Energy-Efficient Mixed-Signal Circuits and Systems for Communication and Signal Processing

by

Jihwan Seol

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2022

Doctoral Committee:

Professor David Blaauw, Co-Chair
Professor Dennis Sylvester, Co-Chair
Associate Professor Ronald Dreslinski
Associate Professor Hun-Seok Kim

Jihwan Seol

jhseol@umich.edu

ORCID iD: 0000-0002-8081-0040

© Jihwan Seol 2022

Table of Contents

List of Figures **iv**

List of Tables **ix**

Abstract **x**

Chapter 1 Introduction **1**

1.1 Explosive Growth of Global Datasphere 1
1.2 Demand for Low Jitter in Communication Systems 3
1.2.1 Low Jitter Clock Generation with Phase Locked Loop 5
1.3 Need for Intelligent Data Processing 6
1.3.1 Keyword Spotting System for Edge Devices 6

Chapter 2 Wide bandwidth Ring Oscillator based Phase Locked Loop with Reference Oversampling Technique **10**

2.1 Introduction 10
2.2 Comparison between the conventional PLL and the Reference Oversampling PLL 12
2.3 Operation of the Reference Oversampling Phase Detector 14
2.4 Circuit Details of Proposed PLL 19
2.5 Measurement Results 25
2.6 Conclusion 26

Chapter 3 Sub-100fs Ultra Low-jitter LC-oscillator based PLL with Reference Oversampling Technique **30**

3.1 Introduction 30
3.2 Modeling and Simulation of the Reference Oversampling PLL 33
3.3 DCO Design for Reference OSPLL 40
3.4 PLL Design and Implementation 48

3.4.1 Overall Architecture 48
3.4.2 Reference Sampling Phase Detector 49
3.4.3 Comparator of the RSPD 52
3.4.4 Multiphase Generator 55
3.4.5 DCO 56
3.5 Measurement Results 58
3.6 Conclusion 65
Chapter 4 A 1.5 μW End-to-End Keyword Spotting SoC with Content-Adaptive Frame Sub-Sampling and Fast Settling Analog Frontend 66
4.1 Introduction 66
4.2 Skip RNN 69
4.3 Proposed KWS System 71
4.4 Implementation Detail 72
4.4.1 Analog Front End 75
4.4.2 Feature Extractor 79
4.4.3 Neural Network Accelerator 81
4.5 Measurement Results 85
4.6 Conclusion 88
Chapter 5 Conclusion 89
5.1 Summary of Contributions 89
5.2 Future Directions 90
BIBLIOGRAPHY 92

List of Figures

Figure 1.1 Global Datasphere forecast (2010-2030) [1]. 1

Figure 1.2 Created and stored data by segments [1]. 2

Figure 1.3 (a) Global data center traffic forecast [2]. (b) Global mobile data traffic forecast [3]. 2

Figure 1.4 ICT electricity demand forecast [4]. 3

Figure 1.5 Required jitter for 5G system according to target EVM [5]. 4

Figure 1.6 Required jitter for PAM4 receiver front-end ADC according to frequency and ADC resolution et EVM [6]. 4

Figure 1.7 Typical keyword spotting system pipeline. 6

Figure 2.1 (a) Operation of a conventional PLL. (b) Operation of the reference oversampling PLL. 13

Figure 2.2 Comparison between a conventional PLL and the OSPLL. 14

Figure 2.3 (a) Simplified block diagram of the OSPLL. (b) Conceptual diagram of the sinusoidal reference and the RSPD sampling points. 15

Figure 2.4 (a) RSPD with the ac coupling. (b) Operation of the RSPD with ac coupling scheme. (c) Pivot RSPD. (d) Operation of the Pivot RSPD. (e) Assignment of the pivot RSPD and ac coupled RSPDs according to the RSPD index. 16

Figure 2.5 Conceptual diagram that the static multiphase error, denoted as Δta, Δtb, Δtc, does not affect the comparator operation. 19

Figure 2.6 (a) Schematic of the RSPD. (b) Schematic of the offset cancellation comparator. 20

Figure 2.7 Operation of the RSPD and PGAIN. 21

Figure 2.8 Detailed schematic of the proposed OSPLL and the multiphase generator. 23

Figure 2.9 Schematic of DCO of the proposed OSPLL..... 24

Figure 2.10 Chip photograph and the measurement setup. 27

Figure 2.11 Measured phase noise plot of the proposed PLL..... 27

Figure 2.12 Measured phase noise plot with different loop bandwidth. 28

Figure 2.13 Measured reference spur of the proposed PLL..... 28

Figure 2.14 Spur vs. FOM2 of ring oscillator-based integer-N synthesizers 29

Figure 3.1 Noise source of the RSPD. 34

Figure 3.2 (a) Different sampling gains due to the different reference slope. (b) Simplified structure of the proportional path with the gain profile and the integral path. 35

Figure 3.3 Discrete time domain model of the proposed OSPLL..... 36

Figure 3.4 Simulated phase noise plot of the proposed OSPLL (N = 20) and the conventional BBPLL (N = 1) at 4-GHz FOUT and 200-MHz FREF. 36

Figure 3.5 (a) Simulated noise amplification due to sinusoidal reference when cosine PGAIN profile is used. (b) Tested PGAIN profile when N = 20. (c) Simulated noise amplification compared with flat profile when N = 20. 37

Figure 3.6 (a) Simulated PLL Jitter according to the loop latency. 39

Figure 3.7 Conceptual operation of the LC-DCO for the OSPLL. 40

Figure 3.8 (a) Conceptual operation of the unit varactor. (b) Simulated VBOT when VSTEP is swept from 0 to 0.6 V. (c) Simulated FSTEP when VSTEP is swept from 0 to 0.6 V. (d) Detailed structure of the unit cell of the proportional DAC. 41

Figure 3.9 (a) Structure of the proportional DAC. (b) Structure of the sub-DAC of the proportional DAC. (c) Conceptual operation of the proportional DAC. 43

Figure 3.10 Conceptual diagram that describes the phase disturbance due to the varactor switching. 44

Figure 3.11 Simulated phase disturbance due to the varactor switching (simulation condition: TT/25 ˚C/VSTEP = 0.3 V/PGAIN = 111). (a) DCO waveform. (b) Voltage disturbance due to the varactor switching according to the switching position. (c) Normalized ISF of the DCO. (d) Phase disturbance of the DCO according to the varactor switching position. 46

Figure 3.12 (a) Proposed DCO tuning pulse timing control scheme. (b) Timing diagram. 47

Figure 3.13 Overall structure of the proposed OSPLL. 48

Figure 3.14 Detailed schematic of the RSPD with its noise source. 49

Figure 3.15 (a) Simulated PLL jitter according to CS with CTOT = 1.4 pF. (b) Simulated PLL jitter according to CTOT. Optimal CS/CC ratio is used for each point, which was obtained using the method in (a). 51

Figure 3.16 (a) Simulated PLL phase noise plot with different HF corner of the ac coupling path. 52

Figure 3.17 (a) Detailed structure of the comparator in an RSPD. (b) Early reset scheme and its operation. 53

Figure 3.18 (a) Proposed multiphase generator. (b) Operation of the multiphase generator. (c) Timing diagram of each unit cell. 56

Figure 3.19 Overall structure of the DCO. 57

Figure 3.20 Die photograph. 58

Figure 3.21 Measured phase noise spectrum of the proposed OSPLL at 4 GHz. 59

Figure 3.22 Measured output spectrum of the proposed OSPLL. 60

Figure 3.23 Measured output jitter of the PLL with difference delay code of the DCO tuning pulse control scheme. 61

Figure 3.24 Measured power breakdown of the PLL. 62

Figure 3.25 Measured PLL jitter across the supply voltages. 62

Figure 3.26 Measured PLL jitter and reference spur of five chips. 64

Figure 3.27 (a) Jitter variance versus power comparison. (b) Reference spur versus PLL FoM comparison plot of prior arts. 64

Figure 4.1 (a) Conventional RNN. (b) Skip RNN. (c) Skip RNN when skip signal is 0. (d) Skip RNN when skip signal is 1. 70

Figure 4.2 Detailed implementation of the Skip RNN. (a) when skip is 0. (b) when skip is 1. . 70

Figure 4.3 Proposed KWS system and conceptual operation. 71

Figure 4.4 Detailed block diagram of proposed KWS SoC. 72

Figure 4.5 Flow Chart of Skip Control FSM. 74

Figure 4.6 AFE additional power according to $T_{SET}+T_{FE}+T_{NN.}$ 74

Figure 4.7 Proposed fast settling analog frontend. 76

Figure 4.8 Structure of DC-coupled Gm-ratio amplifier. 76

Figure 4.9 Simulated KWS accuracy according to high pass corner. 77

Figure 4.10 Proposed two step frequency control scheme. 77

Figure 4.11 Measured AFE performance. 78

Figure 4.12 Structure of the feature extraction block. 79

Figure 4.13 Detailed structure of the FFT. 80

Figure 4.14 Timing diagram of the FFT. 80

Figure 4.15 Power saving with sprint and power breakdown of the FFT. 81
Figure 4.16 Structure of the skip RNN accelerator. 82
Figure 4.17 Conceptual diagram of GRU cell. 83
Figure 4.18 Detailed structure of the proposed skip RNN accelerator. 83
Figure 4.19 Structure of proposed weight memory. 84
Figure 4.20 Scheduling table of RNN accelerator. 84
Figure 4.21 Chip photograph. 85
Figure 4.22 Measured accuracy of proposed KWS according to different skip ratio. 86
Figure 4.23 Power breakdown of proposed KWS with and without skip operation. 86
Figure 4.24 Measured waveform of proposed KWS. 87

List of Tables

Table 2.1 Comparison with recent ring oscillator-based integer-N synthesizers. 29

Table 3.1 Comparison with recent low-jitter PLLs. 63

Table 4.1 Comparison with existing KWS system for edge devices. 88

Abstract

In recent years, there has been an explosive increase in digital data generated globally, driven by various online applications such as video streaming, social media, shopping, gaming, etc. With the rise of deep learning and the Internet of Things (IoT), this trend is likely to be maintained for the foreseeable future. The major portion of the massive amount of data will be generated from endpoint devices such as PCs, phones, wearables, and connected cars. In contrast, the data will be stored in the cloud, creating capacity and energy issues in the data centers and networks. Therefore, to sustain this rapid growth of data, it is critical to develop faster and more energy-efficient communication and network system while reducing the data traffic with efficient processing.

In modern communication systems, clock generation, typically done by phase-locked loops, is an important task. For the next generation high speed communication systems that are being developed these days, extremely clean clock sources with less than 100fs jitter are required. However, the generation of a clock with such low jitter involves a lot of power. Therefore, developing an energy-efficient way to generate low jitter clocks is critical.

As deep learning is getting popular and the number of IoT devices is increasing, we have more data movement between edge devices and clouds. Recently, this increased data movement raised several issues, such as privacy, latency, and network bandwidth congestion. Edge computing is getting attention as a solution to these problems by moving data processing from

clouds to edge devices. However, for edge computing, we need energy-efficient hardware since the battery capacity of edge IoT devices is limited.

In this dissertation, energy-efficient mixed signal circuits and systems are proposed that addresses the aforementioned issues. In chapter 2, low jitter energy efficient ring-oscillator based PLL is proposed. With the proposed reference oversampling concept, the proposed PLL achieves wide bandwidth that can exceed the reference clock frequency so that phase noise from the ring oscillator is greatly suppressed. The proposed PLL achieves a 500fs rms jitter with 3.6 mW power consumption, which translates to 240dB FOM. The PLL also achieved a -80 dBc reference spur thanks to the reference oversampling.

Chapter 3 proposes a 67fs rms ultra-low jitter LC-oscillator based PLL using the reference oversampling techniques. The reference oversampling technique simultaneously achieves low in-band noise and wide bandwidth at the same time, achieving sub-100fs low jitter with excellent power consumption. LC-DCO structure for the reference oversampling and optimal DCO tuning word control scheme is also proposed. This work achieved 67fs rms ultra-low jitter with 5.2 mW power consumption, -256 dB FoM power efficiency.

Finally, in chapter 4, a low power keyword spotting system with a dynamic neural network is proposed. The proposed work adopts the skip RNN algorithm to adaptively power gates the entire system, including the analog front (AFE) end and the digital back end. The proposed AFE structure with a dc coupled first stage LNA and switched capacitive feedback resistors achieves less than 1ms settling time, enabling fast on/off. The proposed work demonstrates 1.5 μW power consumption.

Chapter 1 Introduction

1.1 Explosive Growth of Global Datasphere

In recent years, there has been an explosive increase in the amount of digital data generated globally, driven by a variety of online applications such as video streaming, social media, shopping, gaming, etc. With the rise of deep learning and the Internet of Things (IoT), this trend is likely to be maintained for the foreseeable future. Figure 1.1 shows the forecast of the amount of data created, captured, and replicated worldwide, which is also called the 'global datasphere' [1]. According to this forecast, the amount of yearly generated data worldwide will be doubled in only three years from 2022, reaching 175 zettabytes in 2025. It is also expected that the major portion of this huge amount of data will be generated from endpoint

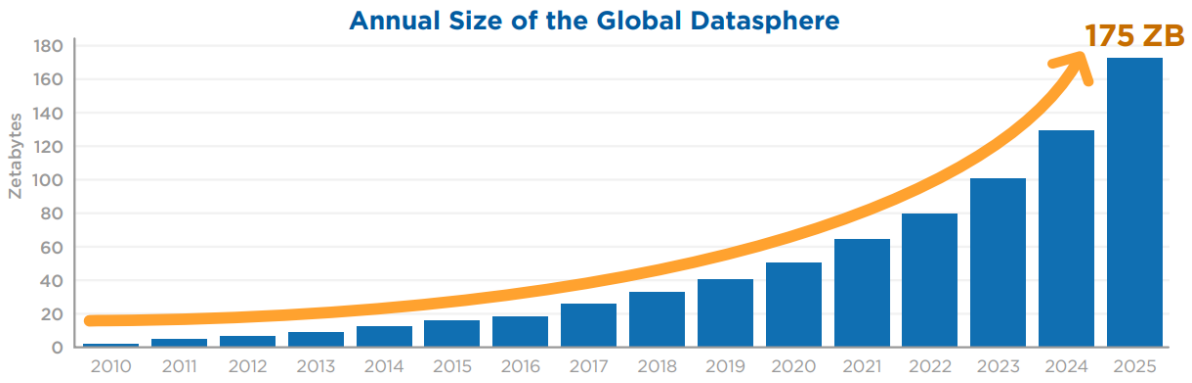


Figure 1.1 Global Datasphere forecast (2010-2030) [1].

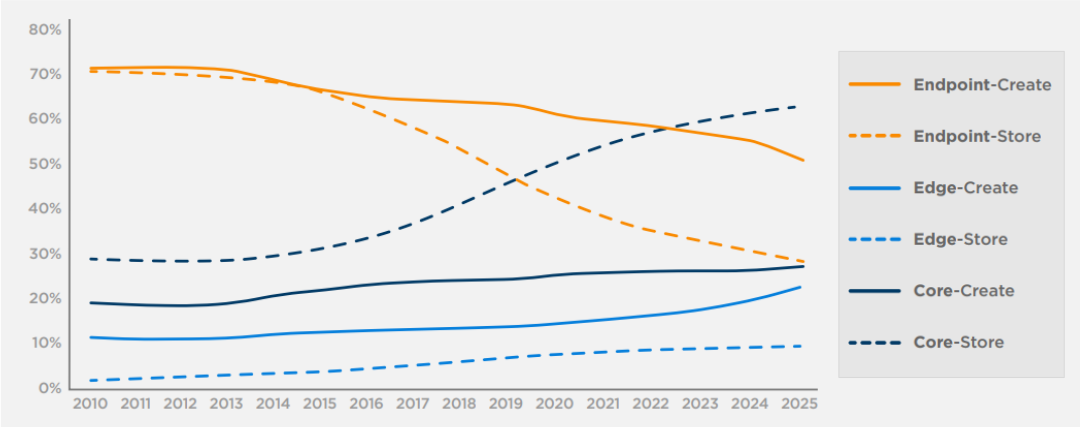


Figure 1.2 Created and stored data by segments [1].

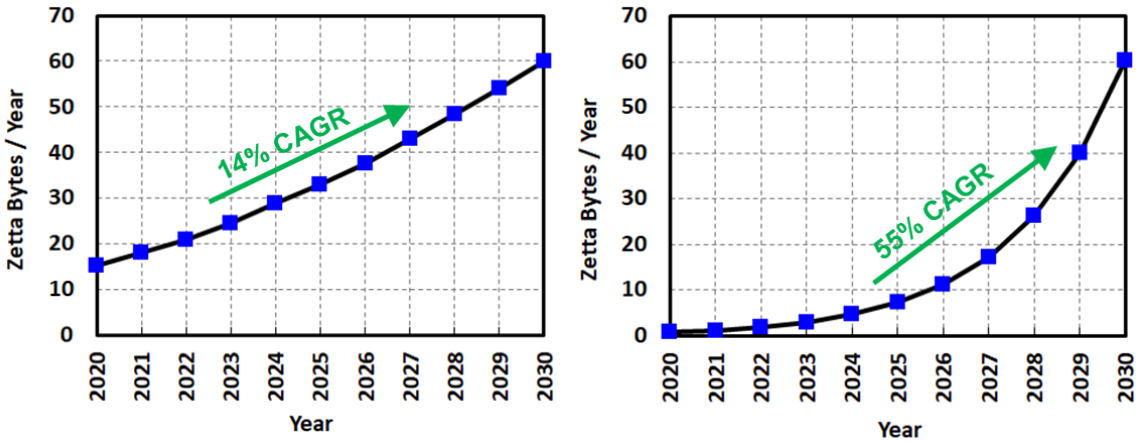


Figure 1.3 (a) Global data center traffic forecast [2]. (b) Global mobile data traffic forecast [3].

devices such as PCs, phones, wearables, and connected cars, while this data will be stored in the clouds, as shown in Figure 1.2. Figure 1.3 supports this outlook. Figure 1.3(a) shows the forecast of the data traffic in global data centers. With a compound annual growth rate of 14%, the data traffic from the data center is expected to be doubled in less than five years. The mobile data traffic, shown in Figure 1.3(b), exhibits an even higher increase rate of 55%, and as a consequence, a fivefold increase is expected in 8 years from 2022.

The aforementioned prediction implies that we need a large increase in the capacity of the data centers and the communication network in the near future, which poses significant challenges. Not to mention the network capacity and cost issues, energy consumption of data

centers and networks is emerging as an important global issue. Figure 1.4 shows the expected electricity demand for information and communication technology [4]. While the electricity demand will increase exponentially, data centers and networks are expected to consume 70% of the total electricity. Therefore, it is critical to develop energy efficient and fast communication technology.

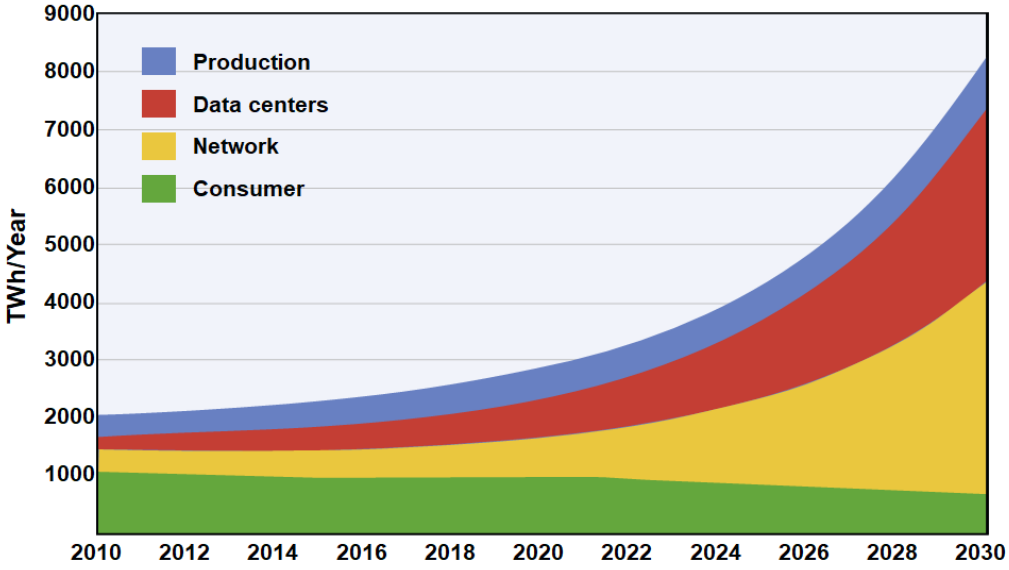


Figure 1.4 ICT electricity demand forecast [4].

1.2 Demand for Low Jitter in Communication Systems

To meet the ever-growing demands for higher data rates, there has been a lot of effort in developing new communication systems such as 5G wireless communication systems and PAM4 ultra-high speed wireline transceivers. In those systems, one necessary requirement is low jitter clock sources, and as the data rate increases, the requirement for the clock jitter is getting extremely stringent [5]-[6],[38]-[40]. Figure 1.5 shows the required integrated jitter of 5G wireless system according to different EVMs. With the 5G target EVM of -35 dB, the required integrated jitter is lower than 100fs [5]. Figure 1.6 shows the required clock jitter for the PAM4

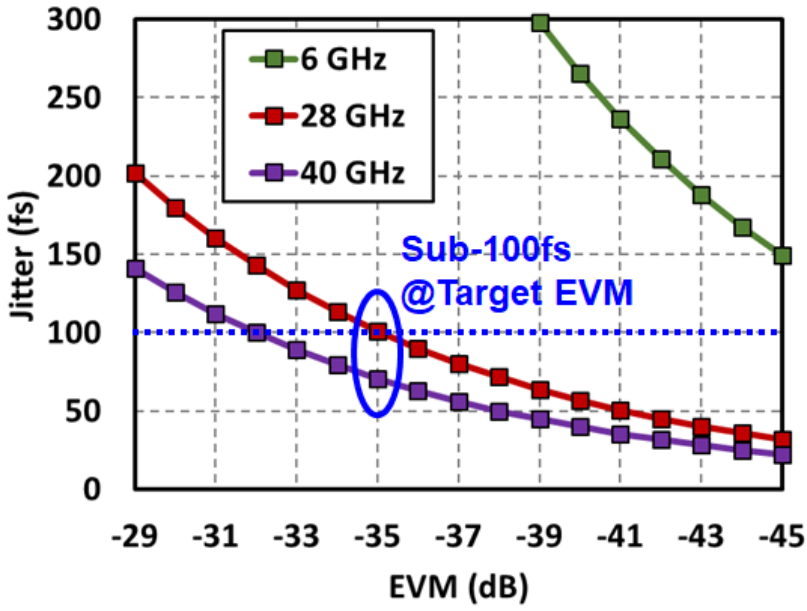


Figure 1.5 Required jitter for 5G system according to target EVM [5].

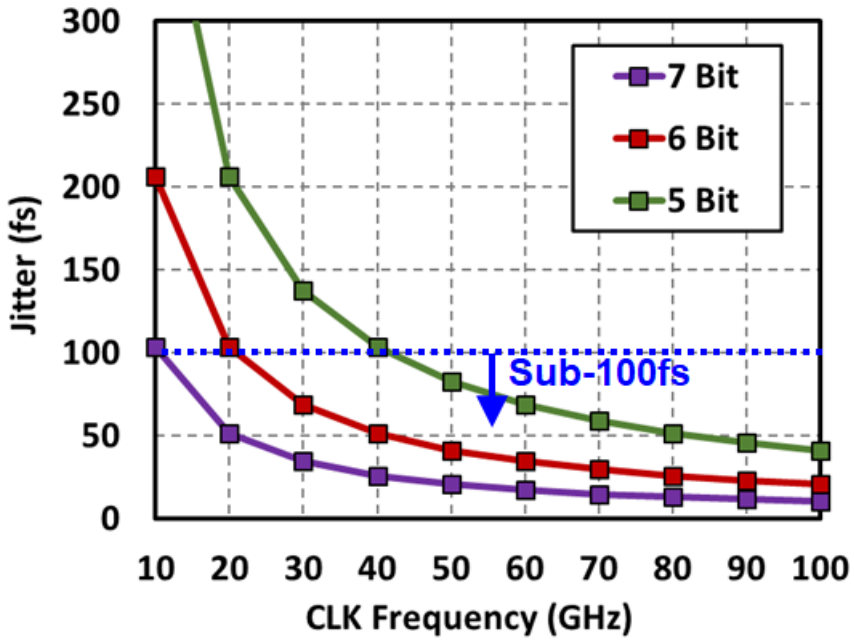


Figure 1.6 Required jitter for PAM4 receiver front-end ADC according to frequency and ADC resolution et EVM [6].

receiver frontend ADC at different frequencies and resolutions [6]. As the frequency increases, the required jitter can be lower than 50fs, which is of great challenge to implement such a low jitter clock source efficiently.

1.2.1 Low Jitter Clock Generation with Phase Locked Loop

For the generation of the clock sources, phase locked loops (PLLs) are primarily used. However, it is challenging to design a PLL with a clock with a jitter lower than 100fs. There have been considerable efforts to reduce the jitter of frequency synthesizers over the years. Charge pump PLLs (CPPLLs) are widely used due to their robustness. However, a large charge pump current is required to achieve low in-band phase noise [15]. Also, the charge pump nonidealities-induced reference spur limits the PLL bandwidth [8], degrading its power efficiency. Sub-sampling PLLs can attain low in-band phase noise by sampling the voltage-controlled oscillator (VCO) output with the reference clock [18]. A large phase detector (PD) gain can be obtained by utilizing the high VCO output slope, which in turn greatly suppresses the noise from the loop components. However, the sampling operation can disturb the VCO output, resulting in a high reference spur. Also, the narrow capture range of the sub-sampling PD makes the loop susceptible to disturbance. Therefore, an additional monitoring block is required. Injection-locked clock multipliers (ILCMs) offer excellent VCO noise suppression by directly injecting the reference clock to the VCO [10]. Thanks to the instantaneous phase correction, the noise suppression bandwidth can be greater than that of conventional PLLs while low in-band phase noise is achieved in a power-efficient manner. However, due to the direct injection of the reference clock into the VCO, ILCMs are susceptible to a large reference spur, necessitating complex calibration. Bang-bang PLLs (BBPLLs) are promising since they offer excellent power

efficiency by making use of a low power 1-bit time-to-digital converter (TDC) as their phase detection block [19]. However, large quantization noise from the bang-bang phase detector (BBPD) can limit their in-band phase noise level. Therefore, energy efficient method to realize both low in-band phase noise and oscillator noise is key to addressing this challenge.

1.3 Need for Intelligent Data Processing

Along with developing faster communication system, it is also necessary to reduce the data traffic itself with more efficient data processing. Edge computing, which is becoming an important research topic recently, can be a solution to this issue [7]. By moving the data processing from clouds to edge devices, a lot of data traffic can be saved. However, for edge computing, it is crucial to develop energy-efficient algorithm and hardware to execute tasks on the edges, which used to be run in clouds.

1.3.1 Keyword Spotting System for Edge Devices

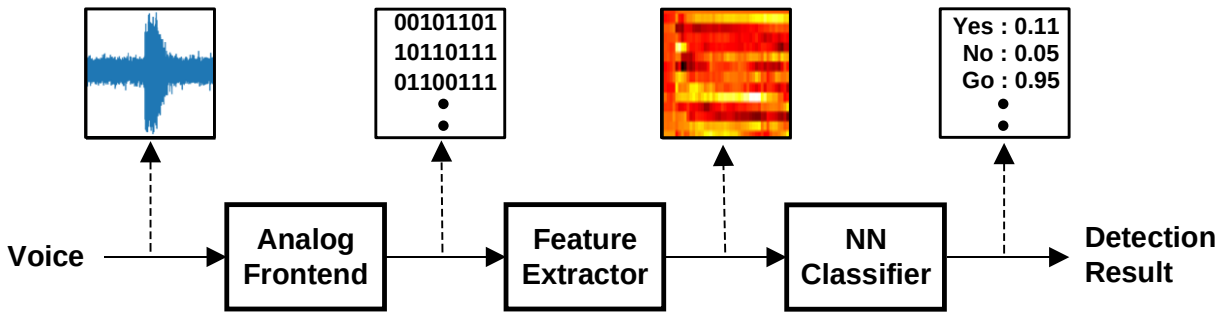


Figure 1.7 Typical keyword spotting system pipeline.

One such application is the keyword spotting (KWS) wake-up system, where KWS continuously monitors the audio input and looks for specific keywords, and activates the rest of the system, or enables more complicated speech recognition task when it detects the keywords.

Figure 1.7 shows a typical block diagram of a KWS system. First, the voice signal from the microphone is applied to the AFE. The microphone output signal can be pretty small, therefore, a low noise amplifier first amplifies it for the later stages. The AFE also performs filtering, where it filters out frequency contents except the desired speech signal range to prevent aliasing and reduce noise. Finally, the AFE converts the amplified and filtered signal into digital forms. Typically, 8b~16b is used for speech processing applications.

Next, the feature extractor preprocesses the incoming audio data for the next stages. MFCCs are commonly used feature extraction module, where it consists of a chain of windowing, FFT, Mel filtering, log and DCT blocks. MFCC basically converts time domain signals to a set of frequency domain signals. This reduces the high dimensionality and redundancy of the inputs and therefore reduce the power consumption of the next NN classifier and also improves accuracy.

The last stage is a neural network, which takes the output of FE as its input, and generates probability that each keyword is presented in the audio input. Traditionally, hidden Markov models (HMMs) are widely used for the classifier with Viterbi decoding block. With the rise of the deep learning, it is commonplace that some form of deep learning algorithm is used as a classifier of the KWS system. CNN and RNN are two popular deep learning algorithm that are used for KWS.

Traditionally, KWS task has been primarily executed in the cloud due to the large processing workload of the deep learning algorithms [29]. In this scenario, users are required to

upload their data to datacenters, which raises several issues, such as privacy, security, and network bandwidth. To address these issues, there has been a lot of effort recently in moving KWS task from clouds to edge devices. However, many of the previously proposed system lack analog frontend, which is a necessary component in the actual KWS system running on IoT devices [31]-[32]. Since the microphone signal can be very small, it is important to have low noise, so the power consumption of the AFE can be large. Therefore, a KWS system that include both the AFE and the digital backend with low power is required.

In this dissertation, energy-efficient mixed signal circuits and systems are proposed that address the aforementioned issues. In chapter 2, low jitter energy efficient ring-oscillator based PLL is proposed. With the proposed reference oversampling concept, the proposed PLL achieves wide bandwidth that can exceed the reference clock frequency so that phase noise from the ring oscillator is greatly suppressed. The proposed PLL achieves a 500fs rms jitter with 3.6 mW power consumption, which translates to 240dB FOM. The PLL also achieved a -80 dBc reference spur thanks to the reference oversampling.

Chapter 3 proposes a 67fs rms ultra-low jitter LC-oscillator based PLL using the reference oversampling techniques. The reference oversampling technique simultaneously achieves low in-band noise and wide bandwidth at the same time, achieving sub-100fs low jitter with excellent power consumption. LC-DCO structure for the reference oversampling and also optimal DCO tuning word control scheme is also proposed. This work achieved 67fs rms ultra-low jitter with 5.2 mW power consumption, -256 dB FoM power efficiency.

Finally, chapter 4 proposes a low power keyword spotting system with a dynamic neural network. The proposed work adopts the skip RNN algorithm to adaptively power gates the entire system, including the analog front (AFE) end and the digital back end. The proposed AFE

structure with the dc coupled first stage with the switched capacitive feedback resistor enables settling less than 5ms, achieving fast on/off. The proposed work demonstrates 1.5 μW power consumption.

Chapter 2 Wide bandwidth Ring Oscillator based Phase Locked Loop with Reference Oversampling Technique

2.1 Introduction

Ring oscillator based PLLs offer many advantages over the LC counterparts such as smaller area, wider frequency range, and less frequency pulling [8]. Therefore, they are suitable for the applications where multiple of PLLs are required in a single chip such as microprocessors, multi-lane wire line transceivers. However, the high phase noise of ring oscillators has prevented its wider use in systems where low phase noise is desired. Therefore, it has been of a great interest to lower the phase noise of ring oscillators.

One of the most effective ways to reduce the phase noise of ring oscillator is using wide PLL loop bandwidth. Since the oscillator phase noise is high pass filtered by the PLL loop, the oscillator phase noise located below the loop bandwidth is suppressed. However, there are several challenges in maximizing the PLL bandwidth. The first major challenge is due to the existence of the PLL stability limit, which is often called the Gardner's limit. Because of the discrete nature of the loop dynamics, the bandwidth of the loop is difficult to exceed one tenth of the reference frequency. Therefore, in many cases, loop bandwidth of a PLL is set to a sub-optimal lower value resulting in high oscillator phase noise.

Another major issue when trying to make the PLL loop bandwidth wide is the reference spurious tone that appears at the reference frequency which needs to be maintained at low level for various communication and data conversion applications. However, there is a fundamental trade-off between the reference spur and the loop bandwidth as the periodic disturbance is generated from the phase detection blocks. Increasing the bandwidth to suppress the oscillator noise exacerbates the reference spur. Therefore, a circuit technique to reduce the reference spur is required.

There have been a lot of efforts to reduce the jitter and phase noise of ring oscillator-based phase locked loops over the years. MDLLs [9] and ILCMs [10] inject the reference clock directly to the oscillator and refresh the phase noise preventing the accumulation of the oscillator noise. Thanks to the instantaneous phase correction, the loop bandwidth can reach up to the half of the reference frequency while providing the low in-band noise. However, such a direct injection of reference clock to the oscillator creates a large reference spur typically ranging at -30~50dBc. FPEC [11] was introduced to overcome such a challenge of ILCM. FPEC emulates the phase correction mechanism of ILCM while using the conventional PLL architecture, so it can simultaneously achieve a low noise and low spur. However, in order to accurately feedforward the phase error to the oscillator, an accurate estimation of the phase detector and oscillator gain is required which may demand complex calibration logic.

The aforementioned previous architectures have introduced attractive noise characteristic. But they all have a fundamental limitation of the loop operating at a reference frequency, which is the origin of the reference spur. Also, it limits the bandwidth to the Gardner's limit of F_{REF} over 10. Therefore, in the proposed design, our goal is to simultaneously achieve wide

bandwidth and low reference spur by utilizing the loop operating faster than the reference frequency.

This paper introduces a reference oversampling PLL (OSPLL) that can fundamentally overcome the bandwidth constraint of a PLL [12]. The loop bandwidth can be configured to be even higher than the reference frequency so that the ring oscillator noise is effectively suppressed without the penalty of a high reference spur. Furthermore, the in-band noise can also be reduced drastically as the effective input frequency is multiplied by a factor of N, where N is the PLL frequency multiplication ratio. The 28 nm prototype design consuming 3.6 mW achieved 508 fsrms jitter, 80 dBc reference spur, and 240.3 dB jitter-power figure of merit (FOM) at 2 GHz output frequency.

2.2 Comparison between the conventional PLL and the Reference Oversampling PLL

To better explain the principle of the proposed OSPLL, we compare the OSPLL with a conventional PLL whose loop operates at every reference cycle TREF, as shown in Figure 2.1(a). The reference clock of the PLL is generated from a crystal oscillator (XO). Since the XO output (REF) is a sinusoidal waveform, it is first converted to a square wave by a reference buffer (BUF). The output of the reference buffer CKREF is then applied to a PD. The PD compares the positive edges of CKREF to those from the feedback clock, FB, which is the divided oscillator output clock. The PD operates at every TREF cycle, which is the period of CKREF and FB. The PD output is then applied to the loop filter (LF), and the LF output LFOUT drives the oscillator, suppressing the oscillator noise. In this case, the reference and PD noise are amplified by the frequency division ratio (N). Also, the reference buffer can induce large noise

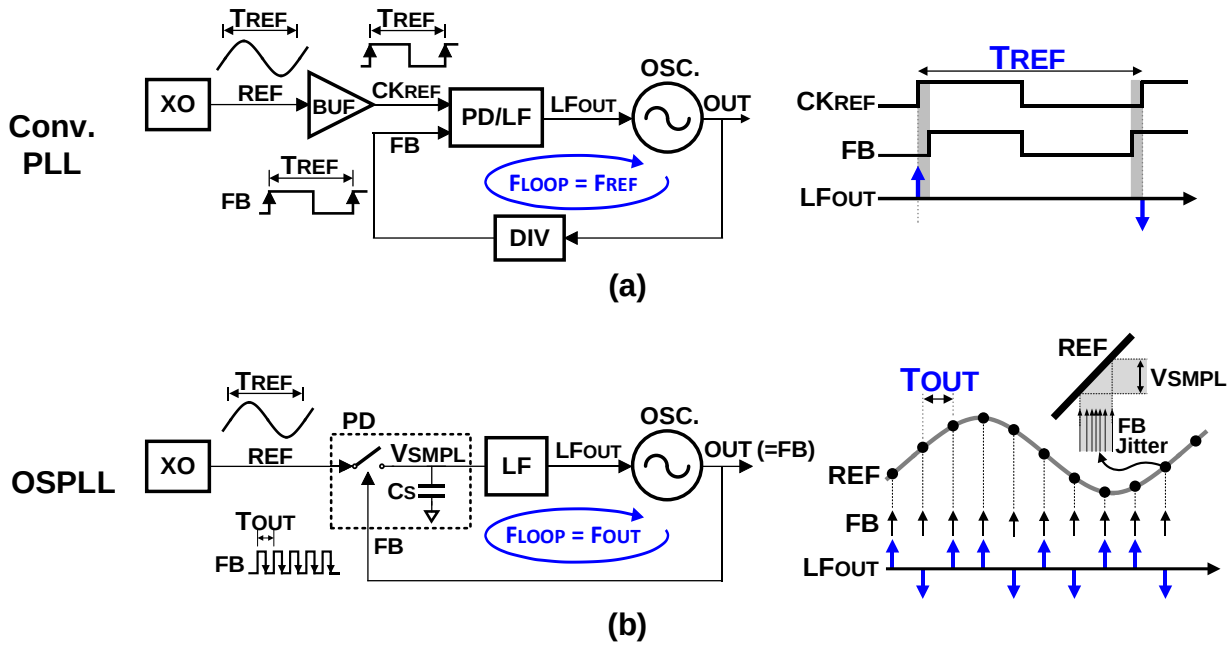


Figure 2.1 (a) Operation of a conventional PLL. (b) Operation of the reference oversampling PLL.

and/or power when it converts the sinusoidal reference, which has a slow slope, to a square wave [13].

Figure 2.1(b) shows the conceptual diagram of the OSPLL. Without a reference buffer, the sinusoidal reference REF is directly applied to the sampling switch input. The switch samples its input at the falling edge of the oscillator output FB, which is the same as OUT, forming a PD. The PD of the OSPLL is based on the sampling circuits whose output voltage is proportional to their input phase error with the gain defined by the slope of the reference clock dVREF/dt. As the PD samples the phase error at every TOUT, the loop now operates at a FOUT rate, noise from the loop components is effectively suppressed by oversampling. This is illustrated in Figure 2.1(b), where the black dots represent the sampling points, and the phase error of FB is converted to the sampled voltage VSMPL. Note that different dc value at the sampling points is addressed using N interleaved PDs implemented with the ac coupling scheme,

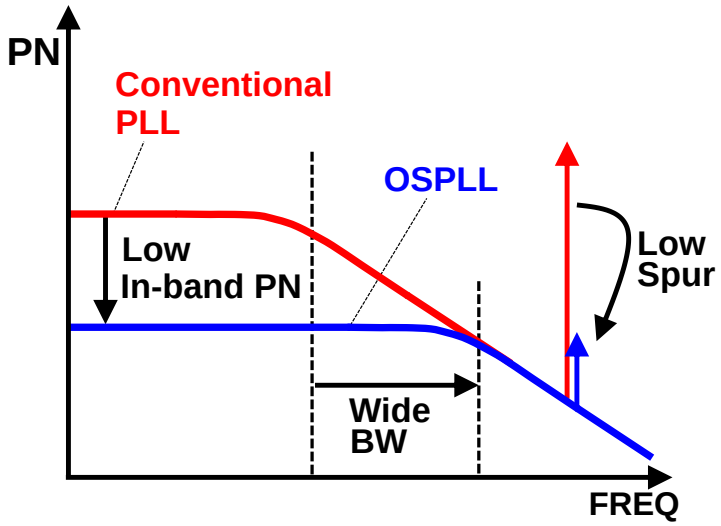


Figure 2.2 Comparison between a conventional PLL and the OSPLL.

which will be explained later in this section. The extracted phase error at every FB edge then drives LF and its output LFOUT suppresses oscillator noise. In other words, the OSPLL loop operates in FOUT not FREF, simultaneously enabling low in-band phase noise, wide bandwidth, and low spur as shown in Figure 2.2. In addition, since the OSPLL directly samples the sinusoidal reference from the XO, no reference buffer is needed, improving power efficiency and output phase noise.

2.3 Operation of the Reference Oversampling Phase Detector

Figure 2.3(a) shows the detailed structure of the proposed OSPLL. N reference sampling PDs (RSPDs) directly sample the input reference sinewave REF from the XO. 1:N denotes the time-interleaved sampling clocks generated from the DCO output (OUT) so that the effective sampling period of the RSPDs is the DCO period TDCO while each RSPDs operates in TREF, relaxing the speed requirements. The RSPD sampling clock 1:N is generated from the multiphase

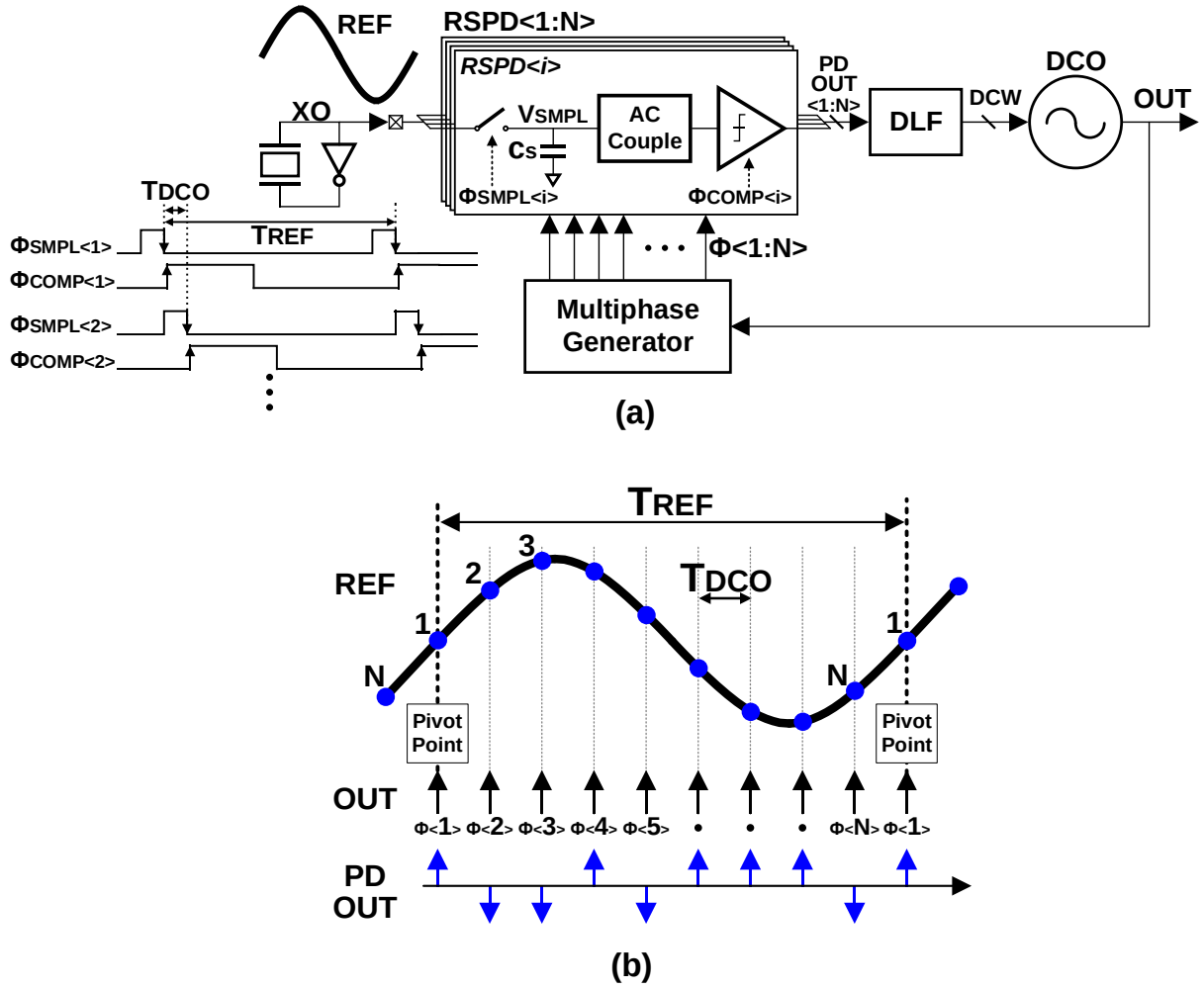


Figure 2.3 (a) Simplified block diagram of the OSPLL. (b) Conceptual diagram of the sinusoidal reference and the RSPD sampling points.

generator, driven by OUT. Each RSPD operating at i receives SMPLi and COMPi to control the sampling switch and the comparator, respectively, in the corresponding RSPDi. At the falling edge of SMPLi, REF is sampled on a sampling capacitor CS. Assuming there is no noise in OUT, each RSPD samples an identical point of the sinewave reference REF, marked with a blue dot with corresponding RSPD index as shown in Figure 2.3(b). Jitter in OUT will then cause a change of the sampled voltage, which represents the phase error. Therefore, the nominal dc value

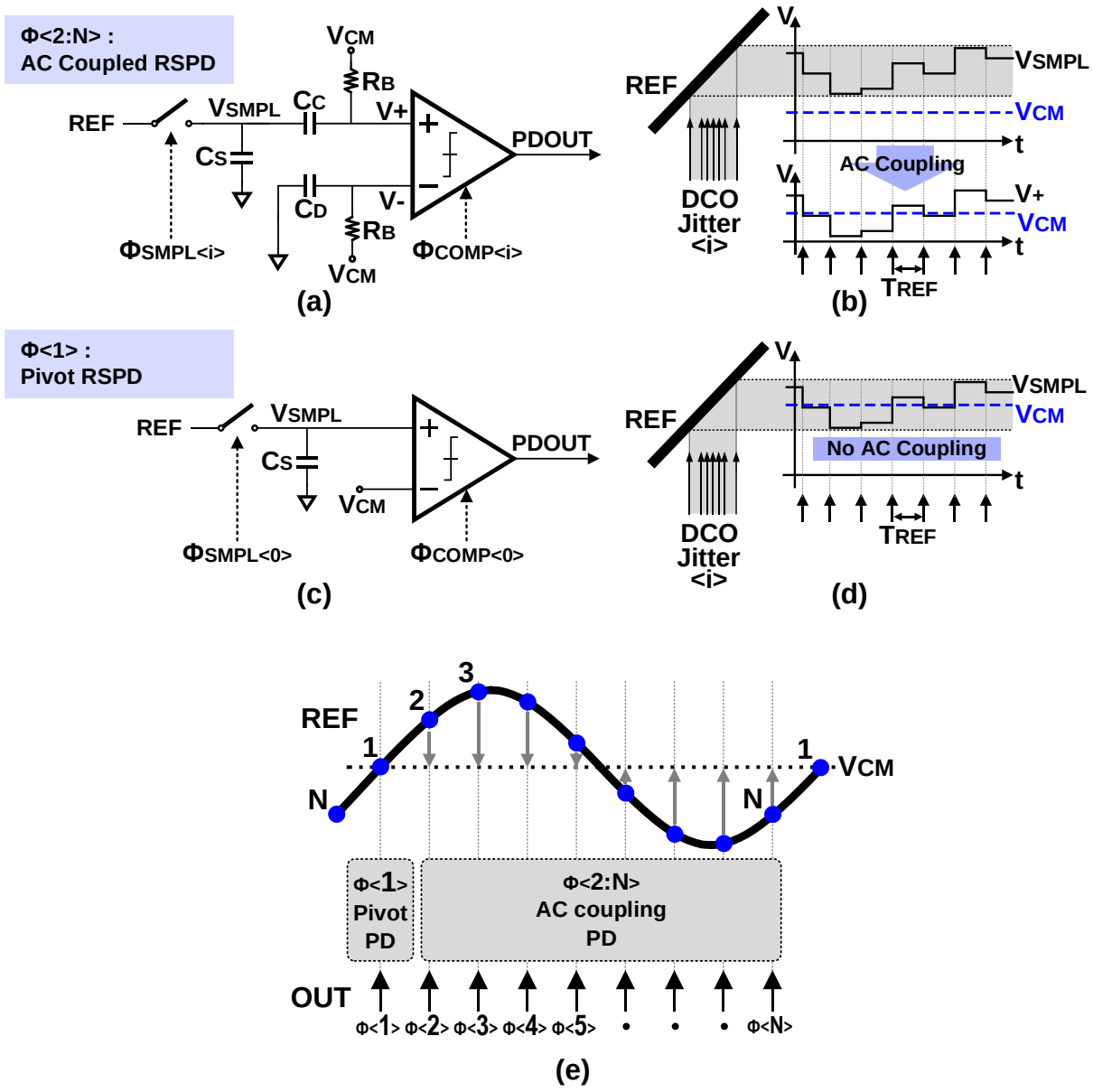


Figure 2.4 (a) RSPD with the ac coupling. (b) Operation of the RSPD with ac coupling scheme. (c) Pivot RSPD. (d) Operation of the Pivot RSPD. (e) Assignment of the pivot RSPD and ac coupled RSPDs according to the RSPD index.

of VSMPL is constant, and the voltage fluctuation due to the DCO jitter is superpositioned on the dc voltage. The sampled voltage VSMPL is then quantized by a comparator at the rising edge of COMPi to produce 1-bit PD output PDOUTi, which indicates the phase of OUT either leads

or lags the phase of REF. An ac coupling capacitor is placed between the sampling capacitor and the comparator to remove the dc voltage of VSMPL. $PDOUT_{1:N}$ is then converted to DCO control word (DCW) by the digital loop filter (DLF) and fed to the DCO, forming a bang-bang PLL (BBPLL) operating at every DCO cycle.

Since each RSPD samples different points of the sinusoidal reference REF, the sampled voltages of the RSPDs have different nominal dc values, which raise several issues. In order for the comparator to accurately quantize the phase error, the sampled voltage must be compared with a reference voltage representing the nominal voltage of the sampling point. As each RSPD samples a different point of the sinusoidal reference clock, N voltage references that accurately represent the sampling points are needed, which is impractical. The different common mode levels of each comparator also make it difficult to optimize the comparator. The ac-coupling scheme shown in Figure 2.4(a) solves this issue by removing the dc of the sampled voltage, so that only a single reference voltage can be used for every RSPD. The dc level of VSMPL is blocked by CC and only the ac fluctuation caused by the output jitter is passed to the comparator input, which is biased at VCM using a large resistor. Note that the other comparator input, V−, is also biased at VCM through the same type of resistor. Therefore, the comparator can operate at its optimal common mode voltage, VCM. This ac coupling operation is illustrated in Figure 2.4(b). The DCO jitter is converted to VSMPL and then its dc level is shifted to around VCM through ac coupling.

Since only ac fluctuations of the phase error can be detected through the coupling capacitor, an RSPD with ac coupling cannot provide absolute dc phase locking. Therefore, one of N RSPDs, RSPD1, is designed as a pivot RSPD, without the ac coupling capacitor, to achieve absolute dc phase locking as shown in Figure 2.4(c). The comparator reference voltage of the

pivot RSPD is set to VCM, which is also the positive zero crossing point of REF, so that it can provide the pivot position of the phase locking at the REF zero crossing while the sampling points of the other RSPDs are defined relative to the pivot position as shown in Figure 2.4(e). Figure 2.4(d) shows the operation of the pivot RSPD where the sampled voltage VSMPL is directly compared with VCM to provide dc phase locking point.

As the ac coupling scheme effectively blocks the dc variation of the sampled voltage, the proposed RSPD is insensitive to the multiphase error between sampling clock 1:N, or the distortion of the input sinewave reference. The sampling clocks, 1:N, are generated from the multiphase generator and distributed to each RSPD. Mismatch in the multiphase generator and the wire delay can create multiphase error between 1:N. This static phase error can move the sampling point slightly, resulting in the dc voltage shift at VSMPL as shown in Figure 2.5. However, the dc voltage shift in VSMPL is blocked by the ac coupling capacitor and it does not affect the 1/0 probability at the comparator output. Likewise, the distortion of the sine reference is blocked by the ac coupling. Therefore, the multiphase error and the reference distortion do not create a reference spur at the PLL output, which is verified by simulation.

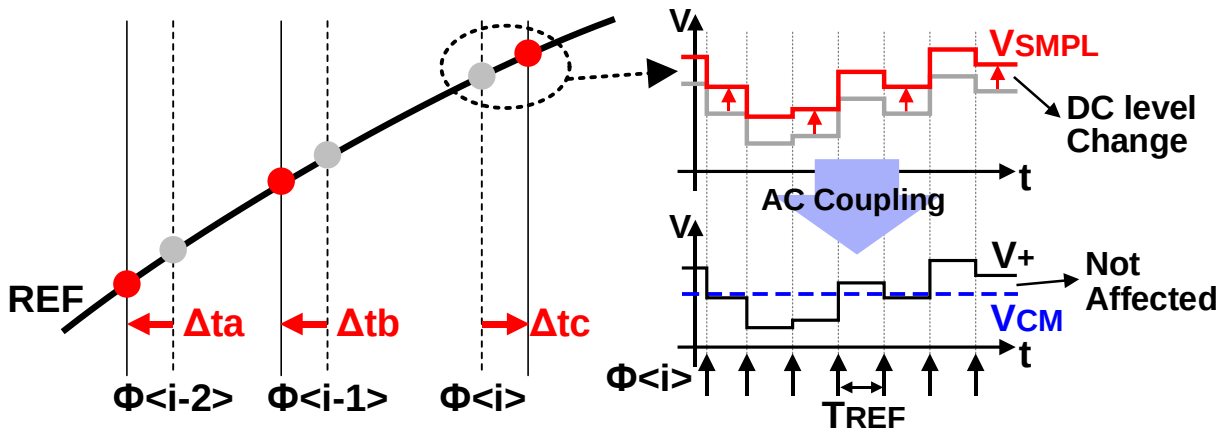


Figure 2.5 Conceptual diagram that the static multiphase error, denoted as Δta, Δtb, Δtc, does not affect the comparator operation.

2.4 Circuit Details of Proposed PLL

Figure 2.6 explains the detailed operation of the RSPD. CLKREF is sampled onto CS at the falling edge of i and transferred to a comparator input, VSMPL, through an AC coupling capacitor, CC, while its DC is defined at VCM through a large resistor, RB. A voltage difference between VSMPL and VCM indicates the output phase error while their DC value is equal to VCM. The polarity of VSMPL-VREF is sensed by a comparator clocked by COMPi whose input offset is cancelled by a charge pump [43] that asserts the voltage of VOC such that PDOUT has equal probability of 1s and 0s. Note that RSPDs with an AC coupling capacitor only sense the AC fluctuation of the phase error. Therefore, one of the 40 RSPDs is designed without a CC so it can provide the pivot position of the phase locking at the positive zero crossing point of

CLKREF, while the sampling points of the other RSPDs are defined relative to the pivot position.

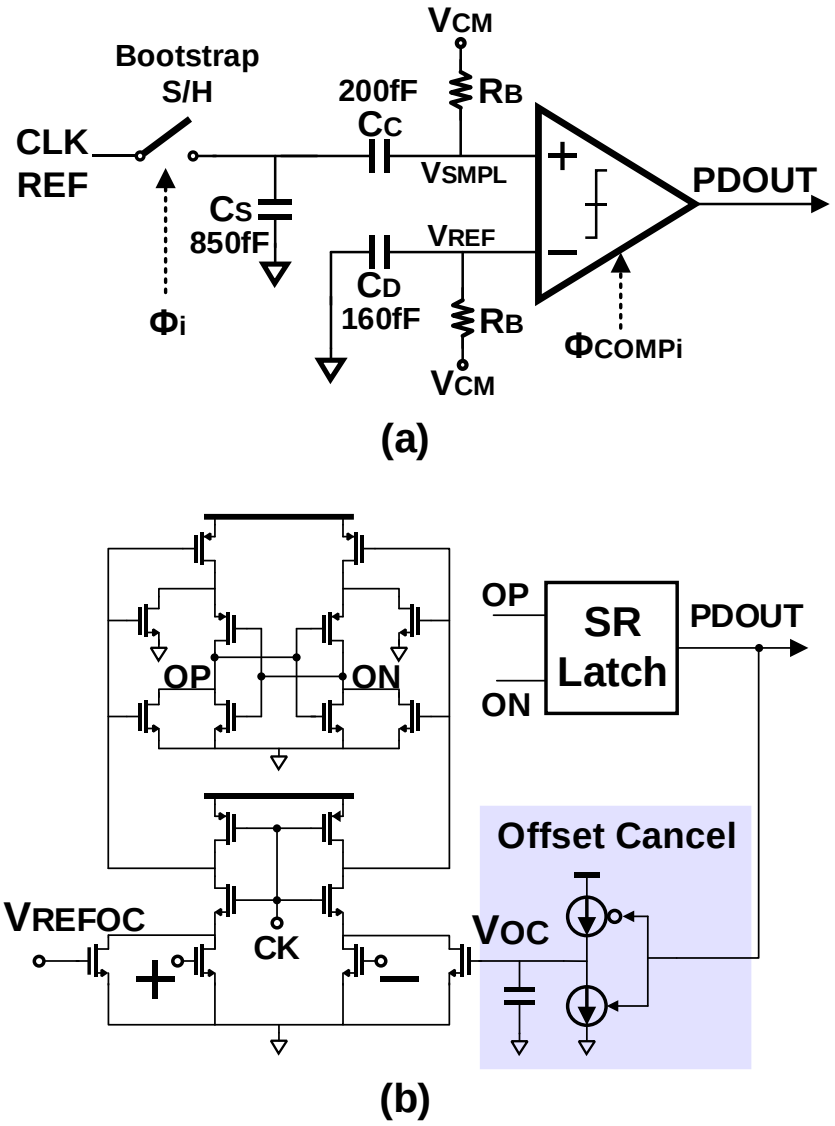


Figure 2.6 (a) Schematic of the RSPD. (b) Schematic of the offset cancellation comparator.

The intrinsic noise of an RSPD consists of sampling noise, kT/CS, and comparator noise, and is the same for every RSPD. On the other hand, the signal of each RSPD, which is the phase error, is dependent on the slope of CLKREF at the point that it samples. For instance, RSPD in

Figure 2.7 samples CLKREF close to its peak point so its output is mostly driven by RSPD noise rather than the phase error. Therefore, the RSPD outputs that sample at steeper points of

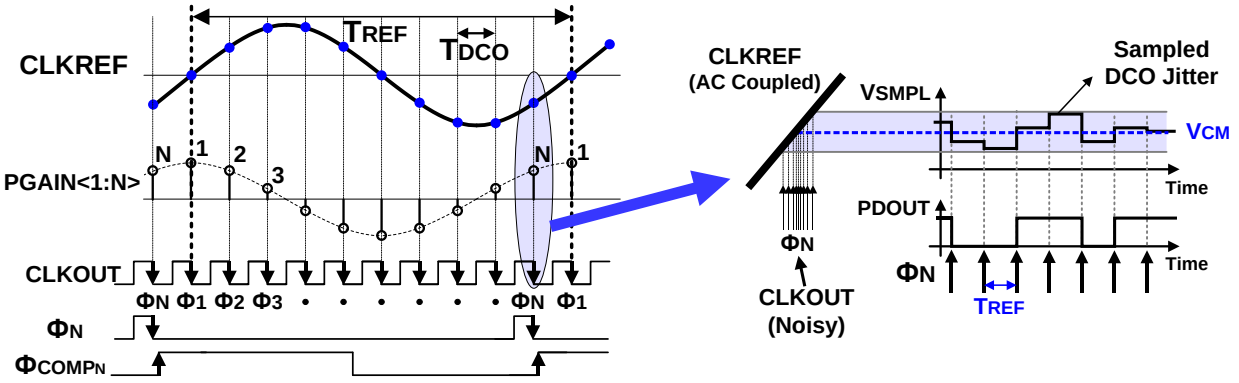


Figure 2.7 Operation of the RSPD and PGAIN.

CLKREF must be emphasized compared to RSPD outputs that sample CLKREF near its peak points. Also, the RSPD outputs that sample CLKREF during its negative slope need to be multiplied by 1. The proportional gain, PGAIN, is set to be quadrature to the input phase to meet the aforementioned two criteria, as shown in Figure 2.7.

Figure 2.8 provides a detailed block diagram of the proposed architecture. In the startup phase, a conventional bang-bang PLL is enabled to lock the output to the target frequency. Then, the startup PLL is disabled, and the outputs of the proposed RSPD, PDOUT, are converted to UP or DN pulses. Each pulse drives DCO for 1 DCO cycle through a PDAC, suppressing the oscillator noise at every DCO cycle. This oscillator noise suppression is done without directly injecting the reference to the oscillator as with MDLLs or ILCMs, minimizing the reference spur. Only PDOUT is connected to the integral path to lower the complexity. The multiphase generator produces 40 time-interleaved clocks using CLKOUT, as shown in the bottom Figure 2.8. It forms a closed chain of 40 unit cells, each of which is composed of a flip-flop (FF) and clock-gating logic. A bypass mux is also included in each unit cell to provide programmability of

the frequency division ratio. Initially one half of the units, U, are set to 1 while the other half, U, are set to 0. Then, the pre-assigned signals propagate in one direction of the chain at every rising edge of CLKOUT. The FF clock is gated when the input and output of the FF are both 0 to save 30% of the multiphase generator power.

Figure 2.9 illustrates the structure of the DCO, which is composed of pseudo-differential delay cells and current DACs including a 3-bit current steering DAC for each RSPD to program a proper PGAIN. UP and DN connections are swapped for the negative PGAIN's. Note that a dummy current path with a unity gain amplifier is used to ensure accurate current steering while UP and DN pulses are toggling.

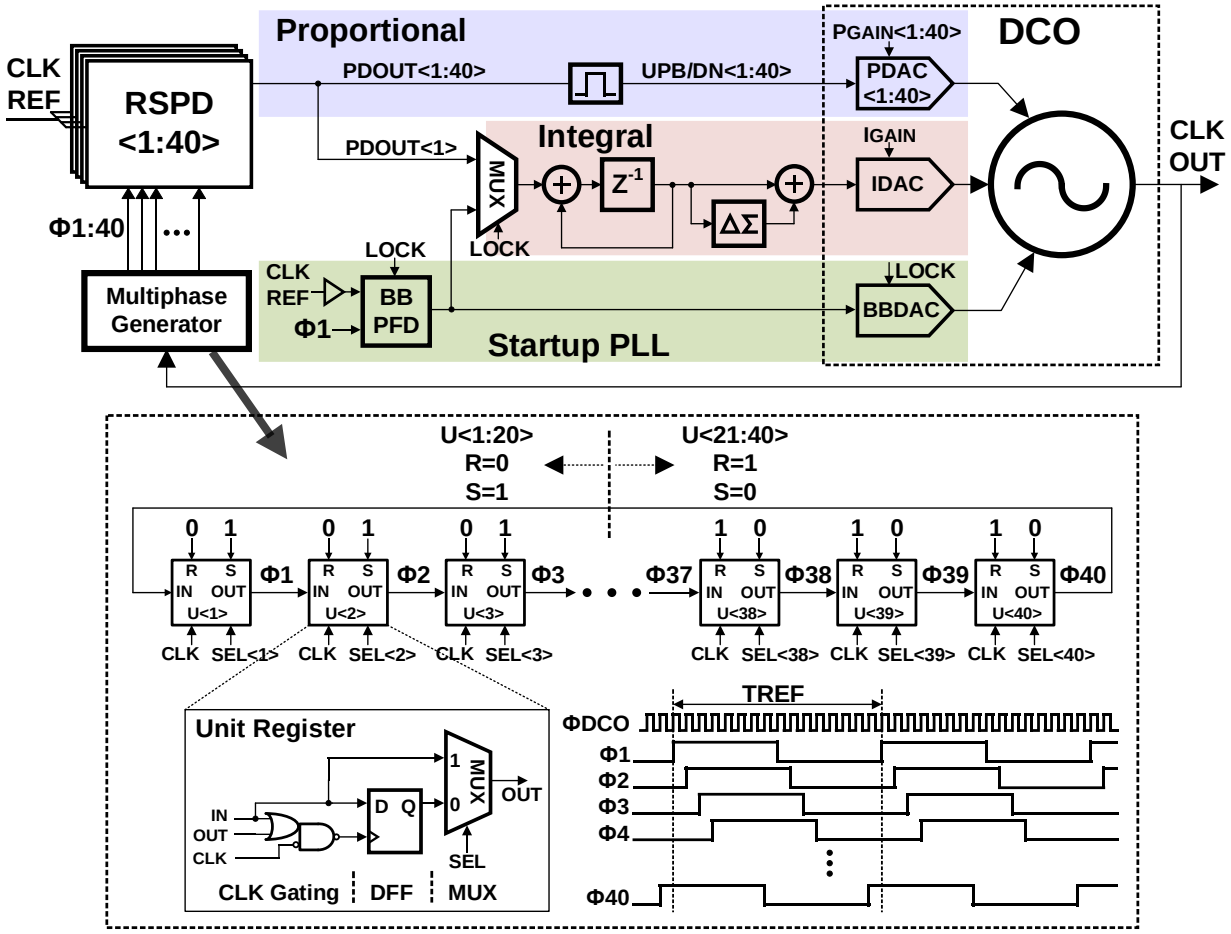


Figure 2.8 Detailed schematic of the proposed OSPLL and the multiphase generator.

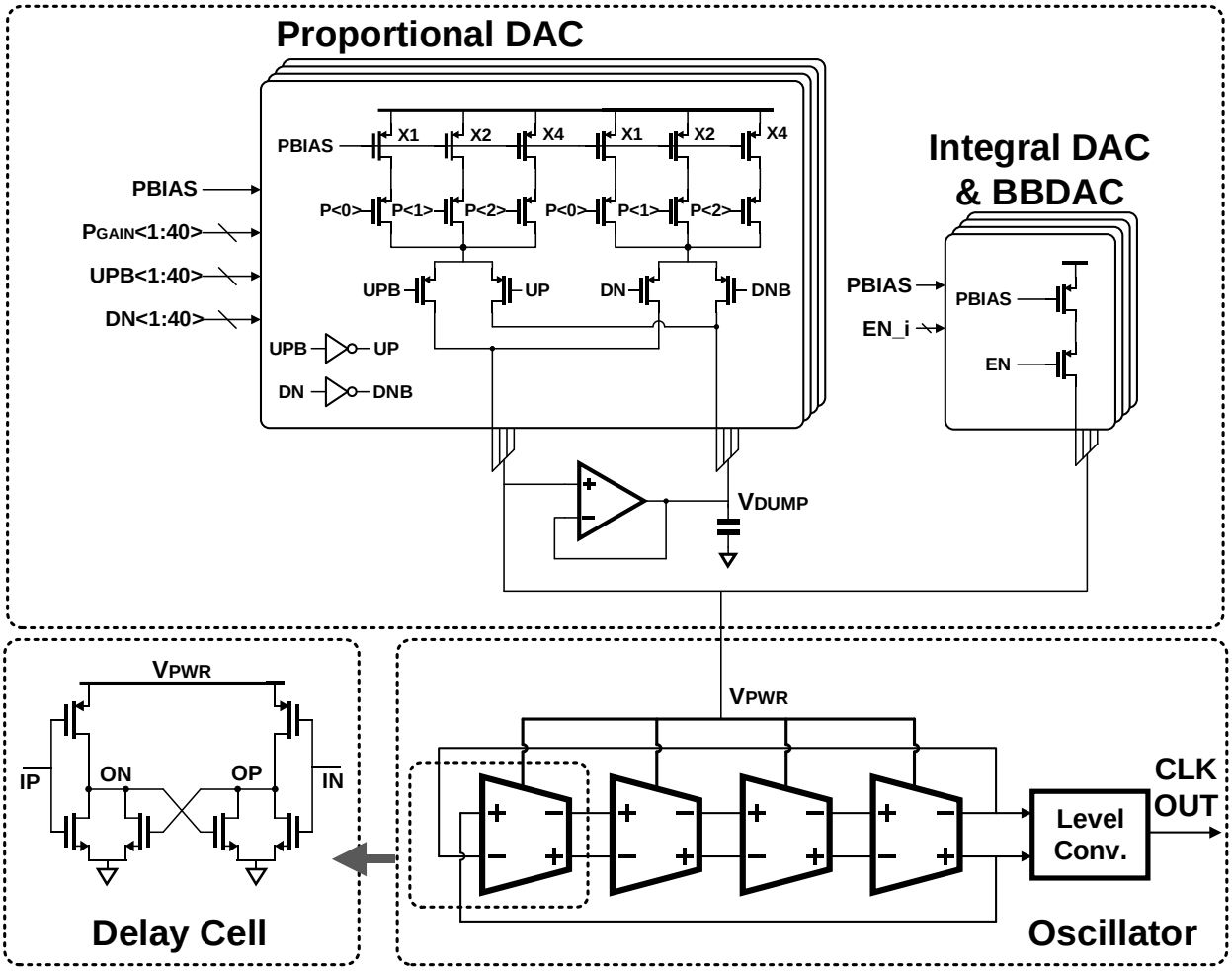


Figure 2.9 Schematic of DCO of the proposed OSPLL.

2.5 Measurement Results

The proposed design was fabricated in a 28nm CMOS technology and occupied 0.07mm2 active area as shown in Figure 2.10. The right figure shows the test setup. Figure 2.11 is the phase noise measurement results of the proposed design at 2GHz output frequency. Integrated jitter was 508fs with 10KHz to 100MHz integration range in the oversampling mode. Compared the phase noise of conventional BBPLL mode the proposed oversampling PLL achieved a significant noise improvement. The in-band noise is reduced from -100dBc/Hz to -123dBc/Hz and the integrated jitter is improved from 3.4ps to 0.5ps.

To test the capability of achieving the PLL bandwidth higher than FREF, we intentionally increased the proportional path gain as shown in the graph on the left. The blue line is when the bandwidth is set for the optimum jitter and the red line is when the bandwidth is set higher than the reference frequency. For the red line, total integrator jitter is large because the bandwidth is not optimal for the integrated jitter. However, this mode can be exploited when the oscillator noise is higher than the proposed design. Also, if this is used as the first stage of a cascade PLL, lower in-band noise can be advantageous while the following PLL filters out the high frequency noise with a narrower bandwidth. Furthermore, such a wide bandwidth characteristic can be highly useful for a fast-tracking application.

A spur measurement result is shown in Figure 1.3. At 2GHz output frequency and 50MHz reference frequency, the measured spur is lower than -80dBc. Even though the figure shows up to twice of the reference frequency, we confirmed all the spurs at integer multiples of the reference frequencies are lower than -80dBc.

Table 2.1 shows the comparison table. 50MHz reference frequency is used and multiplication ratio is 40. Measure phase noise at 1MHz is -120.8dBc/Hz. Integrated jitter is 508fs with 3.6mW power consumption. The jitter-power FoM depicted as FoM1 in the table is a well-known metric for comparing the PLL performance by normalizing jitter using power consumption. For a more precise performance comparison, FoM2 is also calculated by normalizing FoM1 using the reference frequency of 50MHz. The proposed PLL shows the state-of-the-art FoM of -240dB compared with recently published ring oscillator-based frequency synthesizers. Finally, the reference spur is -80dBc which is lower than previous design.

The reference spur versus reference normalized FOM of the recently published ring oscillator-based clock generators is compared Figure 2.14. The proposed work achieved state of the art performance in terms of spur and FOM.

2.6 Conclusion

In conclusion, the proposed reference oversampling PLL operates at F_{OUT}, the proposed PLL has advantages of wide bandwidth, low jitter, low spur. Also, it doesn't require reference buffer. This PLL can be useful in wireless and wireline communication, and also fast reference tracking application thanks to its wide bandwidth. The prototype design with ring oscillator achieved -240.1dB FOM and -80dBc reference spur at 2GHz.

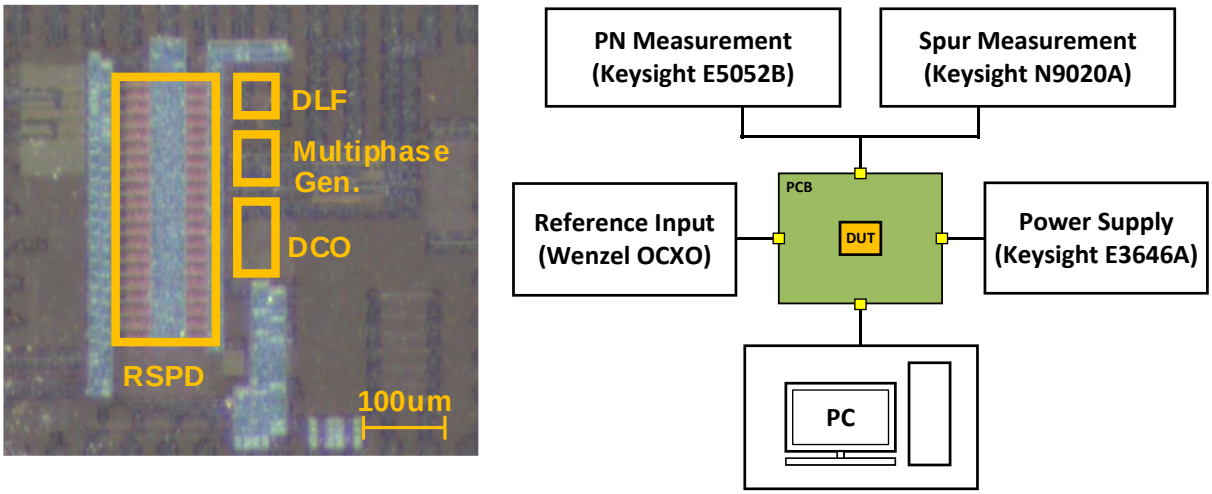


Figure 2.10 Chip photograph and the measurement setup.

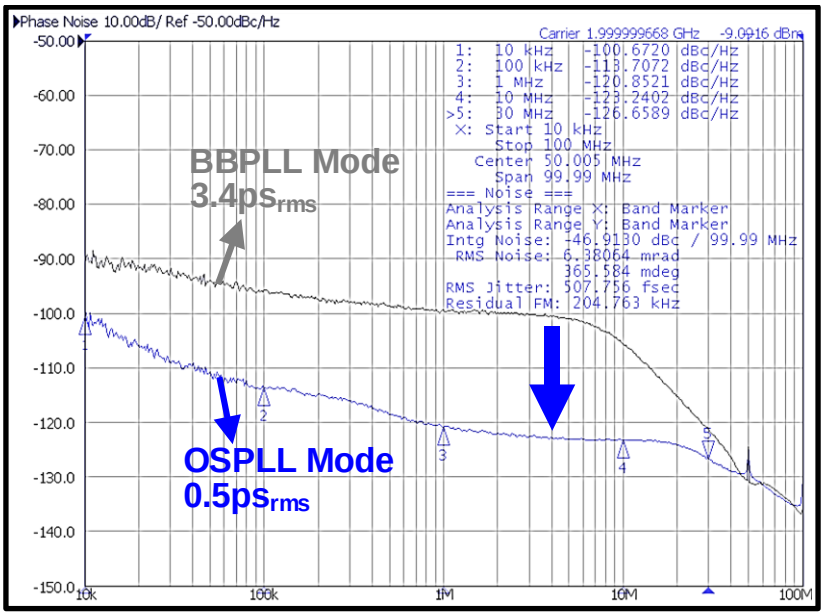


Figure 2.11 Measured phase noise plot of the proposed PLL.

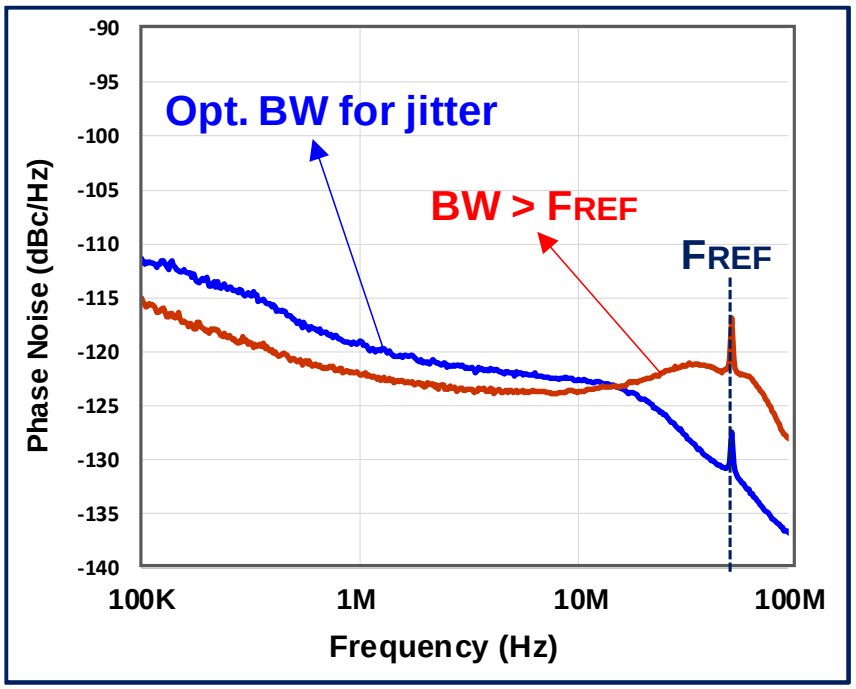


Figure 2.12 Measured phase noise plot with different loop bandwidth.

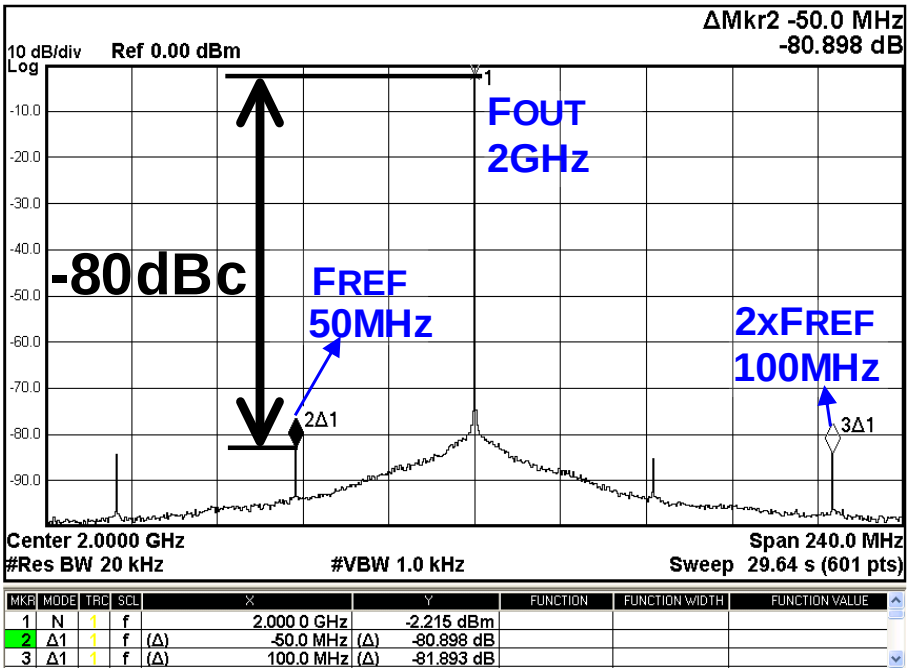


Figure 2.13 Measured reference spur of the proposed PLL.

	This Work	Tsai, VLSI18	Seong, ISSCC18	Kuan, VLSI18	Liu, CICC18	Choi, ISSCC16	Megawer, ISSCC18	Yang, ISSCC18
Type	OSPLL	PLL	PLL	ILPLL	ILPLL	ILCM	ILCM	MDLL
Output Frequency (GHz)	2.0	3.0	2.4	4.0	1.0	1.2	4.752	3.0
Reference Frequency (MHz)	50	200	75	250	100	120	54	200
Multiplication Ratio	40	15	32	16	10	10	88	15
Phase Noise @1MHz (dBc/Hz)	-120.8	-119.0	-119.8	-119.0	-130.0	-134.4	-113.71	-118.0
Integrated RMS Jitter (ps)	0.508 (10K-100M)	0.619 (100K-1G)	0.32 (1K-100M)	0.427 (100K-1G)	0.3 (10K-10M)	0.185 (10K-40M)	0.366 (10K-30M)	0.292 (10K-40M)
Power (mW)	3.6	2.3	6.0	6.3	2.2	9.5	6.5	1.45
*FOM[1] (dB)	-240.3	-240.5	-242.1	-239.4	-246.7	-244.9	-240.5	-249.1
**FOM[2] (dB)	-240.3	-234.5	-240.3	-232.4	-243.7	-241.1	-240.2	-243.1
Reference Spur (dBc)	-80	-52 ~ -41	-75	-61	-45	-53	-53	-44
Area (mm^2)	0.07	0.012	0.055	0.018	0.012	0.06	0.16	0.0056
Technology	28nm	7nm	65nm	7nm	65nm	65nm	65nm	28nm

$$*FOM^1 = 10log\left[\left(\frac{\sigma_{rms}}{1sec}\right)^2 \cdot \left(\frac{Power}{1mW}\right)\right] \quad **FOM^2 = 10log\left[\left(\frac{\sigma_{rms}}{1sec}\right)^2 \cdot \left(\frac{Power}{1mW}\right) \cdot \left(\frac{F_{REF}}{50MHz}\right)\right]$$

Table 2.1 Comparison with recent ring oscillator-based integer-N synthesizers.

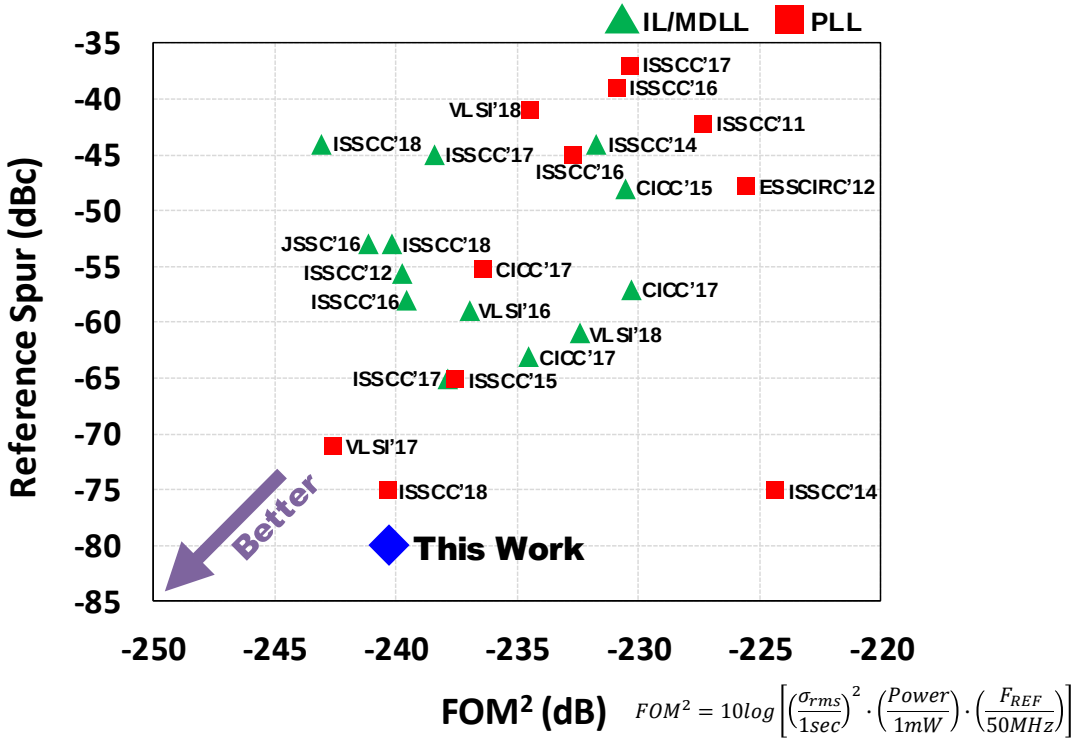


Figure 2.14 Spur vs. FOM2 of ring oscillator-based integer-N synthesizers

Chapter 3 Sub-100fs Ultra Low-jitter LC-oscillator based PLL with Reference Oversampling Technique

3.1 Introduction

Generation of high-purity clock sources is becoming more crucial in today's communication systems. With the advent of advanced communication systems such as 5G wireless radios and ultrahigh-speed wireline transceivers, the required clock jitter is now below 100 fs [5], [6]. For example, the integrated jitter requirement of the local oscillator (LO) in 5G transceivers can be as low as 90 fs to maintain an acceptable error vector magnitude (EVM) [5]. The front-end ADCs in ultrahigh-speed wireline transceivers demand clock jitter less than 100 fs to avoid signal-to-noise ratio (SNR) degradation [6].

However, it is hard to generate a clock with such a low jitter while satisfying other requirements such as power and spur, due to conflicting trade-offs. A wide bandwidth is beneficial since it provides larger suppression of the oscillator phase noise, achieving better power efficiency [14]. However, to achieve wide bandwidth, it is critical that the in-band phase

noise from the loop components is low, which can incur large power consumption [15]. High reference spur [16] and stability issues [17] are another concern that make it difficult to achieve wide bandwidth. Therefore, the bandwidth of a PLL is typically set much lower than the stability limit of FREF/10, increasing the power consumption of the oscillator and degrading overall PLL power efficiency.

One way to avoid the conflicting trade-offs and improve the performance of PLLs is to increase the reference frequency. A higher reference frequency can lower the in-band phase noise by reducing the charge pump noise or TDC quantization noise [20]. Also, the PLL bandwidth can be increased without stability or spur concerns [21].

Consequently, reference boosting schemes, such as a reference doubler, or a quadrupler, have become popular as a way of generating high reference frequencies without using expensive high frequency crystal oscillators [22]. In [22], a delay line with an XOR gate is used, while Megawer et al. [21] employed comparators with multiple voltage references to boost the frequency of the input reference clock. However, both approaches rely on the absolute accuracy of the delay or voltage references to avoid a period error of the output clock, which can lead to a reference spur. Therefore, previous reference boosting schemes require complex calibration logic, making it hard to achieve a multiplication factor greater than four.

In this article, we present a reference oversampling PLL (OSPLL) [23], [24] whose loop operates at the output frequency FOUT, which effectively suppresses the noise from the loop components by oversampling, thereby achieving low in-band phase noise. The low in-band phase noise, in turn, enables wide bandwidth and better oscillator noise filtering, which is achieved with less stability issue and low reference spur as the loop operates at FOUT rather than FREF. This article also presents an LC digitally controlled oscillator (DCO) for the OSPLL with fine

frequency resolution and fast switching capability. Furthermore, a timing control scheme is proposed to optimally position the varactor switching pulse for low phase noise. In combination with the low in-band phase noise and the wide bandwidth, the implemented PLL achieved 67.1-fs integrated rms jitter with 5.2-mW power consumption at 4 GHz, which leads to −256.3 FoM. The measured reference spur of the PLL was −78.1 dBc.

3.2 Modeling and Simulation of the Reference Oversampling PLL

The intrinsic noise of an RSPD is composed of sampling noise (VN,SMPL) and comparator noise (VN,COMP) as shown in Figure 3.1. The noise power generated by the two noise sources is identical for every RSPD. However, the input jitter-to-sampled voltage gain, represented by ΔVSMPL/Δt at RSPDi, depends on the slope (SLOPEi) of the sinusoidal reference where the RSPDi samples, which can be expressed as follows:

$$\begin{aligned} \Delta V_{SMPL} &= SLOPE < i > \cdot \Delta t \\ &= \frac{d}{dt}[A_{REF} \sin(\omega_{REF} t)] \cdot \Delta t \quad , t = \frac{i-1}{N} T_{REF} \\ &= A_{REF} \omega_{REF} \cos\left(2\pi \frac{i-1}{N}\right) \cdot \Delta t \end{aligned} \tag{1}$$

where AREF is the amplitude of the input sinusoidal reference. The main implication of (1) is that due to the different slopes of the sinusoidal reference that each RSPD samples, the gain from the timing jitter Δt to the sampled voltage differs for each RSPD, as does the SNR of each RSPD. For example, near the zero-crossing point of the reference clock, the input slope is steeper, thus the gain from Δt to ΔVSMPL is higher as shown in Figure 3.2(a). This high gain suppresses the subsequent noises VN,SMPL and VN,COMP, providing excellent Δt detection. On the other hand, when the sampling points are near the peak, the gain is small, so the RSPD output is mostly dominated by VN,SMPL and VN,COMP rather than Δt.

For the optimal noise performance of the PLL, it is beneficial to emphasize the RSPDs that sample steeper points of REF compared with the RSPDs that sample REF near its peak points. For this, we set a different proportional path gain PGAIN1:N for each RSPD, so that the gain of each RSPD is proportional to the slope of the sampling points, which is quadrature to the reference clock, REF, as shown in Figure 3.2(a). Note that it is even possible to remove the

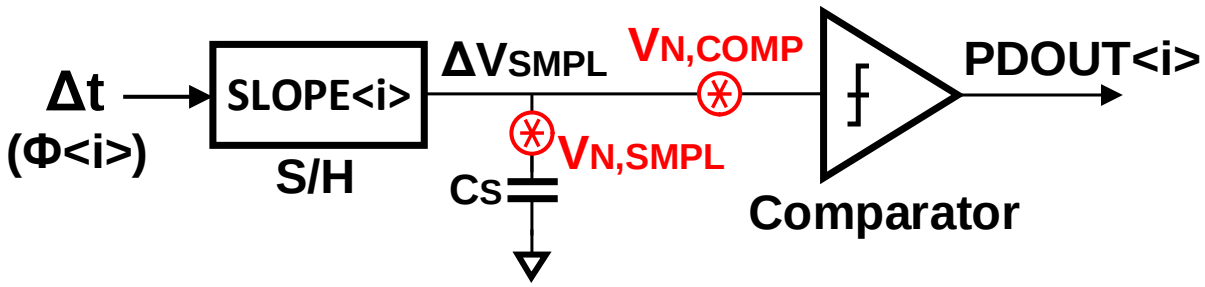


Figure 3.1 Noise source of the RSPD.

RSPDs near the peak/bottom which has near zero PGAIN value to save area and power. Figure 3.2(b) shows the simplified structure of the proposed DLF, where the RSPD output PDOUT1:N is multiplied by the corresponding proportional path gain PGAIN1:N, realizing the gain profile in the proportional path output UP/DN1:N. Note that for the integral path, only one of the PDOUTs, PDOUT1, is used to lower the operation speed and save power.

A discrete time domain model [41], shown in Figure 3.3, is adopted to simulate the PLL noise and optimize its design parameters. The periodically time variant parameters such as SLOPE[i] and PGAIN[i] are modeled as shown in Figure 3.3, while the index i is calculated as MODN [k] + 1. MODN is a modulo-N operation, and k is the index of the timestamps of the simulation, which increases by 1 at every TDCO. In other words, the entire model operates in a single DCO clock domain. The reference sampling operation is modeled with a gain block with the gain of SLOPE[i] as derived in (1), and the sampling noise is added at its output. The comparator model is composed of a 1-bit quantizer and a latency block, Z^{-D}, that models the comparator evaluation time (D = 6 in the implementation, where D represents the number of DCO cycles). Also, the input referred noise of the comparator VN,COMP[k] is added to the input of the quantizer. The DLF consists of a proportional path and an integral path. The proportional path is modeled with the periodically varying gain PGAIN[i], which is proportional to SLOPE[i] with the amplitude of β. The integral path sub-samples its input when MODN [k] =

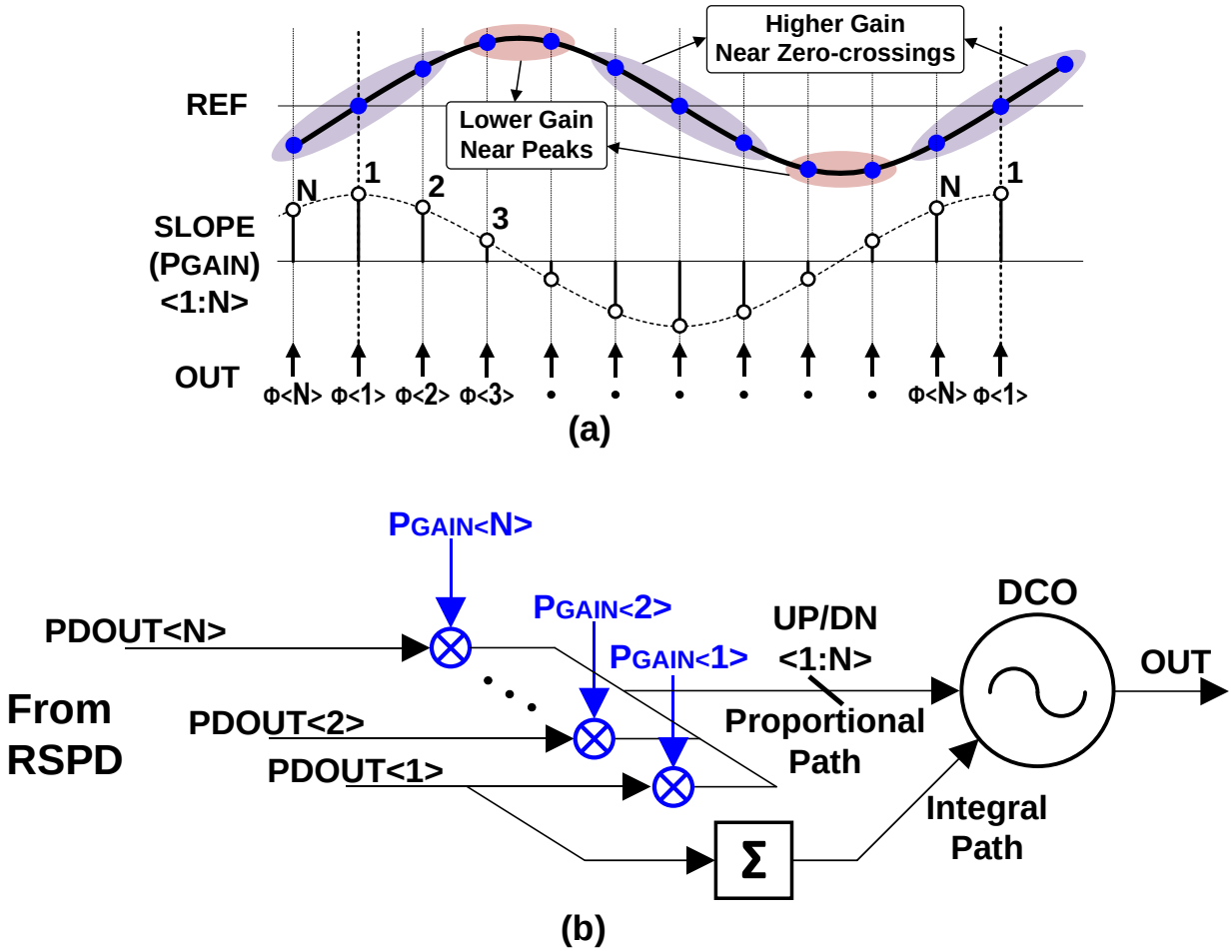


Figure 3.2 (a) Different sampling gains due to the different reference slope. (b) Simplified structure of the proportional path with the gain profile and the integral path.

0 because the integral path only utilizes PDOUT1. The sampled output is then applied to the integrator and the integral gain alpha. The output of the DLF is followed by the DCO, which is modeled as an integrator with a gain of KDCO and the DCO noise tN,DCO[k] added at the output. The DCO output is fed back to the input, forming a PLL loop.

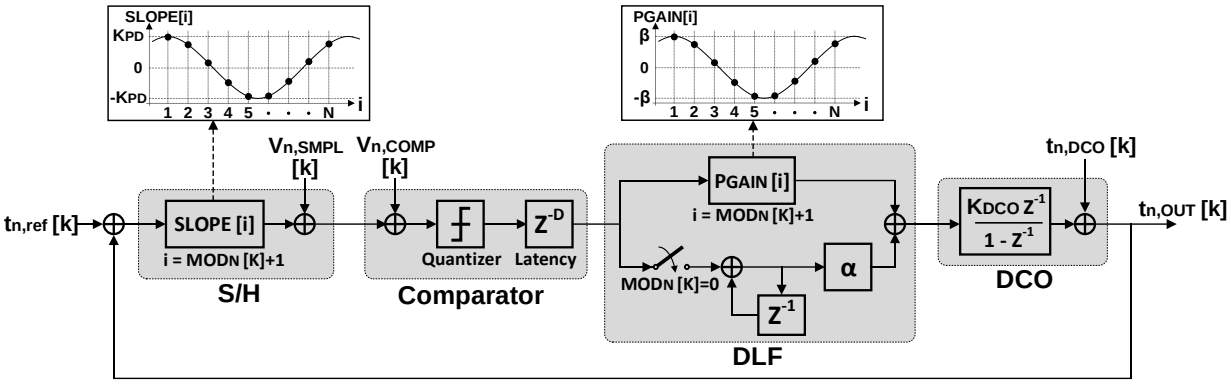


Figure 3.3 Discrete time domain model of the proposed OSPLL.

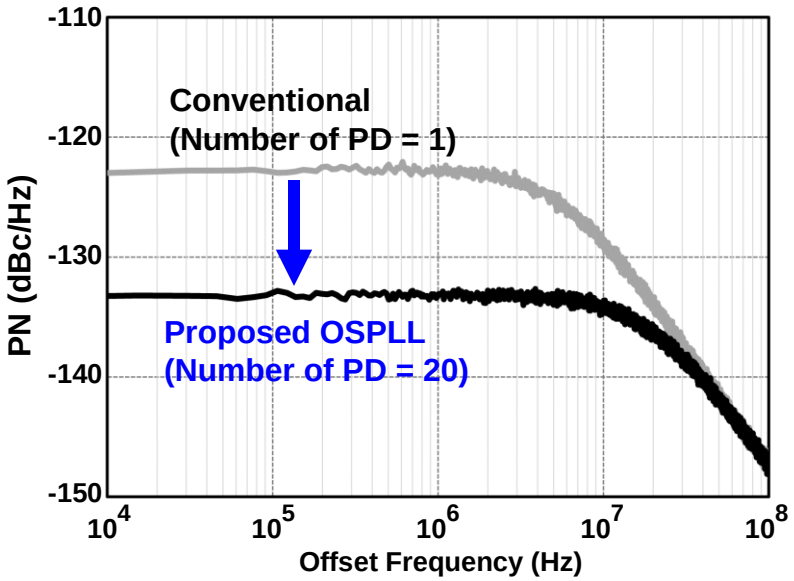


Figure 3.4 Simulated phase noise plot of the proposed OSPLL (N = 20) and the conventional BBPLL (N = 1) at 4-GHz FOUT and 200-MHz FREF.

Figure 3.4 shows the simulated output phase noise spectrum of the proposed OSPLL model. The number of the RSPDs (N) in the OSPLL is 20 and the reference frequency is 200 MHz, resulting in a 4-GHz output frequency. In order to show the advantage of the proposed OSPLL architecture, a conventional BBPLL is also simulated, and its phase noise spectrum is plotted in Figure 3.4. Compared with the BBPLL, the OSPLL achieved 10.1 dB of in-band phase

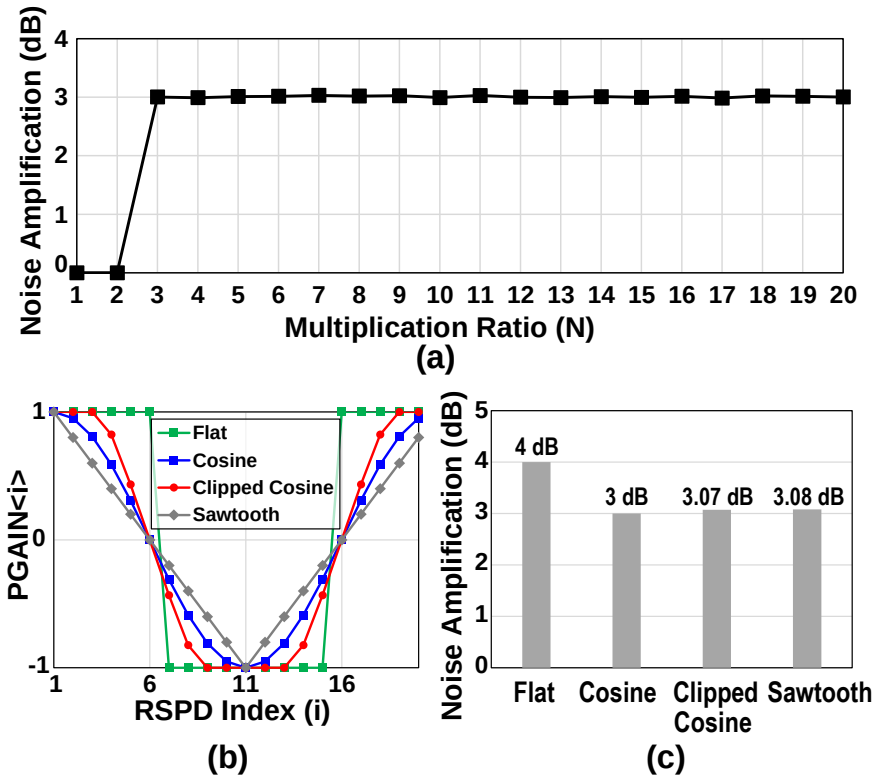


Figure 3.5 (a) Simulated noise amplification due to sinusoidal reference when cosine PGAIN profile is used. (b) Tested PGAIN profile when N = 20. (c) Simulated noise amplification compared with flat profile when N = 20.

noise suppression; the in-band phase noise of the OSPLL and the conventional BBPLL are −132.9 and −122.8 dBc/Hz, respectively. The total integrated rms jitter of the OSPLL and BBPLL are 65 and 122.3 fs, respectively.

In the sinusoidal reference, there exists several low slope sampling points such that the RSPD output is dominated by the loop components noise. As a result, there will be an in-band phase noise amplification compared with when all sampling points have the same highest slope of the sinewave reference. This in turn lowers the in-band phase noise improvement of oversampling from 10logN to 10logN–AN, where AN is the noise amplification. Figure 3.5(a) shows the simulated in-band phase noise amplification with different multiplication ratio N. The

noise amplification is 3 dB regardless of N when $N \geq 2$. This also agrees with the simulation results in Fig. 8, where the improvement from oversampling when N = 20 is 10logN–AN = 13 dB–3 dB = 10 dB. Note that when N = 1, 2, the noise amplification is 0 dB since all sampling points have the same highest reference slope.

The effect of the time-varying proportional gain (PGAIN) on the noise amplification is verified by comparing OSPLLs operating with different PGAIN profiles, as shown in Figure 3.5(b). Flat profile has the same PGAIN value for all RSPD indexes. The cosine is the profile we adopted in this design, which is quadrature to the input sinewave reference. Clipped cosine and sawtooth profiles represent steeper and more gradual slopes compared with the cosine profile. Figure 3.5(c) shows the simulated in-band phase noise amplification when N = 20. The noise amplification of the flat and cosine profile was 4 and 3 dB, respectively, showing that 1-dB in-band phase noise improvement is achieved by using the time varying cosine PGAIN profile. Clipped cosine and sawtooth profile show 3.07 and 3.08-dB noise amplification, which is slightly worse than the cosine profile, but this implies that as long as the profile emphasize the high slope sampling points and deemphasize the low slope sampling points, the improvement is similar.

In addition to the noise of each building block, one important parameter of the OSPLL for jitter minimization is the latency of the loop, which is dominated by the comparator evaluation time modeled as Z^{-D} in Figure 3.3. As the optimal loop bandwidth for the proposed OSPLL is high (~20 MHz) due to its low in-band phase noise level, the minimization of the latency is critical for low jitter [19]. Figure 3.6 shows the simulated PLL jitter with different latency values, increasing at a 0.9-fs/250-ps rate. In this work, the latency is designed with six DCO cycles (1.5 ns) as a result of a trade-off between the latency and the comparator input

transistor size. Note that even though the loop latency is multiples of the effective loop operating cycle (TDCO), there is less stability or limit cycle issue since the loop time constant (1/FBW) is much larger compared with the loop latency [25].

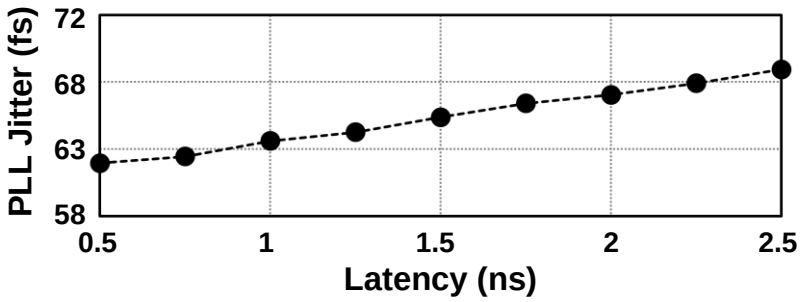


Figure 3.6 (a) Simulated PLL Jitter according to the loop latency.

3.3 DCO Design for Reference OSPLL

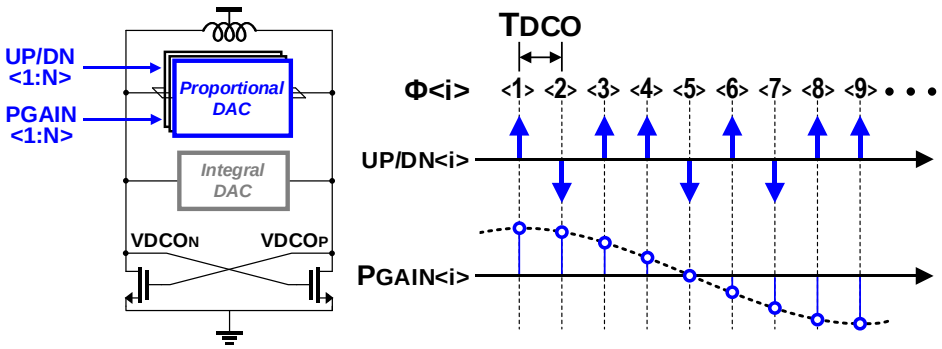


Figure 3.7 Conceptual operation of the LC-DCO for the OSPLL.

The design of an LC-DCO for the OSPLL introduces several requirements that differ from those of a conventional PLL. Figure 3.7 shows the conceptual diagram of an LC-DCO for the OSPLL. Phase detection is performed in every DCO cycle, as is frequency control. However, due to the FDCO rate frequency control, it is challenging to support the fine frequency tuning step that is needed for the LC-DCO. According to simulation, the required frequency step of the proportional path is less than 10 kHz, which requires a sub-10-aF capacitance change. In conventional PLLs, where the loop operates in a reference cycle, this LC-DCO fine frequency step can be easily implemented using delta sigma modulator (DSM) clocked at FDCO or a lower frequency. However, this scheme cannot be adopted for the proportional path of the OSPLL, whose update rate is equal to FDCO. In addition, the DCO needs to support the proportional path gain profile, which is proportional to the slope of the sampled reference point as explained in Section II. Therefore, the DCO should have the ability to set a different gain at each RSPD output as seen in Figure 3.7.

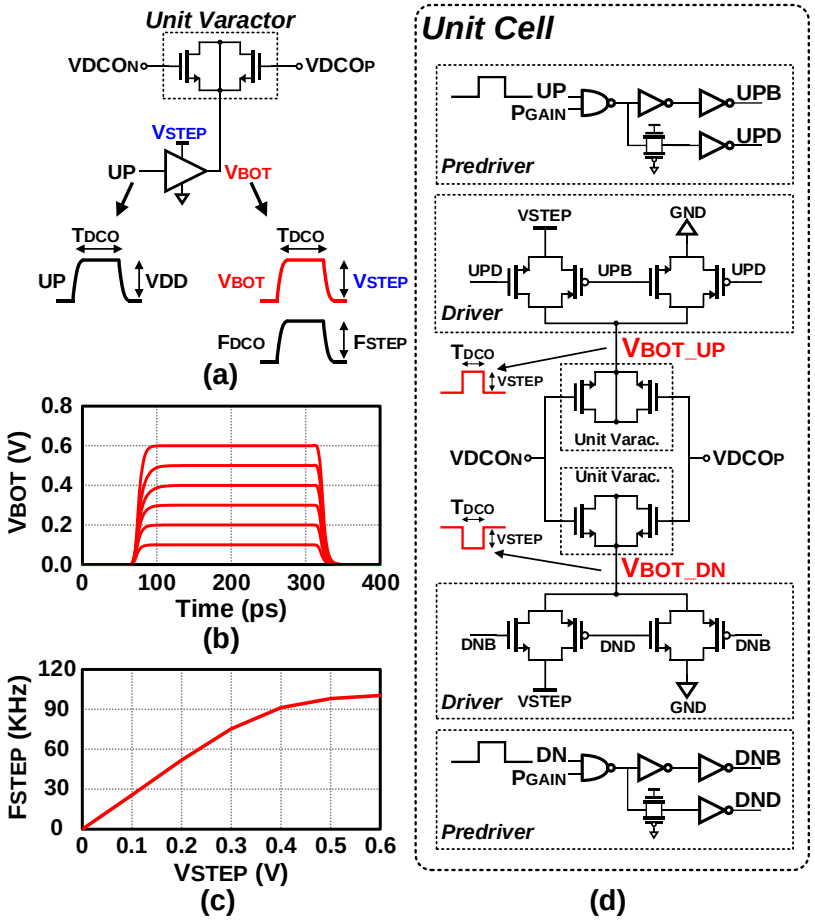


Figure 3.8 (a) Conceptual operation of the unit varactor. (b) Simulated VBOT when VSTEP is swept from 0 to 0.6 V. (c) Simulated FSTEP when VSTEP is swept from 0 to 0.6 V. (d) Detailed structure of the unit cell of the proportional DAC.

Figure 3.8 shows the structure of the unit cell of the proposed proportional DAC used in the LC-DCO to meet the aforementioned requirements of the OSPLL. Figure 3.8(a) shows the principle of the unit cell operation. The unit cell consists of two NMOS varactors whose gates are connected to the VDCOP/N while the bottom nodes of the varactors VBOT are connected to a driver whose supply voltage is VSTEP. This driver generates a pulse with an amplitude of VSTEP, which drives the varactor bottom node for one DCO cycle.

This short pulse enables fast frequency switching so that the DCO frequency can be briefly increased by FSTEP for one DCO cycle. The size of the frequency step, FSTEP, is controlled by VSTEP. By setting VSTEP sufficiently small, we can achieve fine frequency resolution of less than 10 kHz. Figure 3.8(b) shows the simulated waveform of the VBOT node of a unit varactor where VSTEP is swept from 0 to 0.6 V. Figure 3.8(c) plots the corresponding FSTEP, which shows that FSTEP can be continuously tuned from 0 to 100 kHz. Note that the current load from VSTEP is less than 10 μA, which can be supplied from a low power LDO. Figure 3.8(d) shows the detailed schematic of the proposed unit cell which consists of two sets of the unit varactor and its drivers. Transmission gate type drivers are used to provide fast VBOT rising/falling slope across the VSTEP tuning range (0–0.6 V), while delay matched pre-drivers drive the transmission gate type drivers.

Figure 3.9(a) shows the structure of the proportional DAC (PDAC). To realize the proportional path gain profile, the PDAC is constructed with N sub-DACs, whose gain can be separately controlled. A sub-DAC is assigned to the output of corresponding RSPDi. Figure 3.9(b) shows the sub-DAC, which consists of seven unit cells to provide gain programmability as previously described. A 3-bit PGAIN signal controls the number of unit cells that are enabled when UP/DN toggles. This enables each sub-DAC with eight-level programmability of 3-bit binary coding. PGAINi for each sub-DAC is predetermined during the design phase to provide a quadrature profile to the sinusoidal reference shape. Figure 3.9(c) shows the detailed operation of the proportional DAC. The DCO frequency switches according to the UP/DN pulses, and the size of the frequency step is controlled by the PGAIN value of each sub-DAC.

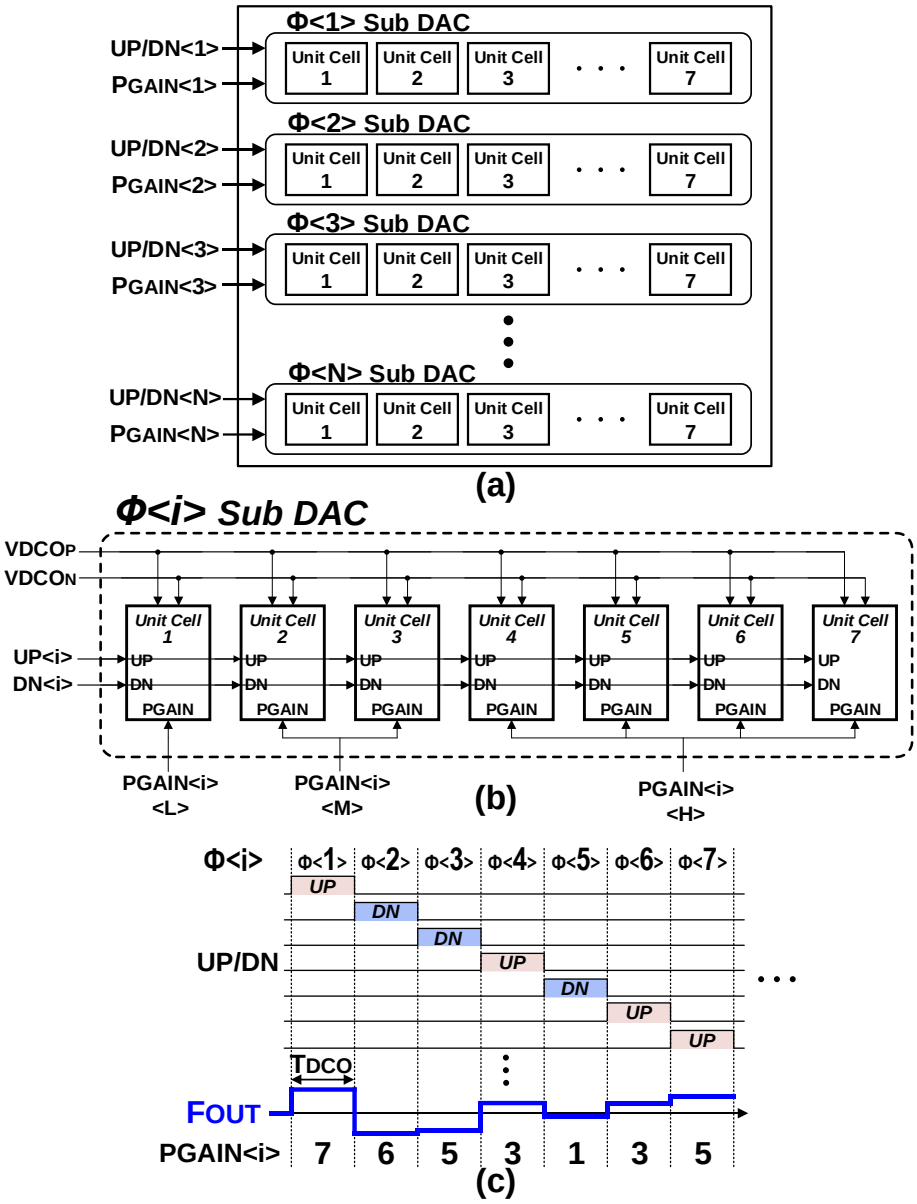


Figure 3.9 (a) Structure of the proportional DAC. (b) Structure of the sub-DAC of the proportional DAC. (c) Conceptual operation of the proportional DAC.

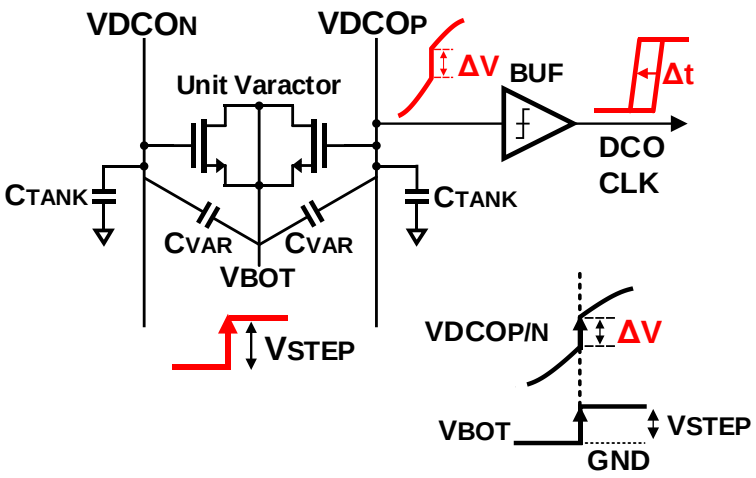


Figure 3.10 Conceptual diagram that describes the phase disturbance due to the varactor switching.

When the UP/DN pulses toggle at every DCO cycle, capacitive coupling from the varactor switching injects deterministic noise to the DCO, which may lead to an increase in PLL output jitter. Figure 3.10 illustrates varactor switching, in which the bottom node of the varactor VBOT rises by VSTEP to increase DCO frequency. This voltage transition injects charge into the DCO tank due to charge feedthrough, causing a voltage fluctuation ΔV at the DCO waveform VDCOP/N. ΔV is determined by the capacitance divider ratio, expressed as follows:

$$\Delta V = \frac{C_{VAR}}{C_{VAR}+C_{TANK}} V_{STEP} \qquad (2)$$

where CTANK is the total tank capacitance of the LC-DCO, and CVAR is the capacitance between VBOT of a varactor and the CTANK, which includes the varactor gate channel capacitance, the gate-to-drain/source overlap capacitance, and the parasitic capacitances due to layout. This voltage fluctuation can cause a phase disturbance of the DCO waveform by directly shifting its zero-crossing, resulting in jitter increase. To minimize such a side effect, we need to: 1) minimize the ΔV upon varactor switching and 2) prevent the created voltage fluctuation from disturbing the zero-crossing of the DCO waveform. The magnitude of ΔV depends on the tank

capacitance CVAR as in (2), which varies according to the varactor switching location with respect to the DCO waveform as shown in Figure 3.11(a) and (b). As an example, the gate capacitance is small when VDCO is low because the varactor is in depletion mode. Therefore, it is preferable to switch VBOT at the valley point of VDCO. On the other hand, the varactor switching at the peak point of VDCO should be avoided as the large varactor capacitance results in a large ΔV.

ΔV then causes DCO output phase disturbance by altering the zero-crossing point of the waveform, and the amount of the phase disturbance depends on the location of the varactor switching with respect to the DCO sinusoid, as described by the impulse sensitivity function (ISF). As shown in Figure 3.11(c), if ΔV occurs near the peak/valley of VDCO, the translation from ΔV to phase disturbance is minimized. However, the phase disturbance due to ΔV is maximum near the VDCO zero crossings.

Therefore, to minimize the phase disturbance, the varactor needs to be switched at the valley point of VDCO where both ΔV and the ISF can be minimized. Near the zero crossing is not the optimal point due to the high ISF. The peak of VDCO is also not the optimal point due to the high CVAR. Figure 3.11 illustrates the simulation results of ΔV, normalized ISF, and phase disturbance, clearly showing that the valley of VDCO exhibits minimum phase disturbance due to the varactor switching.

To ensure the optimal timing for varactor switching at the DCO, the timing that UP/DN pulses reach the DCO should be controlled. Figure 3.12 shows the control scheme for the DCO tuning pulse, which generates UP/DN pulses with a programmable delay line to position them at the optimal varactor switching timing. Similar to [26], a digitally controlled delay line is used to control the timing of the UP/DN pulses. Instead of using N different delay lines for N UP/DN

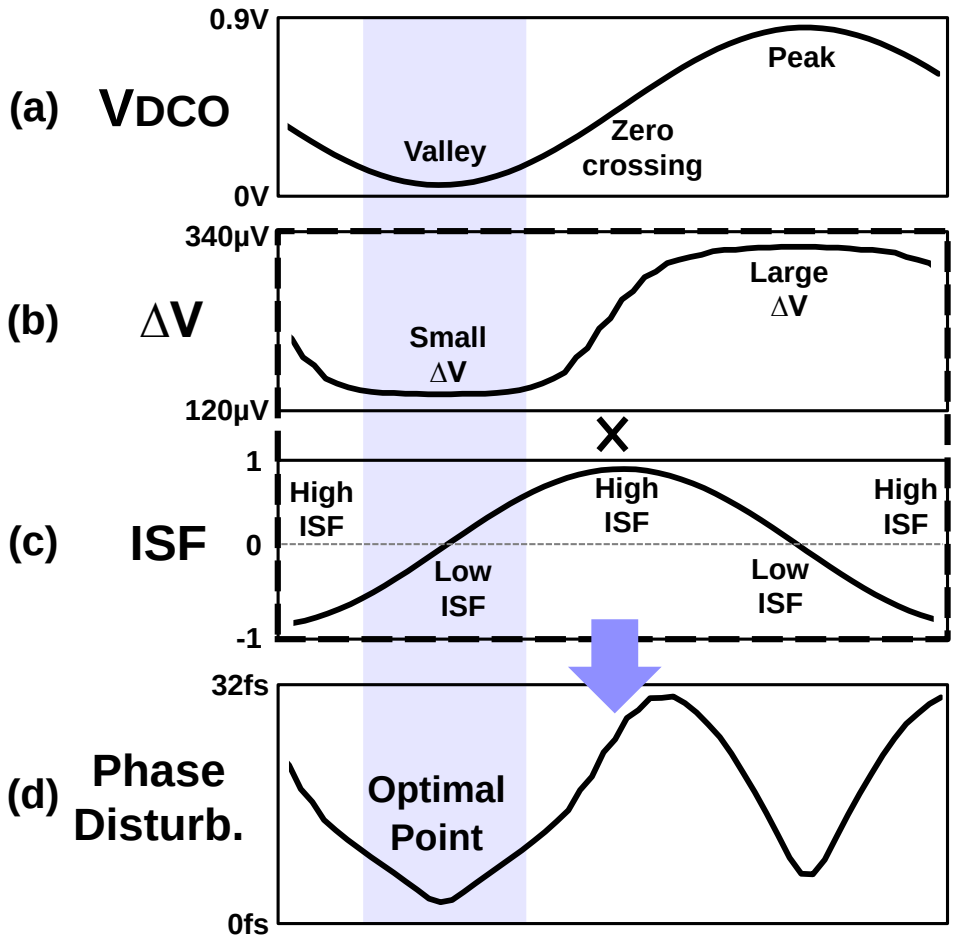


Figure 3.11 Simulated phase disturbance due to the varactor switching (simulation condition: TT/25 ˚C/VSTEP = 0.3 V/PGAIN = 111). (a) DCO waveform. (b) Voltage disturbance due to the varactor switching according to the switching position. (c) Normalized ISF of the DCO. (d) Phase disturbance of the DCO according to the varactor switching position.

pulses, which incurs area overhead, we used a single shared delay line as shown in Fig. 16(a). The DCO output is first divided by two before it is applied to the delay line to lower the power consumption of the delay line. The delay line output, CLK_DLY, then drives the N pulse generator unit, $PULSE_GEN_{1:N}$ where the CLK_DLY is multiplied by $MASK_{1:N}$ to select the corresponding pulse from CLK_DLY and generate $UP/DN_{1:N}$, as shown in Figure 3.12(b). Note

that in the PULSE_GENi unit with odd index, CLK_DLY is inverted before it is multiplied by MASKi to maintain the same UP/DN polarity. In Figure 3.12(b), tDLY is the delay of the delay line, which is controlled with a 5b control word to set the timing of the UP/DN pulses. To cover the optimal range shown in Figure 3.11(d), the delay range of the delay line is larger than 0.5 × TDCO range considering the PVT variation. The delay code is set during the test time for optimal jitter, and the measurement results are discussed in Section V.

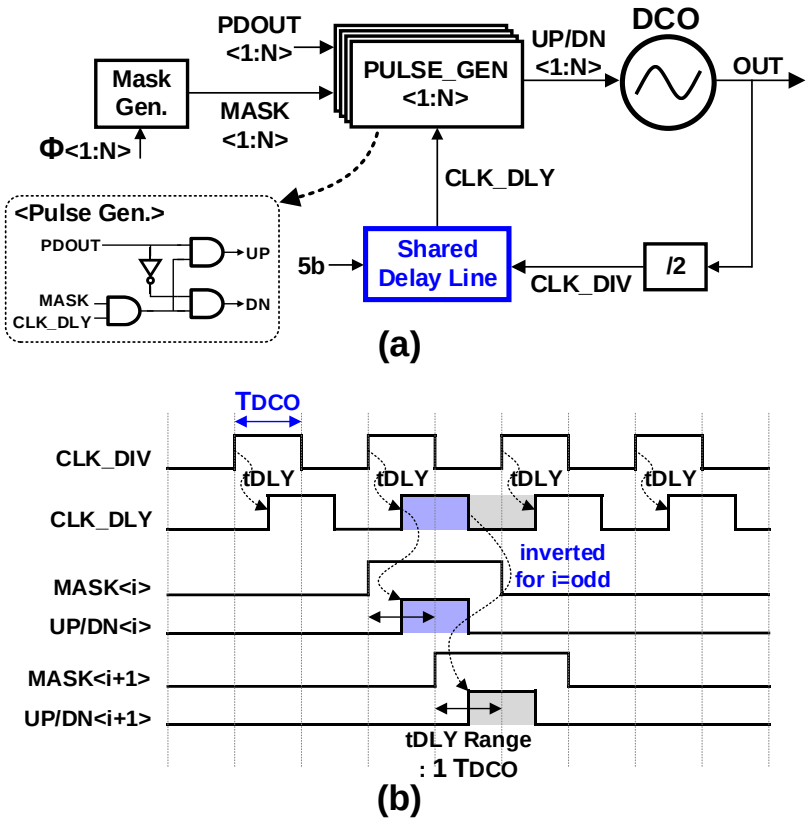


Figure 3.12 (a) Proposed DCO tuning pulse timing control scheme. (b) Timing diagram.

3.4 PLL Design and Implementation

3.4.1 Overall Architecture

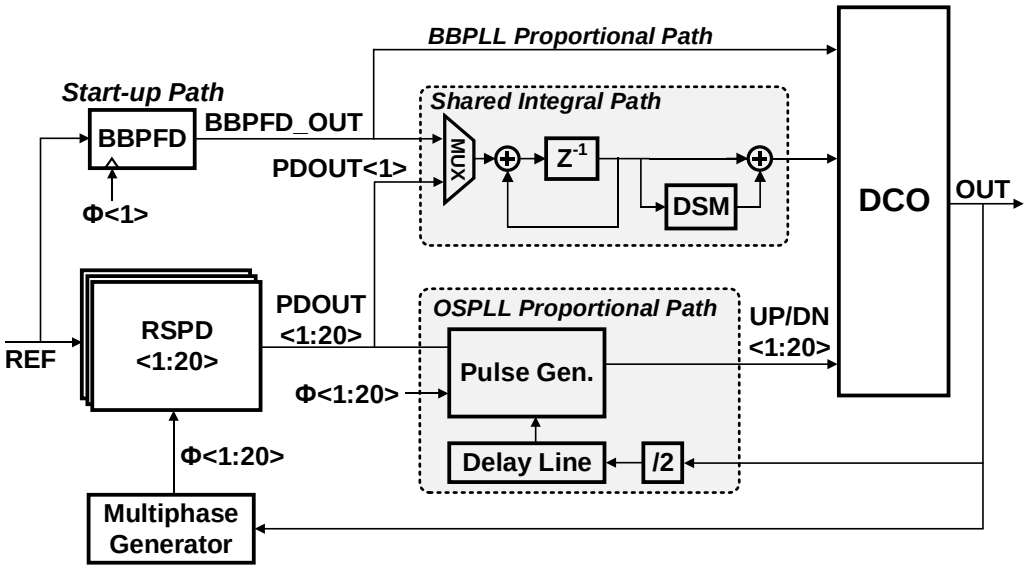


Figure 3.13 Overall structure of the proposed OSPLL.

Figure 3.13 shows the overall architecture of the proposed PLL: 20 RSPDs are used in this design, so that the PLL can generate integer-N multiplication of the reference frequency by 20 or less. The multiphase generator is clocked by OUT and generates the interleaved clocks SMPL1:20 and COMP1:20, which drive the RSPDs. The outputs of the RSPDs, PDOUT1:20, are applied to the DLF, which is composed of a proportional and an integral path. The proportional path contains a DCO pulse generator, and a delay line whose input is a divided-by-two clock of the DCO output. The integral path is used for the type-II operation to provide better DCO flicker noise suppression and a zero phase offset. For the integral path, only one of the RSPD outputs, PDOUT1, is used to reduce its operation speed and power consumption. The 11

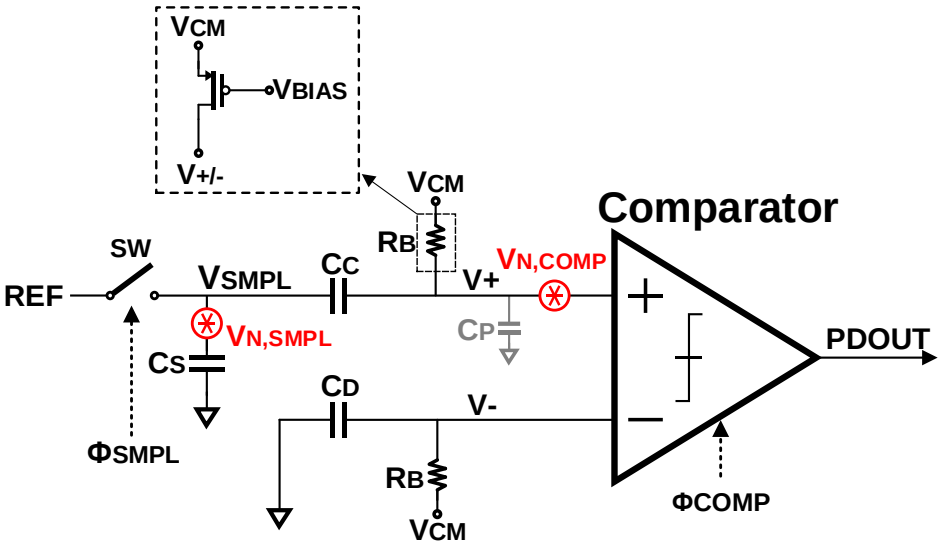


Figure 3.14 Detailed schematic of the RSPD with its noise source.

LSBs of the integrator are applied to DSM to obtain fine frequency resolution. The design also includes a startup path for wide and robust initial locking. When the PLL is turned on, the startup path using a conventional bang-bang phase and frequency detector (BBPFD) achieves initial phase and frequency lock. Then, the startup path is disabled, and the reference oversampling path takes over the loop. The startup path shares the integral path with the reference oversampling path to save area whereas the output of the BBPFD directly drives the DCO for the proportional path.

3.4.2 Reference Sampling Phase Detector

Figure 3.14 shows the detailed schematic of the RSPD. An NMOS transistor driven by bootstrapped SMPL is used as a sampling switch (SW). The sampling capacitor CS, the ac coupling capacitor CC, and the dummy capacitor CD are implemented with MOM capacitors.

It is important to optimize the noise and power of the RSPD, since the in-band noise of the PLL is dominated by the RSPD and its power takes up a large portion of the total PLL power.

The RSPD noise mainly comprises comparator noise (VN,COMP) and KT/CS sampling noise (VN,SMPL). The reduction of VN,COMP requires more comparator power, while a larger area is needed to reduce VN,SMPL. In this design, we first set the noise of the comparator given the PLL total power budget, with the goal of achieving optimal PLL jitter-power efficiency [33]. We then decide the size of the sampling capacitor so that the sampling noise does not dominate the overall RSPD noise. According to the behavioral simulation, the required input referred noise of the comparator is 100 μV. Next, given the comparator noise, we determine the size of capacitors CS and CC. Maximizing the sampling capacitor CS is advantageous for reducing the sampling noise. At the same time, the coupling capacitor CC should also be large to minimize the signal attenuation due to the parasitic capacitance CP at V+ node, which can be represented as follows:

$$\Delta V_{+} = \frac{C_C}{C_C + C_P} \Delta V_{SMPL} \tag{3}$$

Therefore, given a fixed total capacitance, CTOT, there exists an optimal CS to CC ratio. Figure 3.15(a) shows the simulated PLL jitter, when CS is increased, and CC is decreased for a fixed CTOT. When CS is small, the sampling noise dominates the comparator noise. Hence, increasing CS can improve the PLL jitter in this region. However, as CS becomes larger, the signal attenuation effect due to a smaller CC starts to degrade the PLL jitter. Therefore, the optimal ratio between CS and CC can be determined for minimum PLL jitter. Figure 3.15(b) shows the PLL jitter with different CTOT values, where the optimal CS/CC ratio is used for each CTOT. As shown in the figure, the PLL jitter improvement diminishes as CTOT is increased, and eventually saturates as the comparator noise starts to dominate the sampling noise. Thus, we set CTOT to 1.4 pF, the point beyond which increasing CTOT gives negligible jitter improvement considering the area increase (<0.1 fs per 100 fF). The capacitances of CS and CC are set to 1 pF and 400 fF, respectively, yielding the optimal CS/CC ratio. Note that a matching

capacitor, CD, which is equivalent to the series capacitance of CS and CC (300 fF), is attached to the negative terminal of the comparator to minimize the comparator offset induced by the impedance mismatch.

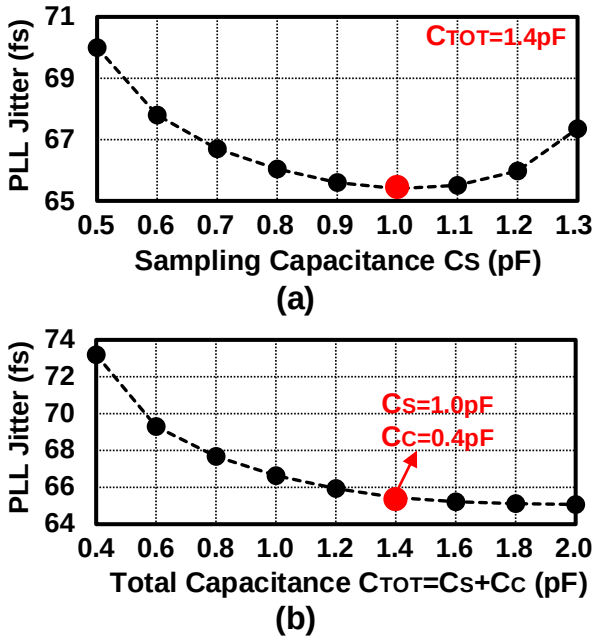


Figure 3.15 (a) Simulated PLL jitter according to CS with CTOT = 1.4 pF. (b) Simulated PLL jitter according to CTOT. Optimal CS/CC ratio is used for each point, which was obtained using the method in (a).

The bias resistor RB sets the high pass corner of the ac coupling path. Its value should be sufficiently low, otherwise it will block the low frequency component of the RSPD input phase error information, degrading the phase noise. As shown in Figure 3.16, the low frequency phase noise increases if the high pass corner frequency is large (≈100 kHz). Note that the phase noise flattens again around 10 kHz due to the pivot RSPD. In this design, the high pass corner is set to a sufficiently low value (below 1 kHz) using a PMOS transistor biased in the sub-threshold region to avoid phase noise degradation at low-offset frequency. Note that VBIAS for the resistor is generated from a diode connected replica transistor with current source load to reduce

the resistance variation with PVT. Since V± node is biased with high resistance, the input transistors of the comparator are designed with thick gate transistors to minimize the gate leakage.

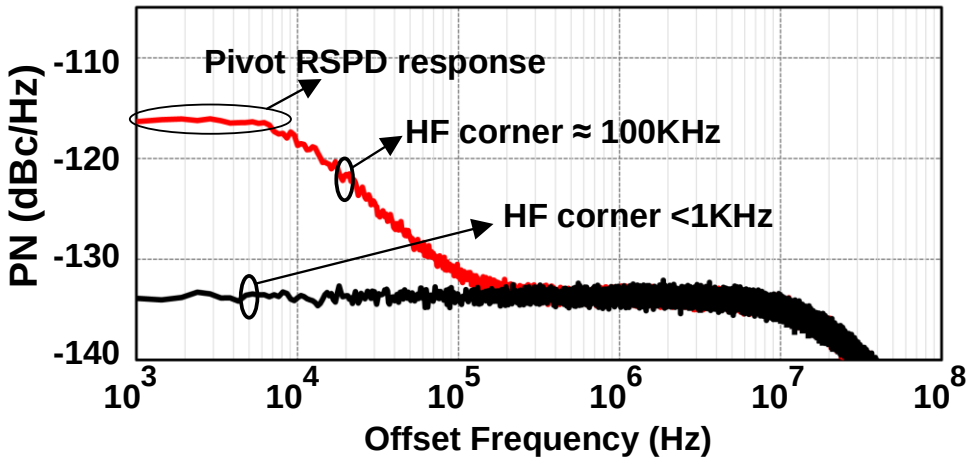


Figure 3.16 (a) Simulated PLL phase noise plot with different HF corner of the ac coupling path.

3.4.3 Comparator of the RSPD

The comparator dominates the noise and power of an RSPD, making its optimization critical. Figure 3.17(a) shows the schematic of the two-stage dynamic comparator adopted in this design. It consists of a dynamic preamplifier followed by a regenerative latch. The comparator is in reset phase when CK is low, pre-charging INT_P/N node to VDD. As CK goes high, the input transistors M1, 2 start to discharge INT_P/N with a differential current in proportion to its input voltage difference, providing a voltage gain. When the voltages reach the latch threshold of the second stage, the latch outputs OP/ON are fully regenerated to the rail-to-rail digital level by the positive feedback. Note that the clock gating transistors M5 and M6 are placed at the drain side of the input transistors M1 and M2, instead of their source side to reduce kickback when CK toggles.

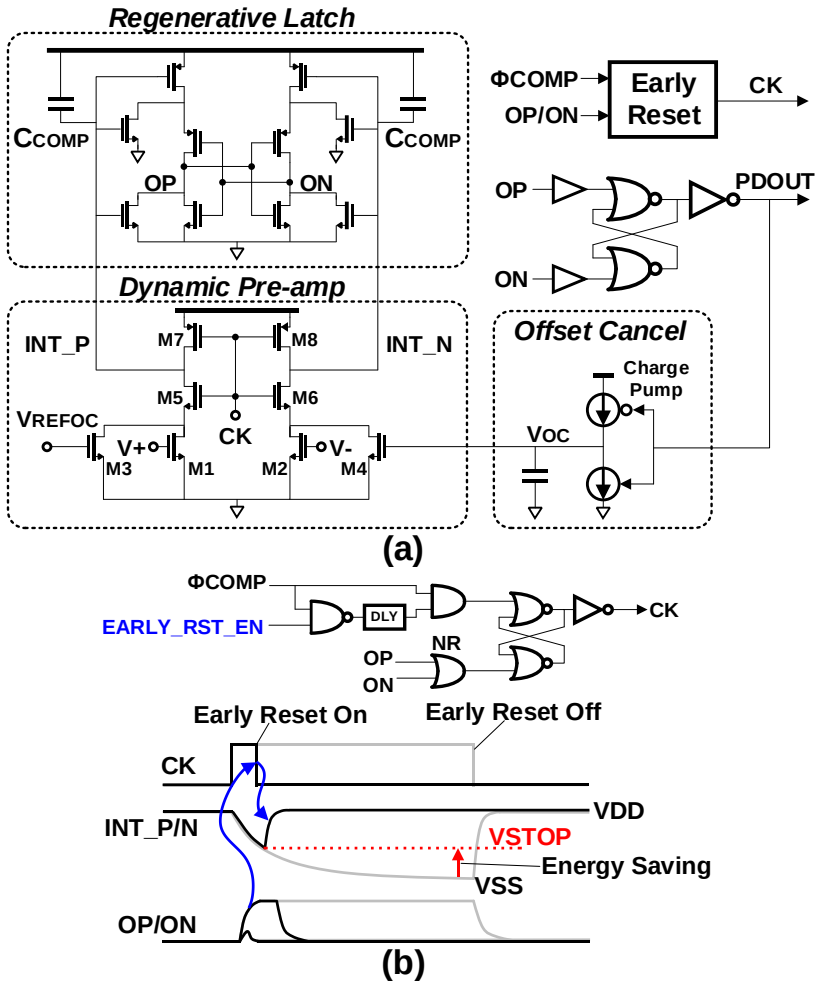


Figure 3.17 (a) Detailed structure of the comparator in an RSPD. (b) Early reset scheme and its operation.

Comparator noise is mainly determined by pre-amplifier noise since the gain of the first stage pre-amplifier suppresses the noise from the second stage latch. Adding capacitor C_{COMP} at the integration node (INT_P/N) reduces the noise level by narrowing the noise bandwidth. In this design, a 430-fF C_{COMP} was used to meet the target input-referred comparator noise of 100 μV. Adding C_{COMP} inevitably increases comparator power consumption, since C_{COMP} is fully charged to VDD and then discharged to VSS in every cycle, consuming $2\times f_{REF}\times C_{COMP}\times V^2_{DD}$. To reduce comparator power consumption while maintaining its noise performance, we adopted the early

reset scheme [42] as shown in Figure 3.17(b). The early reset scheme can reduce the energy dissipated in the capacitor by stopping capacitor discharge once comparator evaluation is finished. As shown in Figure 3.17(b), the NOR gate NR resets the S-R latch when either OP or ON is asserted, thereby stopping the discharge of the INT_P/N node at the V_{STOP} level. As a result, power consumption is reduced to $2\times f_{REF}\times C_{COMP}\times(V^2_{DD}-V^2_{STOP})$. Note that signal EARLY_RST_EN is added to turn on/off the early reset scheme for test purposes. The measurement result confirmed 25% power saving without phase noise degradation, when the early reset scheme is turned on.

In addition to its noise, the offset of the comparator should be minimized to ensure proper operation of the RSPD. If the offset is large, the output of the comparator can be dominated by the offset rather than the input phase error. In this case, the comparator output can be skewed to either one or zero, generating a spur at the PLL output and degrading phase noise. To remove comparator offset, we adopted continuous background offset cancellation using a charge pump as shown in Figure 3.17(a). The charge pump adjusts VOC based on the comparator output. VOC is then applied to M3, forming a feedback loop so that offset is removed and equal probability of 1s and 0s is ensured at the output. Note that it is important to match the charge pump UP and DN currents to prevent the reference spur. The UP/DN mismatch will alter the 1/0 probability of the comparator output and create the reference spur. In this work, the UP/DN mismatch is suppressed to less than 2%, using the charge pump mismatch reduction technique [27]. We verified with simulation that the 2% charge pump mismatch creates less than −100-dBc reference spur. It is important that the bandwidth of the offset cancellation loop is sufficiently low so that the low frequency component of the PD input is not attenuated by the offset cancellation loop, degrading PLL phase noise performance. We set the bandwidth of the offset

cancellation loop lower than 1 kHz, similar to the high pass corner of the ac coupling path. The charge pump is biased with a small current to realize such a bandwidth.

3.4.4 Multiphase Generator

The multiphase generator produces interleaved clocks to drive the RSPDs. Figure 3.18(a) shows the structure of the multiphase generator, which is based on a ring counter design [38]. Twenty unit cells form a closed loop in which each unit cell is composed of clock gating logic, a flip-flop, a mux, and control signal generation logic. The DCO output clock is applied to all of the unit cells as a common clock and a bypass mux is included in each cell to provide programmability of the frequency division ratio [12]. Initially, half of the unit cells are set to one while the other half are set to zero. Then the initialized signals circulate the loop when CLK toggles, generating interleaved clocks as shown in Figure 3.18(b). To reduce the power consumption of the multiphase generator, the clock input CLK_G for the FF is gated so that CLK_G toggles only when the input and output of the unit cell are different, as shown in Figure 3.18(c). The simulation result shows that the gating logic saves 20% power. The unit cell also contains a control signal generator to generate SMPLi and COMPi from i as shown in Figure 3.18(c).

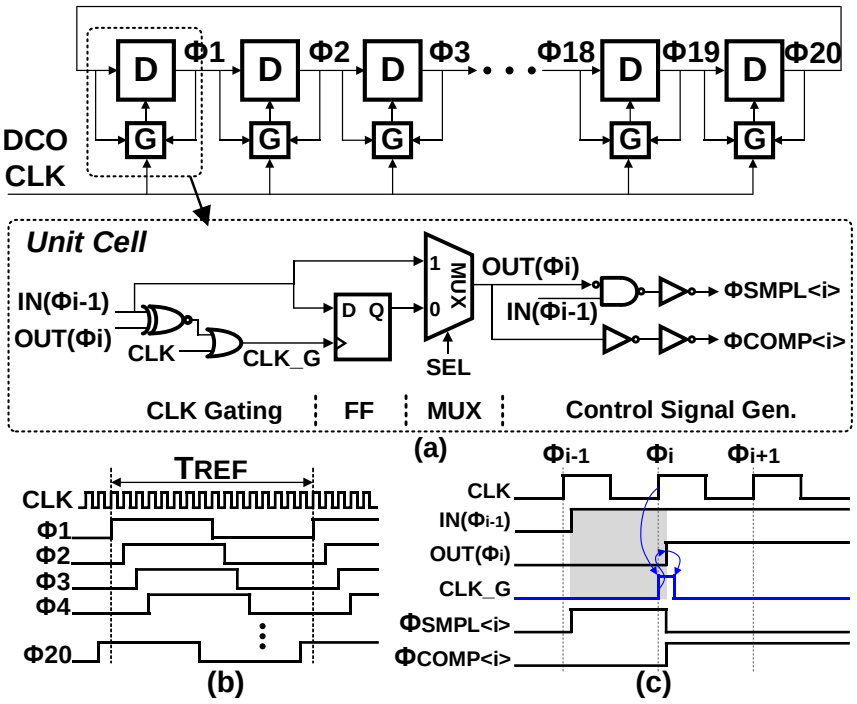


Figure 3.18 (a) Proposed multiphase generator. (b) Operation of the multiphase generator. (c) Timing diagram of each unit cell.

3.4.5 DCO

Figure 3.19 shows the detailed structure of the DCO. We used digitally controlled resistors to set the DCO current instead of the current source to mitigate the flicker noise upconversion. The DAC of the DCO consists of a proportional DAC, an integral DAC, and a PVT DAC. The proportional DAC is driven by UP/DN1:20 generated from the DCO pulse generator. The integral DAC is composed of a DSM DAC, a fine DAC, and a coarse DAC. Of the integral path output, 25 bit is divided into three groups and controls each integral DAC. The LSB part (0–10 bit) is applied to the DSM for the fine frequency step. The middle part (11–16 bit) is converted to

the thermometer code and drives the fine DAC, whereas the MSB part (17–24 bit) is applied to the coarse DAC. The DSM and the fine DAC use the varactor-based unit cell, which is the same as the one used in the proportional DAC, whereas the coarse DAC uses a MOM capacitor as its unit cell. The PVT DAC sets the coarse frequency of the DCO using 8-bit binary control code. An ac coupled inverter with resistive feedback is used to convert the DCO waveform to rail-to-rail digital output, which drives the multiphase generator, and a divide-by-4 circuit driving the DSM logic.

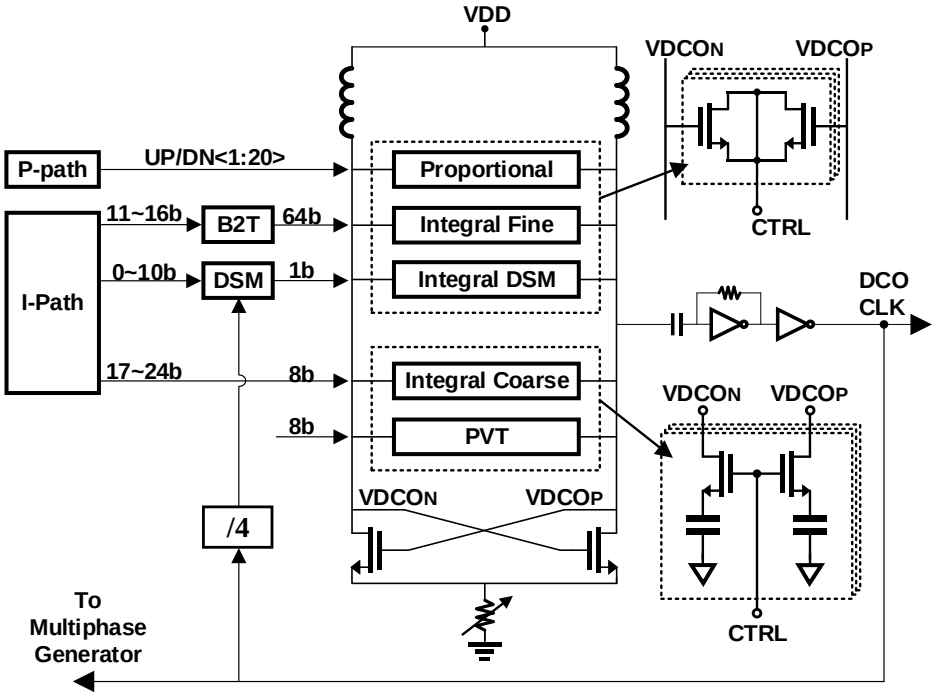


Figure 3.19 Overall structure of the DCO.

3.5 Measurement Results

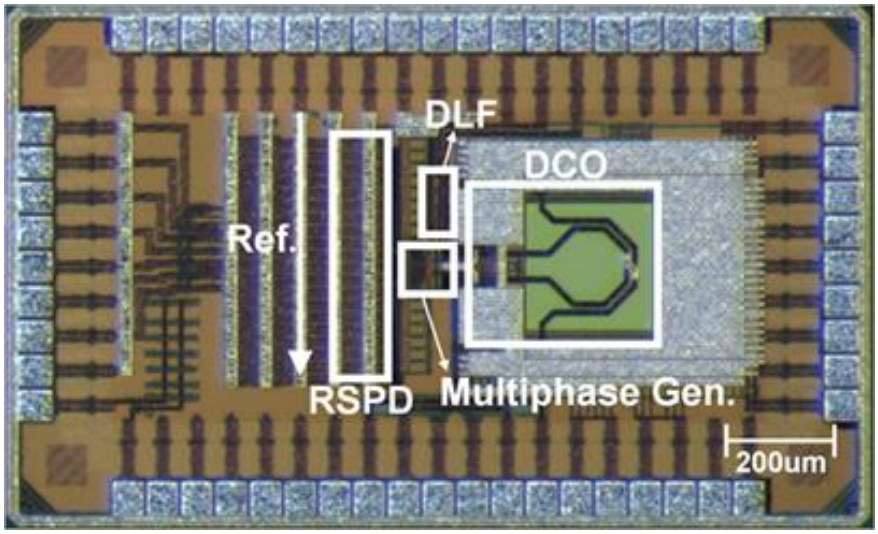


Figure 3.20 Die photograph.

The proposed OSPLL is fabricated in 28-nm CMOS. Figure 3.20 shows the die photograph and the PLL core occupies 0.17 mm^2. To prevent the unwanted noise coupling to the sinusoidal reference, the reference distribution line is carefully shielded with the ground. Also, we used the thick and wide metal layer for the reference line to reduce the resistance and minimize the sinewave attenuation. The PLL generates 4-GHz output frequency using 200-MHz reference input, while consuming 5.2 mW. The reference sinusoidal clock is generated from an external crystal oscillator (Sprinter model from Wenzel Associates). A peak-to-peak 1.2-V reference signal is applied to the chip to avoid any reliability issues of the sampling switches.

Figure 3.21 shows the measured phase noise spectrum of the PLL and the free-running DCO measured from a Keysight E5052B signal source analyzer. The in-band phase noise is −129.2 dBc/Hz at 1-MHz offset and −132.5 dBc/Hz at 5-MHz offset. The integrated rms jitter of

the PLL is 67.1 fs, and the integration range is from 10 kHz to 100 MHz. The PLL achieved −256.3 dB of PLL jitter-power FoM. The DCO phase noise in free-running mode at 4 GHz was −113.6 dBc/Hz at 1-MHz offset.

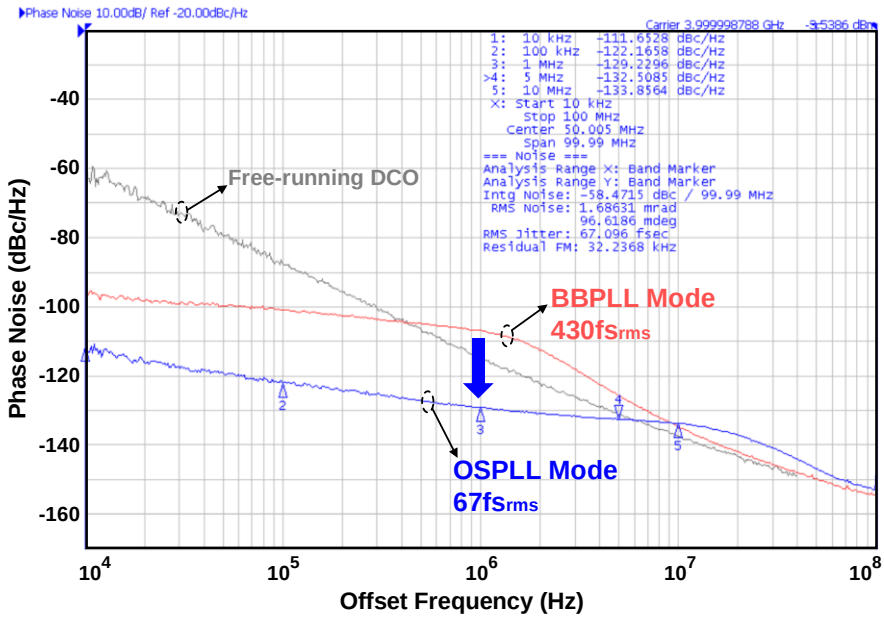


Figure 3.21 Measured phase noise spectrum of the proposed OSPLL at 4 GHz.

Figure 3.21 also shows the measured phase noise spectrum of the proposed the conventional BBPLL mode for comparison. Note that for a fair comparison, both the OSPLL and BBPLL mode are measured with the optimal bandwidth which gives the lowest jitter. The in-band phase noise improvement of the OSPLL was larger than 20 dB compared with the BBPLL, where the in-band phase noises of the OSPLL and the BBPLL were −130 and −105 dBc/Hz, respectively. The integrated rms jitter was improved from 430 fs (BBPLL) to 67 fs (OSPLL), a greater than 5× improvement.

The measured jitter performance of the BBPLL (430 fs) is worse than the simulated jitter of the conventional BBPLL in Figure 3.4 (122.3 fs). We verified with simulation that this is due to the phase noise contribution from the inverter-based reference buffer of the BBPLL, which is necessary to convert the sinusoidal reference from the crystal oscillator to the rectangular pulse for the BBPFD. Since the input of the reference buffer is a sinusoidal clock with slow slope, the noise generated from the first-stage inverter is large. The inverter chain reference buffer used in this design consumes 100 μW. Increased power could lower the noise contribution, but at the expense of higher overall power. The proposed OSPLL avoids this issue since the OSPLL directly samples the sinusoidal reference from the crystal oscillator so that it does not need a reference buffer.

Figure 3.22 shows the measured output spectrum of the proposed PLL measured from a Keysight EXA N9010B spectrum analyzer. The PLL achieved a reference spur of −78.1 dBc.

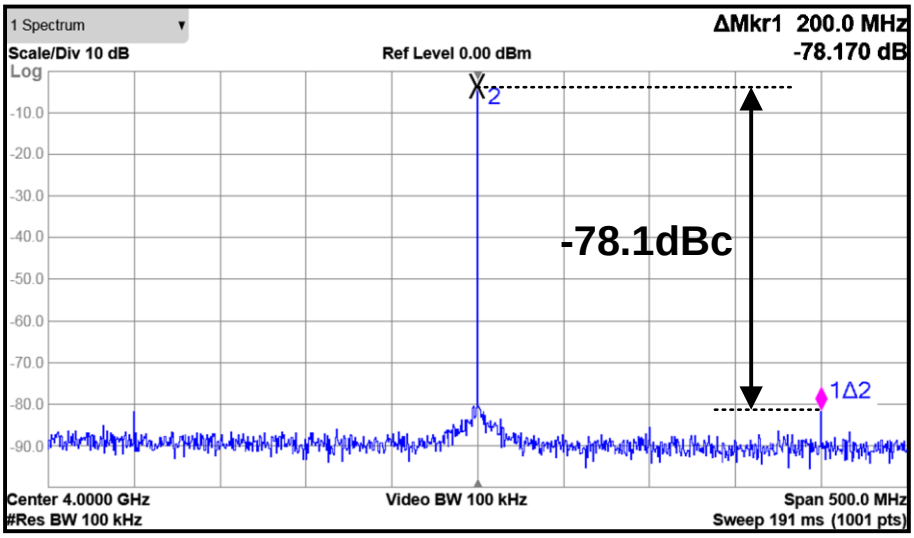


Figure 3.22 Measured output spectrum of the proposed OSPLL.

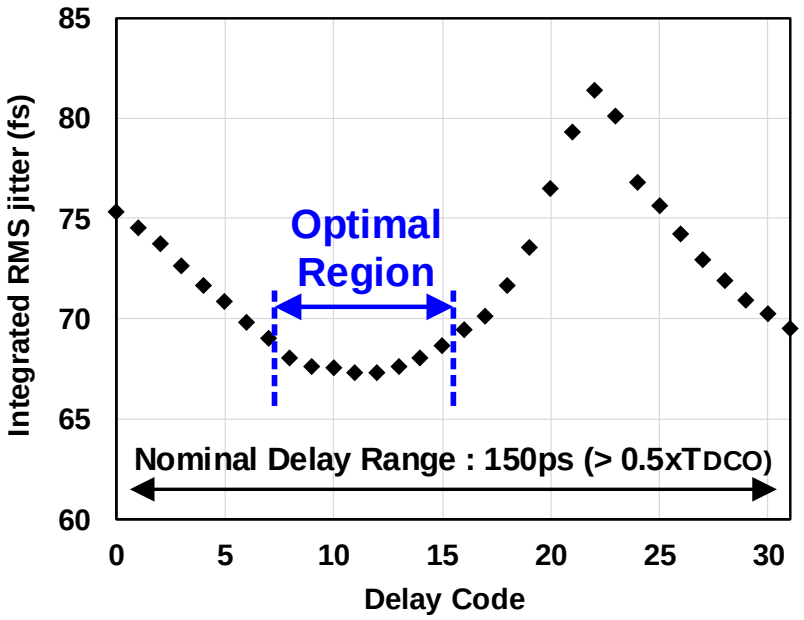
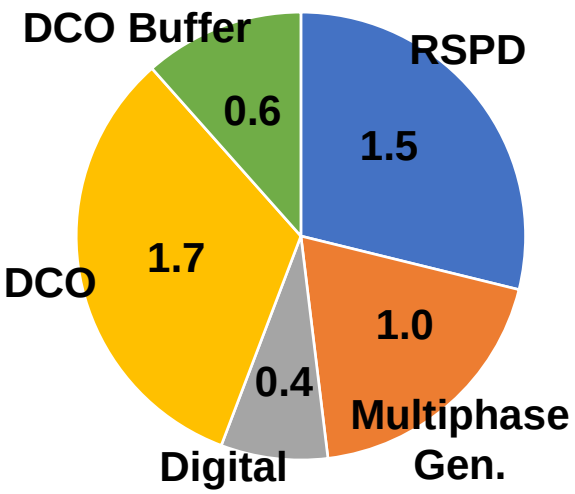


Figure 3.23 Measured output jitter of the PLL with difference delay code of the DCO tuning pulse control scheme.

Figure 3.23 shows the measured PLL output jitter according to the delay code of the proportional path timing control scheme that controls the varactor switching timing of the DCO proportional path. Across all 5-bit delay codes, the measured PLL output jitter changes from 67 to 82 fs, demonstrating the importance of the varactor switching timing control. The optimal code can be set around 11 for minimum output jitter. The range of the delay line is larger than 0.5 × TDCO to cover the optimal region given PVT variation.

Figure 3.24 shows the measured power consumption of each building block. The RSPD, DCO, DCO buffer, multiphase generator and digital logic consume 1.5, 1.7, 0.6, 1.0, and 0.4 mW, respectively, resulting in 5.2 mW of total power. Note that three RSPDs near the peak and three RSPDs near the bottom are turned off to save power.



Block	Power (mW)
RSPD	1.5
DCO	1.7
DCO Buffer	0.6
Multiphase Gen.	1.0
Digital	0.4
Total	5.2

Figure 3.24 Measured power breakdown of the PLL.

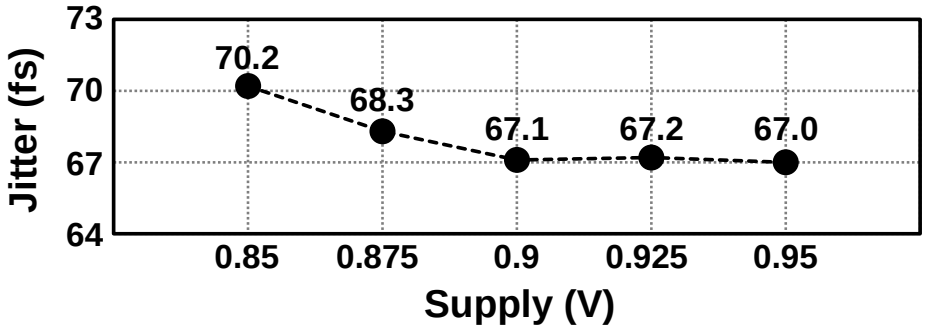


Figure 3.25 Measured PLL jitter across the supply voltages.

Figure 3.25 shows the variation of the PLL output jitter across the supply voltages from 0.85 to 0.95 V. The jitter variation was less than 5%. Figure 3.26 shows the jitter and reference spur measurement of five different chips. Note that the same supply, bias, and control codes are used for all chips. The jitter performance varied from 67.1 to 72.2 fs, while the spur performance measurement results range from −79.3 to −75.0 dBc.

	This Work	RFIC 16 Ravinuthula [41]	ISSCC 18 Turker [40]	ISSCC 18 Sharkia [11]	ISSCC 19 Kim [13]	ISSCC 19 Zhang [12]	ISSCC 18 Sharma [29]	ISSCC 20 Mercandelli [39]	ISSCC 16 Elkholy [17]
Type	**OSPLL**	Charge-pump PLL	Charge-pump PLL	Sub-sampling PLL	Sub-sampling PLL	Sub-sampling PLL	Sampling PLL	Sampling PLL	ILCM
	Type-II	Type-II	Type-II	Type-I	Type-II	Type-II	Type-I	Type-I	Type-I
Process (nm)	**28**	65	16	65	65	40	65	28	65
Reference (MHz)	**200**	500	500	100	100	200	50	500	125
F_{BW} (MHz)	**20**	0.2	3	6	2	4	2	7	1
F_{BW}/F_{REF}	**1/10**	1/2500	1/166	1/17	1/50	1/50	1/25	1/71	1/125
F_{OUT} (GHz)	**4**	8	12.5	5	3.8	14	2.55	12.5	8
Int. Jitter (fs)	**67.1 (69.3*)**	110	53.6	185.3	72	56.4	110	51.7	104.7
Int. Range (Hz)	**10K-100M**	10K-20M	10K-10M	10K-50M	1K-30M	1K-100M	10K-100M	1K-100M	10K-30M
In-band PN (dBc/Hz) (Normalized to 4GHz)	**-130**	-104	-127.1	-122.3	-123.6	N/A	-121.3	-129	-115.6
Ref. Spur (dBc)	**-78.1 (-77.5*)**	N/A	-75.5	-64.1	-75	-64.6	-63	-73.5	-43
Power (mW)	**5.2**	140	45	1.1	19.1	7.2	3.7	18	2.7
FOM (dB)	**-256.3**	-237.7	-246.8	-254	-250.1	-256.4	-253.5	-253.2	-255.4
Area (mm^2)	**0.17**	1.48	0.35	0.01	0.21	0.23	0.36	0.16	0.27

*Averaged across 5 chips

Table 3.1 Comparison with recent low-jitter PLLs.

Table 3.1 compares the performance of the proposed PLL with recent low-jitter integer-N PLLs. Thanks to the oversampling, the proposed PLL achieved the low in-band phase noise of −130 dBc/Hz. The measured bandwidth was 20 MHz, much larger compared with state-of-the-art PLLs. In combination with the low in-band phase noise and the wide bandwidth, the proposed PLL achieved a sub-100-fs low-jitter performance of 67.1 fsrms while maintaining excellent power efficiency, yielding a state-of-the-art jitter-power FoM of −256.3 dB. The proposed PLL demonstrated the best performance in terms of reference spur compared with the other PLLs. The performance comparison of the proposed PLL and recent published PLLs in terms of the reference spur performance versus the PLL FoM is shown in Figure 3.27. The proposed OSPLL achieved state-of-the-art performance in jitter, power, and spur among recent published work.

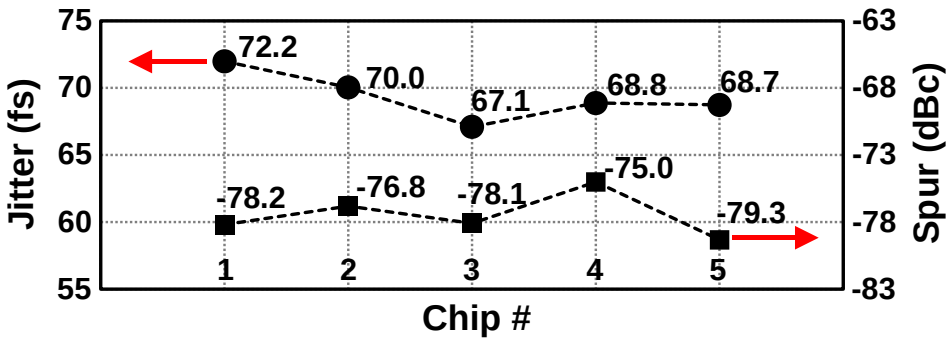


Figure 3.26 Measured PLL jitter and reference spur of five chips.

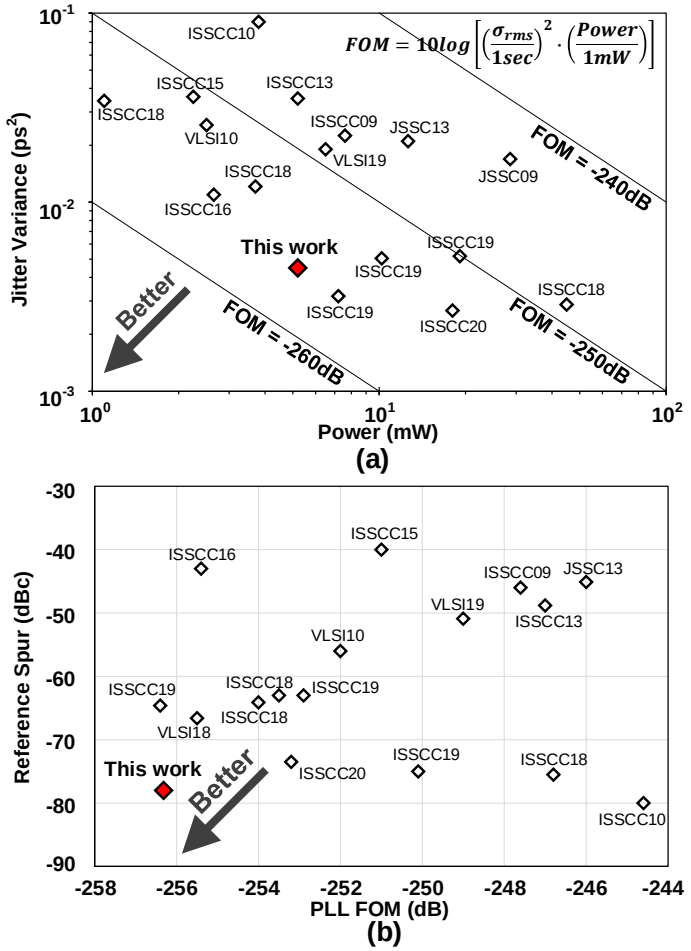


Figure 3.27 (a) Jitter variance versus power comparison. (b) Reference spur versus PLL FoM comparison plot of prior arts.

3.6 Conclusion

In this article, we presented an ultralow jitter LC-DCO-based reference oversampling PLL. The reference oversampling technique effectively boosts the reference frequency to the output frequency, lowering in-band phase noise and achieving a wide bandwidth at the same time. The ac coupling technique in the RSPD removes the effect of the offset and timing mismatch of the RSPD, enabling low reference spur. Since the sinewave reference from the crystal oscillator is directly utilized, the noisy and power-hungry reference buffer in conventional PLLs can be removed. The varactor-switching noise of the LC-DCO is minimized by the proposed DCO tuning pulse timing control scheme, therefore the optimal PLL jitter performance is maintained. The proposed PLL, fabricated in 28-nm CMOS, achieved 67.1-fsrms measured jitter performance and 5.2 mW of power consumption at 4 GHz, resulting in −256.3-dB PLL FoM, while maintaining −78.1-dBc reference spur.

Chapter 4 A 1.5 μW End-to-End Keyword Spotting SoC with Content-Adaptive Frame Sub-Sampling and Fast Settling Analog Frontend

4.1 Introduction

Speech is regarded as one of the most natural means to interface between humans and machines, and it is becoming increasingly important in the era of Internet of Things (IoT) for better user experience [28]. With the recent advances in deep learning, the accuracy of speech recognition has been greatly improved, and as a result, more and more devices are adopting deep learning-based speech recognition as their interface. Traditionally, due to the large processing workload and complexity associated with deep learning algorithms, speech recognition task has been primarily executed in the cloud [29]. In this scenario, users are required to upload their data to the central data center, which raises several issues. Firstly, as the number of the connected devices to the cloud is increasing, too much data will be generated from the connected edge devices, taking up significant portion of the bandwidth resources. Secondly, there is a security issue when users upload their private data to the cloud.

To address these issues, there has been a lot of effort recently in moving data processing from cloud to edge. One such application is keyword spotting (KWS) wake-up system on the edge, where KWS system continuously monitors the audio input looking for specific keywords, and activate the rest of the system or enable more complicated speech recognition task when it detects the keywords. Amazon Alexa and Apple Siri are commercialized examples where KWS system is utilized for the wake-up system.

Since the KWS system should be running continuously, its energy efficiency is of primary concern to maximize the battery life of the system. At the same time the accuracy of the KWS system is also important for the better user experience. However, designing the KWS system with lower power is not easy task due to the complexity of the neural network.

Recently, a lot of work dedicated to the KWS have been proposed, especially targeting low power consumption for running on energy constraint IoT edge devices. [30] can executes real-time KWS with 300 μW power, by using a compact memory storage along with a programmable processor and on-chip weight of 270 KB. SIMD processor proposed by [31] utilizes DNN to run a small vocabulary recognition task with 172 μW. [32] utilized binary neural network and depthwise convolution to achieve 510nW of power consumption. However, the power consumption of the existing works is not low enough for small IoT system, where the battery capacity is limited. For example, assuming the standard SR63 battery (5.5mm X 1.2mm) with 1.55V supply and 17mAh, the power consumption of the KWS should be lower than 2uW, to guarantee the battery life longer than two years.

In the actual KWS system running on edge IoT devices, one necessary component is the analog frontend (AFE). AFE receives incoming raw audio signal from the microphone, and converts it to the digital form so that the backend digital block can process it. However, many of

the previously proposed KWS system lacks the AFE. Since the microphone signal can be very small, and it is important to add as little noise and distortion to the audio signal as possible, the power consumption of the AFE can be large. For example, standard off-chip microphone pre-amplifiers usually consumes hundreds of μWs. Custom designed integrated amplifiers dissipate lower power but they also consume several μW to tens of μW. Therefore, even if digital part of the KWS system consumes little power, the analog part can dominate the system power, and it is important to optimize both the analog and the digital part of the system.

This work proposes a self-contained keyword spotting sensor system where analog frontend and the digital backend (feature extraction + neural network classifier) are included, targeting low power IoT system [37]. The proposed work adopted Skip RNN algorithm [33] to reduce the energy consumption of both the AFE and the digital backend. Using the skip signal from the Skip RNN, both the AFE and digital backend are adaptively power gated depending on the input signal, achieving 3x power reduction. The proposed AFE structure with the dc coupled first stage with the switched capacitive feedback resistor enables settling less than 1ms, achieving fast on/off. The proposed KWS system consumes 1.5 μW (reduced from 4.47 μW with an average of 76% frame skipping) in 28nm CMOS while achieving 92.8% accuracy on five-word GSCD KWS.

4.2 Skip RNN

This work adopted Skip RNN algorithm [33] to save the system energy in both analog and digital blocks. Skip RNN is kind of a dynamic neural network which can adaptively change its computational graph depending on the input. It can dynamically allocate fewer computing resources to unimportant parts of the input data to save computation energy. Specifically, Skip RNN generates a skip signal every time step, in addition to the classification result, to determine whether or not the next time step computation should be skipped. If skip signal is true, then the computation of the next time step can be saved.

Figure 4.1 compares the conventional RNN with the Skip RNN. Figure 4.1(a) is the conventional RNN, where the output of the network is fed back, becoming the next time step input. Figure 4.1(b) shows the conceptual block diagram of the skip RNN. It is consisted of the main RNN and the skip controller, which is also a form of neural network.

At every time step, the skip controller generates a probability of performing a state update at the next time step. When skip is 0, which is shown in Figure 4.1(c), the RNN operates in the same way as the conventional RNN. However, when skip is 1, which is shown in Figure 4.1(d), RNN skips its operation and simply feedforward its input, which is the output from the previous time step. In this way, skip RNN can save power. Figure 4.2 shows more detailed implementation of the Skip RNN. As shown in the figure, the skip operation is simply done with a mux, which is controlled by the skip controller. Note that the controller network can be trained using the backpropagation as opposed to the other dynamic neural network where reinforcement learning is required, and it takes long time to train.

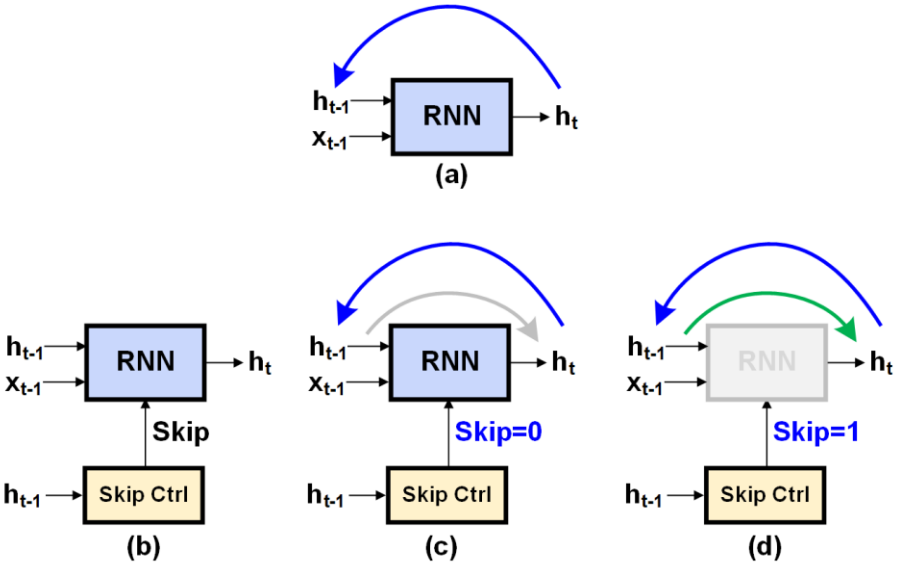


Figure 4.1 (a) Conventional RNN. (b) Skip RNN. (c) Skip RNN when skip signal is 0. (d) Skip RNN when skip signal is 1.

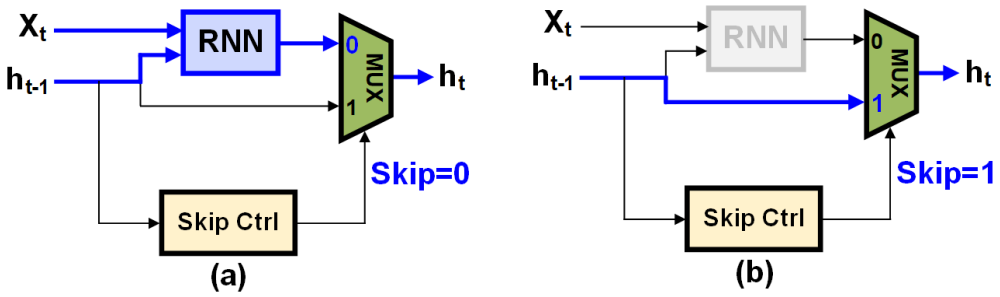


Figure 4.2 Detailed implementation of the Skip RNN. (a) when skip is 0. (b) when skip is 1.

4.3 Proposed KWS System

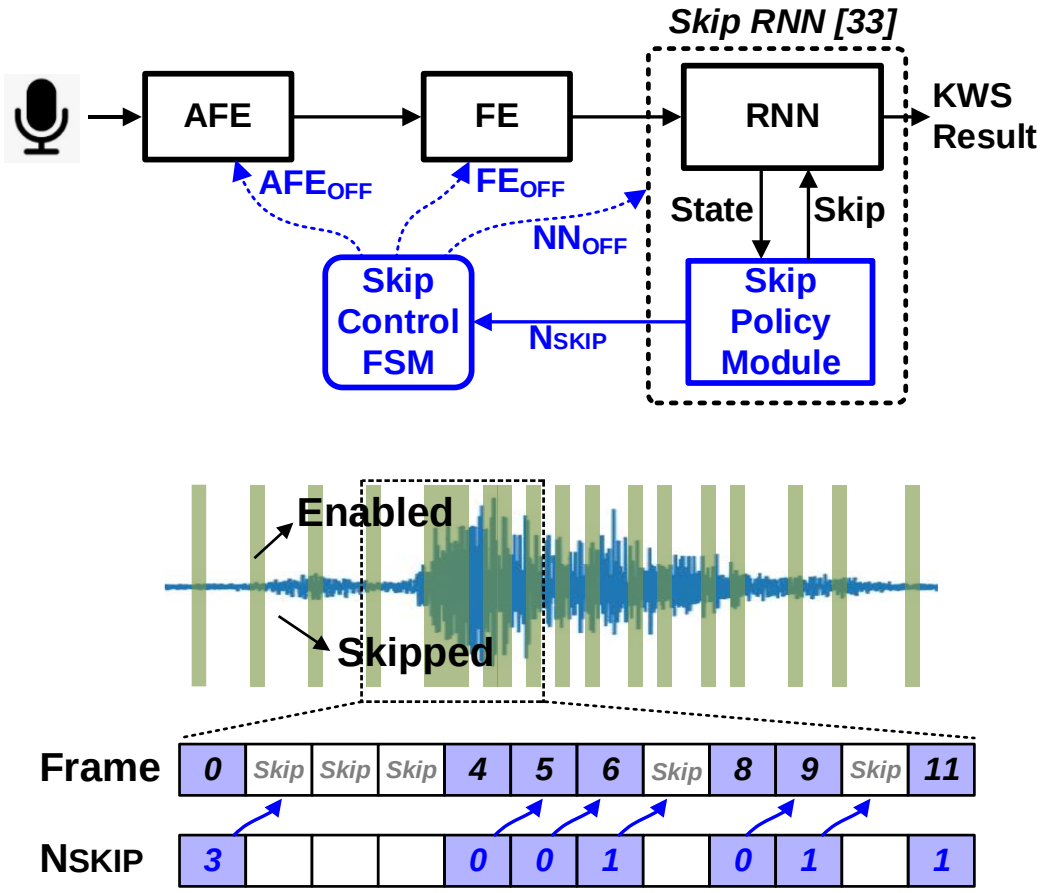


Figure 4.3 Proposed KWS system and conceptual operation.

Figure 4.3 shows the operating principle of the proposed KWS system. The AFE generates an audio frame (16 ms) from the input audio waveform, which is then converted by the FE into a log-Mel feature vector, fed to the RNN. Unlike a conventional RNN, where every incoming frame is processed and updates its state, the skip RNN is augmented with a skip policy module that adaptively determines whether the RNN state is updated or skipped. The skip policy module also produces an N_{SKIP} indicating how many frames to skip. The skip RNN is end-to-end

trained with the skip policy module by having it observe a sequence of frames to simultaneously learn both keyword classification and skip control. Following a decision from the skip policy module, we power down the system including AFE and FE for N_{SKIP} frames. Power down is controlled by the skip control FSM, which is always on.

4.4 Implementation Detail

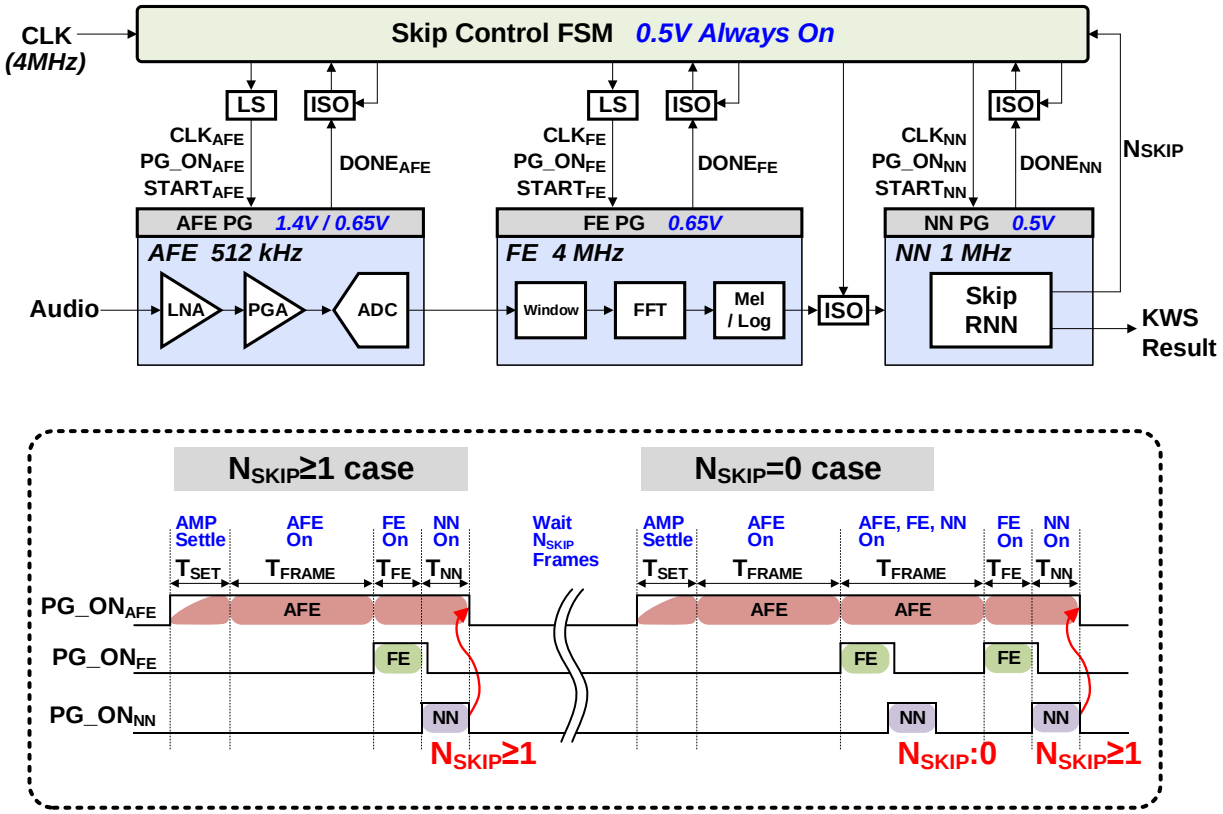


Figure 4.4 Detailed block diagram of proposed KWS SoC.

Figure 4.4 shows the overall block diagram of the proposed KWS system, comprised of an AFE, FE, NN, and skip control FSM. The skip control FSM provides power gate (PG) signals for each of the other three blocks, as well as each block's clock, start, and isolation control signal. Figure 4.4 also shows the detailed operation of the proposed system. KWS begins with PG_ON_{AFE} that turn on the AFE PG, followed by $START_{AFE}$ to enable the AFE control logic (not shown). The FSM inserts a programmable timing margin (up to 250 µs with 15.6 µs step) between AFE PG_ON and START for proper supply stabilization. After the amplifier settling time (T_{SET}), an ADC starts to sample the incoming signal. Upon conversion of one audio frame, the FE and the RNN process the data. The AFE therefore remains on for an additional $T_{FE} + T_{NN}$ to avoid losing any data from the next frame while the RNN processes the current frame. When re-enabling the AFE after one or more skipped frame, the FSM pre-starts the AFE for T_{SET} to settle it before the ADC is again enabled to digitize the new frame. Figure 4.5 shows the flow chart of the skip control FSM. In addition to T_{FRAME}, the AFE also consumes power during T_{SET}, T_{FE}, and T_{NN}, which represents an overhead (Figure 4.6). Therefore, to reduce AFE energy overhead, our design focuses on achieving fast settling time for the AFE, and short operation times for the FE and RNN.

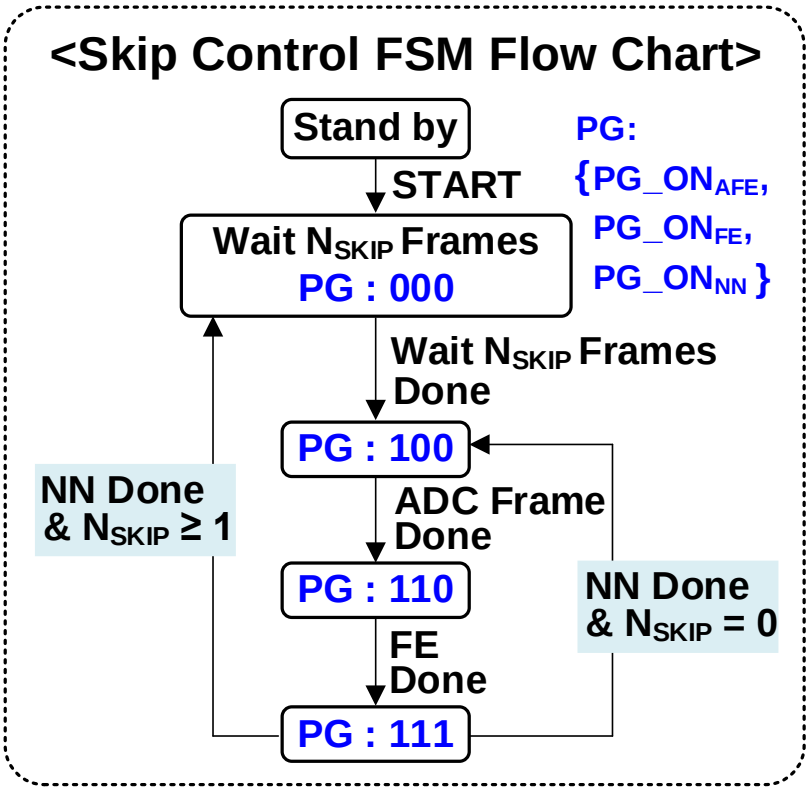


Figure 4.5 Flow Chart of Skip Control FSM.

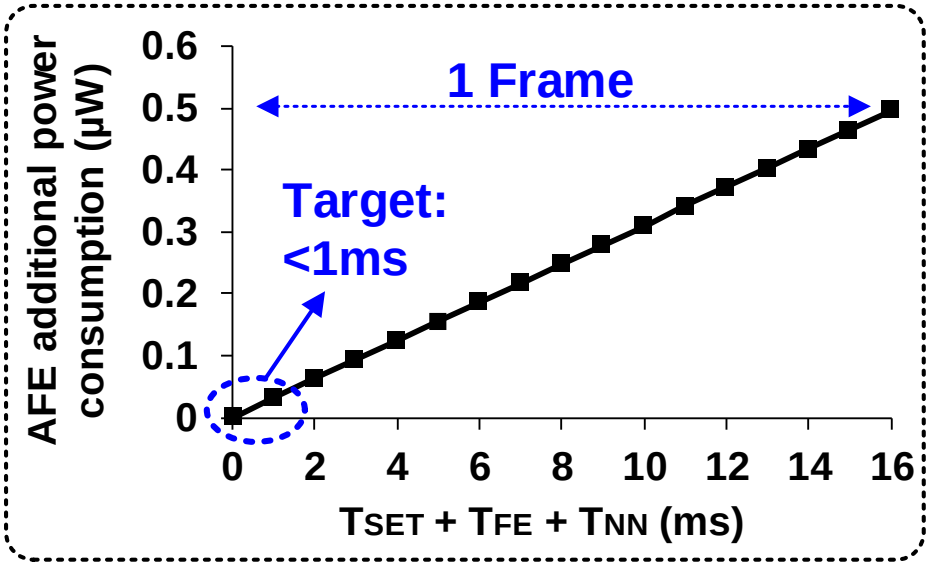


Figure 4.6 AFE additional power according to $T_{SET}+T_{FE}+T_{NN}$.

4.4.1 Analog Front End

Figure 4.7 shows the proposed fast settling AFE. A capacitive feedback (CF) amplifier is attractive for its low power consumption but has a high-pass corner (f_{HP}) that is inherent low, resulting in a long settling time (e.g., 37ms when f_{HP}=50Hz), which is comparable to T_{FRAME}. We therefore designed the first stage LNA with a DC-coupled Gm-ratio structure [36], as shown in Figure 4.8, while the second stage PGA uses a CF structure. Using a DC-coupled structure for the first stage reduces settling time by 54% due to its lack of a high-pass corner. To improve the settling time of the CF PGA, its f_{HP} must be set as high as possible, without filtering out voice content that would degrade KWS accuracy. Based on software KWS simulations (Figure 4.9), fHP can be increased to 140Hz (compared to a conventional value of < 50Hz) but must be carefully controlled across PVT conditions to avoid inadvertent excursions into higher frequencies that would impact KWS accuracy. Traditionally, the corner in a CF amp is set using a "pseudo-resistor", but its high PVT sensitivity (> 10×) cannot achieve the required corner accuracy. Instead, we employ a switched capacitor resistor with equivalent resistance set by its capacitance and the switching clock ($\Phi_{0,1}$), which are insensitive to PVT. Further, we also adopt a two-step frequency control, where $\Phi_{0,1}$ operates momentarily at a higher frequency to achieve fast settling before switching to its final value (Figure 4.10), reducing T_{SET} by 6×. By combining all techniques (Gm first stage, increased f_{HP}, and two-step frequency control), settling time is reduced from 37 ms to <1 ms, yielding a 1.1 µW average power reduction.

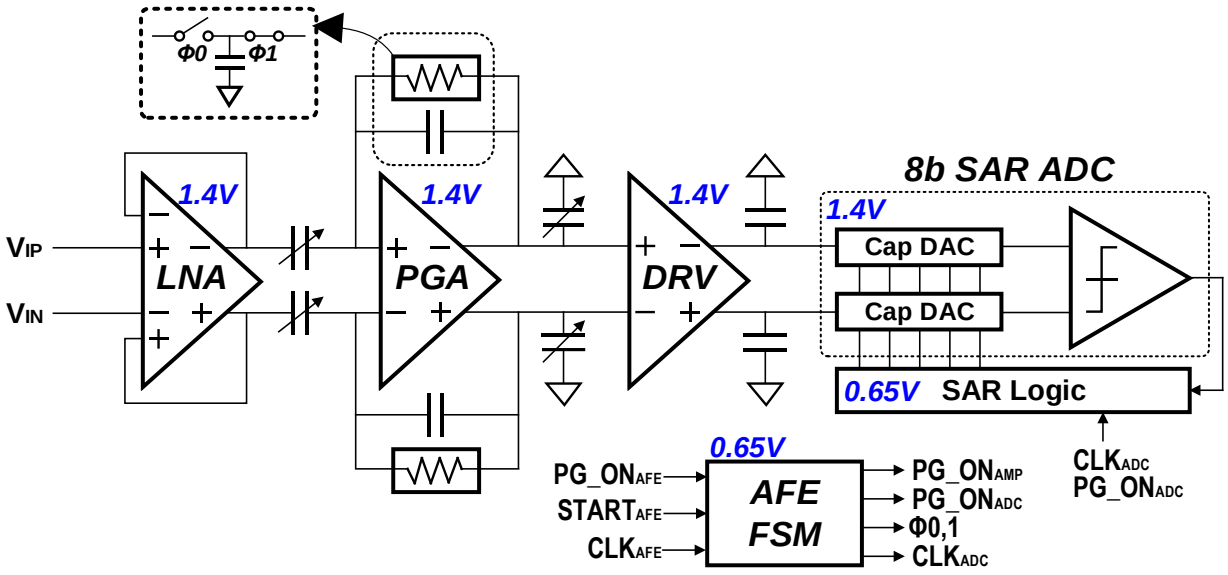


Figure 4.7 Proposed fast settling analog frontend.

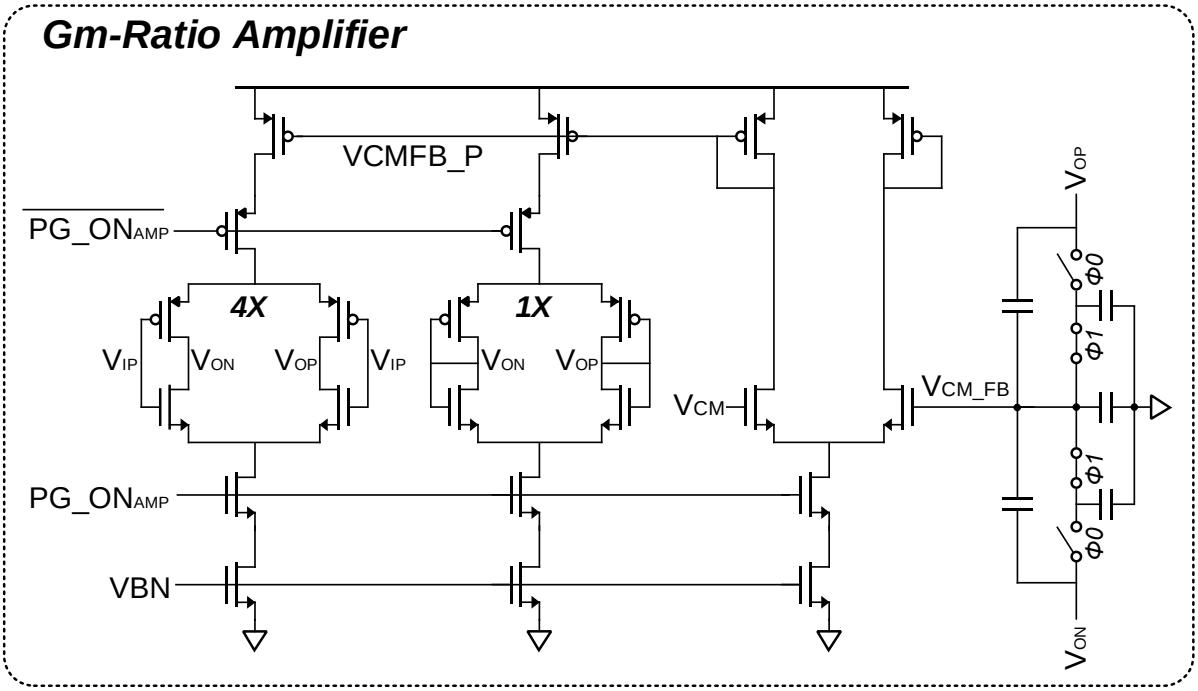


Figure 4.8 Structure of DC-coupled Gm-ratio amplifier.

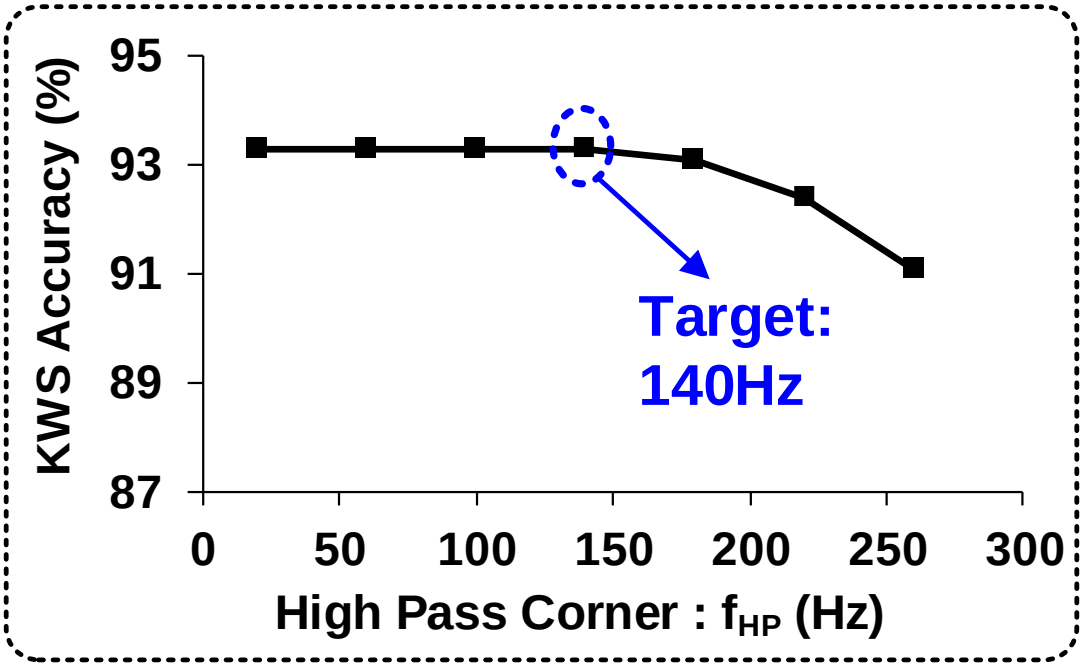


Figure 4.9 Simulated KWS accuracy according to high pass corner.

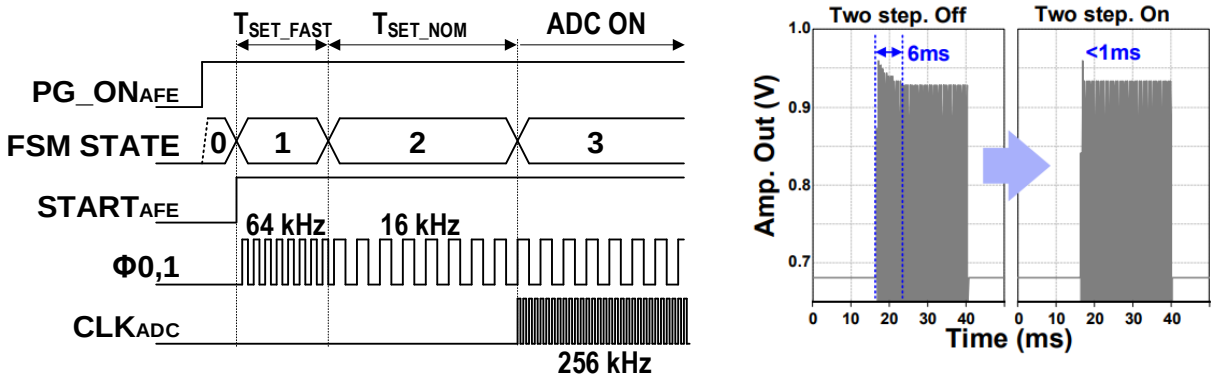


Figure 4.10 Proposed two step frequency control scheme.

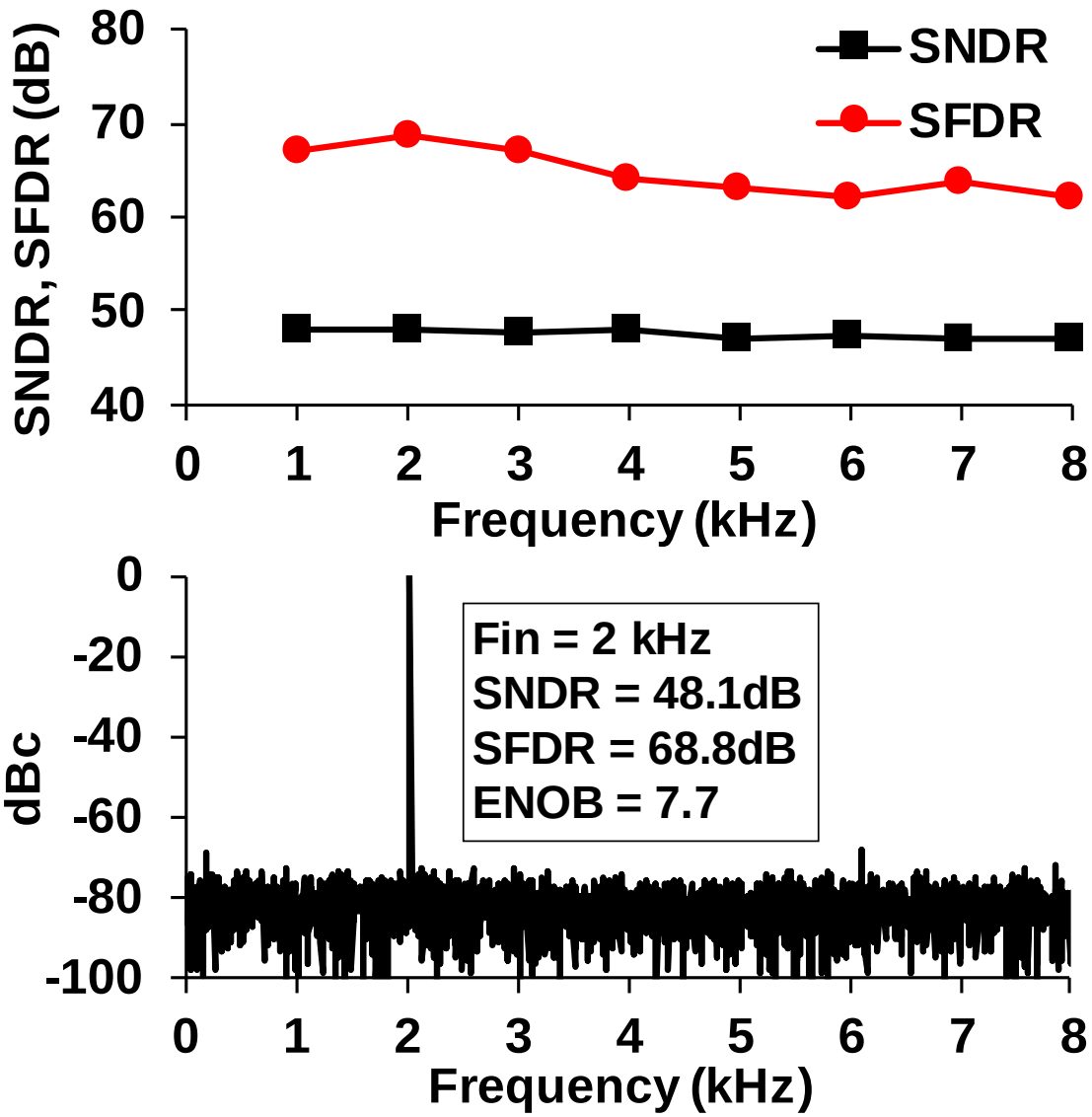


Figure 4.11 Measured AFE performance.

The ADC is an 8-bit synchronous SAR architecture with 16 kS/s sampling frequency. The amplifiers, the ADCs capacitor DAC, and the ADC comparator operate using 1.4V while the AFE FSM and the SAR logic operate at 0.65V. When PG_ON_{AFE} goes low, all AFE blocks except the capacitor DAC are power gated. Figure 4.11 details the measured AFE performance.

4.4.2 Feature Extractor

The FE, shown in Figure 4.12, consists of windowing, FFT, Mel filter, and log units. A 16ms 256-point 8b nonoverlapping audio frame is multiplied by the Hanning window and sent to the FFT module and further processed by power calculation, Mel filter, and log module to generate an 8b 26-dimension feature vector. Because the power dominating FFT must process non-consecutive frames, it cannot use a conventional pipelined architecture. We therefore employ a memory-based fully-folded structure with a single butterfly unit (BFLY, Figure 4.13). Instead of a conventional single bank dual port SRAM we use four banks of a single port SRAM, reducing memory power by 80% without reducing the data access rate. One BFLY calculation requires 2 cycles, where one additional cycle resamples the BFLY input data to reduce glitch power, saving 22% BFLY power. The architecture requires 1152 cycles to finish one FFT, and overlaps the window, FFT power, Mel, and log operations to reduce overall operating time, as shown in Figure 4.14. The FE sprints to complete its computation quickly (290 μs) to reduce AFE power overhead and is then power gated to minimize its leakage power. The FE sprinting frequency is set at 4MHz, which marks the point of diminishing returns as seen in Figure 4.15.

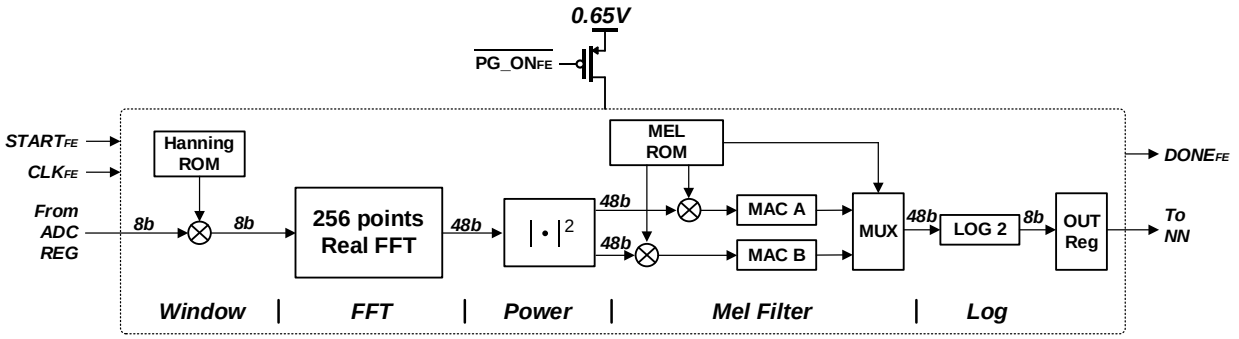


Figure 4.12 Structure of the feature extraction block.

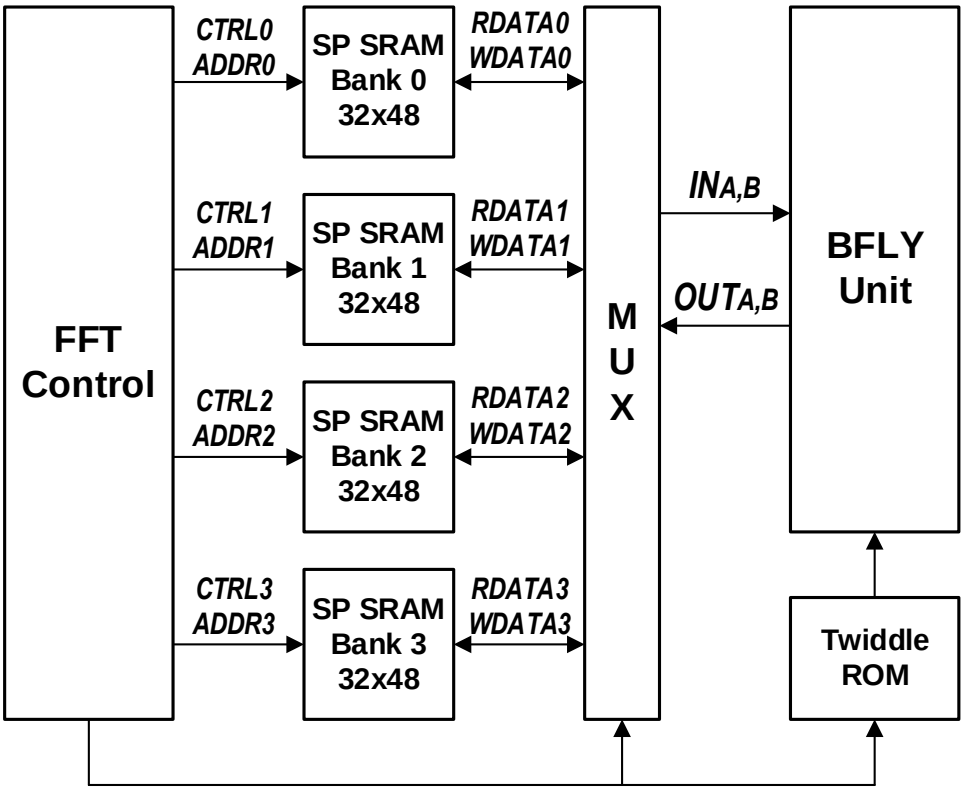


Figure 4.13 Detailed structure of the FFT.

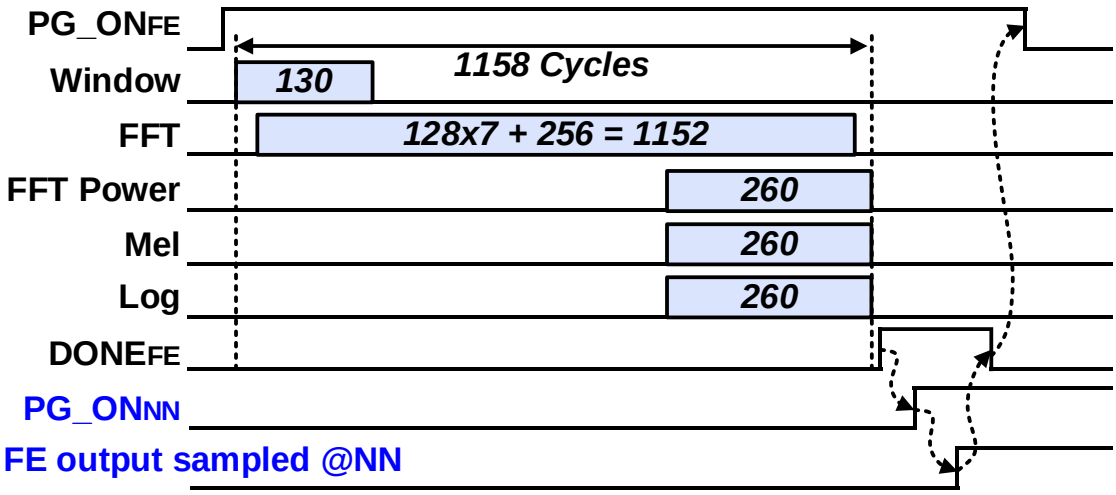


Figure 4.14 Timing diagram of the FFT.

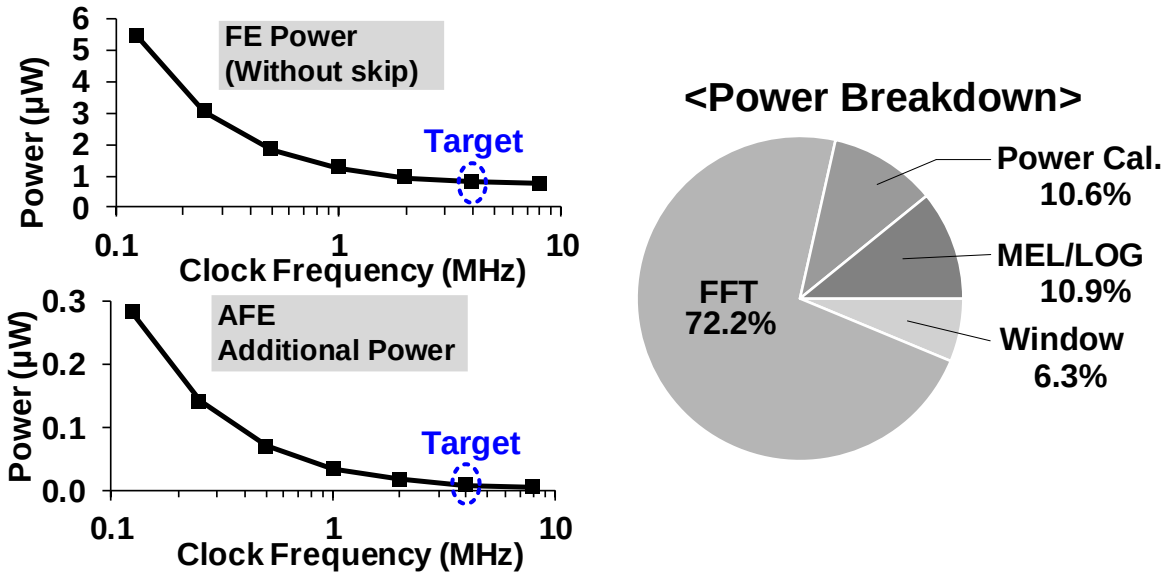


Figure 4.15 Power saving with sprint and power breakdown of the FFT.

4.4.3 Neural Network Accelerator

Figure 4.16 shows the structure of the proposed skip RNN accelerator. The FE feeds the 26-dimension features to the 64 hidden nodes of the RNN. Each RNN cell is designed with a gated recurrent unit (GRU) (Figure 4.17). The 64 hidden states of the RNN are processed by the fully connected (FC) layer with eight output nodes. Seven of these FC nodes generate class outputs (5 keywords, one non-keywords, one noise). The last FC node produces the skip score for the skip policy module, which accumulates the skip score. When the accumulator overflows, it latches a counter to produce N_{SKIP}. Figure 4.18 shows the detailed implementation of the skip RNN accelerator. The RNN accelerator consists of a MAC array, WMEM, functional units (FU), register bank, and a NN FSM. The MAC array is designed with 64 parallel MACs. We adopt an

output stationary dataflow for the MAC array, where the partial sums for the 64 GRU gate vectors are accumulated in the 64 MACs until the processing of each GRU gate vector is completed. For the FC layer, 8 MACs process the 8 FC output nodes in parallel. All activations are scaled and quantized to 12 bits. The always-on WMEM stores the 8-bit weights for the RNN model and is custom designed using latches with high Vth transistors, achieving 20× leakage power reduction compared to the foundry SRAM (Figure 4.19). After each frame, the RNN hidden states are stored in WMEM before the NN block is power gated. The operation of the MAC array, FU, and the register bank are scheduled to overlap their operations reducing a single frame processing time to 495 cycles (Figure 4.20). The RNN sprints with a 1MHz clock to reduce the leakage power and optimize overall energy efficiency.

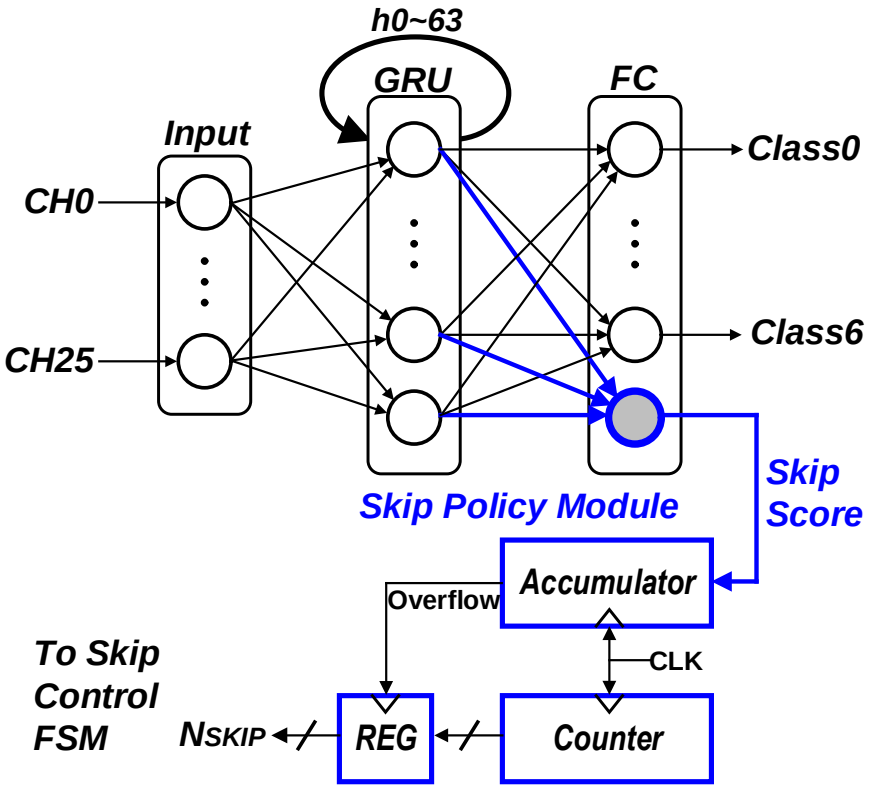


Figure 4.16 Structure of the skip RNN accelerator.

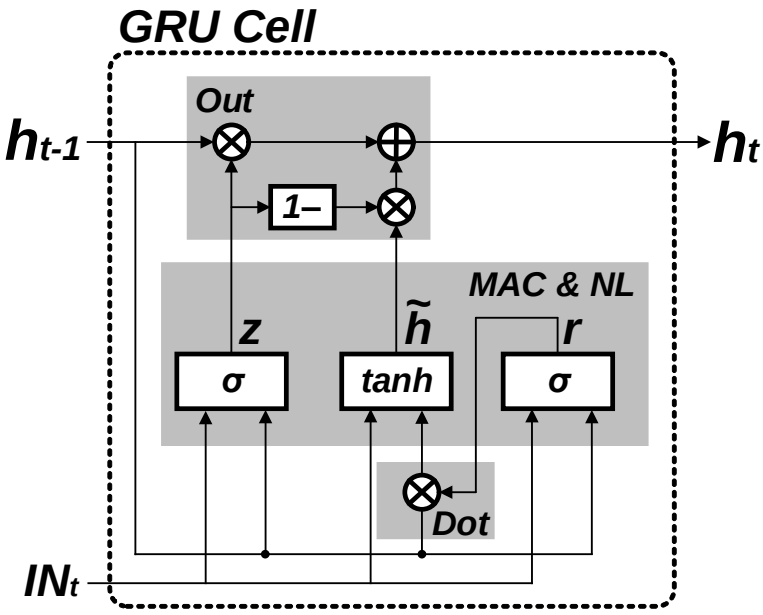


Figure 4.17 Conceptual diagram of GRU cell.

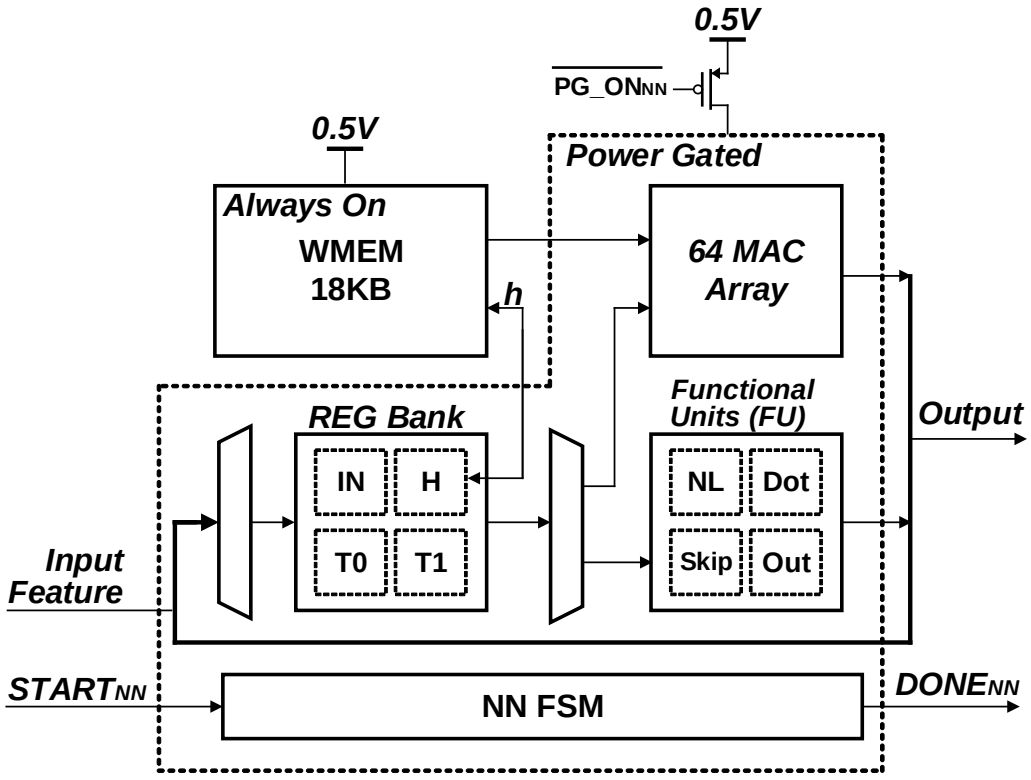


Figure 4.18 Detailed structure of the proposed skip RNN accelerator.

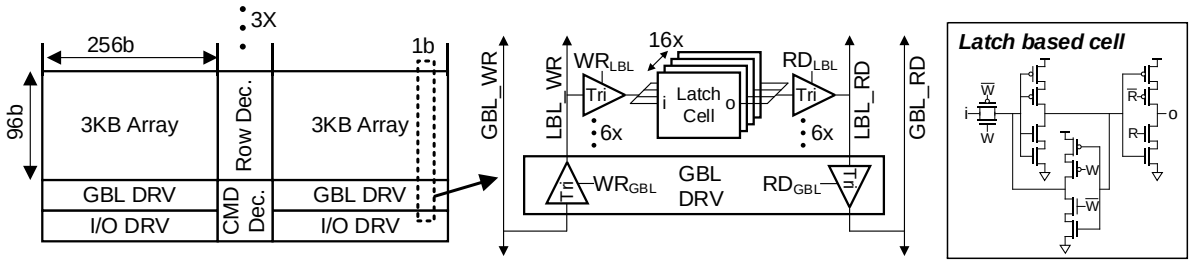


Figure 4.19 Structure of proposed weight memory.

Time

NN #Phase / Operation		0 : r		1 : z		2 : $\tilde{h}$		3 : Out	4 : FC	5 : Skip
#Cycle		95		92		92		65	84	67
Processing Unit (Dest. Reg.)	MAC	T0		T0		T0			T0	
	NL			T1		T1		Out		
	Dot					MAC				
	Out							H		
	Skip									Done
Register (Dest. P.U.)	IN		MAC		MAC		MAC			
	H	MAC		MAC		Dot		Out	MAC	
	T0			NL		NL		NL		Skip
	T1					Dot		Out		

Figure 4.20 Scheduling table of RNN accelerator.

4.5 Measurement Results

The proposed chip is fabricated in 28nm CMOS (Figure 4.21). We trained the model with different skip ratios and tested KWS accuracy (Figure 4.22), where the GCSD dataset with 5 keywords task is used for the training and testing. Skip ratio can be changed using a single hyperparameter of the loss function [33]. The tested accuracy drop is less than 1% at the skip ratio used for the deployed model (=0.76). Figure 4.23 compares the measured power consumption of each block of the proposed KWS system when the skip ratio is 0 and 0.76. Measured power reduces from 4.47uW to 1.48uW, achieving 3× power reduction. Figure 4.24 shows the measured amplifier output and frame enable control for an example audio signal, showing the adaptive enable/skip pattern according to input data. Figure 4.24 also shows the measured dynamic power of the AFE, FE, and NN.

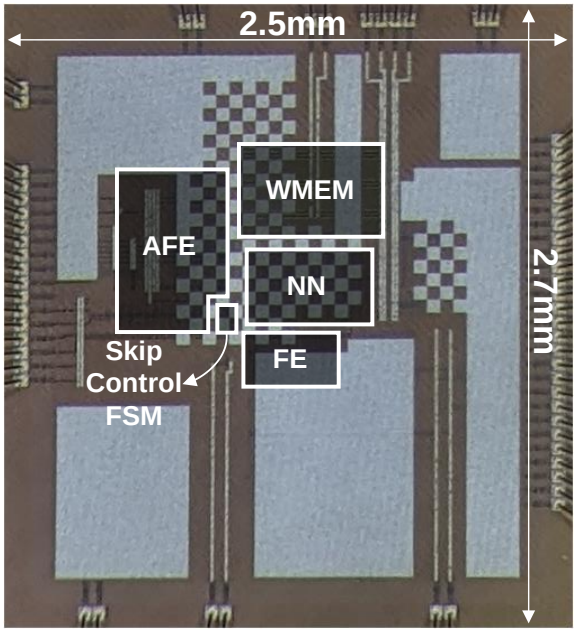


Figure 4.21 Chip photograph.

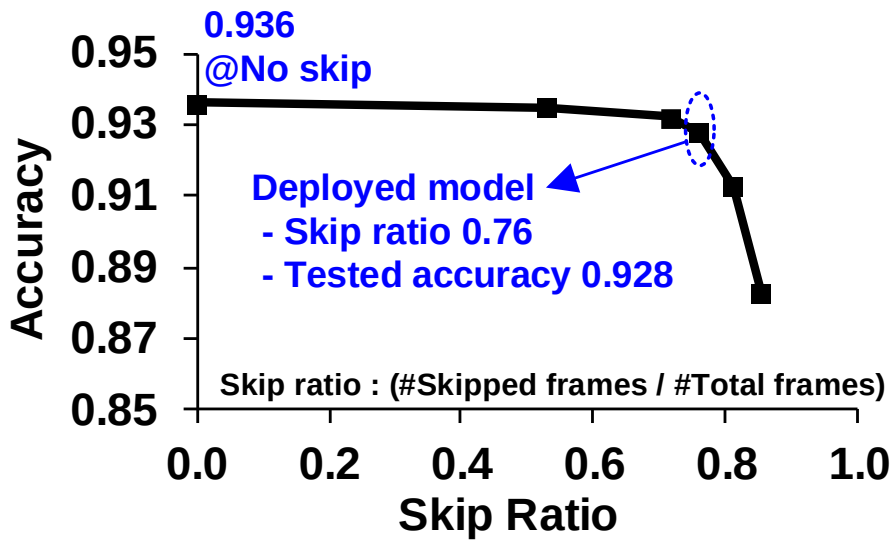


Figure 4.22 Measured accuracy of proposed KWS according to different skip ratio.

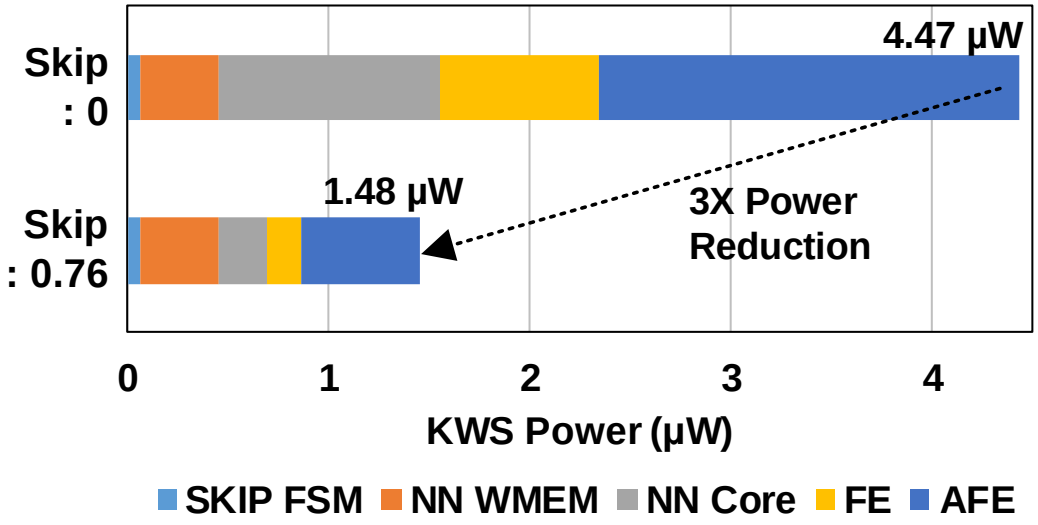


Figure 4.23 Power breakdown of proposed KWS with and without skip operation.

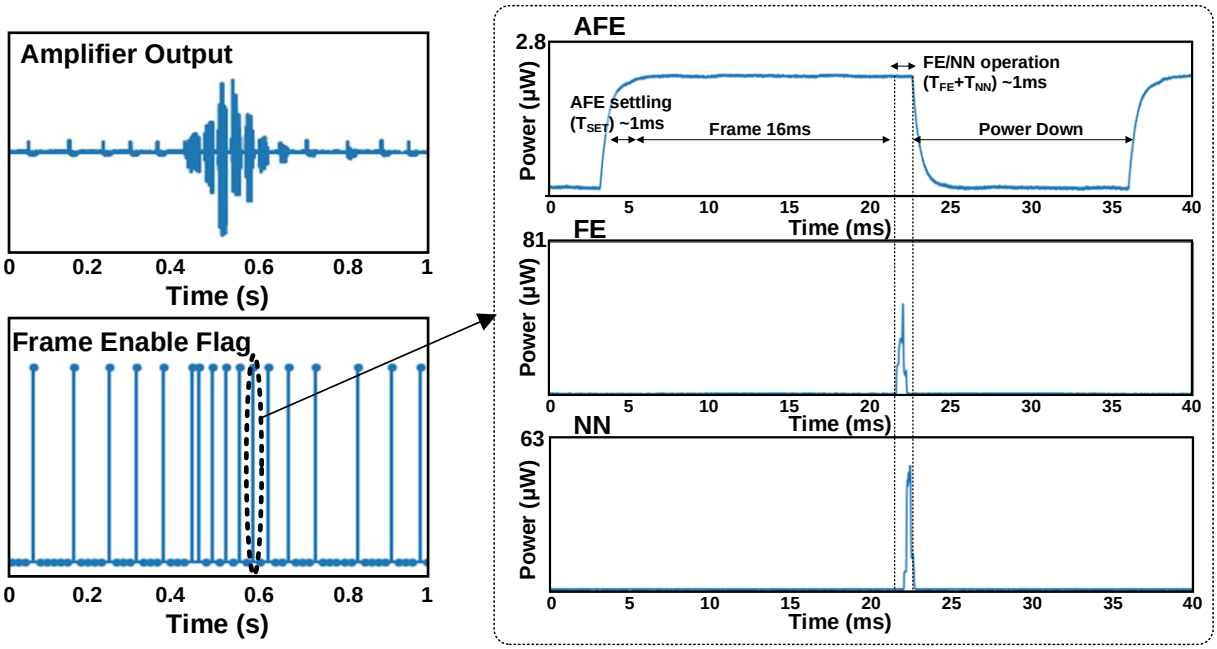


Figure 4.24 Measured waveform of proposed KWS.

4.6 Conclusion

Table 4.1 provides a comparison table of recently published KWS chips. The proposed work achieves 92.8% 7 keywords KWS accuracy with GSCD dataset. The power consumption of the AFE, FE, and NN is 0.59μW, 0.18μW, and 0.71μW, achieving 1.48μW total power consumption. The proposed design is the lowest power KWS system that fully integrates AFE, FE, and NN.

		This work	Kim ISSCC 2022	Shan ISSCC 2020	Giraldo VLSI 2019	Giraldo ESSCIRC 2018
Tech (nm)		**28**	65	28	65	65
Area (mm^2)		**0.8**	2.03	0.23	2.56	1.04
Algorithm		**Skip RNN**	RNN	DSCNN	RNN	RNN
Memory (KB)		**18**	27	2	105	32
Latency (ms)		**16**	12.4	64	16	16
Analog Frontend		**Yes**	Yes	No	Yes	No
Feature Extraction		**Yes**	Yes	Yes	Yes	No
Neural Network		**Yes**	Yes	Yes	Yes	Yes
Dataset		**GSCD**	GSCD	GSCD	GSCD	N/A
# Classes (Keywords)		**7 (5)**	12 (10)	3 (2)	12 (10)	4 (4)
Accuracy (%)		**92.8**	86.0	94.6	90.9	91.2
Power (μW)	**AFE**	**0.59**	13	-	1.8	-
	FE	**0.18**		0.34	7.1	-
	NN	**0.71**	10	0.17	3.4	5
	Total	**1.48**	23	-	16.1	-

Table 4.1 Comparison with existing KWS system for edge devices.

Chapter 5 Conclusion

5.1 Summary of Contributions

This dissertation proposes energy efficient mixed signal circuits and systems for two different area: communication and signal processing. First, low power, low jitter phase locked loops are proposed which is essential parts in emerging communication systems such as 5G wireless system and +100Gbps PAM4 wireline transceivers. Secondly, an energy efficient keyword spotting sensor is proposed which is becoming crucial for edge IoT devices.

In chapter 2, a ring oscillator-based reference oversampling phase-locked loop is introduced. The proposed reference oversampling technique achieves ultra-wide bandwidth which can even exceed the 1/10 bandwidth limit, therefore suppressing ring oscillator phase noise. Due to the oversampling of the reference clock, the wide bandwidth is achieved with a low reference spur. The proposed phase locked loop achieves 240.3 dB Figure of Merit (FOM) and 80 dBc reference spur. The integrated jitter is 508 fs_{RMS} and the power consumption is 3.6 mW at 2 GHz output clock frequency.

In chapter 3, An ultra-low jitter LC oscillator-based reference oversampling PLL is proposed. For the generation of ultra-low jitter with low power, a low in-band phase noise and wide bandwidth are necessary. The proposed reference oversampling technique simultaneously suppresses the in-band noise from the phase detector and suppresses the oscillator phase noise with wide bandwidth. An LC digitally controlled oscillator (DCO) with proportional path gain

profile control and DCO tuning pulse timing control scheme is also proposed for optimal phase noise. Designed in a 28-nm CMOS process, the proposed PLL achieves 67.1-fs RMS-jitter with 5.2-mW power consumption, which translates to −256-dB FoM. The in-band phase noise of the PLL is −130 dBc/Hz at the 4-GHz output frequency. The reference spur of the proposed PLL is −78 dBc.

In chapter 4, an energy efficient keyword spotting sensor system with a dynamic neural network algorithm is introduced to minimize the power consumption of both the analog frontend (AFE) and the digital backend at the same time. The NN adaptively power gates the entire system depending on its input. A fast-settling AFE is proposed to maximize the AFE energy saving from dynamic power gating. Energy optimized SRAM bank structure for the feature extraction block, and also a low leakage custom weight memory for the neural network is proposed. This work achieves 1.5μW power consumption for the KWS task, making it possible to be used for small IoT systems.

5.2 Future Directions

There are many opportunities to improve the works presented in this dissertation. The reference oversampling PLL presented in chapters 2,3 are based on an integer-N divider control scheme. Even though the integer-N based PLLs are used in many important applications such as microprocessors, ADCs, and wireline transceivers, fractional-N PLL is necessary for RF applications. Therefore, the implementation of the fractional-N based reference oversampling PLL can be an important future direction. The large size of the reference oversampling phase detector is one drawback, and it also requires further improvement. Replacement of the analog charge pump based offset cancellation loop to the digital control loop can reduce the area. In

addition, the reference oversampling PLL presented in this work has a low frequency high pass corner due to the AC coupling path in the phase detector, resulting in slow locking time, which needs to be addressed. For the keyword spotting sensor presented in chapter 4, the leakage power from the always-on weight memory is the limiting factor that prevents further power improvement from the dynamic power gating. Further optimization of the leakage power of the weight memory can maximize the benefit of the proposed adaptive frame sub-sampling technique.

BIBLIOGRAPHY

[1] David Rydning, John Reinsel, John Gantz, "The digitization of the world from edge to core." Framingham: International Data Corporation (2018): 16.

[2] Y. Liu, X. Wei, J. Xiao, Z. Liu, and Y. Tian, "Energy consumption and emission mitigation prediction based on data center traffic and PUE for global data centers," Global Energy Interconnection, vol. 3, no. 3, pp.272-282, 2020.

[3] F. Tariq, M. R. A. Khandaker, K. -K. Wong, M. A. Imran, M. Bennis and M. Debbah, "A Speculative Study on 6G," in IEEE Wireless Communications, vol. 27, no. 4, pp. 118-125, August 2020.

[4] Andrae, Anders S.G., and Tomas Edler. 2015. "On Global Electricity Usage of Communication Technology: Trends to 2030" Challenges 6, no. 1: 117-157.

[5] Y. Hu et al., "17.6 A 21.7-to-26.5GHz Charge-Sharing Locking Quadrature PLL with Implicit Digital Frequency-Tracking Loop Achieving 75fs Jitter and −250dB FoM," 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 276-278.

[6] B. Razavi, "Jitter-Power Trade-Offs in PLLs," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 4, pp. 1381-1387, April 2021.

[7] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," in IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, Oct. 2016.

[8] L. Kong and B. Razavi, “A 2.4 GHz 4 mW integer-N inductorless RF synthesizer,” IEEE J. Solid-State Circuits, vol. 51, no. 3, pp. 626–635, Mar. 2016.

[9] B. M. Helal, M. Z. Straayer, G. Wei and M. H. Perrott, "A Highly Digital MDLL-Based Clock Multiplier That Leverages a Self-Scrambling Time-to-Digital Converter to Achieve Subpicosecond Jitter Performance," in IEEE Journal of Solid-State Circuits, vol. 43, no. 4, pp. 855-863, April 2008.

[10] A. Elkholy, D. Coombs, R. K. Nandwana, A. Elmallah, and P. K. Hanumolu, “A 2.5–5.75-GHz ring-based injection-locked clock multiplier with background-calibrated reference frequency doubler,” IEEE J. Solid-State Circuits, vol. 54, no. 7, pp. 2049–2058, Jul. 2019.

[11] T. Seong, Y. Lee, S. Yoo and J. Choi, "A −242-dB FOM and −71-dBc reference spur ring-VCO-based ultra-low-jitter switched-loop-filter PLL using a fast phase-error correction technique," 2017 Symposium on VLSI Circuits, 2017, pp. C186-C187.

[12] J. Seol, D. Sylvester, D. Blaauw and T. Jang, "A Reference Oversampling Digital Phase-Locked Loop with -240 dB FOM and -80 dBc Reference Spur," in Proc. IEEE Symp. VLSI Circuits, Jun. 2019, pp. C160-C161.

[13] X. Gao, E. A. M. Klumperink, M. Bohsali, and B. Nauta, "A low noise sub-sampling PLL in which divider noise is eliminated and PD/CP noise is not multiplied by N2," IEEE J. Solid-State Circuits, vol. 44, no. 12, pp. 3253–3263, Dec. 2009.

[14] A. Elkholy et al., "A 3.7 mW low-noise wide-bandwidth 4.5 GHz digital fractional-N PLL using time-amplifier-based TDC," IEEE J. Solid-State Circuits, vol. 50, no. 4, pp. 867–880, Apr. 2015.

[15] S. Levantino, G. Marzin, C. Samori, and A. Lacaita, "A wideband fractional-N PLL with suppressed charge-pump noise and automatic loop filter calibration," IEEE J. Solid-State Circuits, vol. 48, no. 10, pp. 2419–2429, Oct. 2013.

[16] C. Ko, T. Kuan, R. Shen and C. Chang, "A 7-nm FinFET CMOS PLL With 388-fs Jitter and −80-dBc Reference Spur Featuring a Track-and-Hold Charge Pump and Automatic Loop Gain Control," IEEE J. Solid-State Circuits, vol. 55, no. 4, pp. 1043-1050, April 2020.

[17] F. M. Gardner, Phaselock Techniques. Hoboken, NJ, USA: Wiley, 2005.

[18] Z. Zhang, G. Zhu, and C. P. Yue, "A 0.65 V 12-to-16 GHz sub-sampling PLL with 56.4 fsrms integrated jitter and −256.4 dB FoM," in IEEE ISSCC Dig. Tech. Papers, Feb. 2019, pp. 488–489.

[19] T.-K. Kuan and S.-I. Liu, "A bang bang phase-locked loop using automatic loop gain control and loop latency reduction techniques," IEEE J. Solid-State Circuits, vol. 51, no. 4, pp. 821–831, Apr. 2015.

[20] D. Park and S. Cho, "A 14.2 mW 2.55-to-3 GHz cascaded PLL with reference injection and 800 MHz delta-sigma modulator in 0.13 μm CMOS," IEEE J. Solid-State Circuits, vol. 47, no. 12, pp. 2989–2998, Dec. 2012.

[21] K. M. Megawer, A. Elkholy, M. G. Ahmed, A. Elmallah, and P. K. Hanumolu, "Design of crystal-oscillator frequency quadrupler for low-jitter clock multipliers," IEEE J. Solid-State Circuits, vol. 54, no. 1, pp. 65–74, Jan. 2019.

[22] F. Song et al., "A fractional-N synthesizer with 110 fsrms jitter and a reference quadrupler for wideband 802.11ax," in IEEE ISSCC Dig. Tech. Papers, Feb. 2019, pp. 264–265.

[23] J. Seol, K. Choo, D. Blaauw, D. Sylvester and T. Jang, "A 67-fsrms Jitter, −130 dBc/Hz In-Band Phase Noise, −256-dB FoM Reference Oversampling Digital PLL With Proportional Path Timing Control," IEEE Solid-State Circuits Letters, Invited Paper to the Special Issue on ASSCC 2020, Vol. 3, September 2020, pg. 430-433.

[24] J. Seol, K. Choo, D. Blaauw, D. Sylvester and T. Jang, "Reference Oversampling PLL Achieving −256-dB FoM and −78-dBc Reference Spur," in IEEE Journal of Solid-State Circuits, vol. 56, no. 10, pp. 2993-3007, Oct. 2021.

[25] Z. Xu, M. Miyahara, K. Okada, and A. Matsuzawa, "A 3.6 GHz low noise fractional-N digital PLL using SAR-ADC-based TDC," IEEE J. Solid-State Circuits, vol. 51, no. 10, pp. 2345–2356, Oct. 2016.

[26] R. B. Staszewski, D. Leipold, K. Muhammad, and P. T. Balsara, "Digitally controlled oscillator (DCO)-based architecture for RF frequency synthesis in a deep-submicrometer CMOS process," IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 50, no. 11, pp. 815–828, Nov. 2003.

[27] J. Lee, M. S. Keel, S. I. Lim, and S. Kim, "Charge pump with perfect current matching characteristics in phase-locked loops," Electron. Lett., vol. 36, no. 23, pp. 1907–1908, Nov. 2000.

[28] S. Oh et al., "An Acoustic Signal Processing Chip With 142-nW Voice Activity Detection Using Mixer-Based Sequential Frequency Scanning and Neural Network Classification," in IEEE Journal of Solid-State Circuits, vol. 54, no. 11, pp. 3005-3016, Nov. 2019.

[29] Giraldo, J. S. P., and Marian Verhelst. "Hardware acceleration for embedded keyword spotting: Tutorial and survey." ACM Transactions on Embedded Computing Systems (TECS) 20.6 (2021): 1-25.

[30] S. Bang et al., "14.7 A 288μW programmable deep-learning processor with 270KB on-chip weight storage using non-uniform memory hierarchy for mobile intelligence," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp. 250-251.

[31] M. Price, J. Glass and A. P. Chandrakasan, "A Low-Power Speech Recognizer and Voice Activity Detector Using Deep Neural Networks," in IEEE Journal of Solid-State Circuits, vol. 53, no. 1, pp. 66-75, Jan. 2018.

[32] W. Shan et al., "A 510-nW Wake-Up Keyword-Spotting Chip Using Serial-FFT-Based MFCC and Binarized Depthwise Separable CNN in 28-nm CMOS," in IEEE Journal of Solid-State Circuits, vol. 56, no. 1, pp. 151-164, Jan. 2021.

[33] Victor Campos, Brendan Jou, Xavier Giro-I-Nieto, Jordi Torres, and Shih Fu Chang. Skip RNN: Learning to skip state updates inrecurrent neural networks. In ICLR, 2018.

[34] R. Rothe et al., "A Delta Sigma-Modulated Sample and Average Common-Mode Feedback Technique for Capacitively Coupled Amplifiers in a 192-nW Acoustic Analog Front-End," in IEEE Journal of Solid-State Circuits, vol. 57, no. 4, pp. 1138-1152, April 2022.

[35] J. Giraldo, et al., "18μW SoC for Near-Microphone Keyword Spotting and Speaker Verification," IEEE Symp. VLSI Circuits, pp. 52-53, 2019.

[36] J. Lim, et al., "A 0.19 0.17 mm2 Wireless Neural Recording IC for Motor Prediction with Near Infrared-Based Power and Data Telemetry," ISSCC, pp. 416–418, Feb. 2020.

[37] J. Seol, et al., "A 1.5 μW End-to-End Keyword Spotting SoC with Content-Adaptive Frame Sub-Sampling and Fast Settling Analog Frontend," ISSCC 2023, Accepted.

[38] W. El-Halwagy, A. Nag, P. Hisayasu, F. Aryanfar, P. Mousavi, and M. Hossain, "A 28-GHz quadrature fractional-N frequency synthesizer for 5G transceivers with less than 100-fs jitter based on cascaded PLL architecture," IEEE Trans. Microw. Theory Techn., vol. 65, no. 2, pp. 396–413, Feb. 2017.

[39] W. Wu et al., "A 28-nm 75-fsrms analog fractional-N sampling PLL with a highly linear DTC incorporating background DTC gain calibration and reference clock duty cycle correction," IEEE J. Solid-State Circuits, vol. 54, no. 5, pp. 1254–1265, May 2019.

[40] L. Kull et al., "A 24–72-GS/s 8-b time-interleaved SAR ADC with 2.0–3.3-pJ/Conversion and >30 dB SNDR at Nyquist in 14-nm CMOS FinFET," IEEE J. Solid-State Circuits, vol. 53, no. 12, pp. 3508–3516, Dec. 2018.

[41] I. L. Syllaios, R. B. Staszewski, and P. T. Balsara, "Time-domain modeling of an RF all-digital PLL," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 55, no. 6, pp. 601–605, Jun. 2008.

[42] P.-H.-P. Wang et al., "A near-zero-power wake-up receiver achieving −69-dBm sensitivity," IEEE J. Solid-State Circuits, vol. 53, no. 6, pp. 1640–1652, Jun. 2018.

[43] M. Miyahara, Y. Asada, D. Paik, and A. Matsuzawa, "A low-noise self-calibrating dynamic comparator for high-speed ADCs," in Proc. IEEE Asian Solid-State Circuits Conf., Nov. 2008, pp. 269–272.