Fully Digital Process Variation Sensor for High Performance System-on-a-Chip

Digital process variations-sensor för högprestanda System-on-a-Chip

Viktor Lidholm

Supervisor: Dr. Prakash Harikumar
Ericsson AB, Linköping, Sweden

Supervisor: Dr. Alireza Saberkari
Linköping University, Sweden

Examiner: Dr. Mark Vesterbacka
Linköping University, Sweden
Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/her own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: https://www.ep.liu.se/.

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstå.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnad som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida https://www.ep.liu.se/.

© Viktor Lidholm
Abstract

Integrated Circuit (IC) designers have always faced the problem of small deviations in parameters of their designs in manufactured ICs. This variation due to the manufacturing process (Process Variation) in device parameters from their nominal values result in altered performance and can cause integrated circuits to be malfunctioning. With Moore’s Law going strong, the device dimensions of integrated circuits continue to shrink. This is exacerbating the issue with deviations from nominal design values. Several methods and techniques are employed to reduce the impact of this type of undesirable variation. This thesis explores a method of measuring the variation inside an IC. An accurate measurement of the variation has the potential to enable compensation techniques, tuning performance to desirable levels or even turning a malfunctioning ICs into functional ones.

This thesis proposes a method to obtain an accurate digital representation of the variation, a Process Sensor. The proposed sensor is capable of detecting the variation in P-type Metal Oxide Semiconductor (PMOS) and N-type Metal Oxide Semiconductor (NMOS) devices independently. In addition, a theoretical method to disentangle the temperature dependence from the variation measurement is explored. And conversely via the same method, temperature dependence can be increased, effectively turning the process sensing circuitry into a temperature sensing circuitry.
Acknowledgements

There are a number of people I want to thank for their involvement in this thesis. In no particular order, thanks to my examiner Mark Vesterbacka and supervisor Alireza Saberkari at Linköping University for their feedback and proofreading. Everyone at Ericsson, starting with my main supervisor, Prakash Harikumar for the continuous guidance, feedback, and proofreading. Erik Backenius & Christer Jansson for answering technical questions and giving advice. Niklas Andersson for the opportunity to work in collaboration with Ericsson on this thesis. And finally, Pontus Thomson for helping out with the report template & scripts for generating plots.
Contents

Acronyms

1 Introduction 1
  1.1 Motivation 1
  1.2 Problem Formulation 2
  1.3 Purpose 3
  1.4 Delimitation 3

2 Theory 4
  2.1 Inverter Delay & Ring Oscillators 4
    2.1.1 Inverter Delay Modeling 4
    2.1.2 Frequency of Ring Oscillators 5
  2.2 Variability 5
    2.2.1 PVT 5
    2.2.2 Process Variation 6
    2.2.3 Carrier Mobility 7
    2.2.4 Threshold Voltage 8
    2.2.5 Temperature Dependence 9
    2.2.6 Adaptive Control 10
  2.3 Threshold Voltage Estimation Based on Ring Oscillator 10
3 Method

3.1 Design of Process Sensor Utilizing Ring Oscillators
  3.1.1 Simplified Overview
  3.1.2 Oscillator
  3.1.3 Time to Digital Converter
  3.1.4 Estimating Error of Time to Digital Converter
  3.1.5 Detailed Overview

4 Results

4.1 Sensor Performance
  4.1.1 Temperature
  4.1.2 Adjusted Supply Voltage for Altered Temperature behavior
  4.1.3 Random Mismatch
  4.1.4 Corner Maps and Mismatch
  4.1.5 Power consumption

5 Discussion

5.1 Results
  5.1.1 Temperature
  5.1.2 Mismatch
  5.1.3 Power Consumption
  5.1.4 Limitations
  5.1.5 Method
  5.1.6 Source Criticism
  5.1.7 Ethical Aspects

6 Conclusions and Future Work

6.1 Conclusions
6.2 Future Work
List of Figures

2.1.1 Inverters connected in a Ring Oscillator configuration .......................... 5
2.2.1 Map of the different corners, SlowSlow (SS), SF (SlowFast), TT (Typical- Typical), FS (FastSlow), and FF (FastFast). The first letter indicates NMOS speed, and the second letter indicates PMOS speed. .......................... 7
2.2.2 I-V behavior at three different temperatures. ................................. 9

3.1.1 Simplified System Overview mapped onto corner map. ......................... 11
3.1.2 Ratioed inverters which are the core of the delay cells inside the Ring Oscillators. 12
3.1.3 P-type oscillator architecture. .................................................. 13
3.1.4 N-type oscillator architecture. .................................................. 13
3.1.5 Time to digital converter consisting of DFFs with inverted outputs, inverted edge trigger, and reset. .................................................. 14
3.1.6 Process Sensor system overview. ............................................... 16

4.1.1 N-type Oscillator Frequency and Code output over temperature. ........... 18
4.1.2 N-type Oscillator Frequency and Code output over restricted temperature range. .................................................. 18
4.1.3 N-type Oscillator Frequency and Code output over temperature. ........... 19
4.1.4 P-type Oscillator Frequency and Code output over restricted temperature range. .................................................. 19
4.1.5 P-type Oscillator Frequency and Code output with adjust supply voltage. 20
4.1.6 Three lower than nominal supply voltages and the corresponding output frequencies of the p-type oscillator for a wide temperature range. ............... 20
4.1.7 N-type oscillators’ frequencies for the different number of delay stages. .... 21
4.1.8  Stage configurations starting with 4 delay stages and increasing by 2 up to 12. Each configuration is simulated with 200 points of Monte Carlo mismatch. Mean and $\sigma$ decrease with an increase in delay stages.

4.1.9  Result from a 1000-point Monte Carlo Mismatch Simulation for 5 different clusters corresponding to 5 different models, shown in green dots, overlayed on a single run for the different corners, shown in diamonds. The relative location of points does not change much from 8 to 11 stages.

4.1.10 Result from a 1000-point Monte Carlo Mismatch Simulation for 5 different clusters corresponding to 5 different models, shown in green dots, overlayed on a single run for the different corners, shown in diamonds.
List of Tables

4.1.1 Coefficient of variation $C_v = \frac{\sigma}{\mu}$, where $\sigma$ is the standard deviation and $\mu$ is the mean, for the data presented in Figure 4.1.8. ........................................... 22

4.1.2 Power consumption for the P and N-type oscillators in ON and OFF mode at 85 $^\circ$C across the process corners. .................................................. 24
Acronyms

ABB  Adaptive Body Bias.
DFF  Data Flip-Flop.
DVS  Dynamic Voltage Scaling.
FinFET  Fin Field Effect Transistor.
IP  Intellectual Property.
LSB  Least Significant Bit.
PDK  Process Design Kit.
PVT  Process, Voltage, and Temperature.
RBB  Reverse Body Bias.
RO  Ring Oscillator.
SoC  System-on-a-Chip.
TDC  Time-to-Digital-Converter.
Nomenclature


displays

\[ V_{ss} \quad \text{Ground Reference Voltage.} \]
\[ V_{dd} \quad \text{Supply Voltage.} \]
\[ V_{sb} \quad \text{Source to Body Voltage.} \]
\[ V_{gs} \quad \text{Gate to Source Voltage.} \]
Chapter 1

Introduction

This chapter covers the motivation, problem formulation, purpose, and delimitation of the thesis.

1.1 Motivation

With ever-scaling technology, transistors are approaching the fundamental dimensions of atoms and wavelengths of light. One of the challenges that come with this is the management of variation due to manufacturing [1][2]. While some sources of variation have been mitigated with the introduction of the Fin Field Effect Transistor (FinFET), enabling transistor scaling to continue. However, it also introduces new sources of variation, in part due to its 3D geometry [3].

Traditionally, variation has been handled by designing around the worst possible cases using simulations and targeting an acceptable yield, typically $3\sigma$, resulting in a rejection rate of 0.26% [4]. More recently, optimized performance has become a major concern. In [5], a multi-core, 5G capable, smartphone System-on-a-Chip (SoC) is utilizing Process, Voltage, and Temperature (PVT) sensors and show a reduction in power consumption by 13% under full load. Over 500,000 of said SoC has been manufactured and is in use, demonstrating a very real, state-of-the-art use case incorporating a process sensor.

However, there is a cost associated with this design paradigm. To meet the desired specifications, large guard bands are used (design margins in voltage supply, timing demands, etc) which can result in excessive power consumption, performance left on the table, and extra circuitry consuming valuable area [6][7].

Another way to alter a circuit’s behavior after manufacturing is by applying a bias voltage to the body of the transistors. This is referred to as Adaptive Body Bias (ABB) and can be used to alter the threshold voltage of transistors [4].
1.2 Problem Formulation

The goal of this work is to investigate different methods of characterizing the performance of the on-chip transistors’ deviation from their nominal values using process sensing circuits. The underlying causes of variation are dependent on the manufacturing process and the exact steps are often a closely kept secret by the semiconductor fabricators. Therefore, the variation is characterized by lumped parameters such as threshold voltage, charge mobility, and effective geometry. For a sensor to be useful it needs to satisfy several criteria, these are:

- Simplicity
- Accuracy
- Portability
- Power efficiency
- Small area

Simplicity is desired because of how the sensor will interface with the rest of the SoC or even a different Intellectual Property (IP). This implies a digital output as opposed to an analog sensor which requires extra circuitry to convert a voltage or current into a digital value.

Accuracy should be determined based on the use case of the sensor. As shown in [5], a fairly crude sensor is all that is needed to get a significant improvement in power efficiency. Typically, a more accurate sensor requires more area, highly accurate supply voltage, accurate clocking, or a high-resolution Time-to-Digital-Converter (TDC). Since P and N-type devices vary independently, a sensor capable of detecting variance in both devices independently will impose area cost.

Portability across technology nodes is desirable. Thus, the sensor cannot utilize process-specific devices such as MiM capacitors, special resistors, or MOS devices. In addition, it cannot use circuit designs inextricably linked to a specific Process Design Kit (PDK). The sensor should consist of devices common to most PDKs.

Power efficiency is increasingly becoming a concern in chip design. The power budget of a chip is limited and with each new generation of smaller transistors this is exacerbated by the breakdown of Dennard scaling, meaning that power consumption remains constant, but the
devices are shrinking leading to an increase in power density which in turn is undesirable due to thermal effects.

A small area is desired as the total chip area is limited. Chips are becoming more monolithic and integrating more and more types of circuits on the same chip. Thus, the sensor should be made small so as to leave sufficient space for other types of circuits. This combined with the fact that sensors are often duplicated makes the need for a small area a strong criterion.

1.3 Purpose

The purpose of the thesis is to investigate methods of detecting variability and to design a process sensor that fits the design criterion. Successful implementation of an SoC with a system including a process sensor can give the following benefits:

- Lower power consumption
- Increased manufacturing yield
- Optimized performance in terms of computation

It should be mentioned that the first two mentioned items are directly connected to a sustainable environment. Chips are manufactured on a scale of millions or even billions, a small increase in yield and power efficiency leads to fewer chips being discarded as waste and less power needing to be produced. Extending the battery life in smartphones reduces the need to buy a new device on the one hand, and on the other reduces the power consumption of large computation centers, cloud computation, which has a substantial environmental footprint and continues to increase.

1.4 Delimitation

The device technology used in the simulations of this thesis is protected under a non-disclosure agreement. As such, certain results are not shown with exact values and will instead be omitted or expressed in arbitrary units.
Chapter 2

Theory

This chapter describes the relevant theory for understanding underlying process variation and how it can be correlated to a self-oscillating circuit.

2.1 Inverter Delay & Ring Oscillators

This section describes how inverter delay and Ring Oscillator (RO) frequency can be modeled and understood.

2.1.1 Inverter Delay Modeling

One way to estimate the RC time constant of an inverter, $\tau = RC$, is to use the $\alpha$-power law model for the saturation current [8], combined with the fact that the delay is proportional to $CV_{dd}/I$ [4]. The estimate becomes

$$\tau = k \frac{CV_{DD}}{(V_{DD} - V_{th})^\alpha} \quad (2.1.1)$$

where $k$ is process parameters and $\alpha$ is the velocity saturation index which has a value between 1 and 2. Both $k$ and $V_{th}$ are subject to the effects of process variation, thus the time constant, $\tau$, is as well.

The propagation delay, $t_{pd}$, is the time it takes to charge or discharge the output to 50% of $V_{DD}$ and is described by

$$t_{pd} = RC \ln 2 = \tau \ln 2. \quad (2.1.2)$$
2.1.2 Frequency of Ring Oscillators

An RO is a cascaded combination of inverters connected in a closed loop, see Figure 2.1.1. In addition, a necessary condition for self-oscillation is that the number of inverting stages is odd. The oscillation frequency $f_o$ of an RO is proportional to the time delay $t_{pd}$ of each inverting stage and the total number of inverting stages $n$.

$$f_o = \frac{1}{2t_{pd}n} \quad (2.1.3)$$

If the propagation delay $t_{pd}$ is not equal for the propagation delay for the rise time $t_{pdr}$ and fall time $t_{pdf}$, the ROs frequency can be described as

$$f_o = \frac{1}{(t_{pdr} + t_{pdf})n}. \quad (2.1.4)$$

2.2 Variability

This chapter seeks to explore the parameters of variability and their underlying causes, linking them to the macro behavior of circuits.

2.2.1 PVT

The behavior of circuits is affected by variability, the source of this variability is commonly divided into three categories:

- Process variation
- Supply voltage
- Temperature

These are referred to as PVT. Process variation is mainly a result of manufacturing, but also to a lesser extent due to aging. Voltage and temperature are environmental conditions and vary with time, albeit on a much smaller time scale in comparison to aging.
2.2. VARIABILITY

2.2.2 Process Variation

The following section is more or less accurate depending on the technology being examined. The traditional planar technologies are well studied and understood but this knowledge may not be accurate in regard to the state-of-the-art technologies incorporating FinFETs or Gate-All-Around devices owing to their geometries and different manufacturing steps.

In the planar case, the variations in $V_{th}$ and channel length $L$ impact circuits the most. $V_{th}$ variation is mainly due to dopant variations stemming from the bombardment of atoms during manufacturing and is discussed more in detail in section 2.2.4. Channel length variations are a result of photo-lithography proximity effects, deviations in the optics, and plasma etch dependencies [4].

As transistors scale down, they contain fewer and fewer atoms and as a result, they exhibit less self-averaging of the variability thus exacerbating the impact of variation. The largest source of variation is proportional to the doping concentration; however, some progress has been made with the introduction of FinFETs which generally have lower doping concentration in comparison to planar technologies [9] [10].

FinFETs exhibit a dramatic reduction in random discrete dopants. However, new sources of variability are introduced instead. The dominating parameters of FinFETs are line edge roughness (LER), metal gate granularity (MGG), and interface trapped charges (ITC), and are discussed in [10].

Variability is still a major concern for designers, even though the sources that have traditionally dominated the variation have been significantly reduced. This is due to the fact that other sources’ impact on the variation has increased and now dominate. For the designer, it is not necessary to understand what the causes are but to characterize the circuit’s performance under variation. To this end, simulation models are supplied with the PDKs. These models capture the extremes of the variation and include a lot of different altered parameters. In the more advanced technologies, these models are often encrypted, and their details are a closely guarded secret. These models are referred to as Corner Models and the variation is characterized by transistor speed, see Figure 2.2.1. The global variation of a chip will be found somewhere on the corner map in between the edges of the extremes. However, it is exceedingly rare for the chip to be close to the edges but if a chip simulation can meet the specification under the extreme corners - the chip is likely robust.
Figure 2.2.1: Map of the different corners, SlowSlow (SS), SF (SlowFast), TT (TypicalTypical), FS (FastSlow), and FF (FastFast). The first letter indicates NMOS speed, and the second letter indicates PMOS speed.

### 2.2.3 Carrier Mobility

The average speed \( v \) of charge carriers in the channel of a transistor is proportional to the electric field \( E \) between the source and drain. The relation is scaled by the constant \( \mu \), which is called carrier mobility.

\[
v = \mu E
\]  

Equation 2.2.1 comes with some limitations. Qualitatively this can be understood by examining what takes place in the channel of a transistor. Due to scattering effects inside the channel of the transistor, the carrier mobility is a function of the field strengths present in the channel. Carriers are close to the oxide interface and are scattered by the roughness in the surface. In addition, coulombic interactions with fixed charges in the gate oxide will also occur. All these effects will degrade carrier mobility. Since the resulting roughness in the interface of the channel and the charges inside the gate oxide both are random processes, they suffer from variation due to manufacturing. The net result is that the carrier mobility has been degraded by some random amount [11]. FinFET devices have a relatively larger surface area due to the 3-D structure of the fins compared to planar devices. Since the current flow at the surface of the channel, surface roughness degrades carrier mobility.
2.2. VARIABILITY

2.2.4 Threshold Voltage

One of the parameters that are affected the most by process variation is the threshold voltage. The threshold voltage, $V_{th}$, is defined by:

$$V_{th} = V_{th0} + \gamma \left( \sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right). \quad (2.2.2)$$

Here $V_{th0}$ is the threshold voltage when the source and body are at the same potential, $\phi_s$ is the surface potential, and $\gamma$ is the body effect coefficient.

$$\phi_s = 2v_T \ln \frac{N_A}{n_i} \quad (2.2.3)$$

$v_T$ is the thermal voltage, 26 mV at room temperature, $N_A$ is the doping concentration of acceptors in the channel, and $n_i$ is the intrinsic doping concentration of silicon.

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_{si}N_A} \quad (2.2.4)$$

$t_{ox}$ is the oxide thickness, $\epsilon_{ox}$ is the permittivity of the gate oxide, $q$ is the fundamental charge of electrons, $\epsilon_{si}$ is the permittivity of silicon, and $N_A$ is the doping concentration of acceptors in the channel.

From Equation 2.2.2 it is shown that $V_{th}$ is proportional to $V_{sb}$, implying that $V_{th}$ can be adjusted by applying a voltage to the body. Additionally, Equations 2.2.3 and 2.2.4 are functions of the doping concentration $N_A$, which is subject to variation due to manufacturing [4].

Furthermore, it has been shown in [12] & [13] that the standard deviation of threshold voltage, $\sigma V_{th}$, is

$$\sigma V_{th} = \frac{\sqrt{4q\epsilon_{si}\phi_s}}{2} \cdot \frac{t_{ox}}{\epsilon_{ox}} \cdot \frac{\sqrt{N_A}}{\sqrt{W_{eff}L_{eff}}}. \quad (2.2.5)$$

Here, $W_{eff}$ & $L_{eff}$ are the effective width and length of the transistor, respectively.

Since a low $\sigma V_{th}$ is generally desirable, Equation 2.2.5 shows the issue with shrinking geometry as the $\sigma$ is inversely proportional to the square root of the area. Also, the doping concentration, $N_A$, is increased in the nanometer planar technologies in order to counteract short-channel effects, further exacerbating the variation. As transistors shrink with new technology generations, the $\sigma V_{th}$ increases.

In the case of the FinFET, the doping concentration, $N_A$, is drastically reduced which in turn suppresses the variation. However, threshold voltage variations still affect the transistor due to other sources stemming from the manufacturing process related to FinFETs. The net result is that threshold voltage variation will occur, albeit in a slightly different manner.
2.2.5 Temperature Dependence

The characteristics of transistors are dependent on temperature. The carrier mobility $\mu$ decreases with temperature and is approximated by

$$\mu(T) = \mu(T_r) \left( \frac{T}{T_r} \right)^{-k_\mu}$$

(2.2.6)

where $T$ is absolute temperature, $T_r$ is room temperature, and $k_\mu$ is a fitting parameter.

$V_{th}$ has a nearly linear relationship to temperature and is approximated by

$$V_{th}(T) = V_{th}(T_r) - k_{vth}(T - T_r)$$

(2.2.7)

and $k_{vth}$ is a fitting parameter, typically 1-2 mV/K.

These two parameters counteract one another and there exists a point at which they balance each other out.

![I-V behavior at three different temperatures.](image)

Figure 2.2.2: I-V behavior at three different temperatures.

Shown in Figure 2.2.2 is the intersection point of the I-V characteristics of an NMOS transistor at three different temperatures. This implies that the $V_{gs}$ where the lines intersect, the temperature will not affect the drain current, and the device is temperature independent.

As technology generations progress and devices continue to shrink, the supply voltage to threshold voltage ratio is also shrinking. In deeply scaled technologies, the temperature speed characteristics are completely reversed, and as such propagation delay is decreased with an increase in temperature. It has also been shown that on the circuit level, there exists a supply voltage for which the temperature fluctuations are canceled. However, while it is unlikely that this voltage will be close to the nominal supply voltage and depending on the circuit, it can have adverse effects such as an increase in propagation time and increased electrical stress [14].
2.2.6 Adaptive Control

One way to counter the effects of random variation introduced in the manufacturing process is to measure the operating conditions of the chip and employ different techniques to compensate for non-ideal conditions, which is called Adaptive Control. $I_{off}$ can be reduced by applying Reverse Body Bias (RBB), reducing power consumption in sleep modes.

Dynamic Voltage Scaling (DVS) is commonly used in modern SoCs. DVS adjusts the supply voltage of a sub-circuit based on workload and temperature, traditionally this adjustment is conservative because it assumes the worst-case operating conditions. This can be optimized by also taking into account process variability measurements, leading to a lower voltage set by the DVS.

ABB can be used to adjust the $V_{th}$ across regions of the SoC, as shown in 2.2.2. Systematic variations in $V_{th}$ can be adjusted to more optimal values by adjusting $V_{sb}$ resulting in dynamic control of the threshold voltage where needed. Essentially, the characteristics of the die can be moved to the desired location on the corner map [4].

2.3 Threshold Voltage Estimation Based on Ring Oscillator

This RO structure uses a similar methodology to that introduced by [15] and then further improved by [16]. The underlying variation parameters are related to the oscillation frequency of a Ring Oscillator. A nominal frequency, $f_{Ref}$, derived from an RO with typical parameter values is compared to a measured frequency, $f_M$, with parameters deviating from the nominal values. A simplified model of $\Delta f$ is described by:

$$\Delta f = f_M - f_{Ref} = k_P \Delta V_{THP} + k_N \Delta V_{THN}. \quad \text{(2.3.1)}$$

Here $V_{THP}$ and $V_{THN}$ are the global variations of threshold voltage for p-type and n-type transistors respectively. $k_N$ and $k_P$ are sensitivity coefficients of the RO.

$$k_P = \frac{\Delta f/f_0}{\Delta V_{THP}/\Delta V_{THP0}}, \quad k_N = \frac{\Delta f/f_0}{\Delta V_{THN}/\Delta V_{THN0}} \quad \text{(2.3.2)}$$

The main drawback to using an RO with standard inverters in the chain is that it has low sensitivity coefficients. Therefore efforts have been made to improve the sensitivity by altering the inverter cells. A number of different inverter structures have been explored in [15] and [16]. The RO used in this work is a modified version of the structure with the highest sensitivity in the aforementioned papers.
Chapter 3

Method

This chapter describes the design methodology.

3.1 Design of Process Sensor Utilizing Ring Oscillators

In this section, the system and its sub-systems is described.

3.1.1 Simplified Overview

At the simplest level, the operation of the Process Sensor can be described as follows: A ring oscillator with an output frequency proportional to the underlying process variations is converted into a digital output code corresponding to the frequency.

Figure 3.1.1: Simplified System Overview mapped onto corner map.
3.1. DESIGN OF PROCESS SENSOR UTILIZING RING OSCILLATORS

3.1.2 Oscillator

The sensor is designed to have a high sensitivity to p and n devices. This is achieved by utilizing two different oscillating structures which are sensitive to n and p devices respectively. Both structures consist of delay cells that each invert the signal. Each delay cell only consists of one type of transistor to achieve sensitivity to the device of interest. The fundamental delay cells are shown in Figure 3.1.2.

![N-ratioed Inverter](image1.png) ![P-ratioed Inverter](image2.png)

(a) Ratioed inverter made with only n devices.  (b) Ratioed inverter made with only p devices.

Figure 3.1.2: Ratioed inverters which are the core of the delay cells inside the Ring Oscillators.

The problem with only using one type of device inside the delay cells is that the output of the oscillator will not be able to reach one of the voltage levels of the supply rails which in turn can cause the oscillator to find a stable state which means that the structure is not oscillating anymore. The proposed solution is to use a *Wave Reshaper* at the end of the chain of delay cells. The *Wave Reshaper* consists of a CMOS inverter and a NAND gate to facilitate enable functionality, this extends the functionality of the oscillator and ensures rail-to-rail operation.

Figure 3.1.3 & Figure 3.1.4 shows the oscillator architecture. Since the ratioed inverters cannot be turned off like a CMOS inverter, half of the ratioed inverters will have a static power draw when the NAND gate is preventing oscillation. To remedy this, an additional transistor is connected to the supply voltage in the p-type and ground reference in the n-type, which will act as a power switch for the oscillators. This ensures a proper turn-off mode.

Previously in [16] a delay cell consisting of a ratioed inverter in series with a CMOS inverter is proposed. The structure proposed in this work cascades the ratioed inverters and uses a
Wave Reshaper to ensure oscillation. The result is an oscillator with a higher ratio of n to p
devices or p to n devices for the respective n or p process measurements.

P-type Oscillator

![P-type oscillator architecture](image)

Figure 3.1.3: P-type oscillator architecture.

N-type Oscillator

![N-type oscillator architecture](image)

Figure 3.1.4: N-type oscillator architecture.
3.1.3 Time to Digital Converter

The purpose of the TDC is to convert the natural oscillating frequency of the n and p-sensitive oscillators into a digital code. The signals in the TDC are seen in Figure 3.1.5. The *enable* signal is the duration of the measurement and will effectively forward the oscillating output of the ring oscillator, *ro_freq*. The *reset* signal is used to put the TDC into a known state. During a measurement, *enable* high, each falling edge of *ro_freq* is captured by the first Data Flip-Flop (DFF) and propagated through the chain of DFFs. The inverted output of the DFF is fed back to the data input and will toggle between high and low for each falling edge of the *ro_freq*. The principle is the same for an arbitrarily long chain of DFFs as seen in Figure 3.1.5. The number of falling edges for a given *enable* duration will be found in the binary representation of *D0* to *D(n − 1)* plus one, where *D0* is the Least Significant Bit (LSB). The reason for adding one to the binary output is because a reset will put the internal state of the DFFs to low which in turn will put all *D(n − 1)* bits high. The output code will be valid for a finite amount of time after the enable signal toggles, this is due to the propagation delay of the DFF chain and is proportional to the number of DFFs.

The output result is a digital code that is a function of the oscillation frequency of one of the n or p-type oscillators.

![Figure 3.1.5: Time to digital converter consisting of DFFs with inverted outputs, inverted edge trigger, and reset.](image-url)
3.1.4 Estimating Error of Time to Digital Converter

The measurement time is proportional to the accuracy of the measurement result. An empirical method for deriving the relative error is used by using simulation data. Frequency is derived from the output waveform of the oscillator, $f_{osc}$, and compared to the frequency derived from the output code of the TDC, $f_{code}$.

\[
    f_{code} = \frac{code + 1}{t_{enable}} \tag{3.1.1}
\]

\[
    Relative\ Error\% = \left| \frac{f_{osc} - f_{code}}{f_{osc}} \right| \cdot 100 \tag{3.1.2}
\]

The enable duration can thus be increased until the desired relative error is achieved. Which in turn will dictate the number of DFFs needed to accommodate the largest possible output code without overflowing.
3.1.5 Detailed Overview

The proposed system for the process sensor can be seen in Figure 3.1.6. For a measurement code to be generated, the system uses a reference clock which serves as the time reference for the measurement duration. The control block will enable the counter for a set amount of measurement clock cycles from the reference clock. The control block is implemented in Verilog-A and consists of an internal counter triggered on the rising edge of the reference clock. To generate an output code for each of the oscillator types, the control behavior can be divided into three stages. These are the setup, measurement, and read stages. During the setup stage, the oscillator of interest is selected and the counter is reset. During the measurement phase, the TDC is enabled for a set amount of cycles from the reference clock. And finally, during the read stage, the counter is disabled and the code output is valid.

The reference clock frequency is arbitrary but has some upper limitations. There exists some startup time for the oscillators and a propagation delay before the TDC output is valid. The reference clock frequency was arbitrarily selected to be 500MHz, this is to ensure that sufficient time has passed during one buffer cycle between oscillator startup and counter enable, and also between counter disable and code read.

Figure 3.1.6: Process Sensor system overview.
Chapter 4

Results

In this section, all the simulation results are presented.

4.1 Sensor Performance

Performance of sensor further divided into sections presenting general performance & temperature behavior, effects of mismatch, and compensation technique by adjusting supply voltage.

4.1.1 Temperature

This section presents the simulation results of the frequency of the RO and the output code of the TDC. In addition, each corner of the 5 corners is swept over a temperature range. Differences between the n-type and p-type oscillators can be observed in their temperature behavior.

In Figure 4.1.1 the frequency and the code are shown to strongly correlate. In addition, the separation between the n-type corners’ frequency and code is wide while the p-type corners’ are low. A temperature dependence can also be observed.

In Figure 4.1.2 a realistic temperature range is shown. The narrower temperature range highlights the low temperature dependence.
4.1. SENSOR PERFORMANCE

(a) Frequency vs temperature for the different corners.

(b) Code vs temperature for the different corners.

Figure 4.1.1: N-type Oscillator Frequency and Code output over temperature.

(a) Frequency vs temperature for the different corners.

(b) Code vs temperature for the different corners.

Figure 4.1.2: N-type Oscillator Frequency and Code output over restricted temperature range.
4.1. SENSOR PERFORMANCE

In Figure 4.1.3 the p-type oscillator’s corresponding performance is shown. Notable differences are the shift in frequency for all corners and the steeper slopes owing to the increased temperature dependence of the p-type devices. However, the separation between the p-type corners is maintained while the n-type corners are close. In Figure 4.1.4 a narrower temperature range is shown. In comparison to the n-type oscillator, the slope is steeper.

Figure 4.1.3: N-type Oscillator Frequency and Code output over temperature.

Figure 4.1.4: P-type Oscillator Frequency and Code output over restricted temperature range.
4.1. SENSOR PERFORMANCE

4.1.2 Adjusted Supply Voltage for Altered Temperature behavior

Adjusting the supply voltage can be seen to alter the temperature behavior, suggesting that the effects of temperature can be compensated for by increasing the supply voltage. In Figure 4.1.5 the result of adjusting the supply voltage can be observed. The result is derived in the same manner as described in Section 4.1.1. Conversely, the temperature behavior can be altered in the opposite direction by lowering the supply voltage. In Figure 4.1.6 an almost straight line correlation is shown to be obtained for a certain Vdd, which might be used for separate temperature measurements.

![Figure 4.1.5: P-type Oscillator Frequency and Code output with adjust supply voltage.](image1)

(a) Frequency Vs. temperature with adjusted supply voltage.

![Figure 4.1.6: Three lower than nominal supply voltages and the corresponding output frequencies of the p-type oscillator for a wide temperature range.](image2)

(b) Code Vs. temperature with adjusted supply voltage.

Figure 4.1.5: P-type Oscillator Frequency and Code output with adjust supply voltage.

Figure 4.1.6: Three lower than nominal supply voltages and the corresponding output frequencies of the p-type oscillator for a wide temperature range.
4.1.3 Random Mismatch

This section shows the effect of random mismatch on the sensor circuit. In Figure 4.1.7 each stage configuration shows the result from a 200-point Monte Carlo mismatch simulation. An increase in stages is associated with a lower mean frequency and a lower frequency $\sigma$. In Figure 4.1.8 the mean frequency can be seen tracking the standard deviation and in Table 4.1.1 their ratio is expressed as the coefficient of variation. The data is inconclusive and more Monte Carlo points are needed to make a judgment. However, the indication is that there is no benefit in terms of suppressing mismatch by increasing the number of stages.

![N-type Oscillator Freq. - 200p Monte Carlo Mismatch Sim.](image)

Figure 4.1.7: N-type oscillators’ frequencies for the different number of delay stages.
4.1. SENSOR PERFORMANCE

(a) The mean $\mu$ frequency of an n-type oscillator for different delay stage configurations.

(b) Standard deviation $\sigma$ of the frequency of an n-type oscillator for different delay stage configurations.

Figure 4.1.8: Stage configurations starting with 4 delay stages and increasing by 2 up to 12. Each configuration is simulated with 200 points of Monte Carlo mismatch. Mean and $\sigma$ decrease with an increase in delay stages.

Table 4.1.1: Coefficient of variation $C_v = \frac{\sigma}{\mu}$, where $\sigma$ is the standard deviation and $\mu$ is the mean, for the data presented in Figure 4.1.8.

<table>
<thead>
<tr>
<th>No. Stages</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_v$</td>
<td>0.0190</td>
<td>0.0200</td>
<td>0.0187</td>
<td>0.0199</td>
<td>0.0242</td>
</tr>
</tbody>
</table>

4.1.4 Corner Maps and Mismatch

Here the results are derived from the detailed system description shown in Section 3.1.5. For each simulation point, the TDC code is read twice, once for the result from the n-type oscillator and once for the p-type oscillator. The codes are then mapped onto the corner map where the x-axis is the n-code and y-axis is the p-code.
Figure 4.1.9: Result from a 1000-point Monte Carlo Mismatch Simulation for 5 different clusters corresponding to 5 different models, shown in green dots, overlayed on a single run for the different corners, shown in diamonds. The relative location of points does not change much from 8 to 11 stages.

Figure 4.1.10: Result from a 1000-point Monte Carlo Mismatch Simulation for 5 different clusters corresponding to 5 different models, shown in green dots, overlayed on a single run for the different corners, shown in diamonds.
4.1. SENSOR PERFORMANCE

4.1.5 Power consumption

Power consumption for the P and N-type oscillators, Figures 3.1.3 and 3.1.4, is shown in Table 4.1.2. A drastic reduction in power consumption is observed in the off mode. Some minor variation in power consumption is also observed between the corners.

Table 4.1.2: Power consumption for the P and N-type oscillators in ON and OFF mode at 85°C across the process corners.

<table>
<thead>
<tr>
<th>Corner</th>
<th>SS [µW]</th>
<th>SF</th>
<th>TT [µW]</th>
<th>FS</th>
<th>FF [µW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-type On [µW]</td>
<td>107.6</td>
<td>116.4</td>
<td>147.9</td>
<td>184.6</td>
<td>195.1</td>
</tr>
<tr>
<td>N-type Off [nW]</td>
<td>1.1</td>
<td>3.4</td>
<td>3.9</td>
<td>10.1</td>
<td>16.1</td>
</tr>
<tr>
<td>P-type On [µW]</td>
<td>71.2</td>
<td>136.1</td>
<td>106.4</td>
<td>80.68</td>
<td>146.1</td>
</tr>
<tr>
<td>P-type Off [nW]</td>
<td>4.7</td>
<td>11.6</td>
<td>12.7</td>
<td>42.2</td>
<td>55.0</td>
</tr>
</tbody>
</table>
Chapter 5

Discussion

This chapter contains a discussion regarding the results, methodology, and potential ethical impact.

5.1 Results

The results presented show the operation of a proposed Process Sensor. The different corners specify the range in which the sensor operates. The results show the operation under a variety of different conditions and how it affects the result.

The need for detecting skewed performance of the devices imposes the need for two separate oscillator structures, one sensitive to the N-type devices and one to the P-type devices. The conversion of the oscillator’s frequency into digital codes should see a strong correlation.

5.1.1 Temperature

The result shown in Figure 4.1.1 indicates a strong correlation between the oscillator’s frequency and output code. It also shows how the temperature affects the measurement. Around the TT corner, strong temperature independence can be observed, and this is consistent with the temperature behavior described in Section 2.2.5. For this particular device type and circuit, it was possible to center the point at which the parameters affecting the temperature balance each other out. It should also be noted that the slope of the temperature curves changes depending on the corner, suggesting that $V_{gs}$ alone do not control the behavior. One possibility is that it is the overdrive voltage that will dictate which parameter will dominate behavior, since $V_{th}$ is a function of the process corner. Figure 4.1.2 shows the operation in a narrower, more realistic operating condition and the temperature independence is highest for the TT corner.
5.1. RESULTS

The P-type oscillator performance shown in Figure 4.1.3 shows a strong correlation between the oscillator frequency and the output code, similar to the N-type oscillator. It also indicates that the P-type oscillator is operating slower compared to the N-type oscillator for the same number of delay stages and transistor sizes. The temperature dependence is also a lot stronger in the P-type oscillator as the slopes are steeper for the whole temperature range in comparison the N-type oscillator. Even in the narrower temperature range seen in Figure 4.1.4, the slope can still be observed. Since the slopes are all positive, this suggests that the supply voltage can be increased to find the point where the parameters affecting temperature balance each other out. By modifying the number of stages and increasing the supply voltage by 20%, as seen in Figure 4.1.5, the point of balance can be observed and is centered around the TT corner. However, this potential solution to the temperature dependence behavior does come with a major cost since increasing the supply voltage by 20% requires some type of voltage regulation circuit.

Conversely, the result shown in Figure 4.1.6 suggests that a linear relationship between oscillation frequency and temperature can be obtained by lowering the supply voltage. This suggests that the output code will correlate linearly with temperature which means that the code can be used to measure the temperature of the oscillator. However, the exact supply voltage this occurs at is in itself a function of process parameters, so the straight line behavior will deviate slightly. A potential future solution is to have a tuneable supply voltage, first, take the process measurement and then via a lookup table find the supply voltage for the linear temperature behavior. This means that the exact same circuitry can be used for process and temperature measurements.

5.1.2 Mismatch

Introducing mismatch into the circuit changes the oscillation frequency of the oscillators. The spread of the frequency can be suppressed by using more delay stages, as described in the literature. This effect can be seen in Figure 4.1.7. The $\sigma$ is inversely correlated with the number of stages, but so is the mean $\mu$. Figure 4.1.8 shows the same data as the histogram, represented slightly differently. It can be seen that the $\mu$ and $\sigma$ track each other. Calculating the coefficient of variation shown in Table 4.1.1 shows no real benefit. However, the result is inconclusive because the number of random simulation points used to obtain the result is small.

The final result of the sensor is presented in Figure 4.1.9. First, the diamonds show the results from using the corner models previously presented in Section 4.1.1 at 85°C. In addition, the five clusters of green dots correspond to another set of simulation models. These are Monte Carlo models with a set global variation corresponding to the five different corners. In addition, local mismatch variation is applied and the resulting spread of the clusters can be observed. The Monte Carlo model’s global variation is not as large as the variation from the models used to generate the diamonds. However, even with the mismatch variation, it is
5.1. RESULTS

possible to determine the global variation as the clusters don’t overlap, in a sense this is the resolution.

Comparing the 8 stages to the 11 stages setup the results are inconclusive. As previously discussed, the spread will lower but so does the count. The result looks like a shifted version where everything has been scaled by some constant. More simulation point per cluster is required to determine if more stages can suppress the effects of mismatch on the measurement accuracy. In Figure 4.1.10 the amount of measurement cycles have been doubled, and again the relative position of the points and spread seems to remain the same. Everything has shifted up by a factor of 2. Another conclusion that can be drawn from the increased measurement time is that the quantization error introduced by the glstdc is minimal because of the similarities in the relative positions of the simulation points, suggesting that the measurement ran for enough time in the initial setup.

5.1.3 Power Consumption

Table 4.1.2 shows the power consumption of the oscillators in on and off modes. This illustrates that the oscillators can be turned off, thanks to the power switch transistors shown in Figure 3.1.3 and Figure 3.1.4.

5.1.4 Limitations

Not having direct access to editing the model files restricts the ability to alter the underlying relevant parameters independently and examine the output of the sensor. Also, it proved impossible to generate the Figures of Merit described in Section 2.3 because the variation affecting devices could not be adjusted independently of type. More granular results could be obtained and the whole spectrum between the worst possible cases could be explored, the Monte Carlo models with a slightly different global variation were the best option available.

Without having fabricated a chip, the results are all preliminary as they are based on simulations and as such are subject to parameters not considered in the method used to get the results presented.

It was found during the work that the simulator used to generate the results was unreliable under certain conditions. Two different simulator engines were used for the exact same setup and sometimes yielded significantly different results. Careful examination of the waveforms generated by the simulators determined which one was more reliable. In this work, both simulators have been used, and for the most part, when results have been examined, the deviations between the two were minor. Since every single data point was not compared between the two, there might be outliers that have not been caught.
5.1. Results

5.1.5 Method

The core of this work is the oscillator, and as such, it is the starting point. The oscillator consists of inverting delay cells and many different types of delay cells were considered and briefly examined. The delay cells presented in [16] were found to be the type that exhibited the largest sensitivity to the skewed corners. A slight modification to their suggested topology was made with the removal of the CMOS inverter in their delay cell and replaced with another ratioed inverter. A further modification was the introduction of the wave reshaper to prevent oscillator lock-up.

After the oscillators’ general architecture was developed, a TDC made in Verilog-A was introduced to examine the viability of frequency to code translation. The code was shown to correlate strongly with the oscillator frequency and the Verilog-A block was turned into the TDC shown in Section 3.1.3. A long chain of in total 14 DFFs was used to ensure that the TDC did not overflow during various tests. The total number of DFFs is suggested to be determined by the specification requirements on accuracy, as suppressing mismatch and measurement duration are in opposition to each other when it comes to the required number of DFFs.

5.1.6 Source Criticism

The major influences in the sensor design are attributed to [15] and [16]. The Figure of Merit, $k_p$ and $k_n$, discussed in Section 2.3, is used as the main metric to characterize the performance of their proposed sensors. However, for this to be accurate this assumes that $V_{th}$ is the dominating factor in determining the output frequency. The results observed in this work, when put into the context of varying temperatures, show that $V_{th}$ alone does not dictate performance. It is likely that the technology used in this work have an increased dependence on charge mobility in comparison to the aforementioned works. Also, the temperature behavior of their sensors is not examined in much detail.

A general consideration is that with newer technology, things that once were true in older technologies may not be true anymore due to more complex geometries and different manufacturing processes. As such, the older papers cited in this work are assumed to be true, but they may not be or are only partially true.
5.1. RESULTS

5.1.7 Ethical Aspects

Process sensor technology can lead to increased yield and therefore less waste. This will likely impact the environment in a positive way. Fewer materials will be consumed, and the environmental footprint will decrease.

Power efficiency is another potential aspect where we can see environmental improvement. If large data centers utilize chips with process sensors and dynamic voltage scaling, the power consumed per computation can be reduced. On a large enough scale, this will lessen the demand on the energy grid and depending on the source of the energy will affect the environment.
Chapter 6

Conclusions and Future Work

This chapter contains conclusions and potential future work.

6.1 Conclusions

In conclusion, this thesis presents a process variation sensor, the methodology to design a process variation sensor, and some potential techniques to mitigate the impact of environmental effects adversely affecting the sensor. The properties of the presented sensor worth highlighting are:

- The ability to distinguish skewed variation in n and p devices independently.
- Digital code as the output from the sensor for ease of further processing.
- Small area footprint & low complexity in terms of devices used which enables easy porting into different IPs or PDKs.
- Low sensitivity to temperature variation.
6.2 Future Work

Some of the potential avenues for future work are:

- The digital output code can be used in all sorts of ways to improve circuit performance and tune SoCs. There exist many potential avenues for the digital output code to be used in.

- Design adjustable supply voltage for the oscillators to alter temperature dependence which simultaneously allows the re-use of the circuit to measure temperature and to measure the process variation with low-temperature sensitivity.

- Extend the measurement structure to include independent temperature and supply voltage measurements which could allow for a combined performance measurement enabling further fine-tuning of an SoC.
Bibliography


