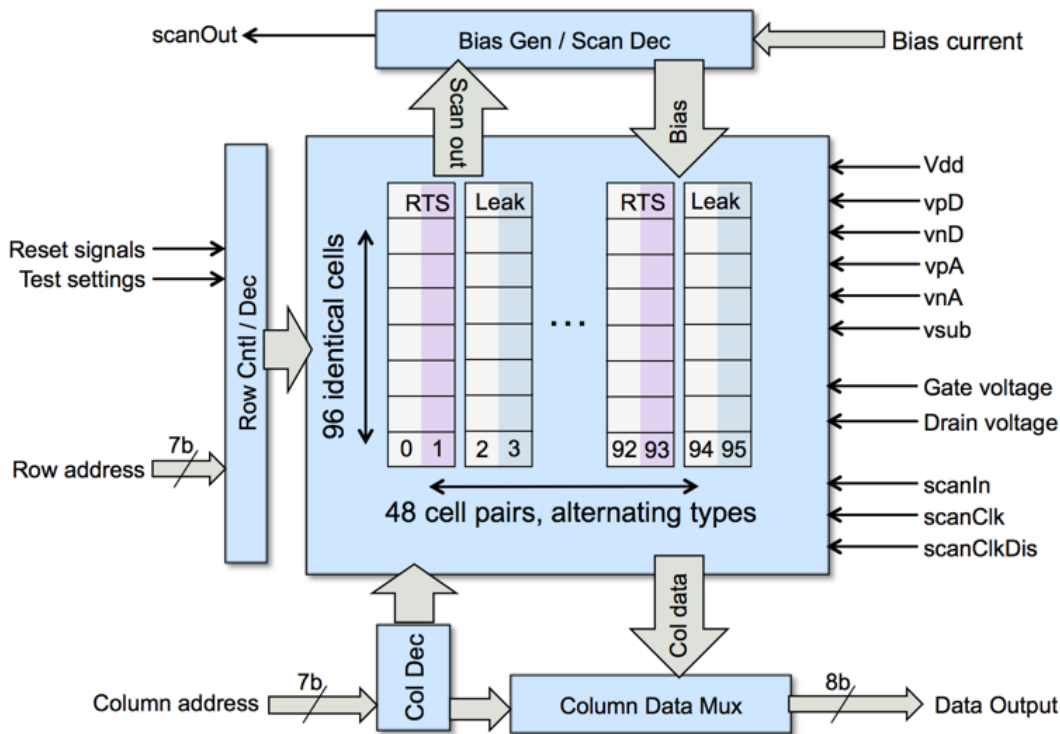
Figure 3.1: High-level schematic of characterization chip system.

DUT V_{TH} flavors	DUT dimensions (W/L)	DUT types	DUT Quantity
HVT, LVT, RVT	80 nm / 30 nm, 80 nm / 86 nm	NMOS	9,216 for each type
HVT, LVT, RVT	80 nm / 30 nm, 80 nm / 86 nm	PMOS	6,912 for each type

Table 3.1: Summary of device under test (DUT) types included on the test chip. Three threshold voltage types, two device dimensions, and thousands of test devices are included for both NMOS and PMOS DUTs.

While traditional ROICs contain an array of identical measurement cells, the variability test chip in this work includes an array of unit cell pairs. Each pair contains one "device under test" (DUT) cell, which uses a digital scan chain to generate a test current from selected test devices, and one measurement cell, which uses a current to frequency converter and counter to generate a digital reading proportional to the test current. The test chip contains 96 cell pairs per column and 48 cell pairs per row. There are two types of unit cell pairs: an RTS characterization pair, in which the V_{GS} of each DUT can be controlled independently to select individual test devices, and a leakage current characterization pair, in which V_{DS} modulation is used to control drain to source leakage current for DUTs which already have $V_{GS} = 0$V. Columns alternate between RTS and leakage test cell pairs to map across-die spatial variation effects for both phenomena.

Table 3.1 summarizes the device types included in the chip. Each DUT cell contains 24 NFETs and/or PFETs, with six unique device types. Only two dimensions are considered: minimum sized transistors for digital gates and longer devices for analog applications. While including further DUT dimensions would provide an opportunity to validate measurement results using known relationships between variability scaling and device area, including fewer dimensions allows more devices of each type to be incorporated within each DUT cell. This provides a larger dataset and improves the
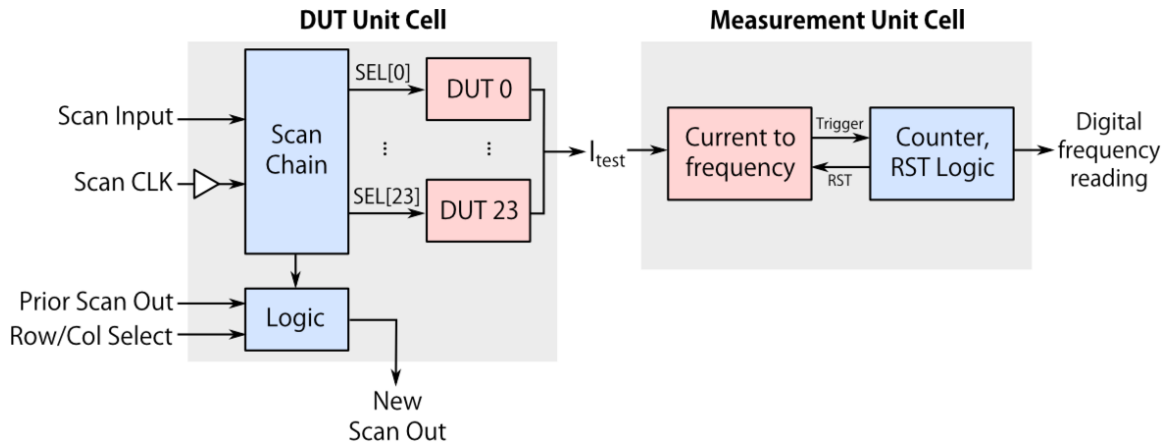
Figure 3.2: General structure of unit cell pair, containing one device under test (DUT) unit cell (left) and one measurement unit cell (right).

likelihood of capturing corner cases of interest, such as extreme leakage levels that would translate to defective pixels in an imager, and provides the opportunity to explore variability as a function of threshold voltage (V_{TH}) by testing different device flavors. The high V_{TH} (HVT), low V_{TH} (LVT), and regular V_{TH} (RVT) devices can be used to validate leakage current measurement trends, as leakage depends directly on V_{TH}.

Because deep submicron technologies are typically intended for digital design applications, understanding the performance variability of minimum sized devices especially prone to mismatch can help define realistic design margins and predict imager array yield. Understanding RTS noise in minimum sized devices is relevant to additional applications, such as SRAM cells that must use small transistors to achieve high memory density. For analog design applications, near minimum-sized devices can be combined in series and parallel to create devices with a larger effective width and length. As a result, performance characteristics at a single size can be used to infer variability characteristics at other dimensions.

Fig. 3.2 illustrates the unit cell pair design. Within the DUT unit cell, a scan chain generates 24 parallel control signals for each test device. An additional control signal provides the ability to select between NMOS and PMOS type DUTs. To test the scan chain for each pixel, row and column select signals generated by the top-level decoders are used to propagate the scan output of a single pixel through to the chip's `scanOut` output. The test current generated by the DUT unit cell is then converted to a digital reading using the measurement unit cell. In this cell, a current to frequency converter is used to generate a digital reset signal with a switching frequency proportional to the input current. As in a typical digital FPA pixel, the digital output is simply a frequency reading obtained by sampling a digital counter at a fixed time interval.

3.2 System-level specifications

To determine the minimum performance requirements needed to obtain accurate RTS and leakage current measurements, we must first understand typical characteristics of the two phenomena. Noise measurements require adequate temporal resolution with moderate absolute measurement accuracy, while leakage current measurements require high accuracy that may be accomplished by long integration times.

RTS noise is typically observed as $\approx 1 - 5\%$ jumps in drain current measurement when a device is operating at or near threshold, drawing current in the microampere range. Because

DUT type	Low leakage corner	Nominal leakage	High leakage corner
HVT 86nm NMOS	0.17	0.25	0.45
RVT 86nm NMOS	0.23	0.38	0.92
LVT 86nm NMOS	1.49	4.06	14.9
HVT 86nm PMOS	0.28	0.34	0.47
RVT 86nm PMOS	0.27	0.33	0.50
LVT 86nm PMOS	0.33	0.54	1.28
HVT 30nm NMOS	0.17	0.24	0.44
RVT 30nm NMOS	0.43	1.00	4.22
LVT 30nm NMOS	0.67	1.89	7.67
HVT 30nm PMOS	0.38	0.63	1.90
RVT 30nm PMOS	0.85	2.02	7.64
LVT 30nm PMOS	5.22	15.2	54.4

Table 3.2: Variation in expected leakage current levels for different device flavors and performance corners. The measurements are normalized to the nominal current generated by a 30nm RVT NMOS device.

much information about RTS is contained in the time domain, the sampling rate dictates the fastest detectable trap time constant. While RTS is typically a low-frequency phenomenon with long trap time constants that may last minutes, short trap time constants can last only fractions of milliseconds. For reasonable temporal resolution, we target a 20 kHz sample rate. Because the counter integrates continuously, the ADC sampling rate can be increased at the expense of resolution. To characterize the magnitude of RTS traps, the measurement system must be able to detect accurate changes in drain current relative to the nominal value. This work targets 0.2% resolution to detect the 1-5% jumps in drain current associated with RTS. To meet the 20 kHz sampling rate target, the measurement unit cell's current to frequency converter will therefore need to generate at least a 10 MHz pulse in response to the smallest measurable test current.

In contrast, sampling rate is of less importance to leakage current measurements. Leakage measurement requires sensing extremely small constant currents down to the picoampere range while minimizing the effect of noise. For leakage current characterization, the minimum detectable leakage current is set by the high-threshold (HVT) 80 nm / 86 nm device. Table 3.2 shows expected room-temperature leakage current levels at different performance corners, normalized to the leakage current of a regular-threshold 80nm / 30nm device. From the table, the measurement system must be able to characterize leakage current varying more than two orders of magnitude. Moreover, minimizing noise is crucial for leakage current characterization. Because integration speed is less critical, however, resolution can be increased by using longer integration times and noise can be mitigated through measurement averaging. In this design, we target an array that can sense signals down to the picoampere range to within 10% absolute accuracy between samples.

From an implementation perspective, the target unit cell area is $12\mu m \times 12\mu m$, with analog measurement unit cell area constrained to $1.5\ \mu m \times 12\mu m$. While power consumption optimization is not a major concern in this design, any source of static power consumption within each unit cell will be multiplied by 5,000 at the chip level. For instance, to keep the power consumption of the unit cell array below 1W at a 1V supply, each unit cell must consume no more than 200 μA of static current. This places restrictions on the amplifier used in any active feedback schemes for maintaining bias voltages.

(a) General schematic of measurement unit cell.

(b) Full schematic of measurement unit cell.
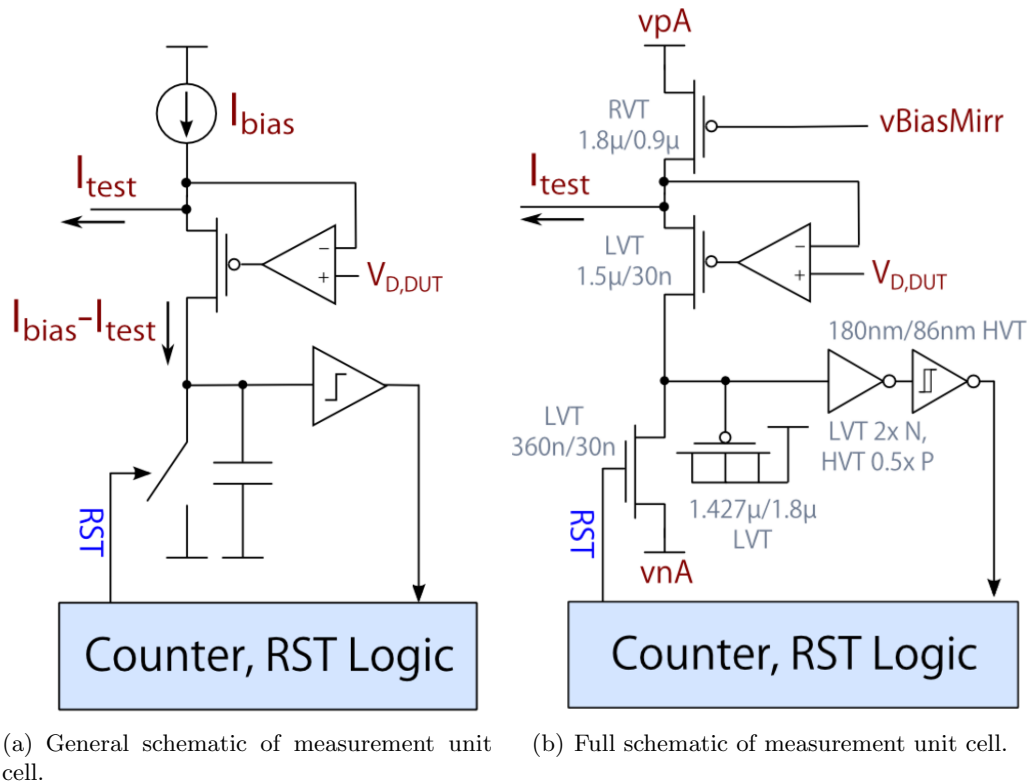
Figure 3.3: Generalized schematic of analog component of measurement unit cell (left) and device-level schematic of measurement cell.

3.3 Implementation details

3.3.1 Measurement cell

In the measurement cell, current from the DUT cell is converted to a pulse with a current-dependent frequency that is digitized using a 16-bit counter. To implement the current to frequency converter, the test current is integrated onto a reference capacitor to generate a voltage ramp, and the capacitor is reset once the ramp voltage exceeds a threshold. As depicted in Fig. 3.3(a), the test current from the DUT unit cell, I_{test}, is subtracted from a reference bias current, I_{bias}, before being integrated onto the test capacitor. To account for variation in the integration capacitance and voltage threshold, each measurement cell can be calibrated using the chip-level bias current reference. An active cascode amplifier is used to ensure that the test current remains constant by fixing the drain voltage of the DUT.

A full schematic of the unit cell, with device dimensions, is shown in Fig. 3.3(b). Here, we can see that the reference current is implemented using a large PMOS device to accurately mirror current from a reference input to the measurement test cell. The active feedback network is implemented using a wide cascode device and a self-biased amplifier to eliminate the need for an additional bias network. The test current is integrated onto the gate capacitance of a PMOS device, and the 'comparator' is implemented using a digital logic gate, which will have a hard-coded trip point that determines its effective reference voltage. As discussed, the input bias reference can be used to generate a current vs. frequency characteristic to calibrate for nonlinearity of the gate capacitance and mismatch in this gate trip voltage. Subsequent sections provide more details about the design, implementation, and expected performance of these subcomponents.
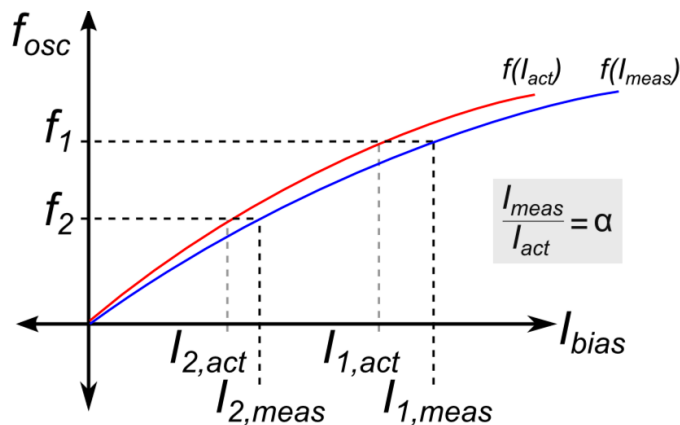
Figure 3.4: Calibration before mirror mismatch nonidealities (red) and after mismatch nonlienarites (blue).

I_{bias}	Monte Carlo σ	3σ % variation
1 a.u.	0.0345 a.u.	10.4 %
0.25 a.u.	0.0125 a.u.	15.0 %

Table 3.3: Predicted variation between current mirror input current (I_{bias}) and output current for a 1.8 μm / 0.9 μm device, determined by 100 point Monte Carlo simulations.

Mirror device

Because the input current mirror must provide bias current that can calibrate for mismatch between the 4,608 measurement unit cells throughout the test chip, the mirror device must be large enough to introduce less variability than the array is designed to detect. To understand the effects of mismatch within the mirror device, we can consider a simplified case in which the current mirrored into the test cell ($I_{bias,cell}$) is off from the desired bias current (I_{bias}) by a given amount α_{OS}:

$$I_{bias,cell} = \alpha_{OS} I_{bias}$$

During calibration, the I_{bias} vs. frequency curve will be measured so that a given change in output signal frequency can be translated back to a known change in current, as shown in Fig. 3.4. If the actual $I_{bias,cell}$ differs from I_{bias}, the calibration curve will be incorrect, resulting in an incorrect measured change in current. In Fig. 3.4, this distortion is accomplished by scaling the horizontal axis by a factor of α. For the same pair of measured frequencies (f_1 and f_2), the actual current will be distorted by an equivalent factor of α if the frequency vs. bias current characteristic is perfectly linear. Therefore, mismatch between the two current characteristics will cause the measured current ($I_{1,meas} - I_{2,meas}$) to differ from the actual current ($I_{2,act} - I_{1,act}$), and this mismatch must be kept low to maintain accurate readings.

Systematic mismatch is reduced by using identical measurement unit cells in the array, and by surrounding the array with identical fill cells. However, the large overall die area (1.33 mm^2) increases the design's susceptibility to within die variation. Monte Carlo simulations incorporating global variation effects suggest that a device of this size will result in 3 σ variation from the input mirror to the test devices of 10-15%, depending upon the magnitude of the bias current. Two sample mismatch results, presented in arbitrary measurement units, are given in Table 3.3. Variation increases at smaller current levels which require high sensitivity.

Bias current mirror variation is expected to be a major source of measurement inaccuracy. RTS noise characterization should be relatively insensitive to current mirror mismatch errors because the main RTS noise parameters of interest are relative trap magnitude ($\Delta I_D/I_D$) and temporal characteristics. Leakage current statistics, however, will be much more strongly impacted. However, including 48 total test devices (and four of each test device type) within most DUT cells can allow these errors to be reduced in post-processing.

Comparator

The comparator is implemented using an inverter skewed to transition at a low voltage. A differential input, low-offset comparator using a fixed reference voltage is not necessary because the input bias current can be used to calibrate the current vs. frequency characteristic of each measurement cell. Therefore, uncertainty in the absolute value of the inverter's trip voltage does not restrict the chip's ability to characterize current measurements. This skewed metastable point is accomplished by including two parallel LVT devices in the pull-down network and two series HVT devices in the pull-up network to ensure that the effective pull down strength of the inverter is much stronger than its pull up strength.

Keeping the trip voltage low ensures that the voltage on the integration capacitor and at the drain of the cascode device is kept small, which has two main advantages. First, this maintains the linearity of the integrating capacitor, which is implemented as LVT PMOS gate capacitance, by reducing the applied voltage swing and ensuring the PMOS device used to generate the gate capacitance will remain saturated. Second, maintaining a low ramp voltage signal will allow the cascode device to be saturated for an extended range of $V_{D,DUT}$ voltages, because the cascode device must have a minimum V_{DS} applied to keep it in saturation. If the cascode device were not fully saturated, the feedback loop maintaining the DUT drain terminal voltage would have a small loop gain, and $V_{D,DUT}$ would not match the expected value.

Simulation at the highest-skew corner suggests that the maximum inverter trip voltage using the skewing technique described is 369 mV, assuming $V_{DD} = 1$V. Monte Carlo simulation suggests an additional 3 σ variability of around 35 mV ($\sigma = 11.56$ mV), indicating that the highest possible trip voltage to expect is around 400 mV.

Integration Capacitance and Reset Device

To measure currents in the microampere range, the measurement cell must use a relatively large integration capacitor and low-resistance switch. As shown in Fig. 3.5, there is a finite reset pulse duration required for the reset switch to fully discharge the integration capacitance. This requirement can be modeled by the time constant τ of the RC network introduced due to the finite on-resistance of the switch (R_{ON}). To ensure that V_{ramp} has decayed to 1% of its initial value during the reset phase, we must set t_{RST} such that:

$$t_{RST} \geq -\ln(0.01)\tau \quad \rightarrow \quad t_{RST} \geq 4.6 R_{ON} C_{int}$$

The integration time t_{int} is directly proportional to the measured current and the integration capacitance C_{int}, such that:

$$t_{int} = \frac{C_{int}(V_{RST} - V_{residual})}{I_{bias} - I_{test}}$$

The reset pulse duration is also limited by the finite gate delay introduced by the comparator and reset control signal generation logic. When the reset pulse width (t_{RST}) is comparable to the integration delay (t_{int}), the linearity of the frequency vs. current characteristic suffers because t_{int}
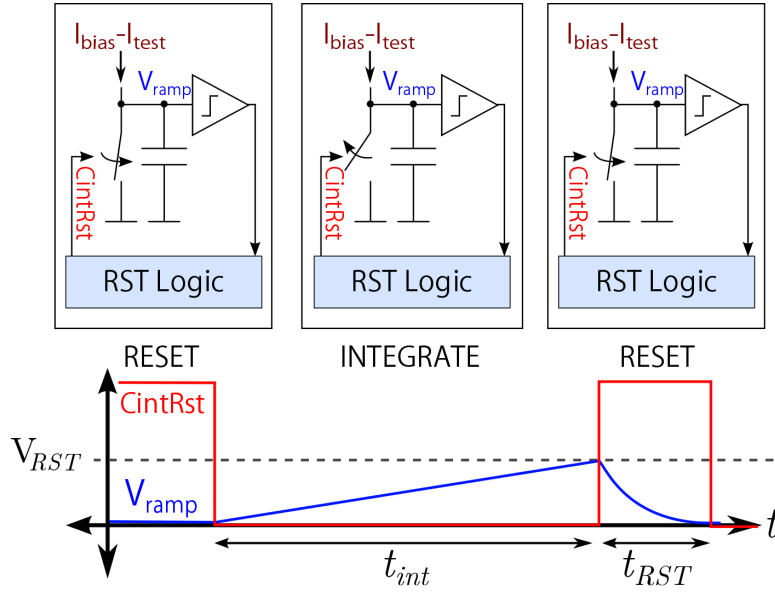
Figure 3.5: This diagram illustrates the reset pulse (CintRst) and voltage at the integration capacitor (V_{ramp}) as a function of time, and the relative state of the reset switch. A finite reset pulse delay is necessary to ensure that the reset capacitor will discharge fully before beginning the next integration phase.

varies with current but t_{RST} remains fixed. If a small reset capacitor is used, the component of t_{RST} allocated to resetting the capacitor may be small, but t_{RST} becomes limited by finite gate delays. Keeping the integration capacitor large ensures that t_{int}/t_{RST} is predominantly determined by the reset switch resistance.

To obtain high capacitance density, the integration capacitor is implemented using MOS gate capacitance. To ensure that the device will always be in saturation for low ramp voltages, thereby improving linearity, we use an LVT PMOS device.

The reset device simply needs to be large enough to discharge the capacitor in the time allotted to the reset pulse duration. A larger device is desirable to allow the integration capacitor to discharge more quickly. However, the built-in logic delay for generating the reset pulse introduces a finite reset delay that cannot be reduced by increasing the reset device size, placing a finite limit on the necessary ON resistance of the switch. Because devices with large dimensions will contain more parasitic capacitance that will couple the sharp reset pulse to internal nodes within the measurement cell, larger devices may also generate large internal current spikes within the circuit.

To minimize these effects, the reset device width is limited to 360 nm for a minimum length (30 nm) transistor. To reduce the switch resistance without incurring additional area penalty, a low-threshold voltage (LVT) device is used. While this will increase the leakage and noise introduced by the reset device, these factors should not impact the characterization measurements. Noise is not a concern for leakage measurements, and the additional leakage current of the reset switch will be accounted for by measuring leakage with a correlated double sampling approach. In comparison, the magnitude of the reset switch current should be small relative to the RTS current levels so that it will not contribute noise to the measurement.

Active cascode

The main purpose of the active cascode is to fix the drain terminal of the DUT to a known voltage so that its drain current can be measured as $I_{ON,DUT} - I_{OFF,DUT}$. This is most relevant to leakage current measurements, in which drain to source voltage control is used to force DUTs into the OFF state. The DC gain of the amplifier sets the static gain error of the feedback loop forcing $V_{D,DUT}$ to the target drain voltage. As shown in Eqn. 2.4, $I_{DS,leak}$ is proportional to $e^{\frac{V_{GS}}{mkT/q}}(1 - e^{\frac{-V_{DS}}{kT/q}})$. With the DUT V_{GS} fixed at 0V, setting the source voltage to the target drain voltage will exponentially reduce the leakage current to ensure a high I_{ON}/I_{OFF} ratio. For the ON-state DUT, $V_{DS} \gg kT/q$ and further changes in V_{DS} will not significantly alter the bias current. Simulations suggest that a maximum V_{DS} change of 2% is sufficient to keep I_{DS} fluctuation below 10%. As a result, the absolute DC gain of the amplifier is around 50 V/V to keep the static gain error within this bound.

Fig. 3.6 presents the schematic for the self-biased amplifier used to maintain $V_{D,DUT}$ at a fixed voltage, designed in part by Vladimir Milovanovic based off of the design in [40]. A self-biased design is chosen to reduce routing complexity and eliminate the need to distribute bias voltages throughout the chip. Because the maximum ramp voltage at the drain of the cascode device is 400 mV, the minimum common mode input of the amplifier should be around 500 mV to keep the cascode device in saturation. To maintain voltages in this range, the cascode gate voltage must be fairly low, but the self-biased amplifier topology does not provide sufficient gain for low common-mode output voltages. As a result, a source-follower level shifter is included at the output of the amplifier. Using LVT PMOS devices and HVT NMOS devices shifts the common mode input range of the amplifier up to around 450-650 mV, as shown in Fig. 3.7, to accommodate the 500 mV minimum input signal required to keep the cascode device in saturation.

The active cascode must maintain a steady drain to source voltage across the test devices while the drain voltage of the cascode device — set by the voltage across the integration capacitor — ramps linearly up to 400 mV. To keep the cascode device in saturation while its V_{DS} fluctuates, a wide, LVT PMOS transistor is chosen. The cascode device is designed to have a low overdrive voltage and threshold voltage to minimize the V_{GS} drop required to maintain a given branch current. Because the cascode device's source voltage is fixed at $V_{D,DUT}$, reducing the V_{GS} of this cascode device increases the amplifier output, extending the range of $V_{D,DUT}$ values that may be used in testing before forcing the amplifier output to the negative supply rail. Moreover, while the wide cascode device generates a significant amount of noise, the majority of this current will recirculate within the device itself instead of propagating to the output of the integrator.

All in all, the bandwidth of the amplifier is around 20 MHz, for a gain-bandwidth product of 1 GHz. The first stage of the amplifier consumes 4.4μA on a 1V supply, while the second stage consumes 3.1μA, for a total static power consumption of 7.5μW. Table 3.4 presents the expected DC gain and phase margin of the amplifier design with layout parasitics across design corners.

3.3.2   Device under test (DUT) cell

The DUT unit cell generates a test current to be characterized by the adjacent measurement cell. Within the DUT cell, digital scan chain data sets local control signals to ensure that only the drain current of desired DUTs contributes to the test current. As shown in Fig. 3.8, the serial scan data is translated to 24 parallel select signals that control either 48 or 24 devices, depending on whether the cell incorporates both NMOS and PMOS, or only NMOS, devices. The scan input and clock are shared between all unit cells in the chip, so the same scan data will be written into each DUT unit cell. However, the scan chain output only propagates to the top level chip output if both row and column enable signals are high.

Table 3.5 summarizes the different types of DUT cells contained within the test array. Three-
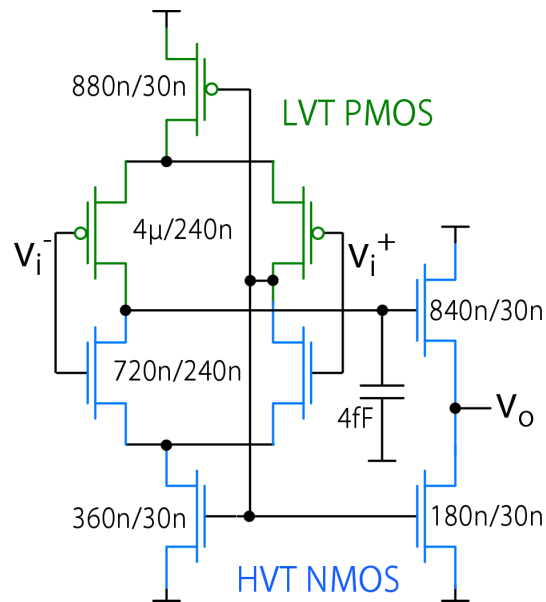
Figure 3.6: Self-biased amplifier topology used in active cascode, based off of a design by Vladimir Milovanovic and [40].
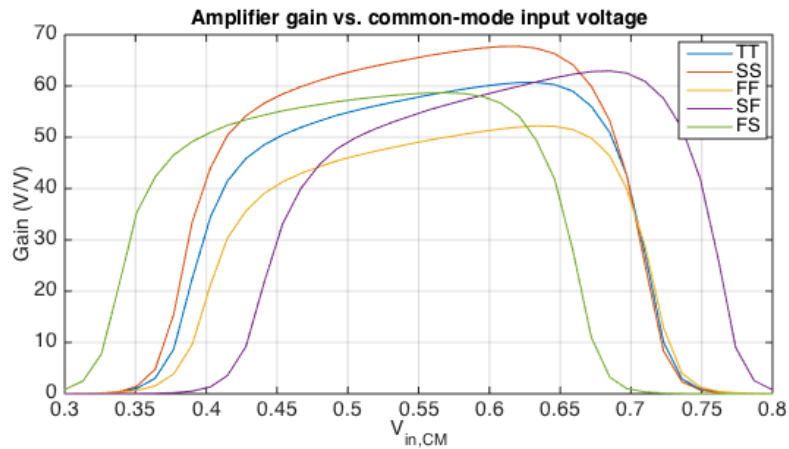
Figure 3.7: Common mode input range of amplifier.

Corner	DC gain (V/V)	Phase margin (degrees)
Nominal	54.0	64.9
FFF	44.1	59.6
SSF	65.0	71.7
FS	56.8	64.4
SF	47.9	66.4
FSF	56.6	63.9
SFF	41.0	68.2

Table 3.4: Mid-range DC gain and phase margin of self-biased amplifier for active cascode.

DUT cell type	Count
PMOS and NMOS, RTS cells	2304
NMOS only, RTS cells	576
PMOS and NMOS, leakage cells	2304
NMOS only, leakage cells	576

Table 3.5: Summary of DUT cell types contained within the test array.
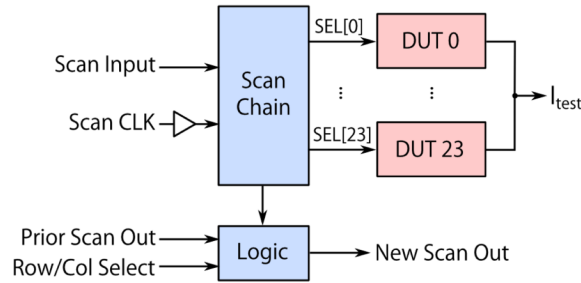
Figure 3.8: Device under test (DUT) cell block diagram.

quarters of unit cells within the test array contain 48 devices (24 NMOS and 24 PMOS), while the remaining quarter of the unit cells contain only 24 NMOS devices. Within these two groups, cells are split evenly between RTS noise and leakage current measurement cells. For leakage current measurement, the $|V_{GS}|$ of DUTs is fixed at 0V while $|V_{DS}|$ is modulated to turn devices ON and OFF. In the RTS characterization test cells, only the $|V_{GS}|$ of selected devices is set to the input value.

Cell control signals

A detailed diagram of the scan chain and select signal generation is presented in Fig. 3.9. This shows the chip-level control signals used to interface with the DUT cell. To preserve signal integrity and reduce fanout, many inverter pairs are used as signal buffers. Inverters in the signal buffers are implemented with minimum length (30 nm) devices with $W/L = 10$, and the clock drivers for generating signals sent to the 24 flip flop inputs are built using $1.5\mu m$ / 30 nm inverters. Figs. 3.10(a) and 3.10(b) show simulations of the clock signals within the scan chain and corresponding flip flop outputs for all process corners, demonstrating that the chosen buffer sizing results in a 100 ps clock rise and fall time. To ensure that timing requirements are met throughout the chip, the input signal clock must be slower than the delay introduced by signal buffering throughout the 96 cells in a column. The simulated scan clock propagation time from the chip input to the final cell in the column is 7.1 ns, while the scan input data propagation time is 2 ns. This indicates that a 10 ns separation between the clock rising edge and data pulse transition is necessary to ensure that each DUT cell scan chain retains the desired information. Because the scan chain results do not need to be updated frequently and the measurement cell integration time will be long relative to the scan chain loading time, scan clock frequencies $\ll$ 100 MHz can be used to build in timing margin without significantly increasing the total testing time.

The CLKdis signal provides an option to disable the scan clock, though to minimize chip pin count, this signal doubles as a means of selecting between NMOS and PMOS test devices. As a result, it can only be used to disable the scan clock for the NMOS-only test devices or for PMOS DUT testing. The remaining logic ensures that the DUT cell scan output will only propagate to the output of the cell's scan chain if the ColEn and RowEn signals are both high.
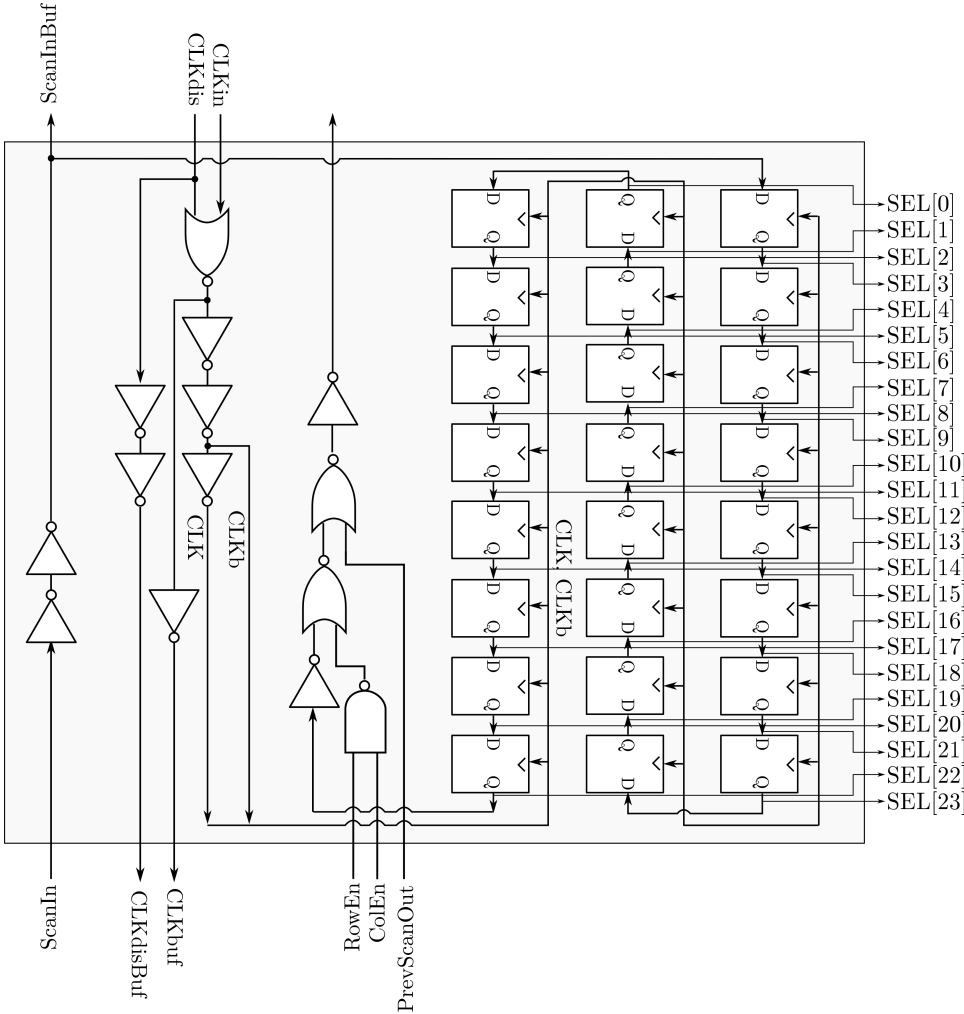
Figure 3.9: Select signal generation within DUT cell.
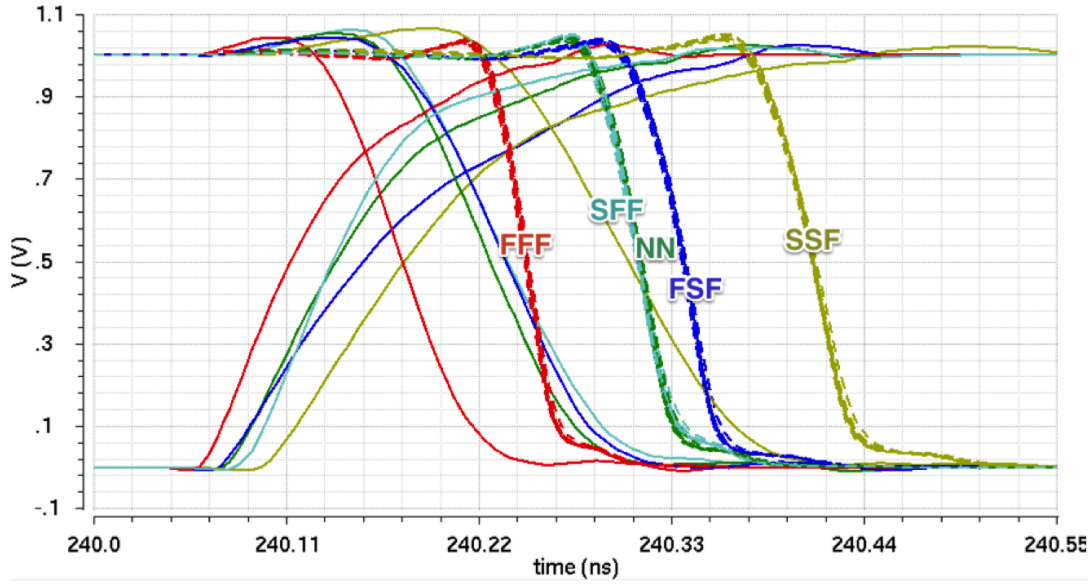
(a) Output of scan chain flip flops (dashed) and input CLK (solid, low to high) and $\overline{CLK}$ (solid, high to low) for each performance corner on a high to low data transition.
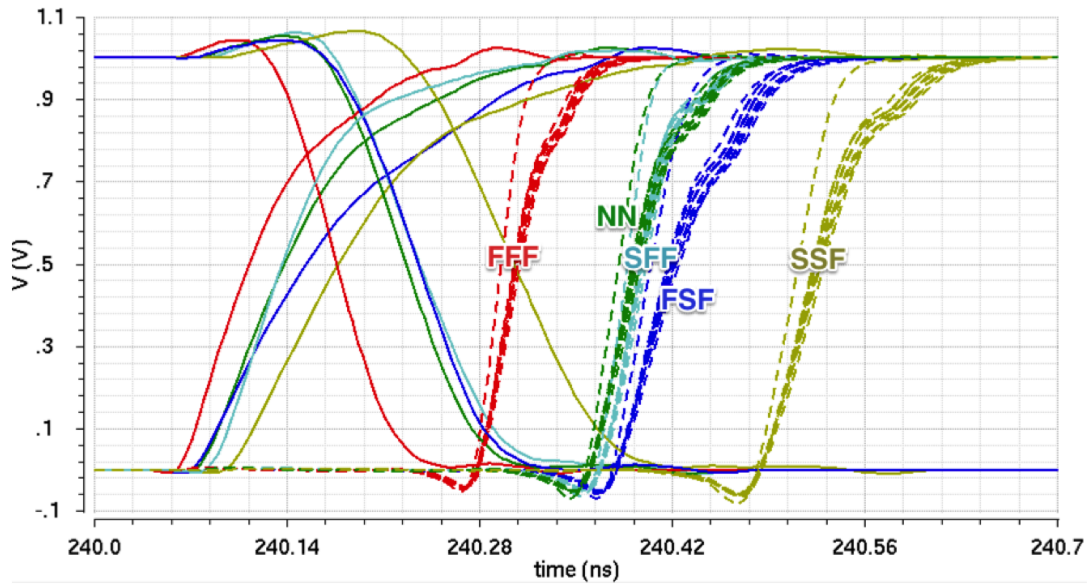
(b) Output of scan chain flip flops (dashed) and input CLK (solid, low to high) and $\overline{CLK}$ (solid, high to low) for each performance corner on a low to high data transition.

Figure 3.10: Scan chain output signals under all process corners.

Combined PMOS / NMOS DUTs
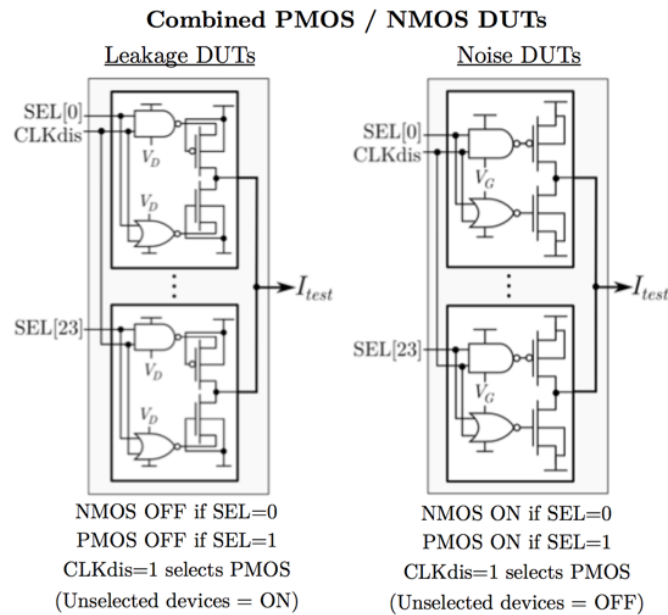
Figure 3.11: Combined NMOS and PMOS DUT cell control signal generation.

DUT selection

Figs. 3.11 and 3.12 show how the V_{GS} and V_{DS} of individual devices are set using the select control signals for the two different cell configurations. Three-quarters of the DUT cells on chip contain both PMOS and NMOS test devices, as shown in Fig. 3.11, while the remaining DUT cells have only NMOS devices, as shown in Fig. 3.12. For the combined PMOS / NMOS cells, NAND and NOR gates are used to switch between a fixed rail and either the gate or drain target voltage, while the NMOS only cells use inverters to select between devices. In RTS DUT cells, logic gates switch the V_{GS} of each test device between a target input voltage and 0V, while for leakage DUT cells, the V_{DS} switches between a target voltage and 0V.

If the applied gate or drain target voltage is near the supply rails, the CMOS logic gates used as voltage switches will operate in the subthreshold regime and the switch output will reach only a fraction of the desired value. Because the inverter, with no stacked devices in the pull-up or
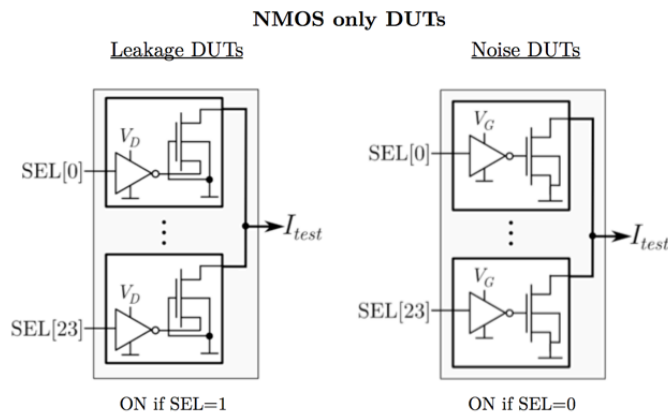
NMOS only DUTs

Figure 3.12: NMOS-only DUT cell control signal generation.
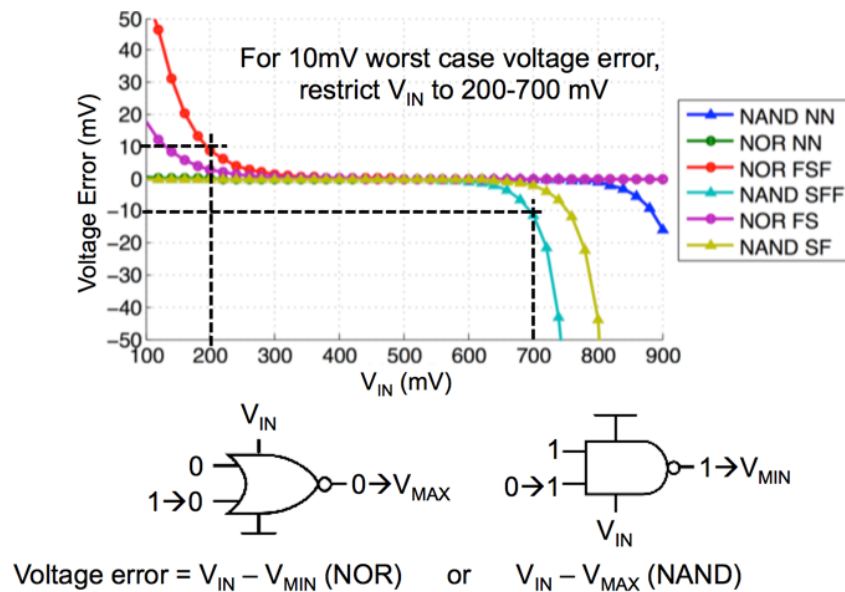
Figure 3.13: Signal swing limitation caused by applying low supply voltages across select gate terminals.

pull-down network, can operate with rail to rail output swing further into the subthreshold range than two-input logic gates, the NMOS-only implementation will be more robust. However, it will yield less data, so only every fourth DUT cell is NMOS-only. If the NAND and NOR gate pull-up and pull-down strength is insufficient to set the DUT terminal voltages, the inverter-based cells will still provide reliable measurements. Fig. 3.13 presents the difference between the maximum (for NOR gates) or minimum (for NAND gates) output voltage and the voltage applied at the logic gate's supply, assuming a maximum voltage of 1 V. Assuming worst-case corners and allowing for 10 mV maximum voltage error, the input swing can reliably range from 200-700 mV.

As mentioned, the CLKdis signal is used to switch between NMOS and PMOS device types. For leakage characterization, unselected devices will default to the ON state, in which the device V_{DS} is set by $V_{drain,targ}$. The NOR gates will generate a logic low if CLKdis=1, while the NAND gates will generate a logic high if CLKdis=0, setting the output of the gate to either the low or high voltage rail, respectively. Because the active cascode feedback in the measurement cell pins the drain voltage of the test devices to $V_{drain,targ}$, this results in a full V_{DS} drop across the DUTs. Devices can be switched OFF using the select signals when CLKdis is 1 or 0 (for PMOS and NMOS, respectively). The opposite is true for RTS noise characterization cells, in which devices default to the OFF state ($V_{GS} = 0V$) when CLKdis=0.

3.3.3 Biasing and scan output logic

The bias network is implemented as a straightforward current mirror, with a fixed factor of 10 reduction in bias current. The biasing network consists of ten diode-connected, unit-sized NMOS devices in parallel to mirror the bias current to a single slave cell for each column in the array. The current generated by the NMOS mirror drives a copy of the PMOS mirror device, cascode, and amplifier in order to set the PMOS bias voltage for each column; Fig. 3.14 shows a diagram of the biasing network. Dummy fill is used in layout to ensure that the surroundings of the biasing circuitry match the surroundings of the measurement unit cell, and the remainder of the analog unit cell replica is used as additional gate capacitance at the NMOS mirror node to reduce noise.

Figure 3.14: Schematic of biasing network implementation. Ten identical n-type mirror devices are placed in parallel to implement a factor of ten reduction in bias current.

Figure 3.15: Chip die photo.

The chip-level scan chain output is the result of a chain of OR gates combining the scan chain output signals from each column. Each OR gate is located in a dummy DUT cell to improve bias circuit matching. When the row and column integration enable signals are configured to select a single cell, only the scan chain output of that specific cell will propagate to the chip-level scan output. However, all-column, all-row, or all-chip integration features can be used to detect errors in the full scan chain array.

3.4 Top-level layout

Fig. 3.15 presents a photo of the fabricated test chip, which was designed in a 28 nm bulk CMOS process. The total die area is 3.24 mm^2, and the area of the chip core is approximately 1.82 mm^2, with the majority of the area dedicated to the 96 × 96 array of unit cells. Digital row and column addressing circuitry, as shown in Fig. 3.1, is contained in the left and bottom sides of the chip, while the bias current mirrors and scan chain output decoding is performed at the top of the chip. The remainder of the chip is filled with dummy unit cells to improve matching.

Chapter 4

Chip Measurements

4.1 Testing procedure

Accurately measuring device current requires calibrating for the mismatch between each measurement cell and sampling the current before and after the DUT is turned on in order to account for parasitic leakage effects. To understand why this is necessary, consider the following model for the time between pulses generated by the current to frequency converter in the measurement unit cell. When measuring NMOS test current generated from the DUT unit cell, the relationship between inter-spike delay and test current is given by the following formula:

$$t_{total} = t_{fix} + \frac{C_{int}\Delta V}{I_{bias} - (I_{test} + I_{extra})}$$

The parameters in the above expression are shown in Fig. 4.1: t_{fix} is the reset pulse duration, I_{test} is the desired test current, I_{extra} is the unwanted current in the test array (gate leakage, additional leakage from non-DUT transistors, etc.), I_{bias} is the measurement cell bias current, ΔV is the voltage swing determined by the delay cell trip point, and C_{int} is the total integration capacitance at the ramp output node. If t_{fix} is small relative to the ramp charging time and C_{int} is constant, the ramp frequency ($1/t_{total}$) will be proportional to I_{test} by a linear scale factor set by C_{int} and ΔV. There will be some measurement offset due to I_{bias} and I_{extra}, but this offset can be eliminated by sampling the frequency before and after the test current is applied. Similar to correlated double sampling techniques used to reduce noise in CMOS imagers, the frequency is first measured when the DUT is deselected (driving the test current to 0 A), and then when the DUT is selected (applying the full test current). The resulting offset cancellation is illustrated in the expression below, where f_1 and f_2 correspond to frequency measurements before and after the test current is applied:

$$
\begin{aligned}
f_1 &\approx \frac{I_{bias} - I_{extra}}{C_{int}\Delta V} \\
f_2 &\approx \frac{I_{bias} - (I_{extra} + I_{test})}{C_{int}\Delta V} \\
\Delta f &\approx \frac{I_{bias} - I_{extra}}{C_{int}\Delta V} - \frac{I_{bias} - (I_{extra} + I_{test})}{C_{int}\Delta V} \\
&\approx \frac{I_{bias} - I_{extra}}{C_{int}\Delta V} - \frac{I_{bias} - I_{extra}}{C_{int}\Delta V} + \frac{I_{test}}{C_{int}\Delta V} \\
&\approx \frac{I_{test}}{C_{int}\Delta V}
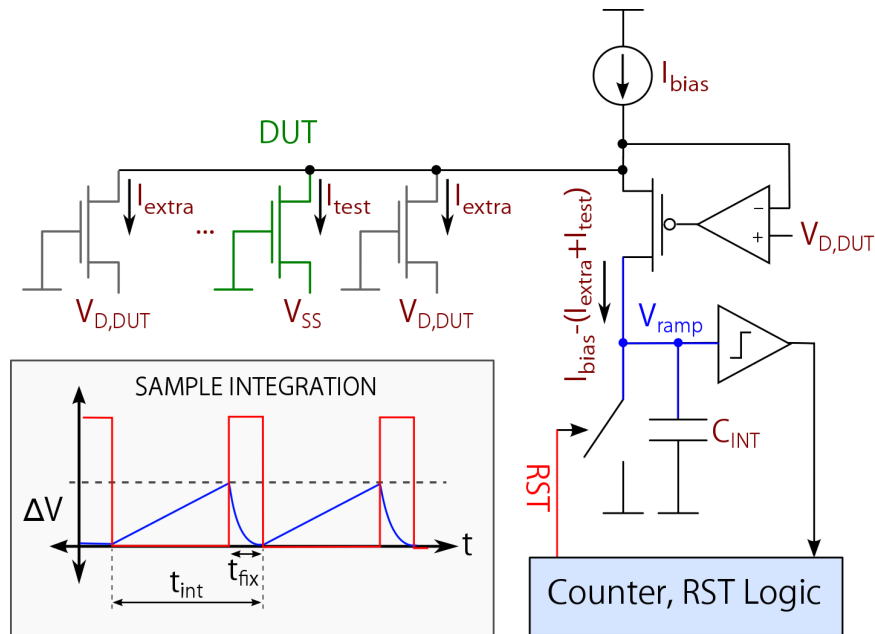\end{aligned}
$$

Figure 4.1: Example of typical integration period and unwanted measurement parasitics.

To first order, the difference in signal frequency is directly proportional to the test current by a factor of $k = 1/(C_{int}\Delta V)$. This scale factor will change between cells due to variability in the integration capacitance amount and delay cell trip voltage. Note that the above analysis assumes that $t_{fix} \ll t_{ramp}$, and that as t_{ramp} gets smaller or t_{fix} shrinks, the linear proportionality constant k becomes inaccurate. Realistically, using voltage-dependent MOSFET gate capacitance to build C_{int} will reduce the linearity of the pulse frequency vs. I_{test} curve. Additionally, large bias current levels will reduce the total integration time and introduce nonlinearities due to t_{fix}. To account for this nonlinearity, each unit cell can be calibrated by measuring a full I_{bias} vs. frequency curve.

As a result, each test cell must be calibrated individually. Moreover, each measurement should be taken differentially to isolate the current of interest from the static bias current and unwanted additional leakage current in the test structure. Generally speaking, the test procedure is as follows:

1. Calibrate the test structure to account for variation in the inverter trip voltage, current mirror offset, and integration capacitance. To do this, measure output frequency as a function of bias current over the desired operating range. This curve can be used as a lookup table to associate a measured Δf with a corresponding ΔI.

2. Turn off all devices using the scan chain and measure the frequency f_{off} to capture parasitic background current levels.

3. Turn on the desired test devices using the scan chain, and then measure the associated frequency, f_{on}.

4. Use the calibration curve to translate $f_{on} - f_{off}$ to a change in current, $I_{on} - I_{off}$. The current difference will be the current associated with the test device.

A detailed description of the measurement procedure is provided in Appendix B. Fig. 4.2 provides a graphical example of how DUT current would be computed from a calibration curve.
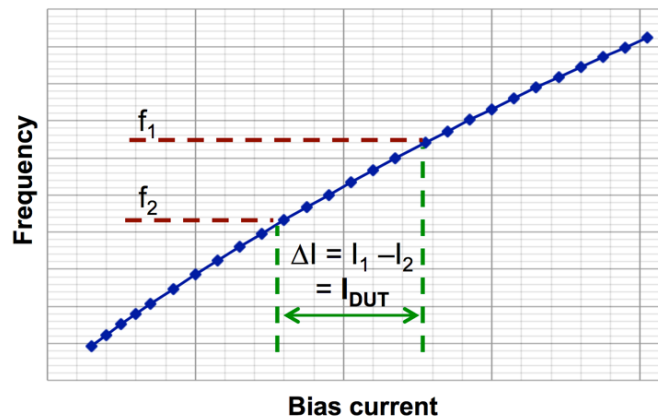
Figure 4.2: The calibration procedure will generate a current vs. frequency characteristic, shown in blue. DUT current can be measured using the device in the OFF state (f_1) and a reading with the device in the ON state (f_2). The DUT current can be determined from the ΔI associated with the change in measured frequencies.

4.1.1 RTS noise measurement

Capture and emission time constants associated with RTS noise are bias voltage dependent, so properly characterizing RTS requires biasing the DUTs near threshold to ensure that traps neither always-occupied nor always-empty. Because this voltage dependence will vary with trap depth, sweeping a wide range of gate voltages is necessary to ensure that multiple traps are captured. For these measurements, around 5-6 gate voltages were measured for the subset of devices tested.

Another major challenge with characterizing RTS noise is obtaining sufficient temporal resolution to capture the full range of RTS time constants. These characteristic time constants can range from fractions of milliseconds to tens of seconds or longer. Sampling each device at 100 kHz over the course of a few minutes, however, is impractical — taking a one minute, 100 kHz sample of drain current for every device on the chip at 6 bias voltages and only 5 temperatures would generate 270 terabytes of test data. As a result, we take a short (1024-point) sample of a subset of test devices using three sampling frequencies: 400 kHz, 4 kHz, and 40 Hz. The DFPA ROIC structure is particularly well-suited to this approach, as the chip can be configured to either sample a single pixel, row, or column at a fast sampling rate or to take many slow (40 Hz and below) measurements of the entire chip. The main limiting factor in readout speed is the rate of data transfer between the Opal Kelly Shuttle LX1 board used for readout and the laptop collecting data, because only a limited amount of information can be stored locally on the Opal Kelly (the FPGA can fit up to twenty-four shift registers holding 1024 16-bit values).

As described in Section 2.2.3, RTS trap state can be forced by applying a large or small DC bias voltage to the DUT gate for a brief time prior to applying the near-threshold gate voltage bias. Because the oxide trap(s) are forced to a known state using this technique, trap amplitudes and mean capture and emission times can be computed much more directly by simply repeating this biasing and measurement procedure. The distribution of time constants is slightly different than that of RTS noise under static biasing conditions, however.

4.1.2 Leakage current measurement

Because the DUT leakage current may be as small as several picoamperes, accurate leakage detection requires long integration times and sample averaging to minimize noise effects. Measurement

resolution is limited by the 16-bit counter, and by the size of I_{bias} relative to the desired test current. Given that I_{bias} must be relatively large in order to be mirrored accurately throughout the chip for calibration purposes, the bias and leakage currents vary by multiple orders of magnitude. The difference between I_{bias} and I_{leak} may be so large that the counter reaches its full level before any impact from the leakage current is noticeable. However, the counter can be allowed to overflow if this is the case. Because each measurement requires sampling both the ON and OFF state DUT currents, and the test signal is encoded in the difference between these two readings, counter overflow due to the large bias current can be canceled. Low-frequency noise generated by the test device, however, can degrade measurements. Leakage current can also be estimated by extrapolating the subthreshold I/V characteristics measured using the noise DUT cells down to $V_{GS} = 0$. Because very small levels of leakage are difficult to detect, and leakage current drops with temperature, we use this technique to generate leakage estimates.

4.1.3   Temperature dependence

In order to characterize the performance of the test chip across temperature, the fabricated chip was wirebonded onto a 100-pin LCC package that could be mounted in a LakeShore Model MTD-125 test dewar. Within the dewar, the test chip is placed against an internal chamber that holds liquid nitrogen to reach cryogenic temperatures. This is surrounded by an open chamber that is evacuated to a pressure of roughly 1×10^{-5} torr to insulate the chip. Hermetically sealed connectors allow for external electrical access to the chip and internal temperature while maintaining the vacuum. Fig. 4.3(a) shows the test dewar with the vacuum pumpout port, and Fig. 4.3(b) shows the packaged chip inside of the dewar.
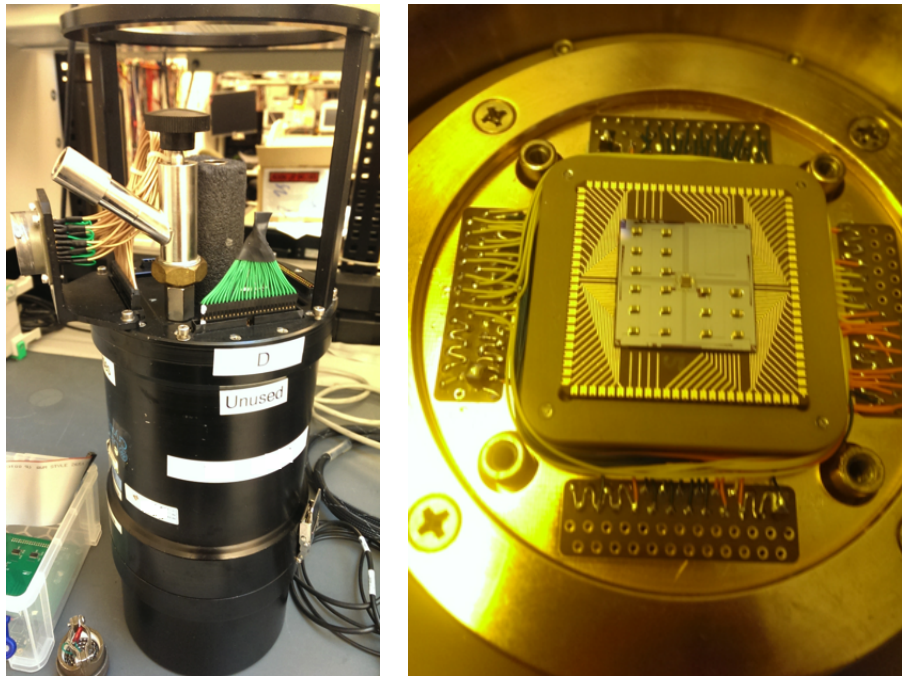
(a) LakeShore MTD-125 dewar used for chip characterization.

(b) 100-pin LCC containing test chip within dewar.

Figure 4.3: Measurement equipment.

4.2   Measurements

4.2.1   Calibration

As previously discussed, relating the counter output to actual current measurements requires measuring pulse frequency as a function of bias current in order to calibrate the system. Fig. 4.4(a) shows the normalized frequency readings as a function of input bias current for all of the 4,608 measurement cells within one test chip. Using the normalized frequency reading at the highest bias current level as an indicator of $\Delta f / \Delta I$ for each test cell, Fig. 4.4(b) shows the distribution of relative frequency values measured.

   As expected, the frequency vs. bias current curve is nearly linear but tapers off at higher frequencies, consistent with nonlinearity caused by a fixed reset time within the current to frequency converter. The relative frequency value, which will depend on mismatch between the bias current mirror transistor, inverter trip voltage, and integration capacitor, displays a roughly Gaussian distribution with $\sigma / \mu = 0.039$.
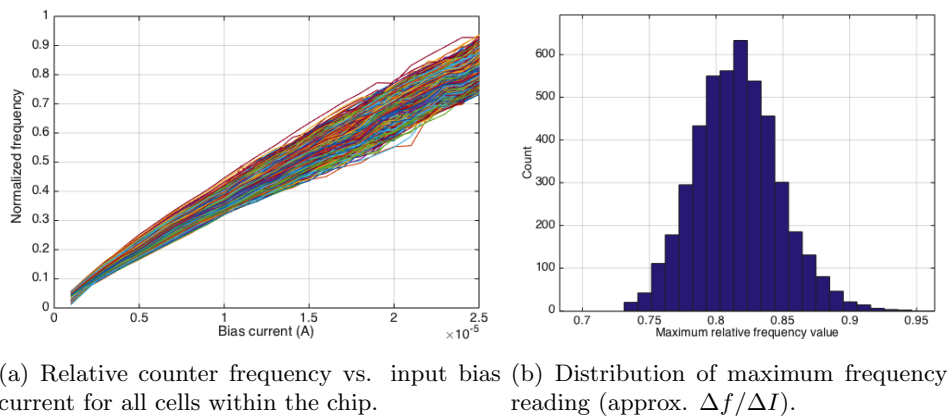
(a) Relative counter frequency vs. input bias current for all cells within the chip.

(b) Distribution of maximum frequency reading (approx. $\Delta f / \Delta I$).

Figure 4.4: Sample calibration curve and cell gain distribution.
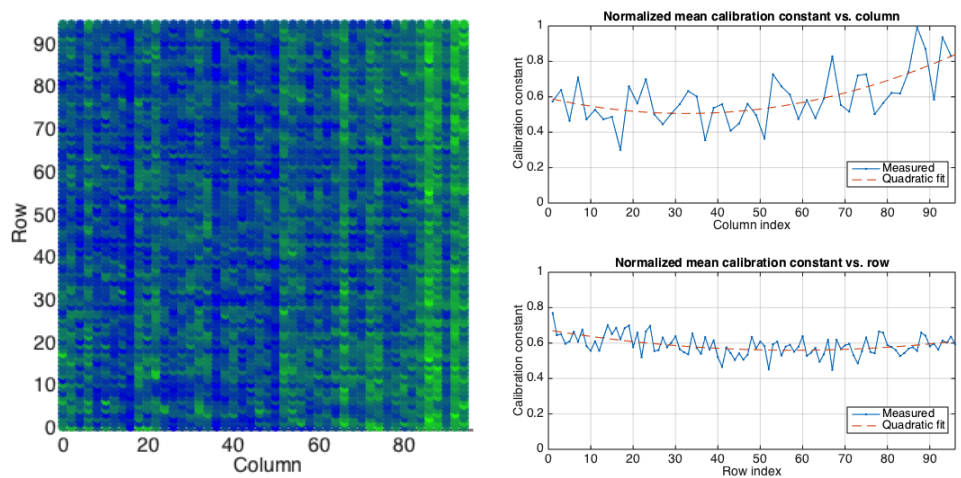
(a) Output count vs. cell index ($I_b = 25\mu A$). Green / blue maps to high / low frequency.

(b) Normalized high-I_{bias} frequency readings per row and column.

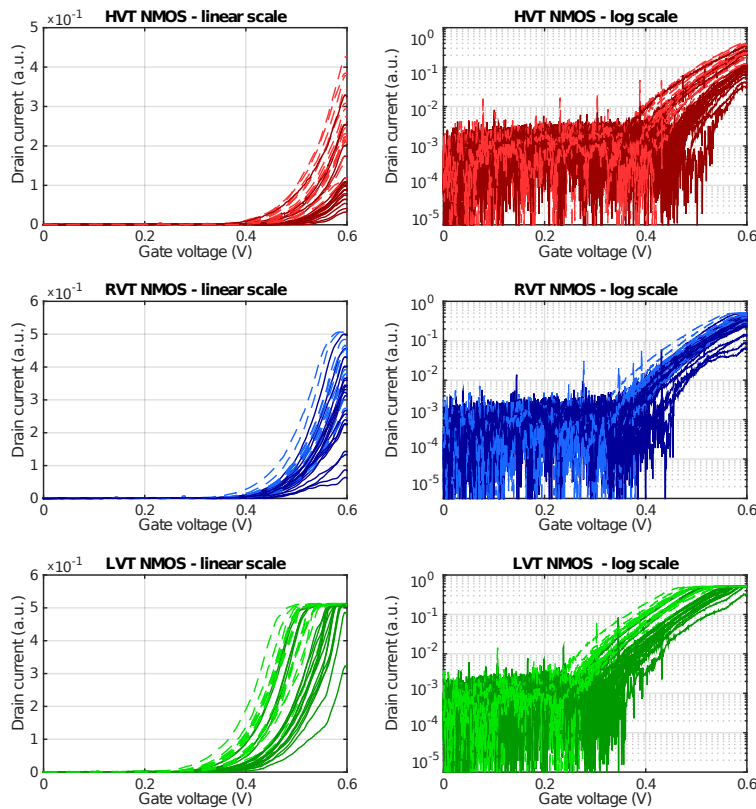Figure 4.5: Spatial distribution of unit cell sensitivity to current.

Figure 4.6: Sample NMOS I/V curves. Measurements from longer-width (86 nm) devices are dashed, while minimum-width (30 nm) devices are solid.

Fig. 4.5(a) maps the maximum relative frequency values presented in Fig. 4.4(b) to their location within the unit cell array. Green cells indicate high $\Delta f/\Delta I$, while blue cells indicate low $\Delta f/\Delta I$. Separate columns can be easily identified by color groupings, likely due to variation in the per-column bias current mirror device. Additional variation within a column can be attributed to mismatch within each cell. The bulk of high frequency readings, which imply high bias current, are clustered towards the rightmost columns. These are the columns in closest proximity to the input current mirror devices (depicted in Fig. 3.14) within the layout, which may be caused by IR drops when a large bias current is applied. This will not impact the fidelity of measurements, provided each cell can be calibrated accurately.

To illustrate the difference between row and column variability, Fig. 4.5(b) shows the normalized average frequency readings within each row and within each column. The per-column average, shown in the top plot, exhibits much more variation than the per-row average, presented in the bottom plot. Within a row, variation due to the per-column bias current mirroring will be averaged out, leaving only variability due to the actual test cell, so this reduction in variability is expected. Fig. 4.5(b) also demonstrates the high frequency readings measured in the right hand side of the array.

4.2.2   I/V characterization

The RTS testing mode can be used to measure I/V characteristics of the test devices, which can be used to extract parameters such as threshold voltage, subthreshold slope, and estimated leakage current. Figs. 4.6 and 4.7 show sample I/V characteristics measured from NMOS and PMOS
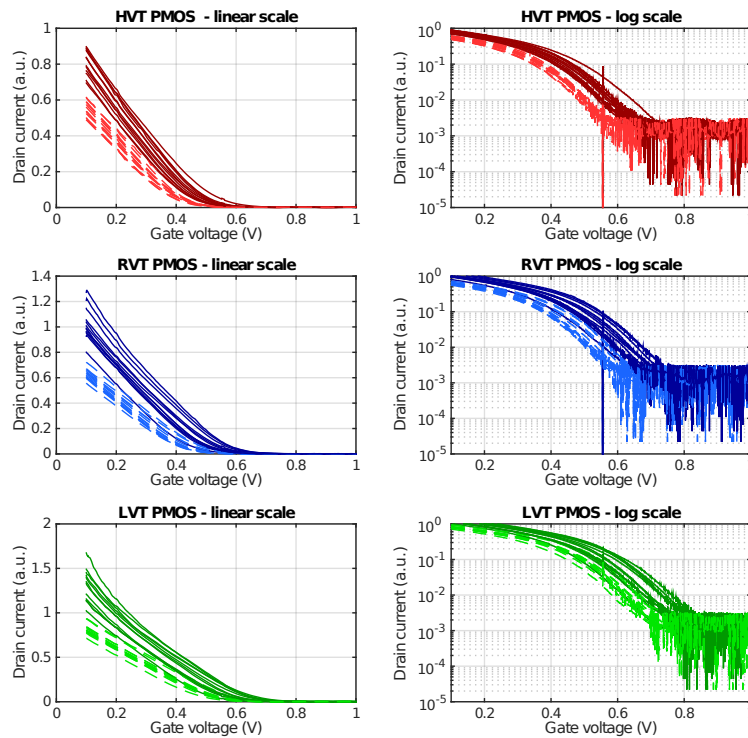
Figure 4.7: Sample PMOS I/V curves. Measurements from longer-width (86 nm) devices are dashed, while minimum-width (30 nm) devices are solid.

devices, respectively, at room temperature. The minimum detectable current, determined by both the testing range and parasitic leakage of all 47 remaining devices in the test array, is clearly visible. Above this point, the exponential subthreshold behavior is evident on a log scale, which can be used to extract the subthreshold slope of the device and estimate leakage.

Because a great deal of noise is present within the measurement cell, long averaging times are necessary to obtain clean, high-resolution leakage current measurements. Moreover, noise generated by the large bias transistor can begin to dominate measurements of leakage current, making it impossible to obtain accurate measurements of leakage in the picoampere range and below. The leakage current of each device can be approximated, however, by extrapolating the subthreshold I/V characteristic to the $V_{GS} = 0$ V point which is otherwise obscured by the parasitic leakage current measurement. Fig. 4.10 summarizes the subthreshold slope measured between all device types at room temperature, and Fig. 4.9 shows the corresponding leakage current.

Another preliminary technique for quantifying variability in leakage current is to characterize variability in the threshold voltage V_{TH} of the set of sample devices. The threshold voltage is estimated using the x-intercept of a linear fit to the above-threshold region of the I_D vs. V_G characteristic. Fig. 4.8 presents the distribution of extracted V_{TH} values for the six device types tested. The extracted V_{TH} values fit the expected trends for this technology, with a wider degree of variability in short channel length and low-V_{TH} transistors.

Temperature dependence

In order to measure device characteristics as a function of temperature, the test chip was cooled to 78 K using liquid nitrogen, as discussed in Section 4.1.3. Higher temperature measurements were taken as the chip gradually warmed to room temperature over the course of roughly 12 hours.
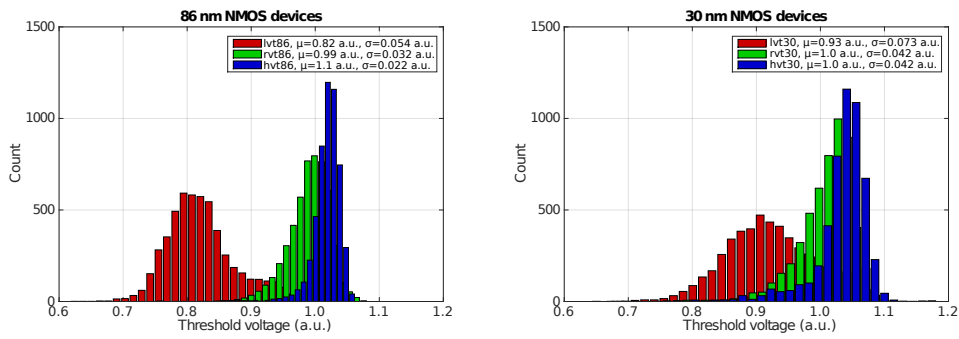
(a) NMOS V_{TH}.
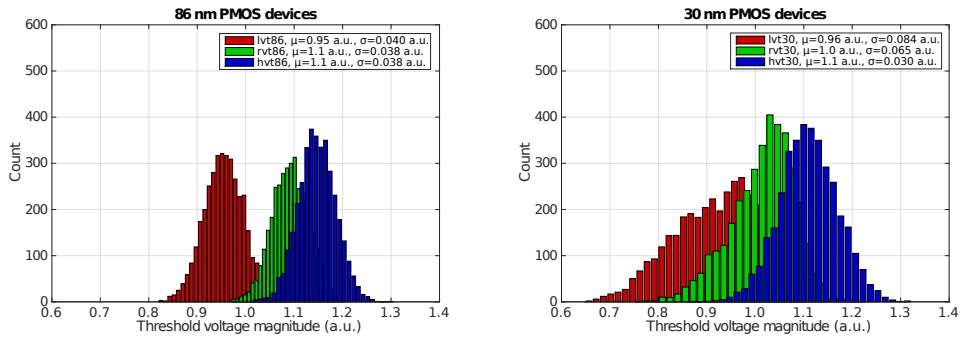
(b) PMOS V_{TH}.

Figure 4.8: Distribution of measured threshold voltages.

(a) NMOS extrapolated leakage.

(b) PMOS extrapolated leakage.

Figure 4.9: Distribution of leakage currents extrapolated from I/V characteristics.

(a) NMOS subthreshold slope.

(b) PMOS subthreshold slope.

Figure 4.10: Distribution of subthreshold slope factors fit to I/V characteristics.

Fig. 4.11 shows I/V characteristics measured for one example of all PMOS device types across temperature. The minimum detectable current level improves as the chip cools, as the parasitic leakage generated from the remaining devices decreases. The subthreshold slope also improves steeply, as expected, but the overall current drops as well due to an increase in effective threshold voltage.
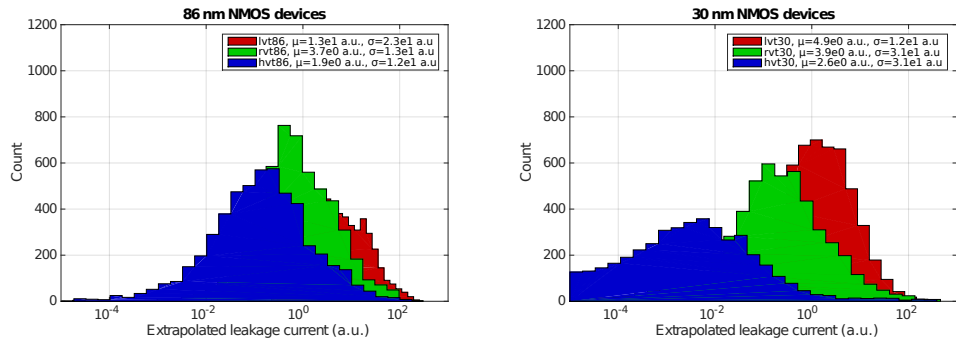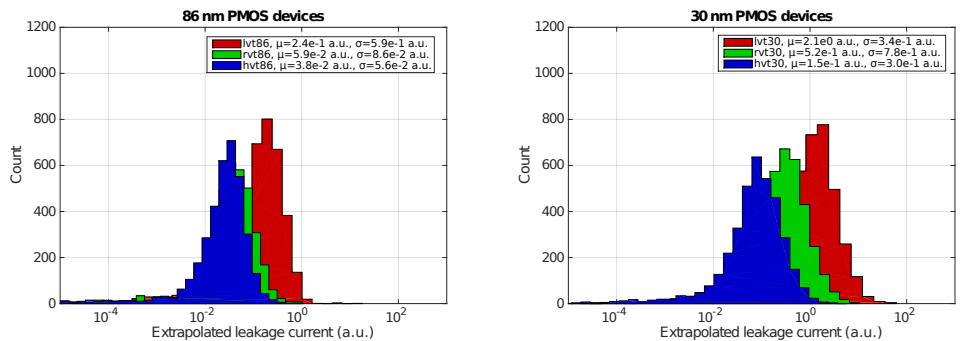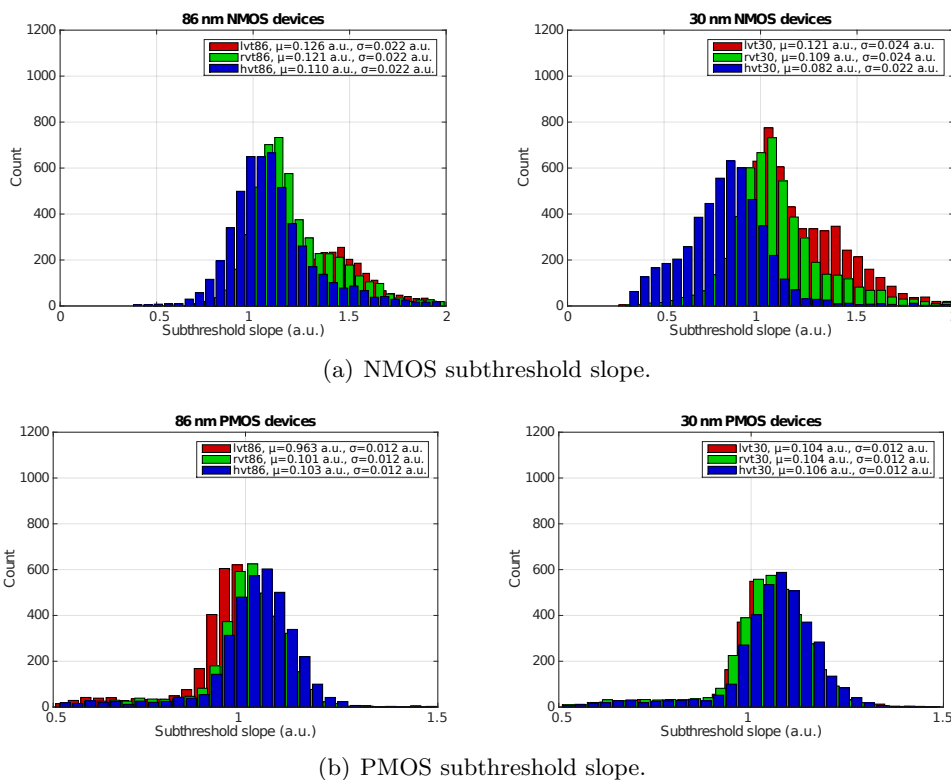
Fig. 4.12 summarizes the temperature dependence of extrapolated leakage, subthreshold slope, and threshold voltage using an Arrhenius plot. The leakage current and subthreshold slope vary exponentially with temperature. The extracted threshold voltage increases by around 12% for most device types as the temperature drops from 300 K to 78 K.

4.2.3 RTS noise

To characterize RTS noise, we sampled DUT drain current at 400 kHz, 4 kHz, and 40 Hz under different bias and temperature conditions. Analysis of the measured data is still ongoing, but this section presents a subset of measurements to highlight some key findings. Ultimately, RTS noise is certainly a concern even at cryogenic temperatures.

Fig. 4.13 shows drain current readings over time for a representative single-trap PMOS device. Current is measured for four gate voltages and six temperatures between 153 K and 200 K. Qualitatively, this plot indicates how RTS time constants increase with temperature, which can be explained by the reduction in thermal energy available to initiate trapping and de-trapping when temperature decreases. Fig. 4.13 also illustrates the gate voltage dependence of RTS noise. When $|V_{GS}|$ is small, the trap is typically empty and the capture times (τ_C, measured as the duration of high current time pulses) are much longer than emission times (τ_E, measured as the duration of low current time pulses). As $|V_{GS}|$ increases, however, the measured drain current is split more evenly
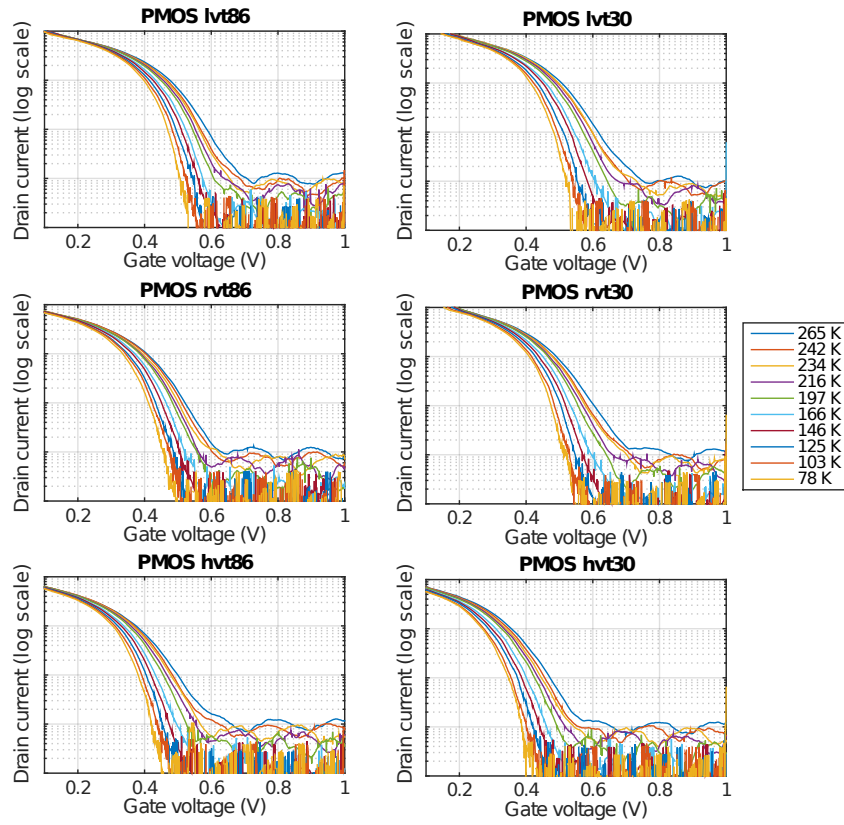
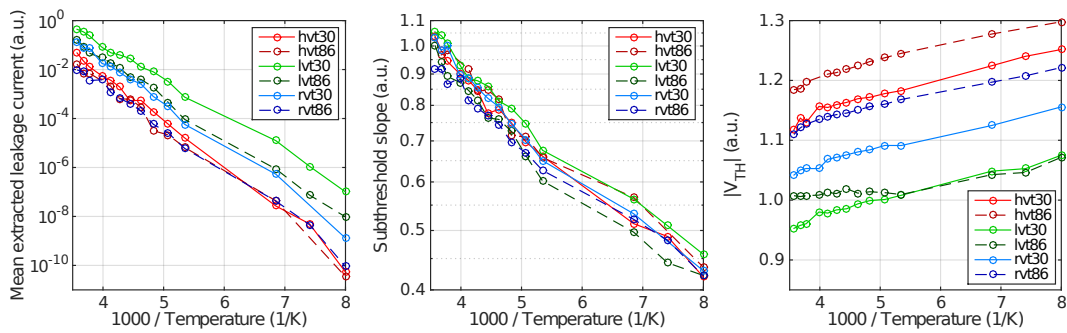Figure 4.11: Sample I/V characteristic of PMOS devices across temperature.

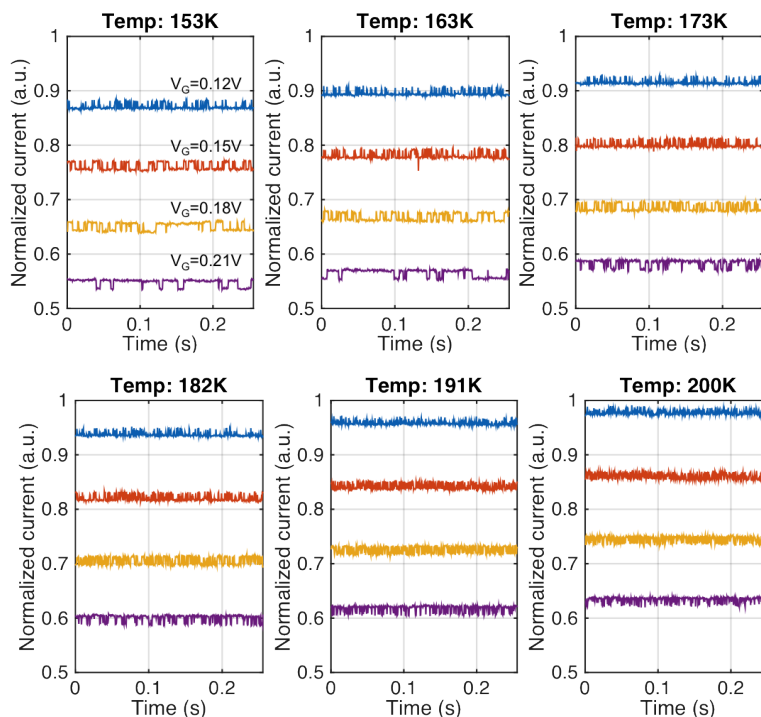Figure 4.12: Sample Arrhenius plot of PMOS device characteristics.

Figure 4.13: Measured I_D vs. time for a single device at various gate voltages and temperatures.

between high and low states, indicating balanced τ_C and τ_E. When $|V_{GS}|$ is sufficiently high, the trap is mostly occupied, and the measured drain current is typically fixed at a low value.

To provide an example of single-trap analysis, Fig. 4.14 shows the distribution of drain current measured for this device at 173 K when $V_G = 0.21$ V. The single-trap behavior of this device is evident in the two peaks in the output sample histogram, and the trap amplitude (ΔI) can be measured as the separation between peaks, marked in green. The number of RTS traps can typically be estimated as $\log_2(N_{peaks})$, where N_{peaks} is the number of peaks in the drain current histogram. For comparison, Fig. 4.14 also shows the histogram of differences between successive current measurements. The two outer peaks correspond to jumps of $+\Delta I$ and $-\Delta I$. The bottom two histograms show the distribution of extracted capture and emission times, which have an exponential fit, as expected.

To provide a more quantitative look at the $|V_{GS}|$ and temperature dependence of τ_C and τ_E, Fig. 4.15 shows the mean τ_C and τ_E extracted from the transient waveforms in Fig. 4.13. From this, it is clear that while $\overline{\tau_C}$ and $\overline{\tau_E}$ vary differently with temperature, the mark-space ratio ($\overline{\tau_C}/\overline{\tau_E}$) dependence on $|V_{GS}|$ is relatively temperature-independent. At high temperatures, the trap exhibits clear "type-I" trap behavior, with negative $\frac{\partial \tau_C}{\partial V_{GS}}$ and positive $\frac{\partial \tau_C}{\partial V_{GS}}$. At lower temperatures, however, the trap behavior is more consistent with "type-III" traps, in which $\frac{\partial \tau_C}{\partial V_{GS}}$ is negative but $\frac{\partial \tau_E}{\partial V_{GS}}$ is closer to zero, indicating that trap mechanics may be somewhat temperature dependent. The "type-I" characteristics are typically associated with carriers trapping from the channel to the oxide and de-trapping back through the channel, while "type-III" characteristics are typically associated with carriers de-trapping through the gate. [22]

One of the main objectives of this work is to determine whether RTS noise is still a concern at cryogenic temperatures. Fig. 4.13 provides one example of RTS noise evident down to 153 K and below, but Fig. 4.16 presents an example of fast RTS measured at 78 K in a PMOS test
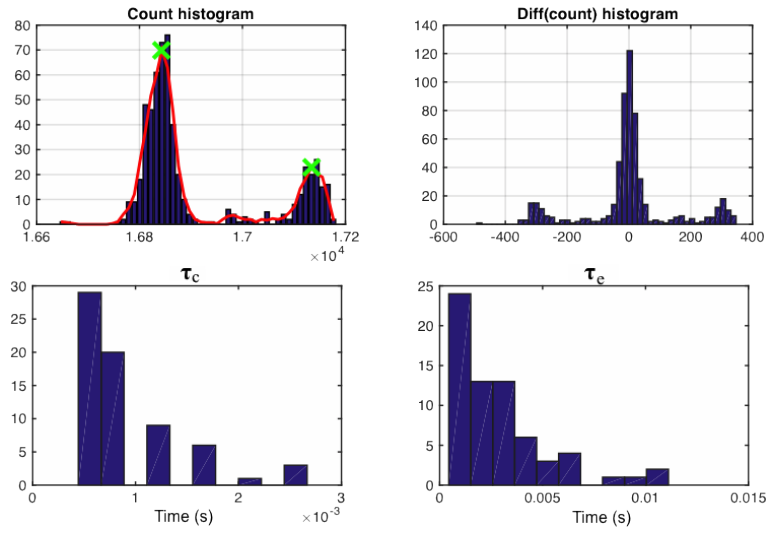
Figure 4.14: Histogram of single-trap RTS noise shown in Fig. 4.13.
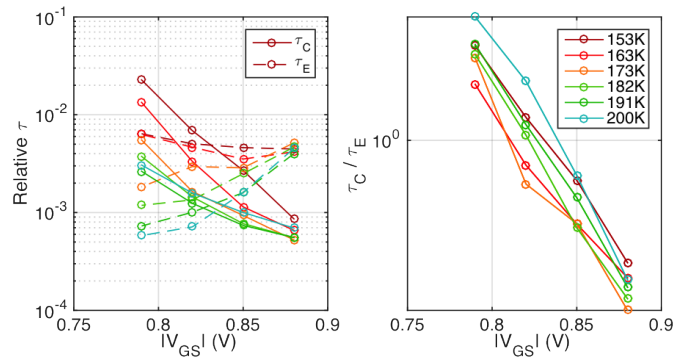

Figure 4.15: RTS time constant dependence on $|V_{GS}|$ extracted from plots in Fig. 4.13.
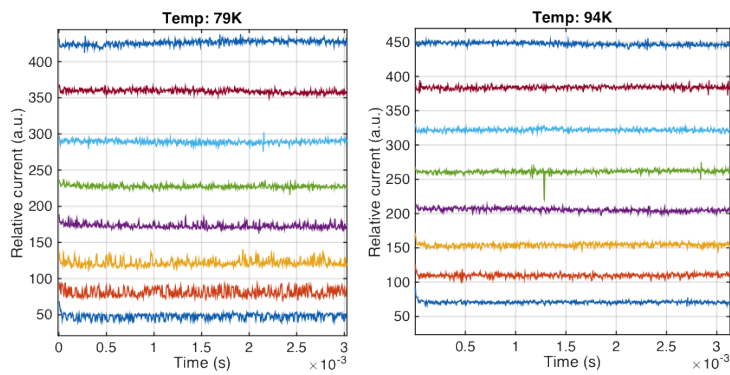

Figure 4.16: Example of RTS measured at 78 K, sampled at 400 kHz.

device. Because the test structure integrates current continuously throughout the sample period duration, any RTS noise switching faster than the chip sample rate will be filtered out. This means that some RTS will "appear" at lower temperatures when the associated RTS time constants are slowed sufficiently to become detectable. In Fig. 4.16, RTS is noticeable at 78 K but "disappears" when the chip heats to 94 K, likely because the time constants are too fast to be detected by the chip. The presence of RTS noise with time constants in the range of tens of microseconds even at cryogenic temperatures, however, indicates that cooling is insufficient to fully eliminate this source of noise in deeply scaled CMOS ROICs.

4.3 Potential sources of error

The measured results generally correlate well with expectation, but some potential causes of measurement discrepancies to be aware of include:

- **Lack of rail-to-rail output swing in DUT bias control gates.** The gate and drain voltages applied to the DUT may not reach the full applied V_{gate} and V_{drain} due to the reduced pull-up and pull-down strength of the bias control switches in the subthreshold regime. Fig. 3.13 summarizes the expected difference between the maximum output voltage and the supply levels as a function of input voltage for different gate configurations. This analysis suggests that measurements should be reliable if V_{gate} or V_{drain} is kept within a restricted voltage range, but measurements may be inaccurate if the applied voltages are too high.

- **IR drops on bias voltage routing**. The current supplied by the V_{gate} and V_{drain} supplies should be small, predominantly set by gate leakage current and drain leakage current (respectively), but substantial parasitic current within the array coupled with parasitic routing resistance could introduce mismatch in the applied bias voltages due to unwanted IR drops. A larger concern would be IR drops due to parasitic substrate resistance at high bias current levels. This may be reflected in the variation in calibration frequency measurements per column, which indicate a drop in mirrored bias current further from the input pin. This will be less of a concern for PMOS device measurements, which do not require a large bias current to be applied throughout the chip. However, this could certainly affect NMOS device measurements, in which the measured DUT current must be subtracted from I_{bias}.

- **Current mirror mismatch.** As discussed in section 3.3.1, mismatch in the current mirror distributing the bias current to each unit cell will be indistinguishable from test current, given that the bias current calibrates the behavior of each unit cell. The mirror devices are large to minimize mismatch effects, but variation may still be manifested as an additional source of random mismatch between each unit cell. Including multiple DUTs of each type per unit cell helps distinguish between DUT mismatch and unit cell bias current mismatch. The spatial distribution of measured frequencies presented in Fig. 4.5(a) indicates that this is definitely a concern.

- **Additional noise sources.** The chip should be characterizing only noise in the desired DUT, not adjacent DUTs in the OFF state or peripheral circuitry in the measurement unit cell. RTS in the measurement unit cell's bias transistor will be small relative to the test device due to its substantial area. The large g_m of the cascode transistor will ensure that most of its noise current recirculates within the device instead of propagating to the output node, and its large size ($1.5\mu m/30nm$) also keeps RTS low. Finally, the large difference between ON-state and OFF-state current within a transistor should ensure that the noise generated by the ON-state DUT is dominant. However, this may need to be verified by including a "trap-forcing" phase for each measurement.

Chapter 5

Conclusions and Future Work

This work aims to investigate the feasibility of utilizing deeply scaled CMOS technology for advanced imaging applications. To accomplish this, a variability characterization platform modeled after a conventional digital focal plane array (DFPA) readout integrated circuit (ROIC) was designed, fabricated, and tested. The test platform investigates variation in two major areas of concern to imager performance: leakage current and random telegraph signaling (RTS) noise. Preliminary measurements of the test chip indicate that RTS noise may be a significant issue in developing high-accuracy sensors even down to cryogenic temperatures, while leakage current levels decrease significantly with temperature. However, while this report presents preliminary findings to illustrate the test chip's functionality and initial results, many additional investigations can be considered:

1. **Examine temperature dependence of RTS noise.** While the relationship between RTS noise time constants and gate voltage is relatively constant across temperature for some traps, other traps exhibit a qualitatively different τ vs. V_{GS} characteristic as temperature changes. We would like to examine this dependence to determine whether it can provide further insight into the physical mechanisms causing RTS noise.

2. **Explore spatial variability of RTS noise characteristics.** Because the test chip is built as a conventional imaging readout circuit, many device measurements can be taken simultaneously across a large chip area. This can allow us to develop an "image" of low-frequency RTS noise effects much more quickly than conventional variability characterization arrays, which address only a single device.

3. **Characterize distribution of RTS noise amplitude and number of traps per device type.** Given the wide range of device types included on the test chip, we would like to correlate key RTS noise parameters — amplitude, number of traps, trap depth and energy — with device parameters in order to identify patterns that can be used to inform future design efforts.

Bibliography

[1] M. Kelly *et al.*, "Design and testing of an all-digital readout integrated circuit for infrared focal plane arrays," *SPIE Focal Plane Arrays for Space Telescopes II*, vol. 5902, 2005.

[2] M. Kelly *et al.*, "Digital focal-plane arrays," *MIT Lincoln Labs Tech Notes,*, 2014.

[3] F. Guellec *et al.*, "A 25 μm pitch LWIR focal plane array with pixel-level 15-bit ADC providing high well capacity and targeting 2 mk NETD," *Infrared Technology and Applications XXXVI*, vol. 78660, 2010.

[4] S. Realov and K. Shepard, "On-chip combined C-V / I-V characterization system in 45-nm CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 48, 2013.

[5] L.-T. Pang and B. Nikolic, "Measurements and analysis of process variability in 90 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 44, 2009.

[6] S. Saxena, "Variation in transistor performance and leakage in nanometer-scale technologies," *IEEE Transactions on Electron Devices*, vol. 55, 2008.

[7] K. Agarwal *et al.*, "A test structure for characterizing local device mismatches," *Symposium on VLSI Circuits*, 2006.

[8] C. M. Mezzomo *et al.*, "Characterization and modeling of transistor variability in advanced CMOS technologies," *IEEE Transactions on Electron Devices*, vol. 58, 2011.

[9] V. Wang and K. Shepard, "On-chip transistor characterization arrays for variability analysis," *Electron Letters*, vol. 43, 2007.

[10] S. Realov, W. McLaughlin, and K. Shepard, "On-chip transistor characterization arrays with digital interfaces for variability characterization," *International Symposium on Quality Electronic Design*, 2009.

[11] X. Yu *et al.*, "Accurate chip leakage prediction: Challenges and solutions," *Symposium on VLSI Technology*, 2012.

[12] J.-C. Sun *et al.*, "Submicrometer-channel CMOS for low-temperature operation," *IEEE Transactions on Electron Devices*, vol. 34, 1987.

[13] Y. Creten *et al.*, "An 8-bit flash analog-to-digital converter in standard CMOS technology functional from 4.2k to 300k," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, 2009.

[14] A. Appaswamy, P. Chakraborty, and J. D. Cressler, "Cryogenic matching performance of 90 nm MOSFETs," *International Semiconductor Device Research Symposium*, 2009.

[15] R. Thusu, "The growing world of the image sensors market," *Sensors Magazine online*, 2012.

[16] A. El Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuits and Devices Magazine*, 2005.

[17] A. El Gamal, "Trends in CMOS image sensor technology and design," *International Electron Devices Meeting*, 2002.

[18] D. J. Brady *et al.*, "Multiscale gigapixel photography," *Nature Letters*, vol. 486, 2012.

[19] H. S. P. Wong, "Technology and device scaling considerations for CMOS imagers," *IEEE Transactions on Electron Devices*, vol. 43, 1996.

[20] M. G. Brown *et al.*, "Digital-pixel focal plane array development," *Quantum Sensing and Nanophotonic Devices VII*, 2010.

[21] T. Matsumoto, K. Kobayashi, and H. Onodera, "Impact of random telegraph noise on CMOS logic reliability," *Custom Integrated Circuits Conference*, 2014.

[22] S. O. Toh, *Nanoscale SRAM Variability and Optimization*. PhD thesis, U.C. Berkeley, 2011.

[23] T. Nagumo *et al.*, "Statistical characterization of trap position, energy, amplitude, and time constants by RTN measurement of multiple individual traps," *International Electron Devices Meeting*, 2010.

[24] Q. Tang *et al.*, "RTN induced frequency shift measurements using a ring oscillator based circuit," *Symposium on VLSI Technology*, 2013.

[25] S. Toh, T.-J. King, and B. Nikolic, "Impact of random telegraph signaling noise on SRAM stability," *Symposium on VLSI Technology*, 2011.

[26] H. Miki *et al.*, "Statistical measurement of random telegraph noise and its impact in scaled-down high-κ / metal-gate MOSFETs," *IEDM*, 2012.

[27] X. Wang *et al.*, "Random telegraph signal in CMOS image sensor pixels," *International Electron Devices Meeting*, 2006.

[28] C. Leyris *et al.*, "Impact of random telegraph signal in CMOS image sensors for low-light levels," *Proc. of the 32nd European Solid-State Circuits Conference*, 2006.

[29] P. Martin-Gonthier and P. Magnan, "RTS noise impact in CMOS image sensors readout circuit," *16th IEEE International Conference on Electronics, Circuits, and Systems*, 2009.

[30] P. Martin-Gonthier and P. Magnan, "Novel readout circuit architecture for CMOS image sensors minimizing RTS noise," *IEEE Electron Device Letters*, vol. 32, 2011.

[31] H. Miki *et al.*, "Voltage and temperature dependence of random telegraph noise in highly scaled hkmg etsoi nfets and its impact on logic delay uncertainty," *Symposium on VLSI Technology*, 2012.

[32] F. Baron *et al.*, "Effect of magnetic field on random telegraph noise in the source current of p-channel metal-oxide-semiconductor field-effect transistors," *Applied Physics Letters*, vol. 83, no. 710, 2003.

[33] H. Bu *et al.*, "Random telegraph signals and low-frequency noise in n-metal-oxide-semiconductor field-effect transistors with ultranarrow channels," *Applied Physics Letters*, vol. 78, no. 3259, 2000.

[34] A. Chimenton *et al.*, "A new methodology for two-level random-telegraph-noise identification and statistical analysis," *IEEE Electron Device Letters*, vol. 31, 2010.

[35] T. Nagumo *et al.*, "New analysis methods for comprehensive understanding of random telegraph noise," *International Electron Devices Meeting*, 2009.

[36] S. Realov and K. Shepard, "Random telegraph noise in 45-nm CMOS: Analysis using an on-chip test and measurement system," *International Electron Devices Meeting*, 2010.

[37] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, February 1989.

[38] K. Nowka, S. Nassif, and K. Agarwal, "Characterization and design for variability and reliability," *IEEE Custom Integrated Circuits Conference*, 2008.

[39] K. Abe *et al.*, "A test structure for statistical evaluation of pn junction leakage current based on CMOS image sensor technology," *IEEE International Conference on Microelectronic Test Structures*, 2010.

[40] V. Milovanovic and H. Zimmerman, "On fully differential and complementary single-stage self-biased cmos differential amplifiers," *EuroCon*, pp. 1955–1962, July 2013.

Appendix A

Test chip pin descriptions

Fig. A.1 shows the internal pin names and associated package pin numbers on the test chip.
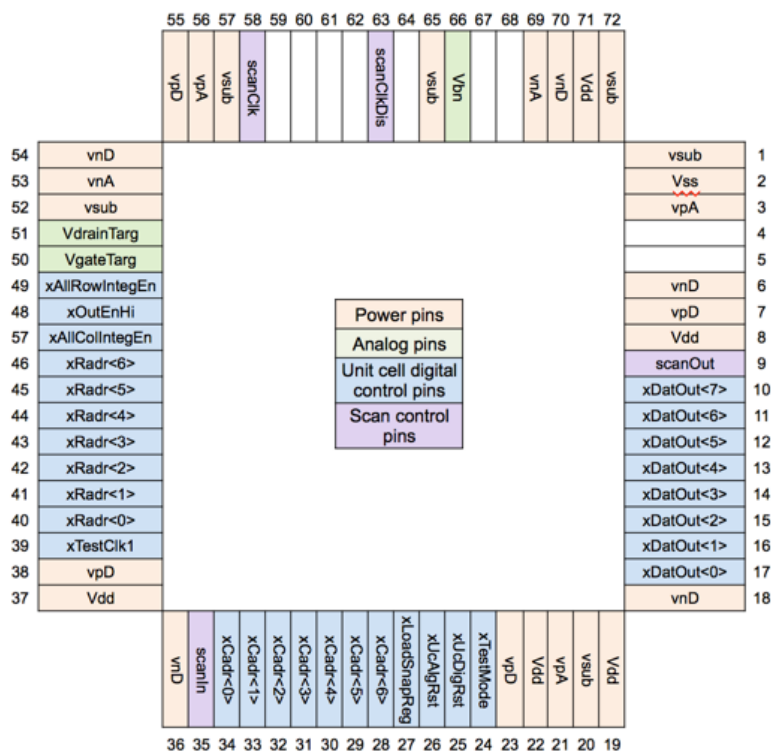

Figure A.1: Chip package pin names.

The functions of the pins are as follows:

- Power pins

  - **vsub**: Chip substrate voltage. All NMOS bulk terminals connect to this node.
  - **Vss**: Negative supply for output driver (negative supply for **scanOut** and **xDatOut<0:7>** pins).
  - **vnD**: Negative digital supply.
  - **vnA**: Negative analog supply.

  - `Vdd`: ESD diode voltage; ensures no input signals exceed this amount.
  - `vpD`: Positive digital supply.
  - `vpA`: Positive analog supply.

- Analog pins

  - `VdrainTarg`: Target DUT drain voltage. Routed to inputs of unity-gain feedback circuit setting DUT drain voltages and positive supplies of leakage DUT select logic gates.
  - `VgateTarg`: Target DUT gate voltage. Routed to positive supplies of RTS DUT select logic gates.
  - `Vbn`: Input current to measurement unit cells. Reduced by a factor of 10x during mirroring.

- Digital pins

  - `scanIn`: Scan input data.
  - `scanOut`: Scan output data.
  - `scanClk`: Scan clock.
  - `scanClkDis`: Disables scan clock and selects between PMOS and NMOS DUTs.
  - `xTestMode`: Configures counter to run in Digital Mode (1) — counts `xTestClk1` pulses — or Analog Mode (0) — counts reset pulses generated in analog portion of measurement unit cell
  - `xTestClk1`: Clocks the counter in the measurement unit cells when `xTestMode` is high.
  - `xUcDigRst`: Resets the digital counters.
  - `xUcAlgRst`: Resets the analog integration node.
  - `xLoadSnapReg`: Digital unit cell register enable.
  - `xRadr<0:6>`: Row address (0 is LSB, 6 is MSB).
  - `xCadr<0:6>`: Column address (0 is LSB, 6 is MSB).
  - `xAllRowIntegEn`: Row select integrate / read out enable.
  - `xAllColIntegEn`: Column select integrate / read out enable.
  - `xOutEnHi`: '0' selects low 8 bits, '1' selects high 8 bits to propagate to data output.
  - `xDatOut<0:7>`: Output data (bits 0-7 or 8-15 if `xOutEnHigh` is 0 or 1, respectively)

Appendix B

Detailed test procedure

B.1 Common test sequences

B.1.1 Enabling a single cell

In order to configure the chip to enable a single unit cell (for measurement or DUT configuration), do the following:

1. Set `xAllRowIntegEn` and `xAllColIntegEn` signals to 0V.

2. Apply the appropriate `xRowAdr` and `xColAdr` codes to select the desired unit cell. Note that columns alternate between DUT unit cells and measurement unit cells, so even integer column addresses correspond to DUT unit cells and odd integer column addresses correspond to measurement unit cells.

3. Set `xOutEnHi` low to ensure the chip propagates the counter LSBs.

4. Set `xUcDigRst` to 1 in order to reset the digital counter values.

5. Set `xUcAlgRst` to 1 in order to reset capacitor in the analog unit cell.

B.1.2 Initializing low DUT current

To ensure that the DUT array does not initially draw a large amount of current, use the following procedure:

1. Set `ScanClkDis` to 0V, which initiates the nMOS device test mode.

2. Set `VdrainTarg` to 0V to provide a low V_{DS} across the test devices.

3. Set `VgateTarg` to 500 mV.

4. Use the scan chain to turn all devices OFF:

    (a) Set `scanIn` to 1V (corresponding to n-type DUTs in the OFF mode)
    (b) Pulse `scanClk` 24 times to load the scan chain within the DUT cells.

B.2   Startup tests

B.2.1   Initial startup

To initialize testing, we can apply the voltages shown in Fig. B.1, and measure the static power consumption to ensure that the chip does not have any unexpected shorts that cause extremely high power dissipation levels.

Figure B.1: Pin voltages for initial startup test.

B.2.2   Scan chain test

To verify the scan chain operation, we can load test data into the scan chain and verify that the corresponding output scan data matches the scan chain input. The following test procedure can be used to load data into the scan chain:

1. Select a known test sequence of 24 binary values.

2. Set `scanClk` and `scanClkDis` low.

3. Set `scanIn` to the first voltage in the test sequence, and after a 10 ns delay, raise the `scanClk` input. After 10 ns, lower the `scanClk` value to 0V.

4. Repeat the above step for each remaining entry in the test sequence.

5. After 24 `scanClk` pulses, the scan chain data will be loaded.

   After the data is loaded, the scan chain output of a single cell can be selected using the `xColAdr`, `xRowAdr`, `xAllRowIntegEn`, and `xAllColIntegEn` signals.

1. Set `xAllRowIntegEn` and `xAllColIntegEn` signals to 0V if testing only a single cell is necessary. Alternately, if both signals are high, the `scanOut` signal will be the logical OR of all `scanOut` signals, providing quick verification that no signals are falsely pulled high. This check can be performed on a per-column or per-row basis by asserting `xAllColIntegEn` or `xAllRowIntegEn`, respectively.

2. If a single cell, row, or column is being tested, apply the corresponding digital address to the `xRowAdr` and `xColAdr` inputs of the chip. Only every other row contains a DUT unit cell, so the `xRowAdr` addresses must correspond to even integers.

3. Pulse `scanClk` 24 times; at each time pulse, the `scanOut` value should correspond to the next value of the test sequence.

B.2.3 Digital unit cell test

The functionality of the measurement unit cell counter may be verified by placing the test chip in a mode that sets the unit cell counters to read an externally-applied test clock signal instead of the reset pulses generated by the analog portion of the measurement cell. To do this, first use the procedure in Section B.1.1 to configure the chip to propagate the output of the desired test cell to the `xDatOut` bits. Note that because this test corresponds to measurement unit cells, `xRowAdr` must be an odd integer. After the chip has been configured, apply the following inputs to the test control pins and measure the counter output readings at the chip output:

1. Set `xUcDigRst` to 0 to re-enable digital counting.

2. Set `xTestMode` to 1V.

3. Apply a 0V to 1V clock to `xTestClk1` (at a known test frequency that can be verified by the output counter).

4. Measure `xDatOut<0:7>`, which should correspond to the 8 LSBs of the counter.

5. When the functionality of the 8 LSBs have been verified, switch `xOutEnHi` to 1V and measure `xDatOut<0:7>`. These should correspond to the 8 MSBs of the counter.

To verify the functionality for all unit cells, this process can be repeated for all of the row and column addresses.

B.2.4 Analog unit cell test

To test the functionality of the analog unit cell integration, we can apply an input current to the chip, measure the counter output after a fixed amount of time, and compare this reading to a similar reading obtained by applying a larger input current to the chip. The counter output should be larger when a larger test current is applied, corresponding to a higher integration frequency. For each test current:

1. First, configure the test devices to provide a minimum current level by initializing the DUT cell selection as described in Section B.1.2 while maintaining 0A input to the `Vbn` pin.

2. Then, enable a specific measurement cell by following the procedure outlined in Section B.1.1. Note that the column address must correspond to an odd integer in order to select a measurement unit cell. After enabling a single cell, configure the chip for analog integration by setting `xTestMode` to 0V. The test clock input can also be fixed low.

3. In order to generate a test current, ensure that `VdrainTarg` is around 500 mV to keep the mirror device saturated, and then increase the current applied to the `Vbn` pin. Increasing the current slowly should minimize rapid current caused by conduction in the non-selected test cells. Once the capacitor in these cells has been charged to the voltage supply, no further static current can be conducted in these cells.

4. After a fixed amount of time, sample the `xDatOut` reading. First, measure the eight output bits with `xOutEnHi` low (corresponding to the LSB readings), and then set `xOutEnHi` high to measure the counter MSBs.

B.3 Calibration

Before variability measurements can be taken, the current vs. frequency characteristics for each cell must be measured. This can be done by sweeping the input bias current over the target measurement range and measuring the corresponding counter output after a fixed time delay. This delay can be made shorter at higher test frequencies to prevent counter overflow and maximize sensing range. This procedure would be identical to the analog unit cell test methodology described above, over a wider range of applied test currents.

To minimize noise and maintain measurement accuracy, the readings should be averaged over multiple samples. In order to calibrate for the leakage current test range, the calibration current input to the `Vbn` node should range from 100 nA to 20 μA, which should translate to 10 nA to 2 μA at the bias current inputs for each unit cell.

B.4 Leakage current tests

Leakage current measurement cells are located every four unit cells at column addresses 1, 5, 9, 13, etc. Fig. B.2 shows an example of the main test waveforms needed for leakage current tests. A sequential procedure is also described below.

B.4.1 PMOS devices

1. Initialize the chip to select a specific measurement cell using the procedure described in Section B.1.1.

2. Configure the PMOS DUT unit cell to turn all devices on:

   (a) Set `VdrainTarg` to the desired PMOS drain voltage. The value of `VgateTarg` can be set to 500 mV.

   (b) Set `scanClkDis` to 0V to enable the scan clock. By default, this selects n-type test devices.

   (c) Set `scanIn` low and then pulse `scanClk` 24 times to set all devices to the ON state.

3. For measurement, set `scanClkDis` high to enable the PMOS test devices.

4. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Take multiple measurements to obtain a stable, low-noise reading.

5. After measuring the baseline frequency that accounts for constant parasitic currents, turn one of the DUT cells off:

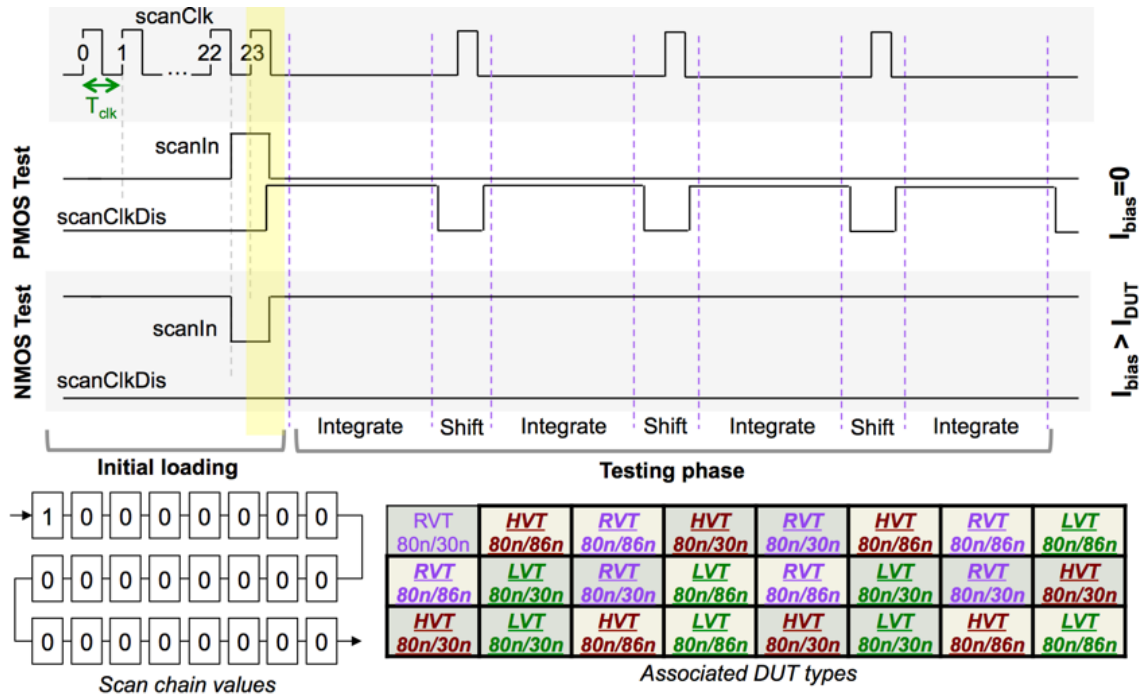   (a) Set `xUcAlgRst` and `xUcDigRst` high to stop integration.

Figure B.2: Sample leakage measurement test waveforms. The scan chain ordering and associated devices are shown below. Bold devices are in the ON state.

   (b) Set `scanClkDis` low to enable scan clock propagation.

   (c) While `scanClk` is low, transition `scanIn` high. Then, set `scanClk` high so to capture the OFF state reading in the scan chain. The first flip flop in the scan chain will now be configured to turn a PMOS device OFF.

6. To begin measurement again, set `scanClkDis` high to enable the PMOS test devices.

7. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Use the frequency vs. bias current plot to determine the current associated with the two frequency readings.

8. To sample the next cell, activate the reset control signals (set `xUcAlgRst` and `xUcDigRst` high) and then lower the `scanClkDis` signal. With `scanIn` low, pulse the scan clock to propagate the DUT control signals and repeat the measurement process.

B.4.2   NMOS devices

1. Initialize the chip to select a specific measurement cell using the procedure described in Section B.1.1.

2. Configure the NMOS DUT unit cell to turn all devices on:

   (a) Set `VdrainTarg` to the desired NMOS drain voltage. The value of `VgateTarg` can be set to 500 mV.

   (b) Set `scanClkDis` to 0V to enable the scan clock. By default, this selects n-type test devices.

    (c) Set `scanIn` high and then pulse `scanClk` 24 times to set all devices to the ON state.

3. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Take multiple measurements to obtain a stable, low-noise reading.

4. After measuring the baseline frequency that accounts for constant parasitic currents, turn one of the DUT cells off:

    (a) Set `xUcAlgRst` and `xUcDigRst` high to stop integration.

    (b) While `scanClk` is low, transition `scanIn` low. Then, set `scanClk` high so to capture the OFF state reading in the scan chain. The first flip flop in the scan chain will now be configured to turn an NMOS device OFF.

5. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Use the frequency vs. bias current plot to determine the current associated with the two frequency readings.

6. To sample the next cell, activate the reset control signals (set `xUcAlgRst` and `xUcDigRst` high) and then — with `scanIn` high — pulse the scan clock to propagate the DUT control signals and repeat the measurement process.

B.5   RTS tests

RTS measurement cells are located every four unit cells at column addresses 3, 7, 11, 15, etc. Fig. B.3 shows an example of the main test waveforms needed for noise tests. A sequential procedure is also described below. This includes an initial OFF state measurement to capture parasitic current measurements; however, the parasitics should be small relative to the ON state device current, so this calibration step can be neglected for coarse measurements.

B.5.1   PMOS devices

1. Initialize the chip to select a specific measurement cell using the procedure described in Section B.1.1.

2. Configure the PMOS DUTs to turn all devices off for calibration:

    (a) Set `VgateTarg` to 500 mV and `VdrainTarg` to the desired PMOS drain voltage.

    (b) Set `scanClkDis` to 0V to enable the scan clock. By default, this selects n-type test devices.

    (c) Set `scanIn` low and then pulse `scanClk` 24 times, setting all devices to the OFF state.

3. For measurement, set `scanClkDis` high to enable the PMOS test devices.

4. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Take multiple measurements to obtain a stable, low-noise reading.

5. After measuring the baseline frequency that accounts for constant parasitic currents, turn one of the DUT cells on:

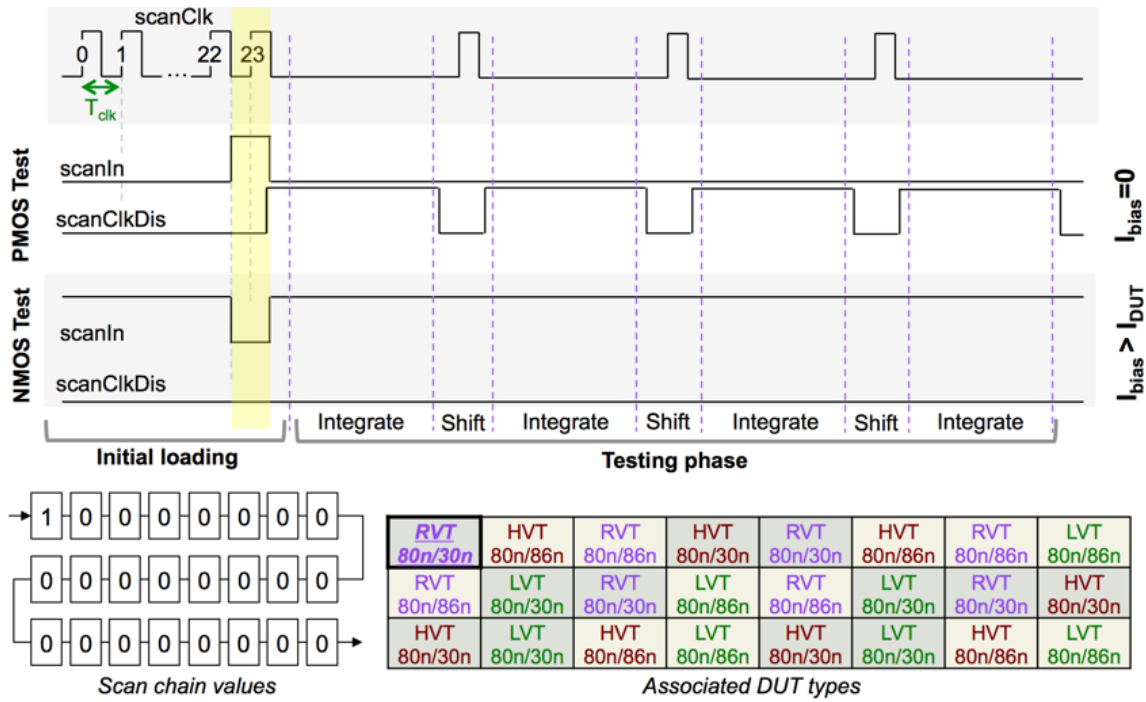   (a) Set `xUcAlgRst` and `xUcDigRst` high to stop integration.

Figure B.3: Sample RTS measurement test waveforms. The scan chain ordering and associated control devices are shown below. Bold devices are in the ON state.

- (b) Set `scanClkDis` low to enable scan clock propagation.
- (c) While `scanClk` is low, transition `scanIn` high. Then, set `scanClk` high so to capture the ON state reading in the scan chain. The first flip flop in the scan chain will now be configured to turn a PMOS device ON.

6. In order to force trapping or de-trapping of test devices, set `VgateTarg` to 0V or 1V (respectively) for 1 ms prior to changing `VgateTarg` to the desired gate bias voltage.

7. For measurement, set `scanClkDis` high to enable the PMOS test devices.

8. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Use the frequency vs. bias current plot to determine the current associated with the two frequency readings.

9. Take samples roughly every 20 μs over a 10 second time interval to obtain a plot of drain current vs. time.

10. To sample the next cell, activate the reset control signals (set `xUcAlgRst` and `xUcDigRst` high) and then lower the `scanClkDis` signal. With `scanIn` low, pulse the scan clock to propagate the DUT control signals. The OFF state current does not need to be re-measured.

B.5.2 NMOS devices

1. Initialize the chip to select a specific measurement cell using the procedure described in Section B.1.1.

2. Configure the NMOS DUTs to turn all devices off for calibration:

(a) Set `VgateTarg` to 500 mV and `VdrainTarg` to the desired NMOS drain voltage.

(b) Set `scanClkDis` to 0V to enable the scan clock. By default, this selects n-type test devices.

(c) Set `scanIn` high and then pulse `scanClk` 24 times, setting all devices to the OFF state.

3. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Take multiple measurements to obtain a stable, low-noise reading.

4. After measuring the baseline frequency that accounts for constant parasitic currents, turn one of the DUT cells on:

(a) Set `xUcAlgRst` and `xUcDigRst` high to stop integration.

(b) The `scanClkDis` signal should already be low, enabling scan clock propagation.

(c) While `scanClk` is low, transition `scanIn` low. Then, set `scanClk` high so to capture the ON state reading in the scan chain. The first flip flop in the scan chain will now be configured to turn an NMOS device ON.

5. In order to force trapping or de-trapping of test devices, set `VgateTarg` to 1V or 0V (respectively) for 1 ms prior to changing `VgateTarg` to the desired gate bias voltage.

6. Set the reset control signals (`xUcAlgRst` and `xUcDigRst`) low to enable integration and sample the counter output after a fixed integration time. Use the frequency vs. bias current plot to determine the current associated with the two frequency readings.

7. Take samples roughly every 20 μs over a 10 second time interval to obtain a plot of drain current vs. time.

8. To sample the next cell, activate the reset control signals (set `xUcAlgRst` and `xUcDigRst` high) and then — with `scanIn` high — pulse the scan clock to propagate the DUT control signals. The OFF state current does not need to be re-measured.