AFRL-RY-WP-TR-2023-0216

RESILIENT ANALOG INSTANCE LANGUAGE (RAIL) ENABLED OPEN-SOURCE MIXED-SIGNAL CIRCUITS

Dr. C-J Richard Shi
University of Washington

NOVEMBER 2023
Final Report

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with The Under Secretary of Defense memorandum dated 24 May 2010 and AFRL/DSO policy clarification email dated 13 January 2020. This report is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RY-WP-TR-2023-0216 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

BOZADA.CHRISTOPHER.A.1131856993 Digitally signed by BOZADA.CHRISTOPHER.A.1131856993 Date: 2023.11.07 08:32:03 -05'00'

CHRISTOPHER A. BOZADA
Program Manager
Aerospace Components and Subsystems Division

WILKINS.GENE.MICHAEL.1116849263 Digitally signed by WILKINS.GENE.MICHAEL.1116849263 Date: 2023.11.07 10:46:44 -05'00'

GENE M. WILKINS, Lt Col, USAF
Deputy Chief, Aerospace Components &
Subsystems Technology Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show "//Signature//" stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
November 2023	Final	11 June 2018	9 December 2022

4. TITLE AND SUBTITLE

RESILIENT ANALOG INSTANCE LANGUAGE (RAIL) ENABLED OPEN-SOURCE MIXED-SIGNAL CIRCUITS

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
FA8650-18-2-7857	N/A	62716E

5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER
N/A	N/A	Y1TK

6. AUTHOR(S)

Dr. C-J Richard Shi

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Washington Office of Sponsored Programs
4333 Brooklyn AVE NE
Seattle, WA 98195

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'SREPORT NUMBER(S)
Air Force Research Laboratory Sensors Directorate Wright-Patterson Air Force Base, OH 45433-7320 Air Force Materiel Command United States Air Forces	Defense Advanced Research Projects Agency (DARPA/MTO) 675 North Randolph Street Arlington, VA 22203	AFRL/RYD	AFRL-RY-WP-TR-2023-0216

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

This material is based on research sponsored by the Air Force Research Laboratory (AFRL) and the Defense Advanced Research Projects Agency (DARPA) under agreement number FA8650-18-2-7857. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory (AFRL), the Defense Advanced Research Projects Agency (DARPA), or the U.S. Government. Report contains color.

14. ABSTRACT

The objective of this research is to develop RAIL, an agile analog and mixed-signal (AMS) design methodology and its demonstration on the implementation of four open-source mixed-signal integrated circuit intellectual properties (IPs): analog to digital converters (ADCs), delay-locked loops (DLLs), low dropout voltage regulators (LDOs), and LPDDR4X memory interface circuits. These four types of mixed-signal circuits are crucial building blocks for complex systems-on-chip (SoC) development, thus are essential for achieving the complete open-source SoC implementation. However, unlike digital RTL design that can be open sourced easily in the form of register-transfer level (RTL) description using the industry-standard Verilog hardware description language and realized by using the industry automated place and route design flow, these analog and mixed-signal designs are designed manually based on analog designers' experience using schematic capture tools and simulators, then laid out manually with layout editor tools. Simulation of these blocks is performed using SPICE-like circuit simulators and is not compatible and hard to be integrated with Verilog simulators. The realization and even the design of mixed-signal blocks are process dependent, rely on proprietary physical design kits (PDKs) from the fabrication foundries, which further hinder the process of open source mixed-signa SoC innovation. The project is built on top of recently digitalized analog design architectures tailed for deep submicron digital-centric fabrication process. RAIL defines a basic set of analog and mixed-signal primitives including delay-sensitive logic gates, switches, passives, differential pairs, and et al. State-of-the-art SAR-ADC, MDLL, LDO and LPDDR4x architectures and circuit implementations are all mapped with this set of primitives and standard digital cells. The standard digital flow is enhanced with analog procedural layout generator, analog constraint generation, and device sizing and topology selection, all in an automated "make" flow. This parametric "make flow" enables process-portable analog and mixed-signal implementation. Furthermore, automated System Verilog model generation from the transistor netlist is integrated in the flow. This, for the first time, provides a complete and trustable digital-compatible flow for the implementation and verification of systems on chip containing analog and mixed-signal blocks, thus makes high-quality open-source hardware feasible and trustable. Four digital-centric mixed-signal integrated circuit intellectual properties (IPs): analog to digital converters (ADCs), delay-locked loops (DLLs), low dropout voltage regulators (LDOs), and LPDDR4X memory interface circuits have been co-developed in this project with the RAIL flow. To achieve the start-of-art silicon performances, these four IPs are analyzed, simulated, implemented in various processes including TSMC 65nm and GlobalFoundries 12nm FinFETs. All four IPs have been tapped out and fabricated successfully. Some measurement results are described.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
Unclassified	Unclassified	Unclassified	SAR	177

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Christopher Bozada	N/A

CONTENTS

1 ACKNOWLEDGEMENTS

This research was conducted as a part of the DARPA POSH (**P**osh **O**pen **S**ource **H**ardware) program. The DARPA program manager who initiated the program is Andreas Olofsson, who vision has shaped this research and enabled system designers, who are not experts on hardware design, can quickly innovate a software-hardware system, and this potentially changes the landscape of the semiconductor industry for years to come. The continued supervision and guidance by DARPA program manager Serge Leef and especially his push towards more automated and full system verification using aggregate real number modeling has provided a solid foundation for verifiable and trustable open-source hardware and software. We thank Christopher Bozada from Air Force Research Laboratory, who served as a government technical engineer on this effort, and Julian Warchall for this constant monitoring of the project status.

The principal investigator for this project is Dr. C-J Richard Shi, a professor in the Department of Electrical and Computer Engineering, the University of Washington, Seattle. The following research staff, graduate and undergraduate students from the University of Washington Department of Electrical and Computer Engineering participated in this project: Dr. Chien-Jian Tseng, Dr. Yuan Cai, Dr. Chang Liu, Dr. Chixiao Chen, Dr, Aili Wang, Rongjin Xu, Ailing Piao, Xindi Liu, Huwan (Frank) Peng, Yi-Hsiang (Phil) Huang, Aaron Chang, Jiayi Wang, Kristi Stefanovicz, and Doruk Arisoy. Professor Michael Taylor from the Department of Computer Science and Engineering at the University of Washington was instrumental in getting us started on this research. In fact, this project was a spinoff of a collaboration between Professor Taylor and Professor Shi, and Professor Taylor's vision of open-source hardware innovation. Professor Visvesh Sathe has collaborated on the digital synthesis flow and all digital PLL design.

2 PROJECT OVERVIEW

2.1 Overall Description

In this project, the University of Washington team has conducted research and development entitled “RAIL: Resilient Analog Instance Language enabled open source mixed-signal circuits” for DARPA’s POSH (Posh Open Source Hardware) program under the Electronic Resurgence Initiative (ERI). The specific project objective is to develop four open-source mixed-signal integrated circuit intellectual properties (IPs): analog to digital converters (ADCs), delay-locked loops (DLLs), low dropout voltage regulators (LDOs), and LPDDR4X memory interface circuits, portable across multiple process nodes from 14nm FINFET to 180nm CMOS.

These four types of mixed-signal circuits are crucial building blocks for complex systems-on-chip (SoC) development, thus are essential for achieving the complete open-source SoC implementation. However, unlike digital RTL design that can be open sourced easily in the form of register-transfer level (RTL) description using the industry-standard Verilog hardware description language and realized by using the industry automated place and route design flow, these analog and mixed-signal designs are often designed manually based on analog designers’ experience using schematic capture tools and simulators, then laid out manually with layout editor tools. Simulation of these blocks is performed using SPICE-like circuit simulators and is not compatible and hard to be integrated with Verilog simulators. The realization and even the design of mixed-signal blocks are process dependent, rely on proprietary physical design kits (PDKs) from the fabrication foundries, which further hinder the process of open source.

A key innovation of project RAIL is an industry-first PDK-independent design methodology and support for open-source mixed-signal integrated circuit IP development. The proposed solution to PDK-independent analog and mixed-signal circuit design includes the development of a novel PDK-independent topology description language, referring to as *RAIL*, and an associated analyzer and compiler to link with existing industry-standard analog and mixed-signal description languages and simulators, PDK-dependent technology mapper and device size optimizer (referred to as netlist synthesizer) and parametric procedural layout generator, thus fully addressing the open-source full portability requirement.

In contrast to mature digital open-source design, open-source analog and mixed-signal (AMS) technology does not exist. Highly PDK-dependent analog design flow prevents the portability between technology nodes. This is due primarily to two major difficulties: First, traditional analog circuit design topologies such as opamps are not generally portable across process nodes; Second, lack of tool support and automation for such experience-intensive analog design.

Therefore, this project advocates the selection of digital-centric process-independent architectures and design methodology for analog design in digital processes. Meanwhile, newly emerging tools and languages, for example, CHISEL for RISC-V, are boosting the traditional Verilog/RTL-dominated digital open-source environment. Therefore, a novel PDK-independent topology description language (TDL) is introduced based on minimal extension of

Verilog/Verilog-AMS language constructs to support a minimal set of analog primitives and their design constraints. On the base of such a language, RAIL parser, synthesizer, and optimizer, have been developed to enable the real open-source AMS design for SoC eco-systems. The RAIL project uses the standard industry tools and/or the OpenRoad tools from the DARPA IDEA program when possible and only develops essential needed additions and scripts for the PDK-independent support of AMS design flow.

Recent digitalized trend in analog design inspires the RAIL project technical approach. Clock generation circuits, data converters, and voltage regulators are experiencing an all-digital evolution. This approach allows a feasible higher-level PDK-independent topology description with a limited set of AMS primitives, and enables the portability of open-source AMS circuits. RAIL supports a basic set of analog primitives including delay sensitive logic gates, switches, passives, differential pairs, and comparators. State-of-the-art digital DLL, LDO and SAR ADCs are all implemented with this set of primitives in this digitalized AMS design flow. Both transistor level and RAIL-level descriptions are systematically analyzed. The automated abstraction of SPICE netlist into System Verilog description---Real Number Models (RNMs)---enables the use of one simulator for the verification and optimization of the entire SoC design. Further the speed up due to the abstraction in real number based signal flow semantics in System Verilog, instead of nonlinear differential equations in SPICE, has lead to multiple orders of magnitude improvement in simulation speed, and thus makes the entire SoC simulation feasible.

2.2 Major Innovations

This project provides ground-breaking technologies composed of six major innovations:

- **Innovation 1: Topology description of digitalized AMS design.** With the digitalized topology trend in modern AMS design, state-of-the-art AMS topologies are becoming much more regular and digital-process-friendly than previous opamp-dominated AMS design. By carefully analyzing these topologies and primitives, we propose a new topology description language by minimally annotating Verilog/Verilog-AMS.
- **Innovation 2: Constraint driven AMS netlist synthesis and parasitic-aware netlist optimization and layout generation**. To complete a feasible design, PDK-independent topology description must be extended with analog design constraint description, as well as PDK-dependent netlist synthesis and optimization. Observing that layout parasitics are essential to device sizing for performance, parametric layout templates are built for each analog circuit primitive as well as subblocks. Performance expressions are derived based on a full characterization of the primitive/subblock in terms of design parameters, where parasitics are accurately captured. This, together with electrical and manufacturability constraints, form the total set of analog design constraints. A robust optimizer is used to solve these constraints for parasitic-aware circuit and layout generation. With this, the vision of no-human-in-the-loop portability among technology nodes for analog design can be achieved.
- **Innovation 3: Dual-loop Digital LDO with fine noise shaping on PVT variation**. Facing the power wall of the silicon, recent SoCs adopt multiples of on-chip digital LDOs, due to their scaling down friendliness. The proposed dual loop D-LDO, working like a multi-stage noise-shaping (MASH) 0-1 modulator, suppresses the PVT-sensitive ripples and instability using an extra $\Sigma - \Delta$ fine loop. Furthermore, the topology, all its primitives, associated design constraints are well described and supported by the RAIL technology.
- **Innovation 4: Bit-wise reference charge reservoir enabled reference-buffer-free SAR ADCs.** Despite excellent energy efficiency and process portability, successive approximation register (SAR) based analog to digital converters (ADCs) still require precise reference buffers for SoC integration; this is known to be

power-hungry, area demanding and not process-portable. Out of our NRO-sponsored 14nm adaptive ADC research project, we have developed a new reference circuit scheme, using a bit-wise charge reservoir capacitor, to fully eliminate the reference buffer and decoupling capacitor. With all the charge-reservoir capacitors buried under the capacitor array, no area penalty is needed. Finite capacitor error is shown to be in the form of sub-radix-2 real number expansion and can be precisely calibrated. This, combined with segmentation, noise shaping, and time-interleaving will be used to generate a portable resolution-speed-adaptive most energy-efficient ADC.

- **Innovation 5: An all-digital MDLL with full digital delay line.** Low power on-chip clock generation is stepping away from phase-locked-loops towards multiplying delay locked loops (MDLLs) due to better noise performance and process friendliness. Previous MDLLs still require a current starving based delay line, which is not preferred in the digitalized trend. We propose a fully digital delay line with digital-to-delay linearity pre-training scheme. The RAIL technology supports the new delay line, while the rest circuit is RTL synthesizable.
- **Innovation 6: Automatic impedance calibration on low-voltage-swing-terminated-logic for LPDDR4X drivers**. Data-intensive tasks drive the high bandwidth and low power memory access for recent SoCs. LPDDR4X, featuring a 0.6V I/O supply voltage, requires extra physical training and calibration. The RAIL technology addresses the digitalized engine with full portability among technology nodes.

2.3 RAIL Open-Source Mixed-Signal Circuit Design Flow

At present, there exists virtually no open-source AMS circuit IPs. The conventional digital open-source flow, releasing a PDK-independent RTL code and synthesizing the netlists on a specific PDK, does not work for the AMS design. The state-of-art analog design requires a great deal of effort to handcraft all IPs on each technology node with substantial design margins (hard IPs). Despite Verilog-AMS (VHDL-AMS) as a description and simulation language standard, there is no abstraction-level hardware description language for AMS design description and synthesis. Sharing PDK-sensitive netlist or a pure Verilog-A(MS) behavioral mode does not contribute to any portability. Therefore, the major problem faced by open-source AMS development is a PDK-independent hardware description language for AMS circuit design and synthesis. However, this is hard to achieve because traditional AMS design involves a great pool of possible topologies and there are no well-accepted high-level abstracted topology description methods, other than netlists, that fit various scenarios. The Verilog-AMS description is only at the behavior level and cannot be synthesized. Circuit size optimizers fail frequently when AMS circuit scales keep expanding.

Observing a recent trend of state-of-the-art AMS design, the RAIL project introduces a PDK-independent design language, built on top of Verilog/Verilog-AMS with a set of predefined analog primitives and design constraints. In fact, it is designed as annotated attributes therefore fully compatible with standard Verilog/Verilog-AMS tools. With this, some minimal addition of design tools like the RAIL parser, synthesizer, and optimizer have been developed open source and enable open-source AMS development.

2.3.1 Digitalized Trend in AMS Design

Moore's law has been transforming AMS design in the past ten years. Operational amplifiers, once the most common components in the analog design, face severe setbacks, including less voltage headroom due to scaling down supply, the low gain due to the reduced output impedance, and worse matching due to the smaller device size. Moreover,

when moving to smaller process nodes, on-chip passive devices are typically not shrinking in size, and become more expensive, compared with transistors. On the other hand, digital equivalents are drawing more and more interest due to their intrinsic scaling-down friendliness. The most well-known example is that a combination of time-to-digital converters (TDCs) and digital filters replaces traditional charge-pump and passive filters in all-digital phase-lock-loop (ADPLL).

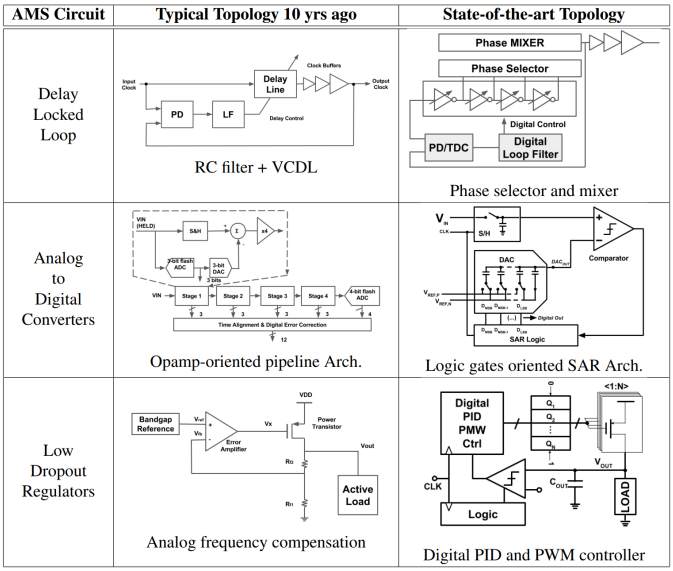


Figure 1. Circuit Topologies of Today vs a Decade Ago for DLL, ADC and LDO

Figure 1 summarizes this trend for three types of circuits DLLs, ADCs. and LDOs. In DLLs, traditional voltage-controlled delay line, consisting of most likely current-starving inverters, has been transformed into a combination of phase selector and phase-mixer. In ADCs, opamp-oriented pipelined architecture has been transformed into the successive approximation register (SAR) architecture dominated by switches, capacitors, comparators, and digital logic gates. In LDOs, traditional analog feedback loop with frequency compensation has been implemented using comparator-based quantizers, and digital proportional-integral-derivative (PID) controllers. Although originally developed for implementing analog functionality on ``lossy" digital processes, it is, in fact, this digitalized trend in AMS design that makes the development of digital-like and digital-compatible design methodology, flows and tool sets feasible for AMS design. It is this trend that makes the open-source AMS circuit IP development and no-human-in-the-loop flow feasible.

2.3.2 RAIL – a PDK-Independent Topology Description Language

With the digitalized trend, this research develops a PDK-independent topology description language for open-source AMS circuit design. Higher level functional abstraction as well as design constraints on the topology, other than netlists, are the RAIL design objectives. Traditional active circuits, such as operational amplifiers and current mirrors are gradually abandoned under the trend. Comprehensive study of the state-of-the-art design has led to an observation that modern analog design uses mostly digital logic gates, together with a few digital-like/digital-process-friendly analog components This is the foundation of our proposed LDOs, ADCs, DLLs, LDO and even high-speed link interface's PHY design. Some typical examples of primitives are as follows:

- Delay sensitive logic gates. Current RTL-based synthesis only focuses on the logic functions within the given time constraints. However, modern AMS design requires an additional accurate delay for time domain operations under certain power supplies, often known as time-based processing. Examples include digital-controlled delay lines in TDCs and DLLs.

- Variation sensitive dynamic transistor pairs. Though active static circuits are rarely used nowadays, differential voltage-input pairs are still common for dynamic mixed-signal design. Matched devices with little variation are desired to convert an analog voltage into a rising/falling edge accurately. Examples include two-stage dynamic pre-amplifiers and latches.

- Switches. Switches of different sizes and types are used for a variety of circuits, such as PMOS switches for LDOs, bootstrapped NMOSs for ADC sampling, transmission gate arrays for phase interpolation, etc. Note that switch drivers need to be co-designed with the switches.

- Passives. Capacitors of various types, such as varactors, MOS, MOM and MIM capacitors are required for different scenarios. Resistive arrays still contribute a lot for post-fabrication tuning. Diodes for ESD, tap cell and proportional to absolute temperature devices are also necessary.

The proposed RAIL topology description language aims for a Verilog/Verilog-AMS compatible environment. It will help circuit designers to easily understand the design intent. It will also simplify the translation into Verilog-AMS behavior models for simulation and verification. A simple example of the RAIL TDL of one switched capacitor sampling circuits is shown in Figure 2.

```
module sampler (vin, vout, vcm, clk);
  // I/O definition
  inout vin, vout, vcm;
  input clk;
  electrical vin, vout, vcm;
  // Param. definition
  parameter real ron=10 ;
  parameter real cs=1e-12 ;

  rail begin
    vout <+ sw.boot(clk?vin:hiZ,ron)
          + psv.momcap(vcm,cs);
  end // '+'=parallel connect

endmodule
```

Figure 2. A Sample Code on a Switched Capacitor Based Sampling Circuit by the RAIL TDL

2.3.3 RAIL Open-source AMS Design Support and Flow

With the RAIL TDL as the basis, the entire design support includes the RAIL parser, RAIL-to AMS translator, RIAL-to-netlist synthesizer, and optimizer. Multi-level verification loops, from the behavior-level to the transistor level are involved in the entire design flow.

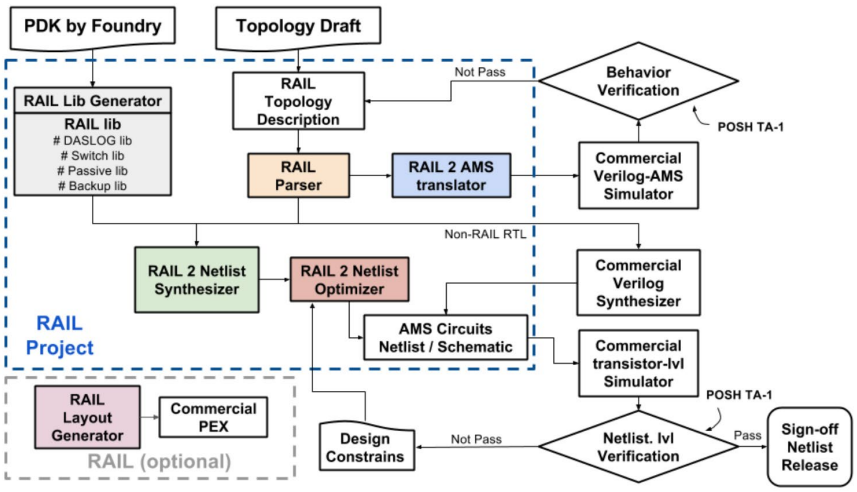


Figure 3. Complete Design Flow for the RAIL Open-source AMS Circuits

Figure 3 demonstrates a complete design flow with the RAIL supported tools, from the initial design intent description to the final schematic/netlist of sign-off quality. Note that the automatic layout generator is optional, as the placement and routing results are not included in a standard open-source digital design. However, parametric cell layout generator is needed with optimized device sizes considering template-driven layouts. The major effort of the RAIL research is boxed by the blue dashed lines in Figure 3. Procedures outside the box are implemented by either commercial tools or open-source tools developed out of the DARPA IDEA program, except the RAIL layout generator. Color boxes in Figure 3 assemble a complete design support EDA tool for RAIL open-source AMS circuit implementation.

The complete RAIL design flow has two entry points. The first path is the RAIL PDK generation, which makes the RAIL AMS library of primitive analog instances for each specific technology node. The other path is PDK-independent, i.e., the critical path for the open-source AMS circuit generation. Designers first code the draft topology into a RAIL topology description file. A RAIL parser reads in the file and generates its corresponding Verilog-AMS model by a RAIL-to-AMS translator. Commercial and open-source simulators are used to verify the topology at the behavioral level. These two paths meet at the RAIL-to-netlist synthesizer. The synthesizer adopts a straightforward instance mapping algorithm to derive the initial netlist without any concerns on parasitics. Very often, the netlist needs several revisions based on the practical circuits. A RAIL-to-netlist optimizer performs trial-and-error by iterations to satisfy all the design constraints. After the optimizer, the generated netlist or schematic is ported into a commercial transistor level simulator. Transistor-level verification shows whether the RAIL design achieves the expected performance or not. If not, a guideline on design constraint revision is suggested. A new optimizing iteration is performed after updating the constraints. Finally, when the design passes all the checklist, it is subject to a sign-off release. For layout-parasitic sensitive analog primitive design, a parametric cell is designed, the performance and parasitics are fully characterized symbolically in terms of design parameters and PDK parameters. This is built into the PDK-dependent RAIL library to be used for netlist mapping and circuit optimization, and layout generation.

2.4 AMS Circuit IP Development with the RAIL Flow

The RAIL flow has been co-developed with, and applied to, the development of four commonly used mixed-signal IPs in today's SoCs, as shown in Figure 4. These are asynchronous SAR ADCs, digital low-dropout regulators (DLDOs), multiplying digital locked loops (MDLLs), and an LPDDR interface PHY. As a spinoff of the SAR-ADC implementation research, a fully synthesizable SAR based compact temperature sensor that can be used for SoC temperature monitoring has been developed in this project. Our investigation on LPDDR4X implementation also leads to a co-development of high-speed IO circuit for the latest industry standard introduced February 2022 for short-reach die-to-die interfacing.

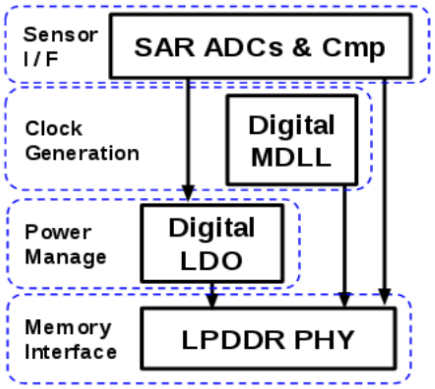


Figure 4. Four Analog and Mixed-signal IPs Developed Supporting Open-source SoC Design

These AMS IPs are ubiquitous in today's system-on-chips (SoCs). ADCs serve as the bridge between the digital process and analog frontend, also often used for digital calibration. MDLLs are key blocks for SoC clock generation and management, also used for phase interpolation and calibration. DLDOs enable power management of different voltage domains. LPDDR PHY solves the critical memory access limitation. Moreover, sophisticated IPs like LPDDR4X and UCIE PHY also involve basic IPs such as DLLs and LDOs. DLLs are used in DLDOs.

2.4.1 RAIL Process-Portable Reference-Buffer-Free Asynchronous SAR ADCs

Analog-to-digital converters, bridging the analog world and digital processing systems, are essential for sensor interfaces of system-on-chips (SoCs). State-of-the-art SAR ADC features high energy efficiency, no static active circuits, and digital design friendliness. Combined with segmentation, noise shaping, and time interleaving, SAR-ADC is becoming a dominating architecture for many applications. Innovations, such as sub-2-radix array and redundancy, further enhance SAR ADCs robustness [1].

Stand-alone SAR ADC tends not to meet the SoC requirement. A major concern is the peripheral reference circuits. SAR ADC of stand-alone designs use hundreds of pico-Fraday on-chip capacitor for reference decoupling, thus circumventing the power-hungry reference buffers. The method is not feasible for SoC design, due to the huge area of the decoupling capacitor. The literature on the solution begins from [2], referring to as charging sharing ADCs, where the capacitive array used for both signals and references. However, it is not adequate for high precision. Low

decoupling reference capacitor with an extra error detection technique was proposed in [4], where only a 3pF decoupling cap is needed.

A new trend of switching reference charge reservoir is drawing a lot of interest from analog design houses such as Analog Devices, because it can fully eliminate the power-hungry buffer or area demanding decoupling capacity without penalties. A sample-wise switched reservoir was proposed in [3]. But it introduces the non-linearity of higher orders. A discrete-time feedback scheme to charge the reservoir is used to optimize the issue in [5].

We have introduced a new reference scheme for SAR ADCs in SoCs, which enhances the reference charge reservoir concept by a bit-wise switching architecture. The proposed bit-wise switch reference charge reservoir not only eliminates the reference buffer or decoupling capacitor, but also forms an intrinsic sub-radix-2 real number expansion, which can be used for precise error correction. Figure 5 illustrates the proposed process-portable reference-buffer free SAR-ADC architecture.

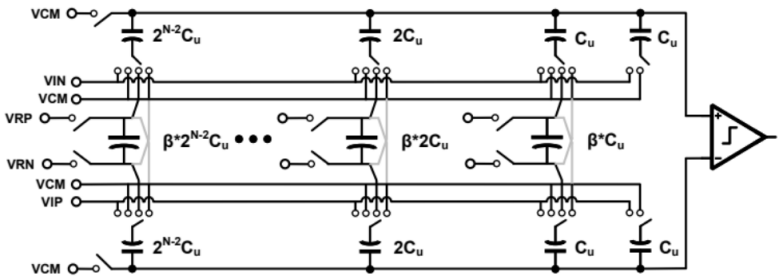


Figure 5. Proposed Reference-Buffer-Free SAR ADC Architecture

Rather than an entire always-on decoupling capacitor, the proposed reference charge reservoir splits the big capacitor bit by bit. During the sampling phase, the charge reservoir capacitor samples the reference voltage, as does the capacitive DAC sample the input signal. During the decision phase, the charge reservoir is disconnected to the reference voltage and co-works with the capacitive DAC. Thus, there is no need for a high-bandwidth reference buffer to track the voltage reference.

Theoretical derivation has been performed to model bit-wise reference charge reservoir. The β factor, a ratio between capacitive DAC and charge reservoir DAC, could express the relationship between analog inputs and digital outputs.

$$\mathrm{V_{in,diff}} = \sum_{i=1}^{N-1} \alpha^{(i)} \frac{C_i}{C_{total}} (2b_i - 1)(\mathrm{VRP} - \mathrm{VRN}) \qquad (1)$$

where *Ci* and *Ctotal* are the capacitive DAC, *bi* is the comparator result of each step, and *α (i)* is,

$$\alpha^{(i)} = \frac{2\beta}{2\beta + 1 - \sum_{j=i}^{N-1} \frac{C_j}{C_{Total}}} = \frac{2\beta}{2\beta + (1/2)^{N-i}} \quad (2)$$

Obviously, the α factor contributes to a built-in sub-radix-2 redundancy by picking a small β. In practice, the reservoir capacitor can be implemented by MOS capacitor/varactor, burying under the capacitive DAC of MOM capacitor. A digital back-end engine computes the final converted digital result by calibrating the capacitor mismatch and calculating a sub-radix-2 summation. Therefore, a compact SAR ADC design with complete resolution and no further area penalty is achieved.

Table 1. AR-ADC Design Specification for a 65nm CMOS Process Node

Spec.	Value
Resolution	12 bit
SNDR	60-dB
SFDR	70-db
Sampling Rate (130/180nm)	125 MSps/ch (65 MSps/ch)
Overall Rate	Up-to 1GSps
Max. intl. Chan	16
Power	<1mW/ch
FoM	<10fJ/conv.sep

To address the wide-range clock rate, from 10MHz to 1000MHz, multiple SAR-ADCs are time interleaved. Each channel ADC is supposed to reach 125 MSPs with more than 500MHz input bandwidth. In old process nodes of 130nm and 180nm, the ADC rate is relaxed to 65 MSPs. Table 1 summarizes the target performance of the proposed ADC for a 65nm CMOS process node. The overall figure of merit (FoM) of the proposed ADC is less than 1fJ/conversion-step with 14nm FINFET, including the reference charge reservoirs. The architecture has been shown to be fully portable from 14nm FINFET to 180nm CMOS by leveraging the RAIL technology, with varying FoMs achieved for a specific process node. Noise shaping can be applied to sacrifice the speed to improve the resolution.

The overall SAR ADC architecture is composed of four parts: (1) capacitive array with switches and drivers, (2) asynchronous logic with embedded accurate delay, (3) sampling network, and (4) dynamic comparators. The RAIL technology covers all the four parts, where (1) and (3) are achieved by switches and passive description and (2) is achieved by delay sensitive logic gates. In (4), the dynamic preamplifier is achieved by variation sensitive pairs, converting voltage into two different edges, and regeneration latch are cross-coupled delay sensitive logic gates.

2.4.2 RAIL Wide-Range Constant-Efficiency Digital LDO Regulators

Voltage regulators with dynamic scaling techniques are frequently utilized in current low power SoC designs. Conventional solutions include analog low-dropout (A-LDO) regulators and switching DCDC converters. However, A-LDO does not follow the voltage scaling-down and switching DC-DC converters requires a low-ESR inductor, whether off-chip or on-chip it is not friendly for general CMOS technology. Digital LDO regulators (DLDO), prevailing recently, do not have these two weaknesses. However, D-LDOs still face some issues for SoC applications. D-LDOs with high-frequency clock usually cannot achieve good energy efficiency [6], while ones with low frequency suffer from slow response and poor stability. To trade-off the frequency issue, an event-driven based architecture with reduced slack latency was proposed in [7], but still requires 100pF output capacitor. To avoid area-demanding capacitors, burst-mode and multi-loop designs have been published [8]. More recently, digital PID (proportional, integral and differential) controllers are adopted in [9] to replace analog passive devices. Another important trend of digital LDO is fast response and search scheme. Conventional linear search is too slow for many applications, though most stable. In [10], a binary search, like a SAR ADC, was proposed to achieve fast responses. A fine-grained D-LDO with multi-step switching and asynchronous adaptive pipeline control was presented to realize nano-second resettling response [11]. By these methods, D-LDOs outperform A-LDOs using newly introduced search scheme and high clock rate. Furthermore, search steps can be in sub-radix-2, instead of binary, using the same technique we introduced for reference-buffer-free SAR-ADC implementation. This sub-radix-2 real number expansion search introduces redundancy that allows self-error correction.

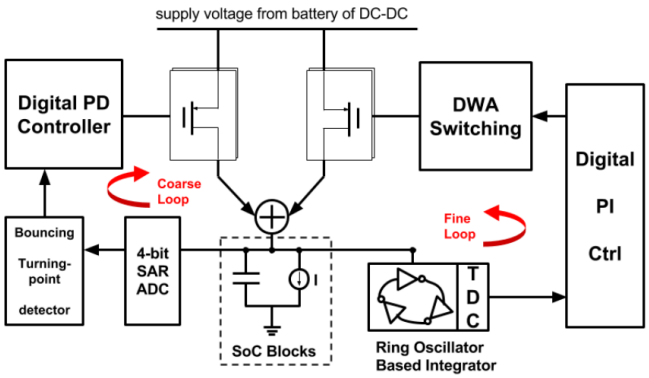


Figure 6. Proposed Digital LDO Regulator IP Architecture

In the project, we developed a RAIL-compatible digital LDO IP solution, as shown in Figure 6. The digital LDO is composed of two loops to achieve both fast response and wide range. The coarse loop is responsible for the rapid load change. It includes a 4-bit SAR quantizer, a bouncing and turning point detector and Digital Proportional-Differential controller. The controller's output drives a PD switch array as power transistors. In contrast with traditional barrel shift controller, the 4-bit quantizer enables a fast and stable step response with digital differential compensation.

The fine loop in the proposed LDO adopts a shifter-based noise-shaping scheme to avoid large output capacitor in the light load mode. The idea is inherited from the hybrid SAR/single slope/noise shaping ADCs. Traditionally, the pulse-width modulation (PWM) scheme in LDOs requires a huge capacitor. In

the design, noise-shaping nature of the structure requires less decoupling capacitance. In the loop, a delay-line based ring oscillator performs voltage integration, serving the first order noise shaping. A TDC converts the oscillation signal into digital signals. A digital proportional-integration controller absorbs the digital signal and shuffling fine switching array by a DWA/DEM scheme, thus forming the second order of noise-shaping. In contrast with pulse width modulation, the loop moves the quantization error into high frequency, thus requiring less capacitor.

To support open-source SoC design, it is highly desirable to have a universal D-LDO topology that achieves constantly high energy efficiency for both heavy load and light load scenarios and further the switch from the heavy load to light load and vice versa can be fast. This is impossible if an LDO is implemented using the analog topology, since the power consumption of the static amplifier is constant. As long as the output power reduces, the overall efficiency reduces unavoidably. However, in the D-LDO, load-aware clock generation can achieve this by digitally adjusting the clock.

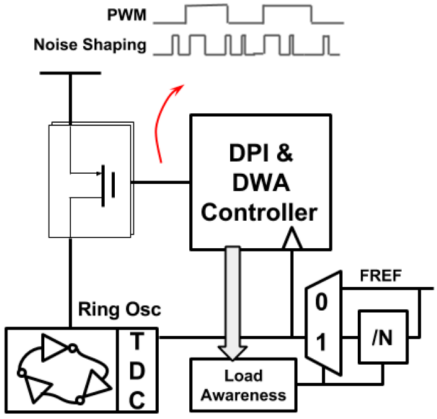


Figure 7: Load-ware D-LDO Clock Generation

Figure 7 shows the diagram of load-aware clock generation. For the light mode, the coarse loop is barely on. Only the fine loop with noise-shaped pulse drives the power switch. In a delta-sigma loop, a constant oversampling ratio, other than the constant clock, is preferred. Thus, the load awareness block detects the allowed minimum clock and outputs a divided clock signal to drive the entire fine loop. Therefore, the power consumed by the D-LDO scales with the load and an almost-constant energy efficiency is achieved.

Table 2. The DLDO Specification for Open-source SoC Implementation

Spec.	Value
Input voltage	1.8*-3.3V
Output voltage	0.25-1.8*V
Measured ripple	<10mV
Ext. Clock Rate	Up-to 10MHz
I_{out}	100μA – 100 mA
Efficiency heavy load light load	 > 85% > 80%

The proposed LDO's specification is summarized in Table 2. This design is fully portable across the process nodes of 14nm FINFET to 180nm CMOS due to the RAIL technology. Previously a fully synthesizable DLDO has been proposed [12], it lacks portability among technology nodes. In the proposed topology, the controllers and detector are described in the Verilog RTL domain, thus realizing by the typical digital open-source flow. The rest, including the SAR quantizer, the switches (with drivers), the ring oscillator and it TDC logic are described with RAIL. RAIL TDL supports switches straightforwardly. Ring oscillators and TDC logics are implemented using the delay-sensitive logic gates optimized for a certain supply voltage range.

2.4.3 RAIL Multiplying Delay Locked Loops

As the technology advances, clock multiplication with phase alignment has become a necessity for all-digital ICs. Dynamic frequency scaling and high-speed interfacing require DLLs for clock recovery. Though PLL was a conventional tool for multiplying the frequency, multiplying delay locked loops are drawing more and more interest recently. High-performance PLL demonstrates its good phase noise with the help of LC voltage-controlled oscillator (VCO). But LC tanks are not affordable in SoC design. When the VCO is transformed to a ring oscillators topology, PLL's noise performance deteriorates dramatically. The key reason is the accumulation of the jitter inside the ring integrator. Multiplying DLLs show much better noise performance because the jitters are cleared in each reference clock edge [13]. Digital MDLLs face several issues. The first concern is the decoupled traded-off between the TDC bandwidth and ring oscillator's noise performance. A joint study in [14] was performed to optimize the design. The second problem is that there is some analog circuitry in the digital MDLL, for example, the current-DAC biased delay line. The third issue is the supply-noise sensitive ring oscillators. Normally, analog LDOs are used to prevent the issue. However, the RAIL project's target, a fully digital solution with D-LDOs, might suffer from the noise. To address these concerns, we propose a digital multiplying delay lock loop (DMDLL) to meet the requirement of all (or most) digital SoC.

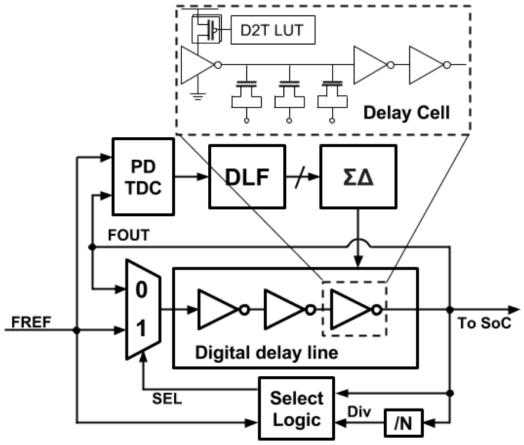


Figure 8. The Digital Multiplying DLL Architecture

As shown in Figure 8, the proposed DMDLL is composed of phase detector (PD) or time-to-digital converters (TDC), digital loop filter, a ΣΔ modulator, digital-controlled delay line and feedback logics. In contrast with the traditional current starving topologies, the design seeks a fully digital solution of the

delay line. Both starving transistors and load reconfigurability methods are adopted. To maintain a good linearity within the certain range, a linear enhanced digital-to-delay lookup table (LUT) is pre-trained before the loop works. The LUT could be implemented by a non-volatile memory. To improve the supply noise tolerance, differential-pair based delay cells are utilized, though only single half circuits are drawn in the figure. In practice, the search engine follows such a procedure: (1) a comprehensive delay measurement over all possible options, (2) select 7-15 possible codes with good linearity, (3) store the codes in the LUT and shuffle the codes during operation. A joint digital DLL and LDO design flow relaxes the constraint on supply noise.

Table 3. The targeted Specification of a MDLL for Open-source SoC Implementation

Spec.	Value
Input f_{REF}	10MHz
Output freq range	10Mhz - 1000Mhz
Power (mW/GHz)	0.5
Integrated RMS Jitter	0.5 ps
Phase noise 10kHz	95 dBc/Hz
Total Power	$< 1mW$

Table 3 summarizes the target performance of the proposed MDLL solution. According to the BAA, the maximum multiplying factor is 50, given that there is voltage doubler. The measured integrated RMS jitter is expected to approximately 0.5ps and the frequency range is from 0.01 to 1GHz. With a 1V supply voltage, the power consumption of the entire MDLL is less than 1mW, making the MDLL a low-power candidate for frequency generation other than power-hungry PLLs in energy efficient SoCs. To deliver the DMDLL IP to RAIL, all the blocks are treated differently. The digital logic block, such as selective logic and digital filter loop, can be directly synthesized via Verilog RTL. These circuits are fully compatible with the current digital open-source flow. There are blocks that could not be achieved by Verilog, need the RAIL project support. Examples include digital controlled delays, which are expected to be implemented by a combination of delay sensitive logic and variation sensitive pairs. Again, the RAIL flow ensures the DLL's portability among technology nodes.

2.4.4 RAIL LPDDR4X Interface PHY

Recent artificial intelligence progress calls for a new round of innovation of computer architecture and implementation of energy-efficient edge-computing devices. As for these SoC designs, a critical bottleneck is the high bandwidth memory access. DRAM interface, under the progress, evolves faster than ever in terms of protocol, package, and topology. A radical type of SDRAM, referring to as low power DDR generation 4 extended (LPDDR4X), features the lowest power consumption and most compact package in the record. It saves more than 50% transmitter power compared with its previous LPDDR4X version by reducing I/O voltage from 1.1V to 0.6V [17].

Known for the high bandwidth, DDR interfaces are hard to design due to the difficulty of edge training compatible PHY. Impedance calibration (ZQ) and duty-cycle correction are the most common training processes. Other training processes include data strobe calibration (DQ2DQS) etc [18]. The RAIL flow enables well the open-source development of LPDDR4X PHY IPs, because most calibration and drivers are working under the digital (linear region of switching) mode.

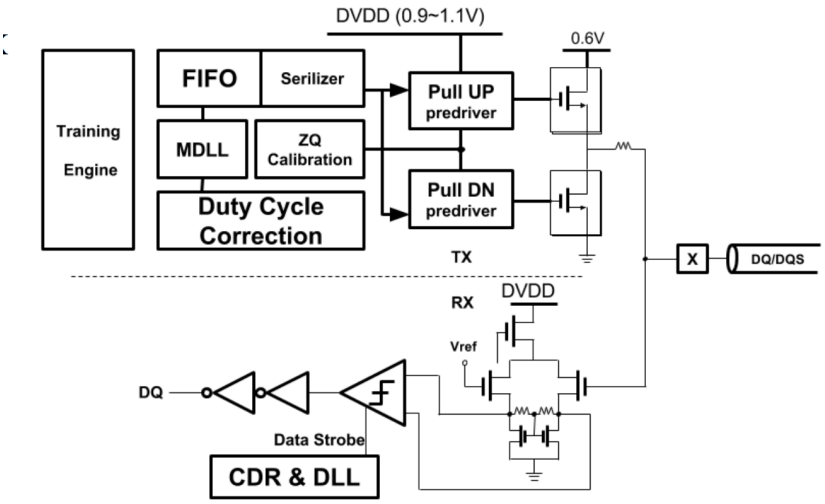


Figure 9: Proposed LPDDR4X Interface PHY IP Solution

Figure 9 shows a complete transceiver channel of an LPDDR4X interface PHY. The output driver adopts a low-voltage-swing terminated logic (LVSTL) with a Nover-N topology under the power supply of 0.6V. The rest circuit' power supply is the common digital supply voltage.

The training and calibration engines have been developed as finite state machines (FSMs) and designed using RTL synthesis, while the implementation of DLL and the comparator parts have been discussed in the previous MDLL and SAR-ADC sections. The LVSTL logic works as two NMOS switches when leveraging RAIL TDL. The duty cycle corrector can be achieved by a cross-coupled latch as the DASLOG regenerative comparator. The static preamplifier in the receiver path is realized by a P-type voltage-to-time converter with self-biasing. Therefore, the overall LPDDR4X PHY can be completely described in RAIL.

During the execution of the project, the LPDDR4X PHY [21] has been extended to support the emerging industry-standard UCIe for energy-efficient high-speed die-to-die interface [22]. The ASIC memory controller and UCIe to the Network on Chip (NoC) adaptor have also been implemented based on the extension of two open-source efforts [23,24].

2.5 Refences

[1]. C. C. Liu, C. H. Kuo and Y. Z. Lin, "A 10 bit 320 MS/s low-cost SAR ADC for IEEE 802.11ac applications in 20 nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 50, no. 11, pp. 2645- 2654, Nov. 2015.

[2]. J. Craninckx and G. van der Plas, "A 65fJ/conversion-step 0-to-50MS/s 0-to-0.7mW 9b charge sharing SAR ADC in 90nm digital CMOS", in *IEEE ISSCC Dig. Tech. Papers*, 2007, pp. 246-600.

[3]. R. Kapusta, J. Shen, S. Decker, H. Li, and E. Ibaragi, "A 14b 80MS/s SAR ADC with 73.6dB SNDR in 65nm CMOS", in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC 2013),* San Francisco, CA, 2013, pp. 472-473.

[4]. C. H. Chan et al., "A 0.011mm2 60dB SNDR 100MS/s reference error calibrated SAR ADC with 3pF decoupling capacitance for reference voltages," *in IEEE Asian Solid-State Circuits Conference (A-SSCC),* Toyama, 2016, pp. 145-148.

[5]. M. Liu, A. van Roermund, and P. Harpe, "A 10b 20MS/s SAR ADC with a low-power and area-efficient DAC-compensated reference," in *43rd IEEE Euro. Solid State Circuits Conf.*, Leuven, Belgium, 2017, pp. 231-234.

[6]. S. B. Nasir, S. Gangopadhyay and A. Raychowdhury, "A 0.13m fully digital low-dropout regulator with adaptive control and reduced dynamic stability for ultra-wide dynamic range," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, San Francisco, CA, 2015, pp. 1-3.

[7]. D. Kim, J. Kim, H. Ham and M. Seok, "A 0.5V-VIN 1.44mA-class event-driven digital LDO with a fully integrated 100pF output capacitor," in *2017 IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, CA, 2017, pp. 346-347.

[8]. M. Huang, Y. Lu, S. W. Sin, S. P. U and R. P. Martins, "A fully integrated Digital LDO with coarse fine-tuning and burst-mode operation," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 7, pp. 683-687, July 2016.

[9]. W. J. Tsou et al., "20.2 Digital low-dropout regulator with anti PVT-variation technique for dynamic voltage scaling and adaptive voltage scaling multicore processor," in *IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, CA, 2017, pp. 338-339.

[10]. L. G. Salem, J. Warchall and P. P. Mercier, "A 100nA-to-2mA successive-approximation digital LDO with PD compensation and sub-LSB duty control achieving a 15.1ns response time at 0.5V," *in 2017 IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, CA, 2017, pp. 340-341.

[11]. F. Yang and P. K. T. Mok, "A nanosecond-transient fine-grained digital LDO with multi-step switching scheme and asynchronous adaptive pipeline control," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2463-2474, Sept. 2017.

[12]. A. Fahmy, J. Liu, P. Terdal, R. Madler, R. Bashirullah and N. Maghari, "A synthesizable time-based LDO using digital standard cells and analog pass transistor," in *ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference*, Leuven, Belgium, 2017.

[13]. 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 sub-picosecond jitter performance", *IEEE Journal of Solid-State Circuits,* vol. 43, no. 4, pp. 855863, Apr. 2008.

[14]. A. Elshazly, R. Inti, B. Young and P. K. Hanumolu, "A 1.5GHz 890W digital MDLL with 400fsrms integrated Jitter, -55.6dBc reference spur and 20fs/mV supply-noise sensitivity using 1b TDC", in *ISSCC Dig. Tech. Papers*, pp. 242243, 2012.

[15]. H. Kim, Y. Kim, T. Kim, H. J. Ko and S. Cho, "A 2.4-GHz 1.5-mW digital multiplying delay locked loop using pulse width comparator and double injection technique", *IEEE Journal of Solid-State Circuits*, vol. 52, no. 11, pp. 2934-2946, Nov. 2017.

[16]. S Kundu, B Kim, and C H. Kim, "A 0.21.45-GHz subsampling fractional-N digital MDLL with zero-offset aperture PD-based spur cancellation and in situ static phase offset detection*", IEEE Journal of Solid-State Circuits,* vol. 52, no. 3, pp. 799811, Jan., 2017.

[17]. C. K. Lee et al., "A 5Gb/s/pin 8Gb LPDDR4X SDRAM with power-isolated LVSTL and split die architecture with 2-die ZQ calibration scheme," in *2017 IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, CA, 2017, pp. 390-391.

[18]. S. M. Lee et al., "A 0.6V 4.266Gb/s/pin LPDDR4X interface with auto-DQS cleaning and write VWM training for memory controller", in *2017 IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, CA, 2017, pp. 398-399.

[19]. RISC-V: The Free and Open RISC Instruction Set Architecture. https://riscv.org/

[20]. Nvidia Deep Learning Accelerator (NVDLA). http://nvdla.org/

[21]. Low Power Double Data Rate 4 (LPDDR4) specification, https:// www.jedec.org.sites.default/files/docs/JESD209-4.pdf

[22]. UCIe Specification 1.0, https://uciexpress.org/team-3.

[23]. Enjoy-digital (2021) litedram [Source code]: https://github.com/enjoy-digital/litedram

[24]. Wavious LLC (2021) wav-lpddr-hw [Source code]: https://github.com/waviousllc/wav-lpddr-hw

3 RAIL: A PROCESS-PORTABLE AGILE OPEN-SOURCE MIXED-SIGNAL CIRCUIT DESIGN FLOW

Digital IC design automation tools have been developed over the past decades and are now mature, while analog IC and high-performance custom digital design is still heavily dependent on designers' experience. This work proposes a fully automated procedural circuit and layout generator in the automated framework that utilizes existing digital electronic design automation (EDA) tools, aiming at open-source analog IP layout automation without designers' manual effort. Analog layout constraints, such as symmetry and current paths, are identified, implemented, and tested on various circuit topologies. Experiments have shown that the proposed method reduces the design time significantly while ensuring the layout quality based on post-layout simulation. This automated RAIL implementation flow also integrates the automated System Verilog (SV) model and test bench generation (using real number modeling) from a transistor-level circuit schematic, thus ensures the correctness and fidelity of the synthesized layout.

3.1 Introduction

Analog and mixed-signal (AMS) circuits are the front end of various real-time integrated circuit systems. AMS circuits, as a core technology, for high-speed IO interface and clocking, PVT sensing, are needed for digital signal processing and computing systems. While digital design can be synthesized from RTL to layout, AMS circuit design is time consuming and heavily depends on the designer's experience, further, layout affects fundamentally the performance of AMS circuits.

Since the early 1990s, academic research has been directed towards automated analog design tools, for examples, KOAN/ANAGRAMII [1]. Those tools consider various constraints such as layout symmetry, device merging and abutment, well merging and bulk contacts, over the device wiring, crosstalk avoidance, and integrated rip-up/reroute. Placement algorithm is mainly based on annealing algorithms. Unfortunately, these efforts have stayed mainly in the academic community and no viable commercial tools exist for automating analog design as of today.

There are also tools developed for procedural layout generation based on automating analog layout steps. However, these tools require designers to be software programmers or software programmers to be designers.

The University of Washington team has developed an alternate approach, and implemented in a tool called IPRAIL, to the automated analog synthesis based on the extraction of analog designer’s knowledge from an existing layout, and then applying these knowledges expressed as constraints to automated layout generation [2-3]. IPRAIL further incorporates advanced analog design and layout constraints and structural hierarchy [4-7]. These techniques have been applied to a number of non-trial analog and radio-frequency (RF) designs with quality comparable to that of human designers.

Very recently, the University of Berkeley team has developed BAG, a process-independent schematic and layout generator [8]. The University of Texas at Austin team has developed MAGICAL, an analog IC layout system to build a layout from an unannotated netlist [9].

In this work, we have developed the RAIL layout generator. Different from our prior effort IPRIAL which was developed as a complete and independent analog layout tool, RAIL realizes the analog layout in a digital APR flow with the standard digital automated place-and-route (PNR) tools such as IC Compiler II from Synopsys and OpenRoad from the DARPA IDEA program. The additional components added to this flow include auto-generated analog standard cell (RAIL) library, procedural analog device and macro generator, analog constraint generator, and analog-constraint-aware ECO changes.

The RAIL flow has several contributions:

- Automatic design flow that takes advantage of current digital layout tools and creates analog circuit layout with layout-effect-aware optimization
- Grid-based analog device generator compatible with digital layout tools
- Built-in RAIL standard cells
- Constraint based placer, including merging/abutment, current flow and symmetry as constraint to optimize the circuit performance and manufacturability
- Incorporation of a circuit optimizer in the loop
- Incorporation of automated generation of System Verilog models for auto-generated analog and custom macros for complete system-level design verification

As a methodology demonstration, a 7-bit high-speed successive approximation ADC (SAR ADC) is realized with various DRC-clean layouts generated for a TSMC 65nm process. Further this SAR-ADC is successfully implemented in a GlobalFoundries 12nm FINFET process. Our flow, library and test cases have been open sourced on github.

The chapter is organized as follows: Section 2.2 describes the RAIL architecture and the RAIL flow. Then Sections 2.3 and 2.4 present respectively the details of analog cell generator and the built-in RAIL standard cells. The constraint-based placer and MAKE flow are described in Sections 2.5 and 2.6. Experimental and simulation results of several designs including comparators, CDAC, SAR ADC in TSMC65nm and GF12nm are presented in Section 2.7. Real number modeling and verification flow are described in Section 2.8. Section 2.9 lists the open-source resource links. Section 2.10 concludes this chapter.

3.2 Fully Automatic Layout Generation

Given a technology physical design kits (PDK) and an annotated parametric Hspice netlist, the RAIL analog layout generator produces the Verilog netlist, device layout, layout constraints and scripts to be used by digital automatic

layout tools. The layout constraints are fundamental to improve the quality of placement and route. Scripts include placement and route constraints, and are compatible with commercial or open-source EDA tools.

The analog-specific components include a grid-based device generator, a constraint-based placer and router. The layout of each individual device is first generated and then the overall solution is sought with placement and route constraints. The results are then written into script files that can be further processed in the digital implementation flow.

3.2.1 Device Generator: Annotated Hspice Netlist

The device generator reads in an annotated Hspice netlist, which includes the descriptions of the sizes (width, length, number of fingers and multipliers) of each device, and creates an initial layout. The device generator can generate the layout with various aspect ratios specified.

3.2.1.1 Constraint Based Placer and Router: the Device Sizes and the Circuit Topology

The input to the placer and router is the Hspice netlist and the physical layout of each device from the device generator. The netlist specifies the connectivity among the devices and is used to construct a graph for the circuit and to find the current-flow paths and symmetric devices and nets, which are then used to generate placement and route constraints.

3.2.2 Outputs

The RAIL tool outputs Verilog netlists, the frame file, and the script files that are needed for industry digital layout tools. With these scripts, then the digital layout tools (commercial or opensource) are used to generate the final layout in the GDSII files. Then the physical validation of the generated layout is performed using Mentor Graphics' Calibre tool for design rule check (DRC) and LVS compliance. Post-layout simulation is conducted to validate the functionality and performance o the generated layout.

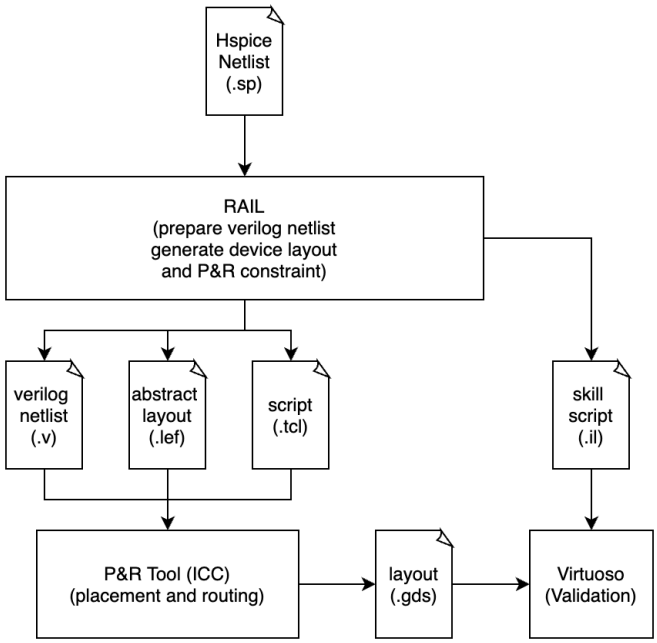


Figure 1. Input and Output of the Flow

3.2.2.1 Standard Cells

Standard cell layouts are generated following some specific set of layout rules and guidelines so that the whole layout can be easy with standard cells as building blocks. Active device layouts are generated using the standard cells. The standard cell is designed using the routing grid. The pins must be placed on the grid, so that the digital PNR tools can place and route without violating DRC rules. Cell width and height must be multiples of a single grid space. All the standard cells have the same height. Some analog transistors use double heights. All the cells are generated as more squared shapes instead of wide rectangles. Metal layer 1 is used in the cell and metal 2 is used only if needed. The substrate is connected to the power or the ground based on its type (p or n).

3.2.2.2 Macro Cells

A macro cell is more flexible in layout styles than a standard cell. Passive devices such as capacitors and resistors, and inductors are generated as macro cells. Macro cells are parameterized with design parameters specified in the annotated Hspice netlist. Then, the abstract view is generated during cell generation. It includes three terminals, and a blockage area in the middle. It prevents routing over the capacitor to reduce crosstalk and noise. The abstract view is used in the standard digital design flow for routing.

3.2.2.3 Module

Standard cell and macro cell are the lowest levels of the layout design. Modules are defined by the design built with standard cells and macro cells. The component to build the module has a fixed width and height.

3.2.2.4 Block

A block is defined as a design built with modules or blocks. The component to build the block has flexible width and height.

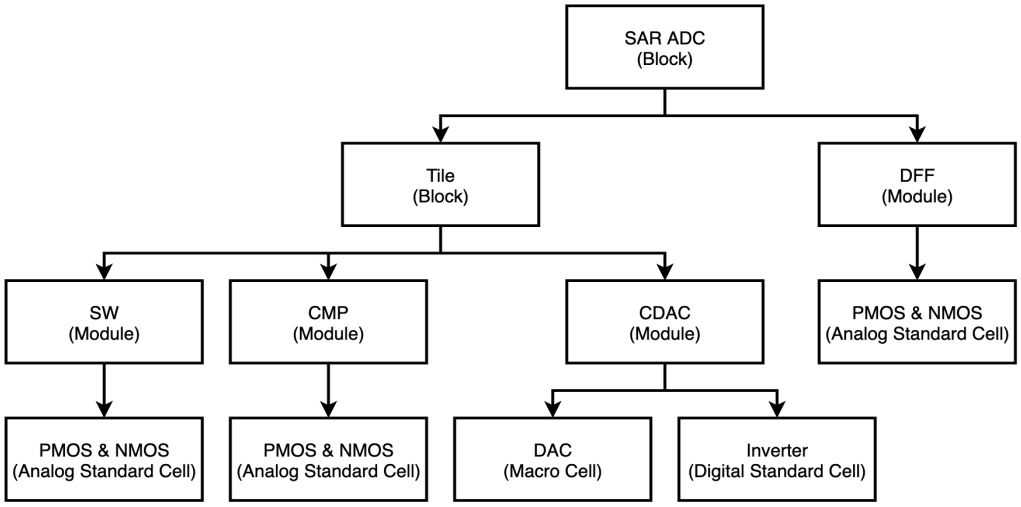


Figure 2. The Hierarchy of a SAR ADC

3.2.3 Active Devices

Device geometry refers to transistor width (W) and length (L), resistor/capacitor/inductor parametric values. The device sizes are typically given by analog designers. In the RAIL flow, device sizes are automatically obtained by the parasitic-aware circuit optimizer.

The layout of an active device follows layout design rules specified for the technology in use. All the design rules such as the minimum spacing between metals on the same layer and its minimum width are specified in terms of a uniform grid.

1) **Routing grid:** All the standard cell pins should ideally be placed on the intersection of horizontal and vertical routing grids with exceptions as abutment type pins (VDD and GND). As shown in Figure 3 (a), metal 1 (light blue) pins are placed on the vertical routing grid and metal 2 (yellow) connects all source terminals and drain terminals and is placed on the horizontal routing grid.

2) Grid spacing: The grid spacing needs to be defined for each routing layer. By default, the grid spacing is defined as two times the minimum metal width. The spacing is increased with wider metal used in the cell.

3)

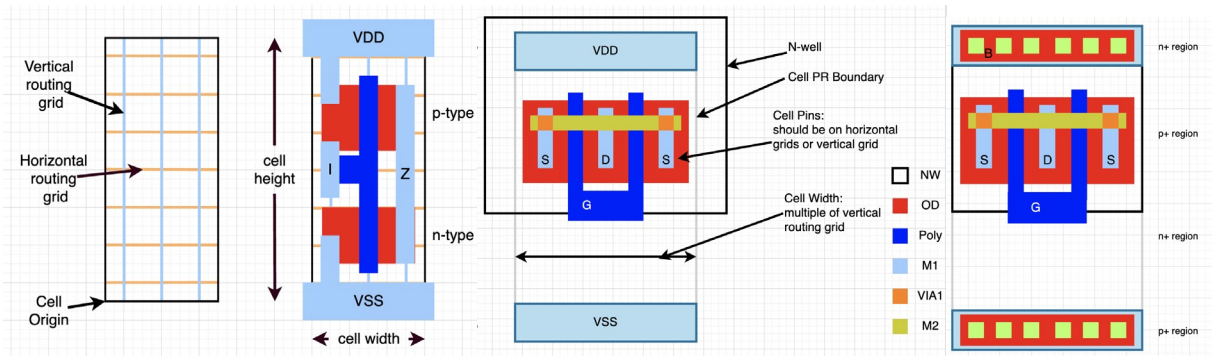


Figure 3. (a) Routing Grids (b) Inverter Standard Cell Layout (c) an Analog device in this Work (d) Biasing

4) **Cell width and height:** The cell width must be a multiple of the vertical grid spacing. The cell height must be a multiple of the horizontal grid spacing. All the cells must have the same height.

5) **Double cell height**: Some complex cells, for example, level shifter LVLLHCD8LVT in TSMC 65, can be designed with the double height. It uses 2 rows, with VDD-VSS-VDD three power rails incorporated from the bottom to the top. Such a cell can be used together with single-cell-height standard cells in a standard digital flow. Figure 4 shows an example of a double-height PMOS device. It has double height and two power rails placed on the top and bottom. The source and drain are placed on two sides and the middle routing rail is reserved for ground routing.

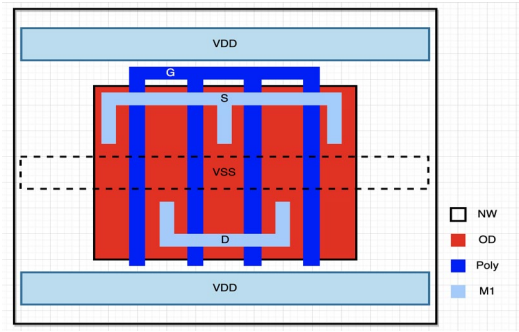


Figure 4: Pmos with Double Height

6) **Metal utilization:** Metal 1 is commonly used in standard cell net connection, and metal 2 is used only if needed. Metal 3 and higher metal layers are not used in standard cells (intra-cell routing) and are reserved for routing to connect standard cells or macros (inter-cell routing).

7) **Biasing**: Standard digital cells always bias a PMOS device to the power net and a NMOS device to a ground net. However, analog device substrate is often biased to a different voltage. For these types of devices, the RAIL tool uses power rail and ground as body pins and is placed isolated from other devices. In figure below, power rail is utilized as a bulk terminal and for this topology, the netlist is updated to include bulk terminals.

8) **Resizing**: With the Hspice netlist, the device generator figures out the best layout topology for each device. If the device has a large width size, it splits the device with more fingers, since total width = width*num of finger. As shown in Figure 5, given a device with total_w = w0 and number of finger(nf) = 2, the device can be resized to w = w0/2, nf = 4, and the total_w is the same.

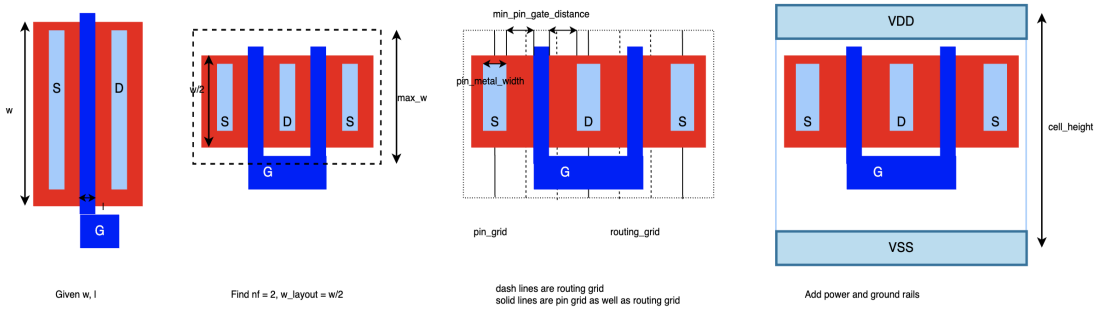


Figure 5. Device Layout with Same Total_w: (a)w = w0, nf = 2 (b) w = w0/2, nf = 4

Device layout is generated based on the design grid. Except for power (VDD) and ground (VSS) pins, metal 1(light blue) pins are placed on a vertical routing grid and metal 2(yellow) connects all source terminals and drain terminals and is placed on horizontal routing grid.

Standalone device: Different from CMOS standard digital cells where each cell is composed of several p devices and n devices, standalone cells have only one type of device, for example P-switch and N-switch. Although one side of the device is missing, to connect other cells in the same row, layers are still mandatory to be used, including NW, RX fin (in GF 12), RC (in GF 12), and power rails.

3.2.3.1 Device Parameters

Parameters of a cell are categorized to two types: process parameters and design parameters. Process parameters are defined by the fabrication process while design parameter describes a design, which are defined by an analog designer, in our flow, computed automatically based on the design specification provided. The symbol, description, default value and functions of the parameters are listed in Table 1.

Table 1. A Summary of Device Parameters

Process parameter	description	Default value	
max_w	Max width within a cell height	PMOS: 0.52 um NMOS: 0.39 um	
cell height	Standard cell height	1.8 um	
co2po	Minimum distance between CO and PO	0.055 um	
Design parameter	description	Default value	Function
w	Device width	0.2 um	(defined by user)
l	Device length	0.06 um	(defined by user)
mw	Metal width of pins	0.1 um	(defined by user)
nf	Number of finger after resizing	1	w/max_w Nf is an integer equal or larger than 1
uw	Unit width	0.2 um	uw = ceil(w/nf)
rgp	Routing grid pitch	0.4 um	Mw+ co2po*2+l Rgp is multiple of 0.2

3.2.3.2 Evaluation of Standard-Cell Active Devices

The behavior and performance of an active device can be characterized based on the simulation of its pre-layout netlist or post-layout netlist as extracted from Calibre. The Hspice netlist is read as the input and a SKILL script is generated and imported to the Cadence environment. After the layout is imported, parasitics are extracted with Calibre PEX.

Figure 6 summarizes the results of comparing a pre-layout device, the same device extracted from a PCell-generated layout, and that extracted from the RAIL layout. The device performance includes its area, threshold voltage, input capacitance and on-resistance. The MOSFET has the width ranging from the minimum size 200nm to 4.16um.

From Figure 6, we can see that the RAIL device area is about 1.2~2x larger than that of Pcell, due to the grid constraint and power/ground rail in the cell. The threshold voltage varies because the source and drain pins must be placed on the routing grid. The length of diffusion is increased, which changes the mobility of carriers. Input capacitance is reduced since the transistors are generated with more numbers of fingers and the drain is shared, and the drain capacitance is thus reduced. This is more significant in devices with larger widths. On- resistance is reduced because the device is split into multiple fingers, which are transistors (equivalent resistor) in parallel.

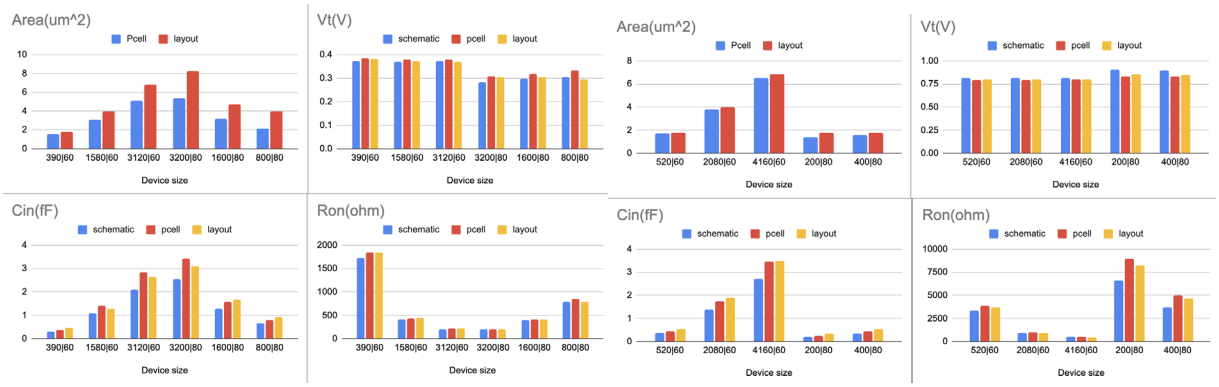


Figure 6. Device Performance of (a)NMOS (b) PMOS on Area, Threshold Voltage, Input Capacitance and on Resistance

3.2.3.3 General Capacitors

The RAIL layout wraps a p+ poly resistor and an MOM capacitor as a macro. The design parameters of MOMCAP are shown in Table 2. It includes the number of vertical fingers, the number of horizontal fingers, finger width and spacing, top, and bottom metal, and these parameters are specified in the Hspice netlist.

Table 2. Design parameters of MOMCAP

Symbol	Design parameter	Default Value in a 65nm CMOS
nv	Number of vertical fingers	1
nh	Number of horizontal fingers	1
w	Width of fingers	0.1 um
s	Space between fingers	0.1 um
stm	Top metal used	M5

The passive device layout is fixed with given parameters, as shown in Figure 7 (a). An abstract frame can be generated and can be incorporated to the digital flow for placement and route. The abstract frame includes three terminals for all passive devices, plus, minus and bulk as shown in Figure 7 (b). Metal pins are specified for over the cell routing and the rest of the metals are specified as routing blockage to prevent any routing over the area, preventing crosstalk and noise.

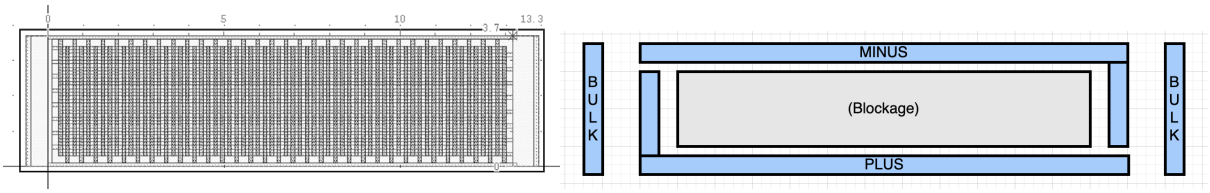


Figure 7. (a) Layout of MOMCAP (b) Frame Illustration of MOMCAP

3.2.3.4 CDAC IP

The schematic of a typical 3-bit capacitive digital to analog converter (CDAC) is shown in Figure 8 (a). There are two generally used capacitor layout implementations, MIMCAP and MOMCAP. However, MIMCAP is not area-efficient for small capacitor implementation and the minimum capacitance it generates is usually larger than 1fF. Thus, this design flow uses MOMCAP to implement CDAC and introduces a unit-switch-plus-cap DAC IP [source], as shown in Figure 8 (b). It minimizes the parasitic routing capacitance and resistance by merging the reference switches with unit capacitors, combining them into a single cell. Hence the settling is more uniform and lesser determined by parasitics and variations.

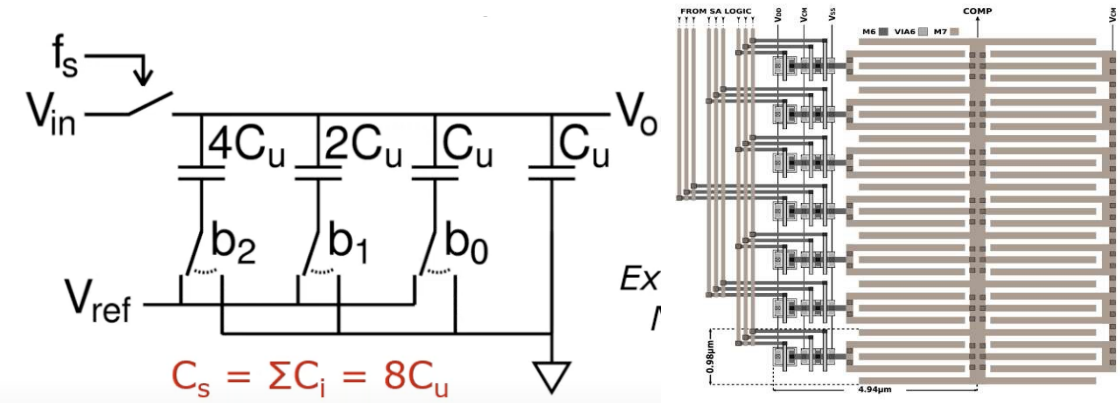


Figure 8 (a) Schematic of a Typical Capacitor DAC and (b) Layout Illustration of DAC

Using this CDAC scheme and the number of digital bits as parameters, layouts are automatically generated. The extracted capacitance on each reference switch node is shown in Table 3. The capacitance discrepancy is small.

Table 3. Capacitance(fF) of CDAC Reference Switches

	Positive side			Negative side		
	a0	b0	Discrepancy	a0	b0	Discrepancy
16Cu	3.953	3.954	0.000	3.951	3.957	0.007
8Cu	1.975	1.977	0.002	1.980	1.977	0.003
4Cu	0.992	0.988	0.004	0.990	0.989	0.001
2Cu	0.495	0.494	0.002	0.498	0.495	0.004
Cu	0.237	0.237	0.000	0.240	0.254	0.013

An alternative way to generate the layouts of capacitors is to follow the standard cell layout rules similar to what is done for active devices. As shown in Figure 9, the blue cap cells are the active unit capacitors, the green cap cells are the dummy unit cap cells and the cell on the bottom row with blue outline are the standard cell capacitors. The routing is preset for this mesh pattern. These standard cells share the same cell height, which eases the routing of power and ground. The standard cell layout property also yields a more compact layout integration.

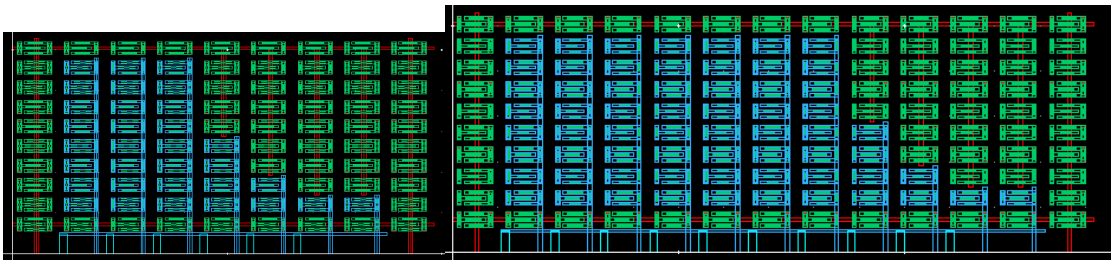


Figure 9. Example of using Unit Cap to Generate CDAC6b and CDAC 7b in GF12

3.3 RAIL Standard Cell Library

This section describes the RAIL analog standard cell library. It includes for each ell its features, functionalities, characteristics, and expressions. Cells in this library are designed using the standard cell metal resources (metal 1 and metal 2); place and route tools can use the layers of upper metals for inter-cell routing.

3.3.1 TGAT: CMOS Transmission Gate

The Transmission Gate is a switch widely used in AMS designs. The unit CMOS transmission gate can be created by truncating the MUX2 cell.

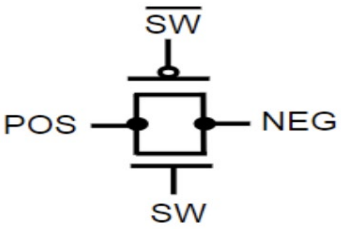


Figure 10. Transmission Gate

3.3.2 PSWX/NSWX: Single Transistor Switch

Single transistor switches are also very common in AMS designs. For example, an NMOS switch can serve as the core transistor for the bootstrapped switch. The symbol of a /PMOS single transistor switch is shown in Figure 11.

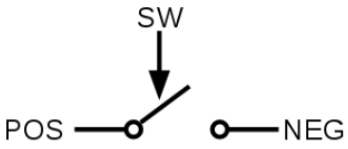


Figure 11. Single Transistor Switch

3.3.3 MX2U: Pull-up Analog MUX 2

A pull-up analog MUX2 is to set a node either set to an analog signal or a logic high. The logic high selection is normally used in reset/pre-charge cases. The analog signal selection is implemented by a typical transmission gate. Figure 12 shows the symbol for a pull-up analog MUX2.

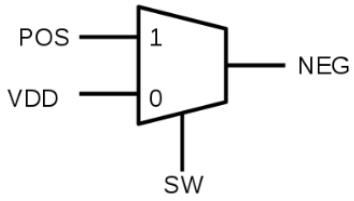


Figure 12. Pull-up Analog MUX2

3.3.4 4MX2D: Pull-down Analog MUX 2

A pull-down analog MUX2 is to set a node either to an analog signal or a logic low/ground. The logic low selection is normally used in reset/pre-charge cases. The analog signal selection is implemented by a typical transmission gate. Figure 13 shows the symbol for a pull-down analog MUX2.

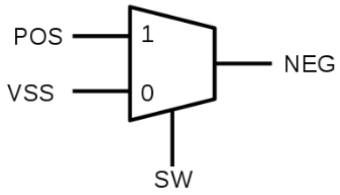


Figure 13. Pull-down Analog MUX2

3.3.5 TBUF: Tri-state Buffer

Tri-state buffers are used for many analog purposes, for dynamic settling circuits, as a switch driver for cap arrays. Its output is either logic high, logic low, or high impedance. The standard cell library provide a reference design, known as BUFT. However, sometimes this may not meet the requirements of analog design.

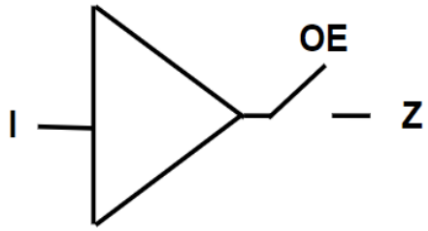


Figure 14. Tri-State Buffer

3.3.6 SWFD: Switch for Driver

Typical tri-state buffers may not meet the requirements for charge domain applications. In other cases, it may be difficult to synchronize the enable signal and data input (especially in asynchronous design). The RAIL library's solution uses a straightforward method that contains one PMOS and one NMOS. These switches serve as drivers for wireline transceivers. A current-mode push-pull driver can be created by connecting the switch with current sources. Alternatively, if voltage sources are connected to the switch, a voltage-mode driver is created. Figure 15 shows the symbol for the Switch for Driver.

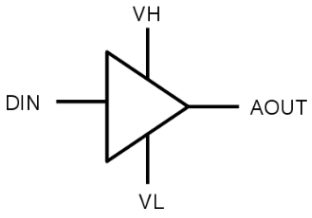


Figure 15. Switch for Driver

3.3.7 PXRO: LDO Power Switch with RO Unit

When designing a digital LDO, it is necessary to have a power switch/transistor to control the regulated voltage. To this end, we modify the tri-state buffer to form a power switch. Note that the regulated voltage needs to be sensed or converted into digital codes. In this case, a ring-oscillator is desired. As a result, the extra NOR gate in the tri-state buffer is replaced with an output voltage supplied RO cell.

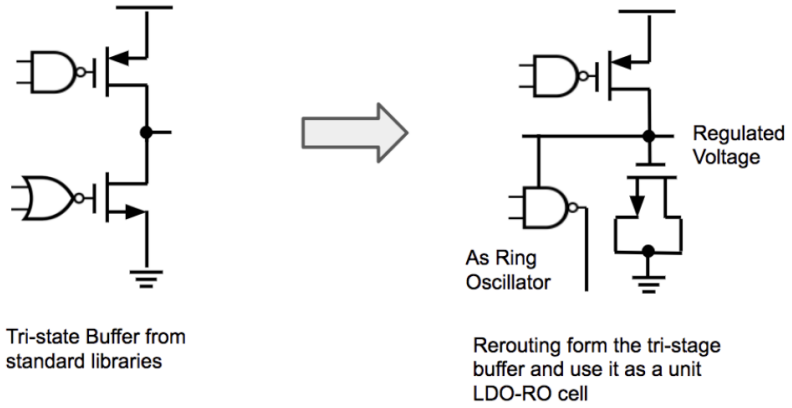


Figure 16. The Change from the Tri-state Buffer Topology to the LDO Power Switch

3.3.8 DLY1: CMOS Delay Cells

Normally, a digital standard cell library provides a variety of delay cells used for timing closure. These cells are also useful when synthesizing a digital controlled delay line (DCDL). When building a RAIL library, it is highly recommended to re-use these cells.

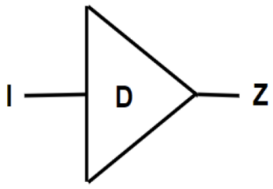


Figure 17. CMOS Delay Cell Symbol

3.3.9 VCDCX/VCDPX: Voltage Controlled Discharge Cell

A significant portion of the digitizing AMS design is the conversion of traditional voltage domain analog signals into pulse width presented time-domain signals. Therefore, a voltage to time converter is important. Figure 18 shows a typical voltage-controlled discharge cell. When CLK is low, Vout is precharged to power supply. When CLK rises, Vout is discharged via the voltage modulated transistor. The lower Vin, the slower Vout falls. This circuit is widely used in ring oscillator designs and can serve as the dynamic preamplifier of latch-based comparators. In addition, with an extra discharge path (without preset transistor), it is transformed into a half circuit for a time-difference amplifier.

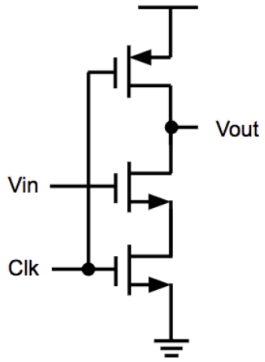


Figure 18. The symbol of voltage-controlled discharge cell

3.3.10 Static Common-Source Amplifier Half Cell

The voltage-controlled discharge cell has the same topology as a typical common source amplifier. (Recall the five-transistor amplifier common in analog textbooks.) Interestingly, this same circuitry can be used in simple static amplifier design if biased properly. The only difference is the additional node, Vtail, used as analog I/O. It is used for AC ground (tail point) / resistive degeneration connections. Though the cell can work in traditional analog designs, we emphasize that the primary focus o the RAIL project is on the digitized AMS design.

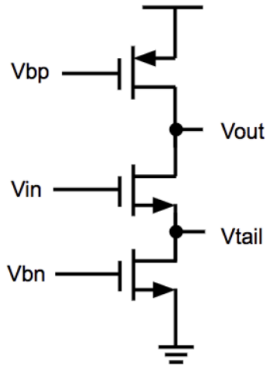


Figure 19. The Symbol of the Static Common-Source Amplifier Half Circuit

3.4 Constraint Based Placer

This section we describe several techniques implemented in the RAIL constraint-based placer. We note that a good placer is essential to a lot of blocks in digitalized AMS design, for example, high-speed SerDes driver and receiver stage, as shown later for LPDDR4X/UCIe designs.

A challenge example used in this chapter is a 7-bit high-sped SAR-ADC, and layout optimization is the key to achieve the required performance. The comparator is a key component in high-speed SAR ADC. The delay of the

comparator is a major influence on the speed and its noise determines the accuracy. This ADC design has in total 7 comparators, which makes the power of the comparator a significant contributor to total power of the ADC. This design flow implements device merging, current-flow and symmetry constraints to optimize the layout performance for such a performance-demanding layout-sensitive design.

3.4.1 Merging

Device merging and abutment is permitted during placement. It not only increases layout density and reduces routing parasitics, but also increases cell performance. The placer can merge complex structures instead of a library of individual devices.

The implementation is illustrated in Figure 20. Two devices share the same source and are placed separately before merging. The devices can be merged during placement and since vertical source pins are placed on the vertical routing grid, other pins are kept on the grid as well after merging.

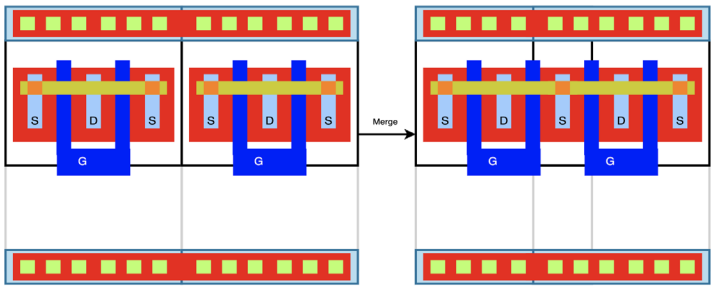


Figure 20. Device Merging

3.4.2 Current Flow

Current flow sets the constraint of the devices along the y axis. This constraint is implemented with monotonic routing from signal paths from the power to the ground, this reduces both the interconnect wire length and routing-induced parasitics and improves the post layout circuit performance.

Given a design schematic as shown in Figure 21 (a), the current flow path is found from the power to the ground to construct current-flow constraints as shown in Figure 21 (b). All current-flow constraints should be monotonic from the power to the ground. Placement constraint relating to the y coordinates of each cell can be represented and one of the layout representations that satisfies such a constraint is shown in Figure 21 (c).

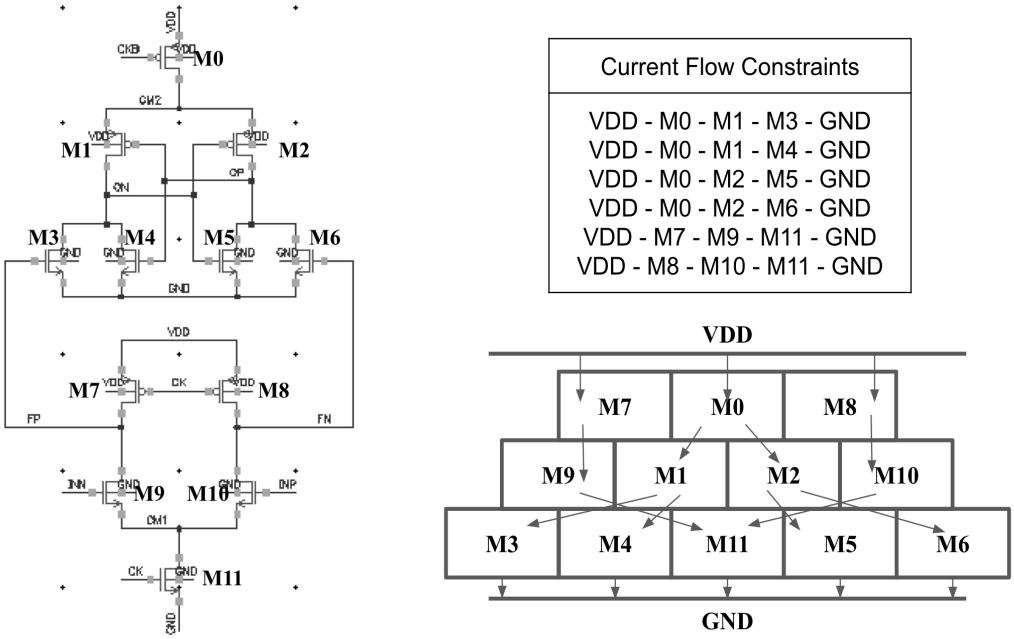


Figure 21. (a) Schematic (b) Current Flow Constraint, and (c) Layout Representation

3.4.3 Symmetry

Symmetry restricts devices to a mirrored placement and sets the device constraint along the x axis. Symmetry is used to offset geometric and electrical issues and helps reduce the sensitivity to on-die thermal gradients and parasitic mismatches between two identical signal flows. This design flow utilizes graph theory to find symmetry pairs. This section describes basic graph concepts, symmetry pair detection using graph theory and a demonstration of a two-stage comparator used in the 7-bit SAR ADC. More advanced and detailed description of this method and other techniques can be found in our early papers published on IPRAIL [4] and Frosty [5].

3.4.4 A Graph-based Method

A *graph* is made up of *nodes* connected with *edges.* The node degree is the number of edges connected with a node. For example, the graph in Figure 22 has 3 nodes and 2 edges, where node B has the degree of 2 and both node A and node C have the degree of 1. Then, for each node, suppose the degree sequence for a node is [degree1, degree2, degree3], the signature S is defined as

$$S = (\pi + \text{degree1})(\pi + \text{degree2})(\pi + \text{degree3})$$

where π can be any transcendental number (e, π, etc.)

In this example, suppose that π is 1 for simple illustration, S(A) = (1+degree(B)) = (1+2) = 3, S(B) = (1+1)*(1+1) = 4 and S(C) = (1+2) = 3. To find equivalence of two nodes, simply compare numerical signatures of nodes. Node A and node C are considered equivalent, since S(A) = S(C)

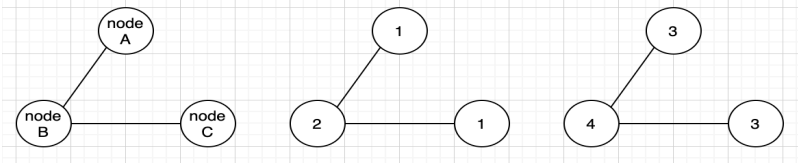


Figure 22. Illustration of Graph, Degree Map and Signature Map

3.4.5 Symmetry detection using a graph-based method

Each device in an analog circuit is labeled as a node and nets that connect devices are labeled as edges. For example, in Figure 23, M4 is connected with M1, M2, M3, M5 and M6, so the degree of M4, degree(M4) = 5. Assuming π = 1 to simplify the calculation, the signature of M4 is calculated with

Signature(M4) = (1+degree(M1))*(1+degree(M2))*(1+degree(M3)) *(1+degree(M5)) *(1+degree(M6))

=(1+6)*(1+6)*(1+6)*(1+5)*(1*6)

= 14406

After completing the signature map, find equivalent nodes by comparing signatures of nodes. In this case, M4 and M5 can be characterized as equivalent, since they have the same signature 14406.

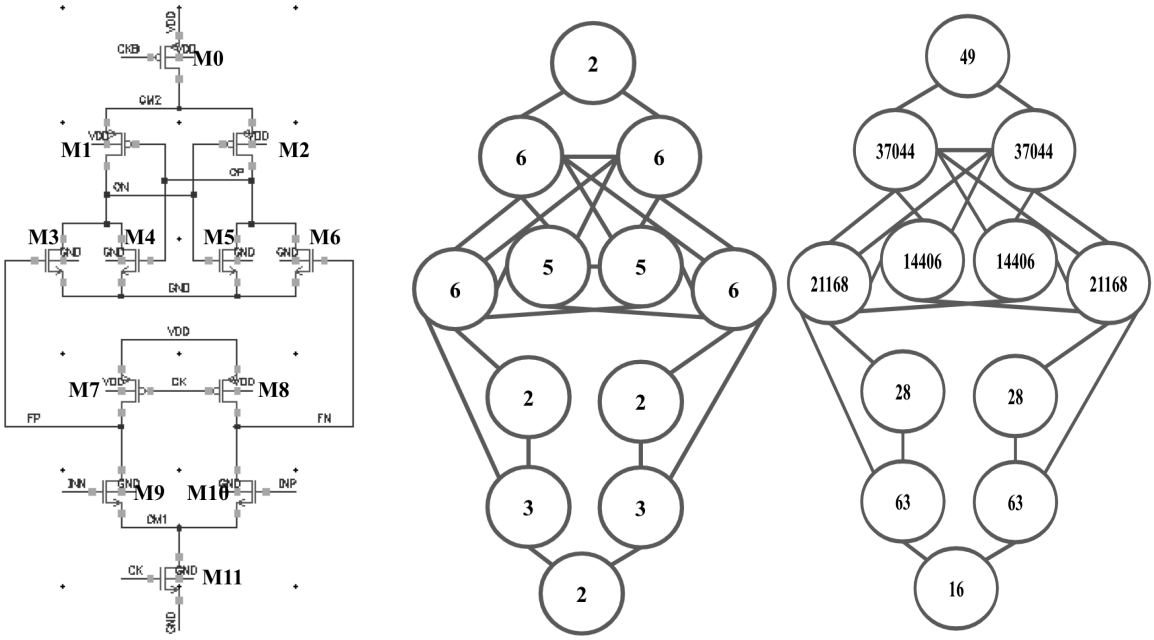


Figure 23. (a)Comparator Schematic (b) Degree Map (c) Signature Map

**assuming π = 1*

Comparing the signature only could lead to false positives. Thus, current flow paths are also used to enhance the algorithm. For two current paths, if (I) number of nodes equals, (II) for all equivalent devices, device sizes (width, length, and number of fingers) are equal and (III) signature S of equivalent devices equal. Then these two current paths are considered as equivalent. In this comparator, VDD - M0 - M1 - M4 - GND and VDD - M0 - M2 - M5 - GND are considered as a pair of equivalent paths.

After equivalent-path pairs are found, symmetry device pairs, self-symmetry devices, and symmetry nets can be labeled. Since VDD - M0 - M1 - M4 - GND and VDD - M0 - M2 - M5 - GND are equivalent, matching devices are considered as symmetry pairs (M1&M2, M4&M5), device connected with symmetry pairs is characterized as self-symmetry device(M0), and matching nets are considered as symmetry nets (ON&OP).

3.4.6 Current Flow Enhanced Equivalent Device Finder

Besides comparing signature values of two nodes to find the equivalent relation, current flow paths are used to enhance the algorithm. For two current paths, if (I) number of nodes equals, (II) for all corresponding devices, device sizes (width and length) are equal and (III) signature S of corresponding devices equal. Then these two current paths are considered as equivalence. In the case below, for example, Vdd-M5-M1-M7-VSS and Vdd-M6-M2-M7-Vss are considered as a pair of equivalent paths.

After equivalent path pairs are found, symmetry device pairs, self-symmetry devices, and symmetry nets can be found. In the following case, since Vdd-M5-M1-M7-VSS and Vdd-M6-M2-M7-Vss are equivalent, corresponding devices are considered as symmetry pairs (M5-M6, M1-M2), corresponding nets are considered as symmetry nets (X, Y) and if the corresponding devices points to the same device, the device is self-symmetry (M7).

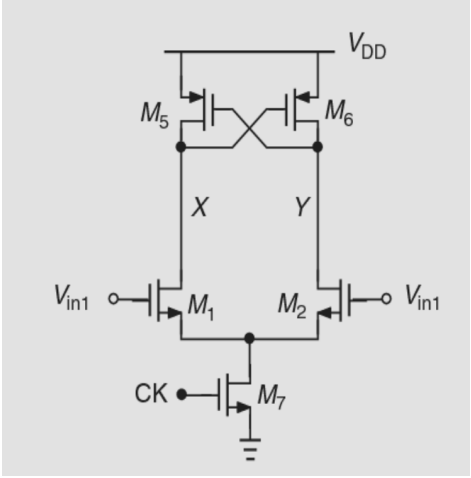


Figure 24. An Example of Current Path Searching

For each symmetry pairs (M1, M2), the constraint is formulated as below:

$$x(M1)+x(M2) ==0,\ y(M1) = y(M2)$$

In this case, the constraint can be formulated as below:

$$yM5=yM6,\ xM5 +xM6 = 0$$

$$yM1=yM2,\ xM1+xM2 = 0$$

$$xM7+xM7 = 0$$

3.4.7 Grid Placement

The blocks are placed on an even number of rows and the block heights are even standard cell height. This prevents short on the power and the ground.

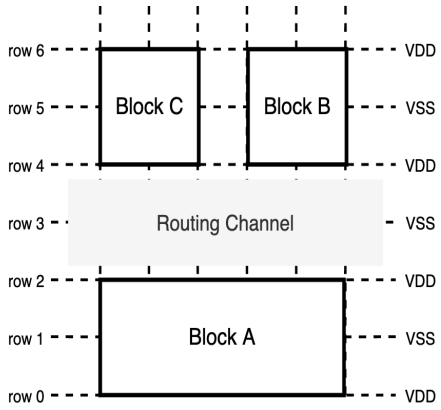


Figure 25. Grid Placement

3.4.8 Proximity and Isolation

Proximity and isolation constraints are used for wells. For standard digital cells, the bulk of each device is connected to either the power rail (p type) or the ground rail (n type). By placing standard cells in rows, the bulks are automatically connected. However, in analog circuits, the bulk is biased to a different voltage and placing such cells could cause short circuit and isolation constraint is used to avoid such case. As shown in Figure 26, the PM2 bulk is biased at a different voltage other than the power and the ground, and the vdd power is occupied as the bulk pin, which is generated from the RAIL device generator.

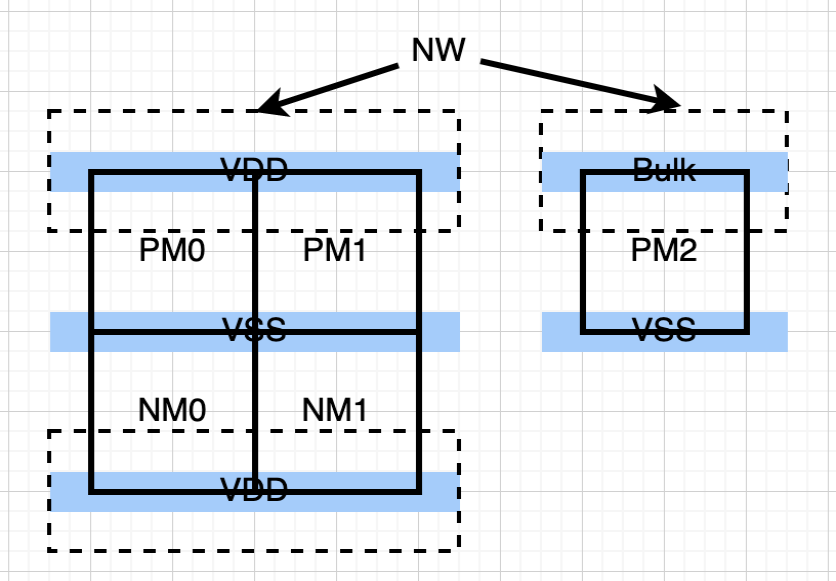


Figure 26. Placement of Standard Cell with Isolation Constraints

3.5 Setup

<technology>.lef / railxxlib.lef:

Include all the metal and via used, but not the physical information below metal layer. It is used for digital CAD tools for placement and route. The technology.lef is the standard cell lef file provided by the vendor and railxxlib.lef includes the physical information of the cells generated by this automated flow.

<technology>.mk

<technology> is the technology used in current design, for example, GF12, tsmc65lvt, etc. This file specifies the library path and the decap/filler cells to be used in the design.

<design>.mk: (required)

Basic design specs, including the core width/height and power/ground name. <design> is the name of a design block that designers are working on, for example, COMP, CDAC, etc. The same naming strategy applies to other file names below.

<design>.v: (required)

Verilog netlists for the target design. This Verilog cell can be exported from Cadence's NC-Verilog or generated with **make netlist** command for certain pre-built circuits.

<design>.json: (optional)

User-defined layout constraints and the constraint is written as a dictionary data type. For example, to define array-like placement constraint, define the constraint type as "rp_group", and define the grid size, the x/y coordinates of the rp group and the instances used in this group.

<design>.tcl: (optional)

User-defined pin constraints. The syntax follows the syntax of the ICC2 script.

<design>.1.tcl: (optional)

Arraylike place and route constraint written in the ICC2 syntax. This script is automatically generated by the flow.

3.5.1 Make Flow: APR

Make library

The IC compiler II reads the library frame file (.lef) as the input to create a cell library for the tool to place and route. Three frame files are read here: digital standard cell, extended standard cells, and IO pad cells. The designer can edit the railxxlib.lef file to add/edit a cell to the library. Then use **make update library** to update the library with updated information.

Make floorplan

The IC compiler II reads in the Verilog file of the design. Designers need to make sure that the syntax of the Verilog file meets the requirement of the IC compiler platform. The digital tool automatically links the design, then reads the width and height included in design.mk file, and creates an empty floorplan.

Make place

The IC compiler II first runs the design.tcl file, which includes the most important constraints of the design, including pin constraints, power-ground route, placement of top-level blocks. Then, run a python script to create additional layout constraints that is specified in design.json file, including grid place and route. The output of python script is a tcl file that is then read into the IC compiler for cell placement and route. After the cells with constraints are placed, run create_placement to place the rest of the cells. Finally place all the pins in the block using the place_pin command.

Make route

This step first routes power and ground rail in the standard cell pattern. For each technology node, the height is different. Then run route_auto to route the rest of nets.

Make finish

This step automatically inserts tap cells, decap and filler cells, which are specified in the technology.mk file. The user needs to define the name of which cells are used in this step. Once all the cells are inserted, the layout in GDSII file and updated Verilog netlist, including decap cells that are inserted in IC Compiler platform, are created.

Make export

Before importing to Virtuoso for further verification, the GDSII layout needs to merge with standard cell layout, since it only contains the placement of cells and metal routing, but not layers below metal layers.

3.5.2 Make Flow: Verification

Make setup

Create a new directory and copy all the tcl scripts to run strmout/drc/lvs/pex and pre/post layout simulation. The design name is set in a makefile and drc/lvs/pex rules are set in the given scripts, the user needs to modify the rule path based on one's own system.

Make run

This step includes drc check, lvs check and run pex to extract parasitics for verification. This command can be separated into four commands: make strmout, make drc, make lvs, make pex.

- **Make strmout**: The user can choose to export the gds from the target library and the design from virtuoso. The library and design name can be set with makefile and running this command will export the gds from the "layout" view. The user can also skip this step if an gds file is already created and saved in the strmout folder.
- **Make drc**: Once the gds is ready in strmout folder, the user can run "make drc" to check if there are any design rule violations.
- **Make lvs**: Before running this command, the user needs to make sure that the Hspice netlist can be found in the lvs directory for verification. Then running the "make lvs" command will compare the gds in the strmout folder with the hspice netlist. Once the comparison is done, the lvs report will be printed on the terminal.
- **Make pex**: Once the user confirms the lvs passes with no error, "make pex" command can be used to run parasitic extraction for post-layout simulation.

Make sim

The user needs to create and save a Spectre testbench in the sim folder, where the user can set up the inputs and choose which output to be saved. More results including calculating delay and power can be set with plot.ocn, which is a ocean file for simulation and analysis. A sample script is included in the flow and the user can modify the script based on one's need.

Make postsim

Once the pre-layout simulation is done, the user can run "make postsim" for post-layout simulation. This command will run the simulation with the same configuration in pre-layout simulation but include the parasitics during the simulation. The parasitic file is generated from "make pex" and the configurations are inherited from pre-layout simulation.

3.6 Experimental result on a two-stage comparator in TSMC65nm

The quality of layouts generated by this design flow is compared with tape-out quality manual layouts by experienced analog IC designers using the same testbench. Current flow and symmetry pairs are detected,2 and constraints are used for layout generation.

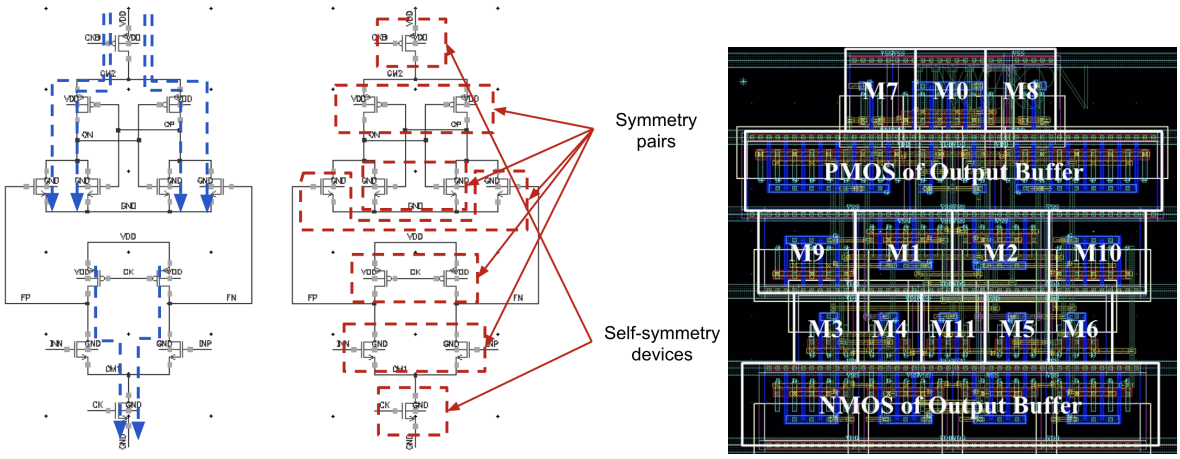


Figure 27. Two-stage Comparator (a) Current Flow Detection (b) Symmetry Detection (c) Auto-generated Layout

With the symmetry routing constraints realized in ICC2, the mismatch of the symmetrical pairs routing length is as low as 3% and the post-layout simulation confirms that the quality of generated layouts is close to that of manual layouts.

Table 4. Symmetrical Nets Routing Length Mismatch

Wire name	Length	mismatch
VON|VOP	10.683 | 10.482	1.8%
INP | INN	6.670 | 6.470	2.9%
NET_OP | NET_ON	19.165 | 19.565	2.08%
NET_FN | NET_FP	7.503 | 7.503	0%

Performance can be further improved by generating more compact device cells, which can reduce routing length and capacitance parasitics.

Table 5. Simulation Comparison of pre-layout, Manual Designed Layout and Auto-generated Layout

	Schematic	Manual	Auto-generated
Area(u^2)	/	88.2	124.2
Power(uA)	71.64	116.4	135.2
Delay(ps)	83.9	155.5	143.2
Noise(m)	1.5	4.2	2.32
Offset(mV)	0.6	2.2	11.83

3.6.1 NAND Comparator

The analog circuit is not always built from devices. In this experiment, the designer designed a NAND comparator with 3-input NAND gates and inverters. All the cells are taken from the standard cell library as shown in Figure 28.

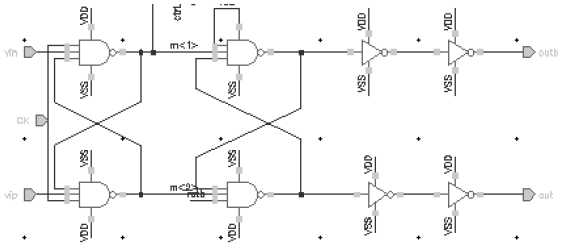


Figure 28. The Schematic of a NAND Comparator

With the automated flow, there are two strategies to create such a design: separate the standard cells into devices, which gives more freedom to place each device, or use the layout provided in the standard cell library. The layout generated with these two strategies are shown in Figure 29 below. The layout built from devices is shown on the left. It applies the current flow and symmetry constraints. The device's layout is wrapped as single standard cells and PMOS are placed on the top rows and NMOS are placed on the bottom rows, which follows the current flow constraint. The layout shown on the right is built with given standard cells and it only follows symmetry constraint since the layout of each cell cannot be modified.

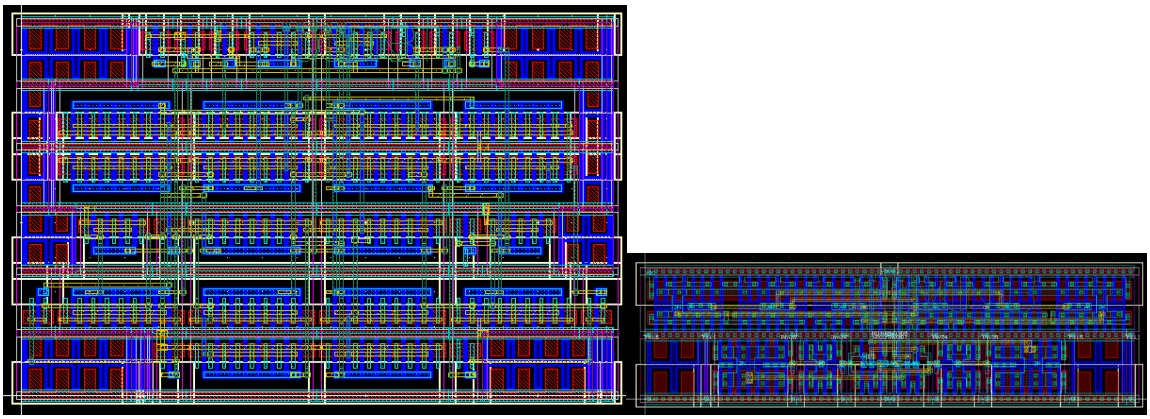


Figure 29. The NAND Comparator Layout Left: RAIL Layout, Right Standard Cell Symmetry Layout

The simulation result is summarized in Table 6. The first strategy occupies more area since each wrapped cell only contains either PMOS or NMOS, which makes using pre-made standard cells more area efficient. However, the first strategy has higher performance on all the dimensions, delay, power, and noise, due to less routing parasitics.

Table 6. Summary of Pre-layout and Post-layout Performances

	Schematic	RAIL	ICC symmetry
size	-	17x11	13.6x3.6
delay	258.9ps	175ps	388ps
idd	206.7u	202.3u	223uA
kickback	409uV	262uV	314uV

3.6.2 Comparator in GF 12

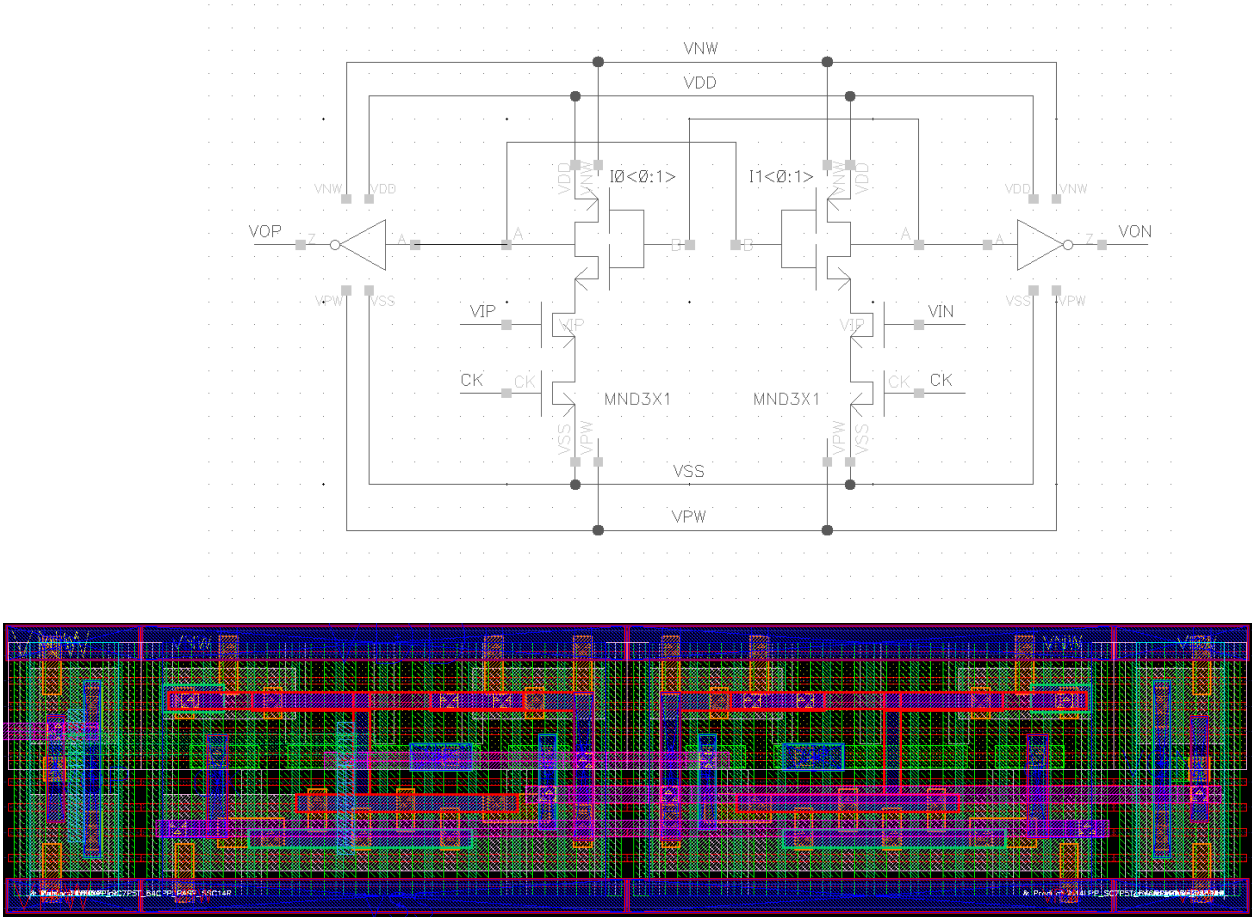


Figure 30. Schematic and Layout of Comparator in Global Foundry 12nm

The cell placement follows symmetry constraints.

Table 7. Summary of Pre-layout and Post-layout Simulation Results

	Power (uW)	Delay (ps)	Kickback (uV)	Diff. Kickback (nV)	fck
Pre-layout	6.92489u	36.0564	276.455	-6.18838u	2.5G
ICC post layout	12.5265u	108.149	1166.47	-27.5088u	2.5G

3.6.3 CDAC with Parameterized Sizes

The CDAC is wrapped as a parametrized IP. Figure 31 shows the two versions of the CDAC generated with increasing finger lengths. The second version increases the unit capacitance and modified the placement, which lowers the mismatch between edbits.

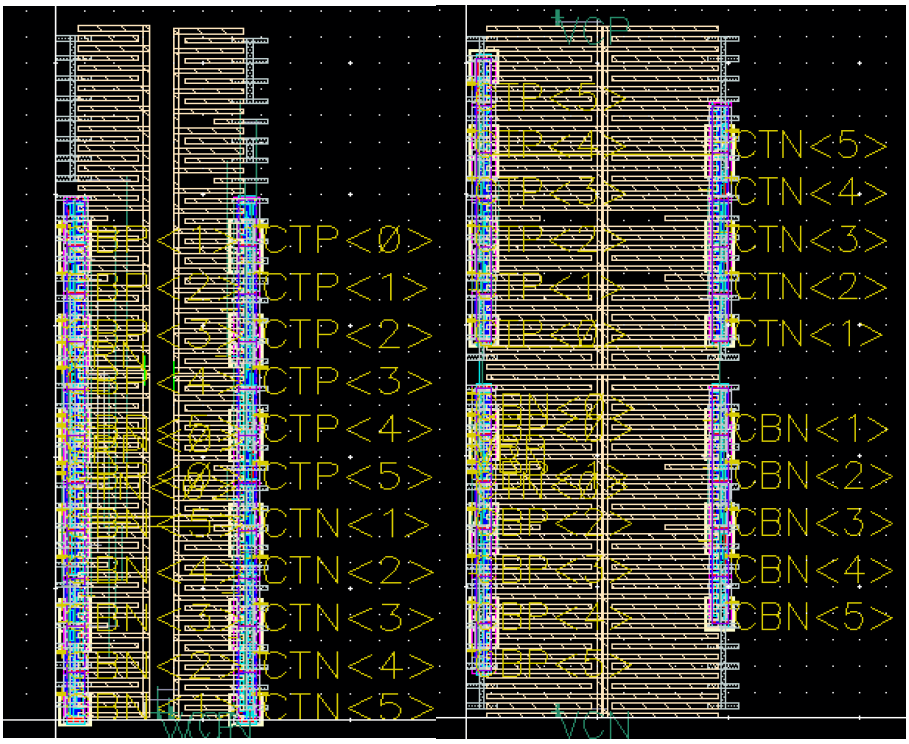


Figure 31. Left: Version 1 with Smaller Finger Length Right: Version 2 with Increased Length

Table 8. Extracted Capacitance of Version 1

Version 1 (fF)	TP	TN	Error	BP	BN	Error
5	1.9882	1.9880	0.000	1.99126	1.99252	0.001
4	0.9.95666	0.993610	0.002	1.00607	0.99411	-0.012
3	0.512206	0.496567	0.031	0.569522	0.49751	-0.145
2	0.2.94015	0.2.51675	0.168	0.246814	0.248014	0.005
1	0.113765	0.118459	-0.040	0.113765	0.118326	0.039

Table 9. Extracted Capacitance of VERSION 2, with the Increased Cap Size & Symmetry Constraint

Version 2(fF)	TP	TN	Error	BP	BN	Error
5	3.95331	3.95361	0.000	3.95059	3.95723	0.002
4	1.97483	1.97711	-0.001	1.98038	1.97691	-0.002
3	0.991875	0.987988	0.004	0.989807	0.988780	-0.001
2	0.495264	0.493565	0.003	0.498435	0.494606	-0.008
1	0.236535	0.236524	0.000	0.240499	0.253509	0.051

3.6.4 SAR ADC in TSMC 65nm

We implemented a 5bit SAR ADC in TSMC 65nm. The layout has a size of 80x20um and pre-layout simulation shows the ENOB as 4.8. The initial post-layout simulation achieved the ENOB of 3.827. Then we increased the capacitance of the CDAC unit cap and the bootstrapped switch capacitor, the layout can be generated rapidly. Post-layout simulation indicates the ENOB of 4.655.

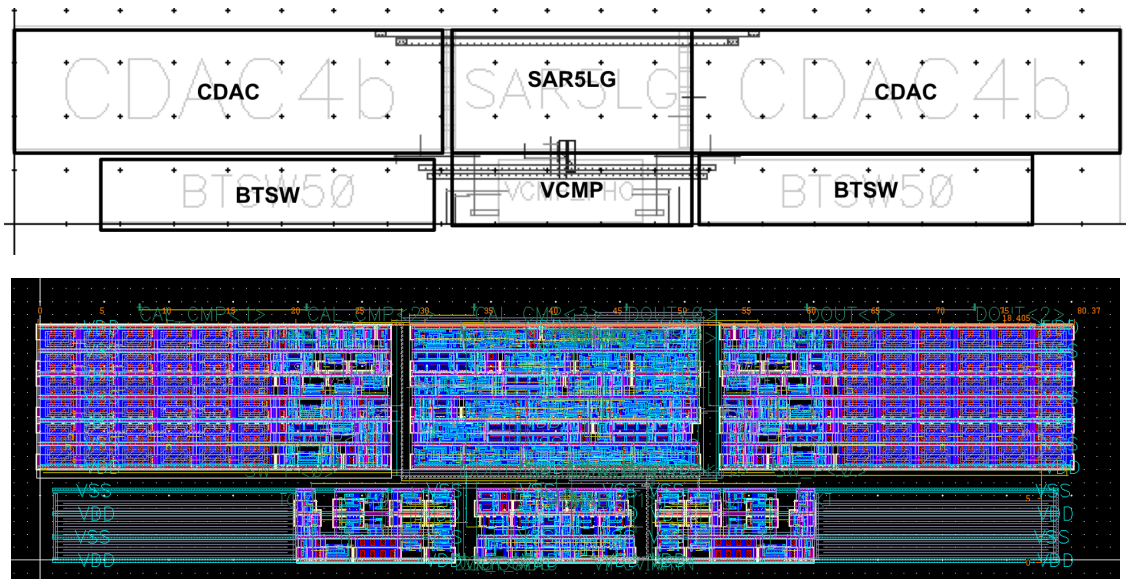


Figure 32. Floorplan and Screenshot of SAR ADC 5b Layout

The bootstrapped switch in this design uses the isolation constraint to avoid short circuits on the body, highlighted in the yellow box below, and the capacitor in the switch is created in ICC2 with specified parameters, for example, metal layer and finger length.

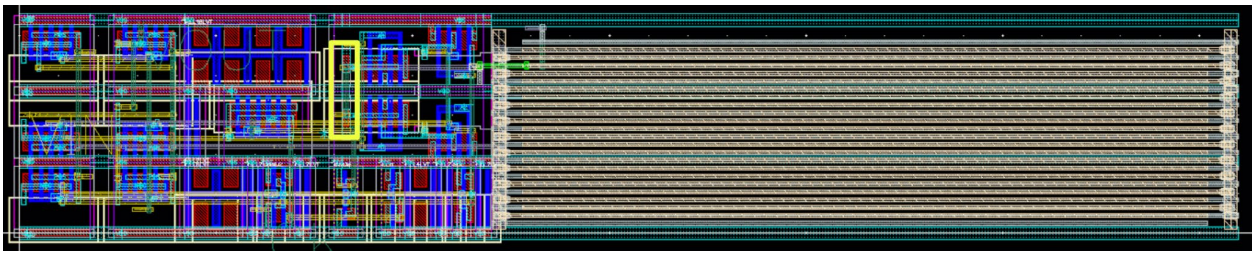


Figure 33. The Layout of RAIL-generated Boot-strapped Switch

3.6.5 Unrolled SAR ADC in TSMC65nm

We implemented the proposed layout design flow in Python. All designs are in a TSMC 65nm technology, extracted for parasitics with Calibre PEX, and simulated with Cadence Spectre.

We have collected the experimental results of this framework with device generation and three designed blocks, including two-stage comparator, CDAC IP and 7bit SAR ADC. Two stage comparator is fully automatically generated with the annotated Hspice netlist and optimized with current flow and symmetry constraints. Capacitor DAC is a passive device IP block, which is widely used in mixed-signal design for signal processing.

We have tested different floorplans of the same SAR ADC design. First, we introduce the floorplan as Figure 34. However, the ready signal passed from tile 2 to tile 2 is significantly large compared to ready signal between other two tiles due to the deficiency of the floorplan. This causes the ENOB to be as low as 3.09.

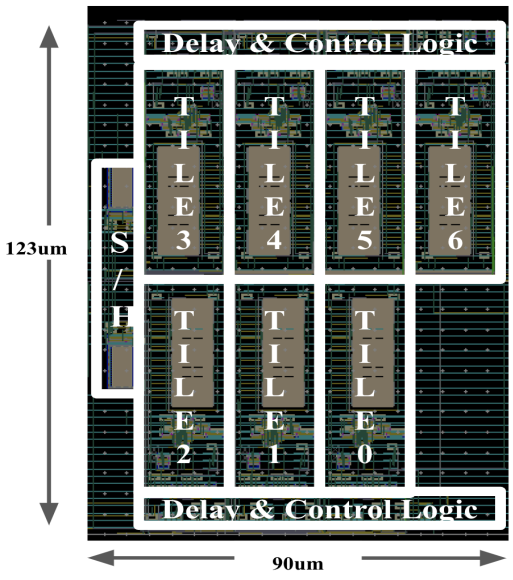


Figure 34. SAR ADC floorplan version 1

Then we came up with the second floorplan as Figure 35. The 7-b SAR ADC occupies an area of 100um x 83um and the layout view is shown in Figure 35. The floorplan of the ADC has been modified to reduce long wires in the critical path to increase speed.

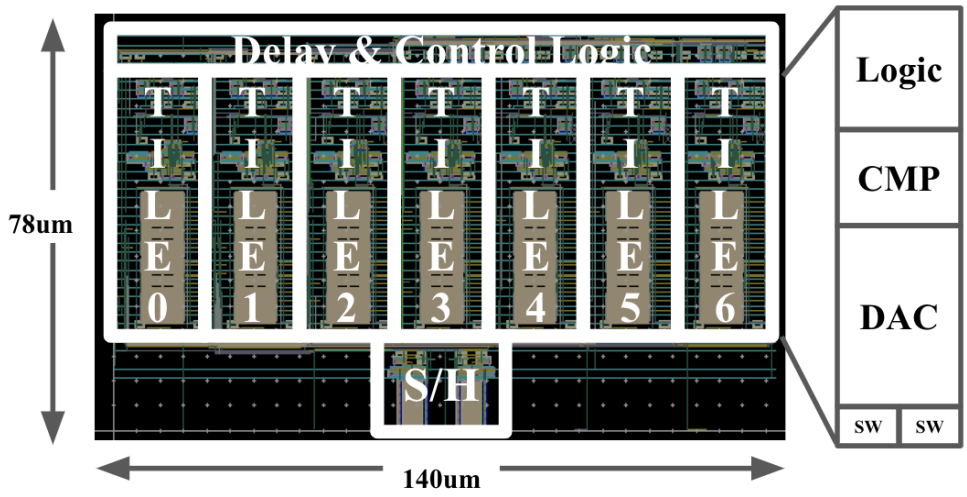


Figure 35. 7-b SAR ADC Layout View with Zoomed-in View of a Tile

The area of auto-generated layout is 10920um^2 and compared with manual layout area, which is 5428um^2, the auto-generated layout has an area overhead of 50%. The main overhead comes from grid-based cell generation, where the device itself has an area overhead to place the pins on the routing grid.

Power and sampling frequency can be improved by increasing the routing width to reduce parasitic resistance. Accuracy can be improved by improving the matching on CDAC after routing.

Table 10 Performance Comparison of Pre-layout, Manual Layout and Auto Generated Layout

Design	Schematic	Manual Layout	Auto-generated
Size (um x um)	N/A	118 x 46	140x78
Power (uW) @ 470MHz	592.08	910.32	1326
Sampling Frequency (GHz)	1	0.65	0.47
ENOB	6.474	6.43	5.32

3.6.6 SAR ADC in GF 12

We implemented an 8-bit SAR ADC in GlobalFoundries 12nm. This design utilized three power domains: VRP/VRN, DVDD/DVSS and AVDD/AVSS and is composed of following blocks:

- CDAC: capacitor DAC
- Logic: SAR Logic
- CKGen: clock generation
- BSW: bootstrapped switch
- Comp: comparator + Calibration
- DataCollector: increase output drivability

The initial floorplan and the layout are shown in Figure 36 below: two bootstrapped switches take differential input signals and require symmetry constraints. All other blocks are placed in the middle, following signal flow constraints.

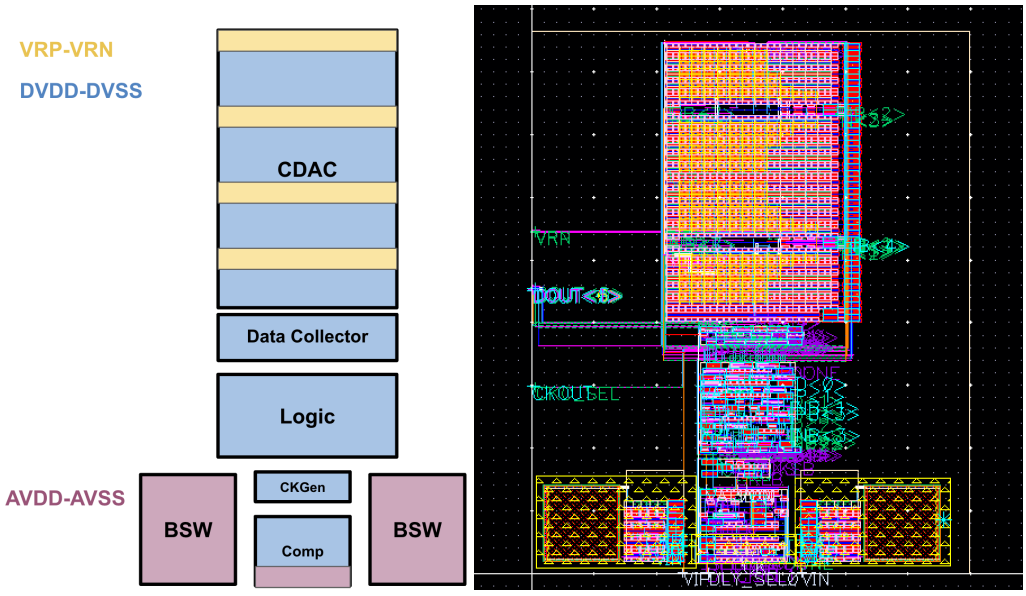


Figure 36. Layout Version 1 size:34x42

Due to the flexibility of the floorplan, we changed the floorplan as shown in Figure 37 below, to make the area more efficient and reduce the routing length.

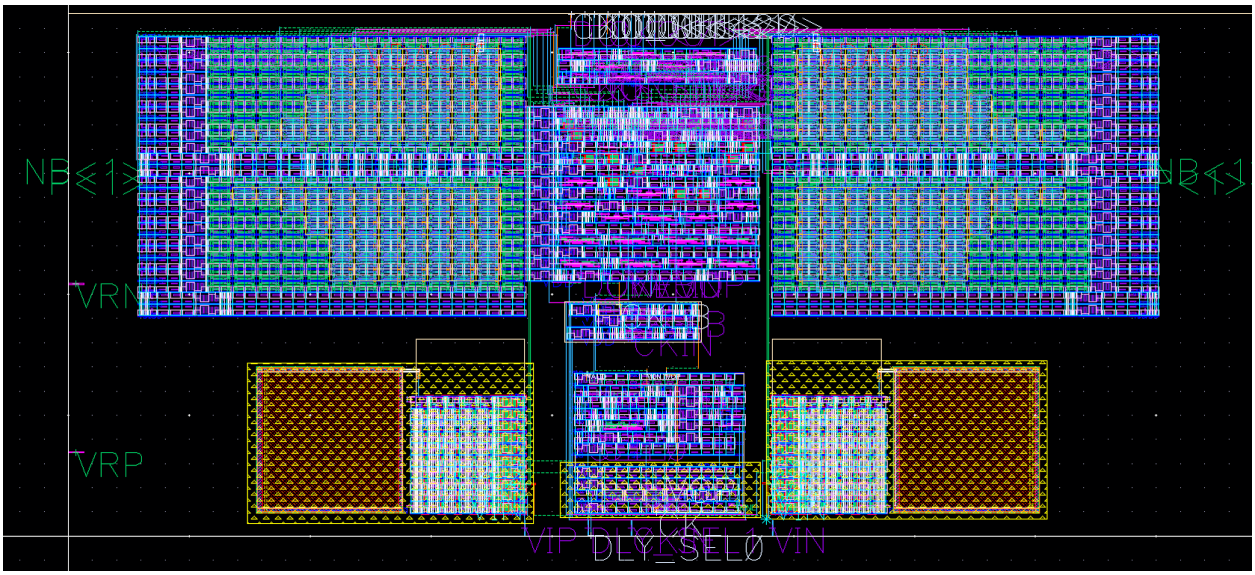


Figure 37. Layout Version 2 size:42x20

3.6.7 Introduction

Perhaps one of largest challenges of mixed-signal system-on-chip (SoC) design is that there is virtually no viable way to fully verify the correctness and performance of a SoC design before tape out. The established methodology for digital verification is to perform System Verilog simulation and emulation, fast discrete-event discrete-value simulation, while analog verification is using SPICE transistor-level simulation, slow continuous-time continuous-value simulation. In this project, we incorporate an innovative signal-driven abstraction of a transistor level circuit in terms of System Verilog real numbers, also referred to as real number modeling. This can be viewed as discrete-event continuous/discrete-value simulation. By representing physical quantities with high precision, real number modeling in System Verilog ensures faster simulations while capturing the continuous nature of analog signals.

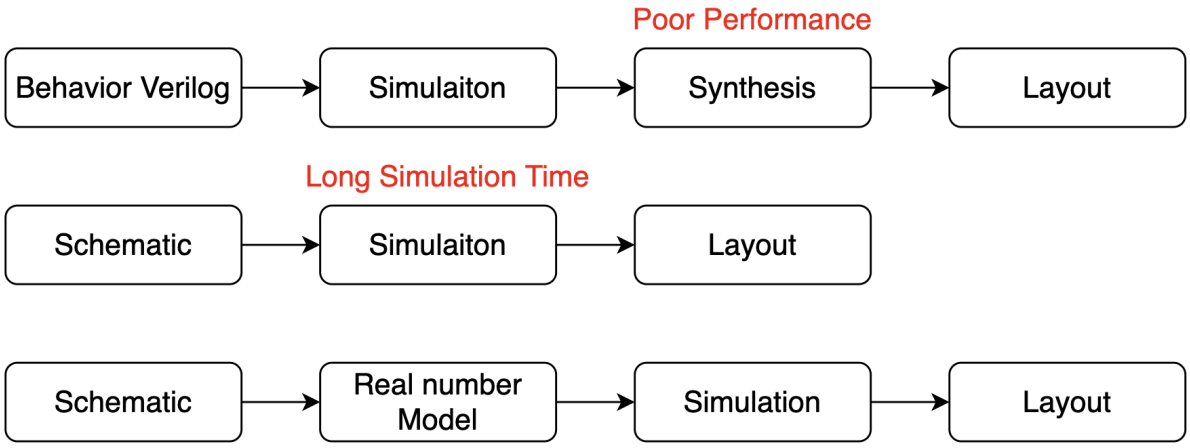


Figure 38. The Design Procedure of Digital, Analog and Real Number Mod

eling

3.6.8 Automated Real Number Model Generation

Real number models can be automatically generated bottom up for a hierarchical AMS design using a recursively-verifying-modeling (RVM) methodology [12]. Once a designer checks in his/her design as a cadence library cell, the verification script will automatically take the checked in cell netlist as the input, combined with the pin specification as defined in a pin table file, then automatically generate all the system Verilog models pin matched with the schematic, and test benches. Then the verification script will automatically run a transistor level simulation on the design under test (DUT) cell schematic also a System Verilog simulation on the extracted System Verilog Model with the System Verilog Test benches generated based on the same transistor level test bench auto-generated. Then a waveform comparison will be performed, and a verification report will be generated.

As illustrated in Figure 39, the RVM methodology starts from a low level primitive (cell). Consider a particular cell, we classify its design correctness constraints to internal constraints and interface constraints, dependent whether constraints are affected by the cell interface connections and operating environments. A cell is characterized and verified using test benches against its internal constraints. If a cell does not meet the specification and satisfy the internal constraints, we report that the cell has a design error. Once a cell is verified to be design correct, then a pin-matched behavioral model that abstracts out the internal details but preserves the interface pins and cell functionalities (performance and power of interest) is generated.

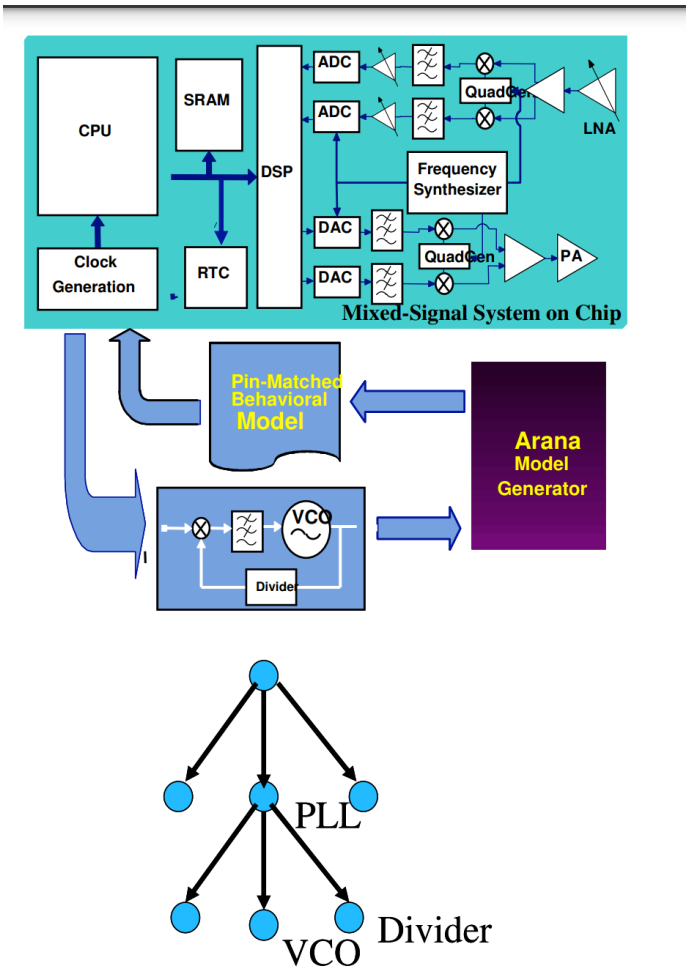


Figure 39. An Illustration o the Pin-matched Hierarchical Verifying-modeling Methodology

The behavioral model is “pin-matched” in the sense that it can be used to replace the original netlist at the interface, the same test bench can be run using the same simulator. We further classify the cell interface constraints to assertable constraints and non-assertable constraints, dependent upon if a constraint can be written in behavioral models that can be checked automatically during simulation. All assertable interface constraints are automatically inserted in a generated behavioral model for a cell. Once a cell is verified to be design correct and its behavioral model with automated assertable interface constraint checking inserted, we can replace the original cell by its behavioral model. Then move to the next level of design hierarchy as shown in Figure 39.

The real number modeling based RVM methodology has several advantages:

1) it is recursive and scalable and is cable of handling the design complexity.
2) It is automated, and there is no room for human-introduced errors (once the tools are verified).
3) Error messages are checked only needed. Internal constraints are verified once at the cell level and interface constraints are verified when a cell is used.

4) Only behavioral models are generated for a cell that is designed correctly. This allows behavioral models to be optimized for efficiency. In the other words, there is no need to generate behavioral models for a bad-designed circuit. This is an inherit advantage over SPICE.
5) The hierarchy can be controlled by a designer and verification engineer at any level of interest, allowing the tradeoff between verification details and speed.
6) A small set of data is handled; this facilitates the debug.
7) The same set of design constraints can be used for optimization and design re-targeting.
8) The methodology uses the same simulator and same test benches.

The key enabler for the RVM methodology is automated generation and validation of behavioral models from a transistor netlist. An automated tool has been developed based on recent progresses on and switch-level logic analysis [5], model order reduction [11], and symbolic circuit analysis [12-14].

The tool can automatically generate behavioral models for analog circuits, digital circuits and RF circuits. It consists of a model generator, a model validator and a model optimizer. The model validator verifies a behavioral model transistor level circuits using designer's test benches and test configuration and process-voltage corners and generates the validation report. Sometimes, for individual models, an optimizer can be used to optimize the model parameters to match the waveform results. The characterization of individual blocks involves of tools to build test benches for custom-IC circuits. One of the critical aspects of custom IC design is how to build a variety of test benches. The integration of System Verilog Models or AMS blocks together with the original RTL models allows the deployment of metric-driven UVM verification methodology that has been developed for large digital design verification for mixed-signal SoC verification. Further as illustrated in Figure 40, the flow allows the development of comprehensive test benches to ensure the full verification coverage with not only a unified SV based UMV verification flow but also a dramatically accelerated simulation based on purpose-based abstraction.

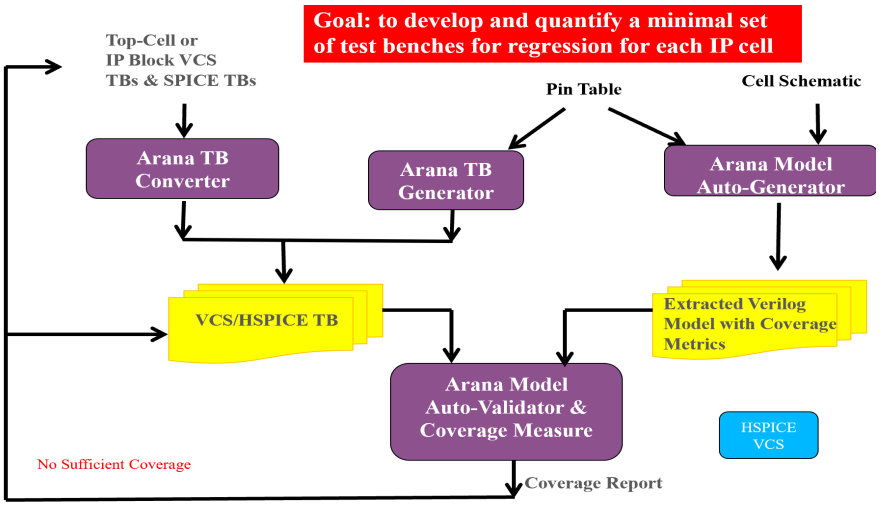


Figure 40. Metric-driven Verification Flow for Mixed-signal SoCs

3.6.9 Experimental Results

Here we show a simulation comparison between Hspice and Verilog simulation, while Hspice took 2 hours to run the simulation and Verilog simulation used 10 seconds. The delay between output and input is 120ps and same as Verilog simulation, which confirms the robustness and precision of real number modeling.

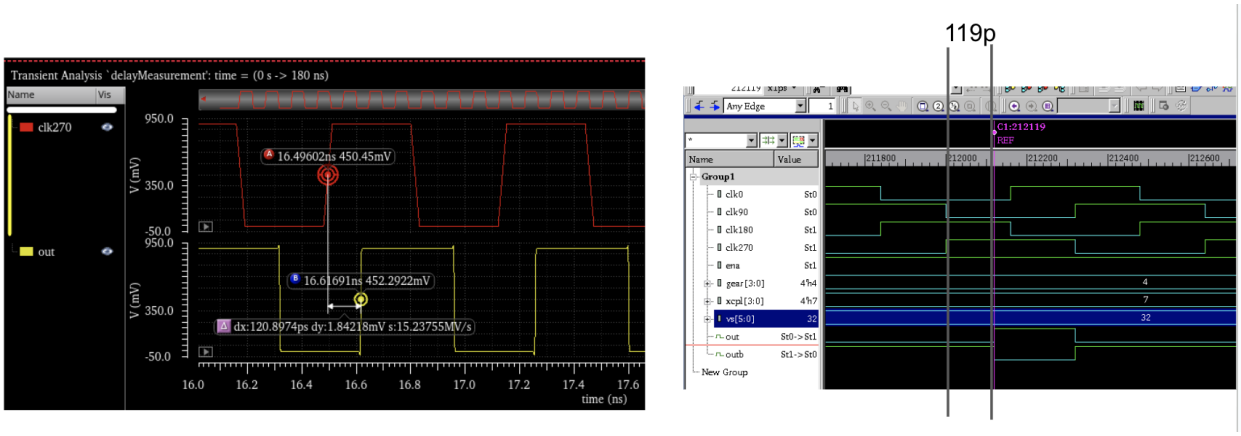


Figure 41. Simulation Waveforms

Figure 41 Left is a simulation result in Spectre and on the right is Verilog simulation using VCS. The output is a delayed signal of clk270 and the delay time are 119p in both simulation runs. This implies that the simulation result matches between Spectre and VCS.

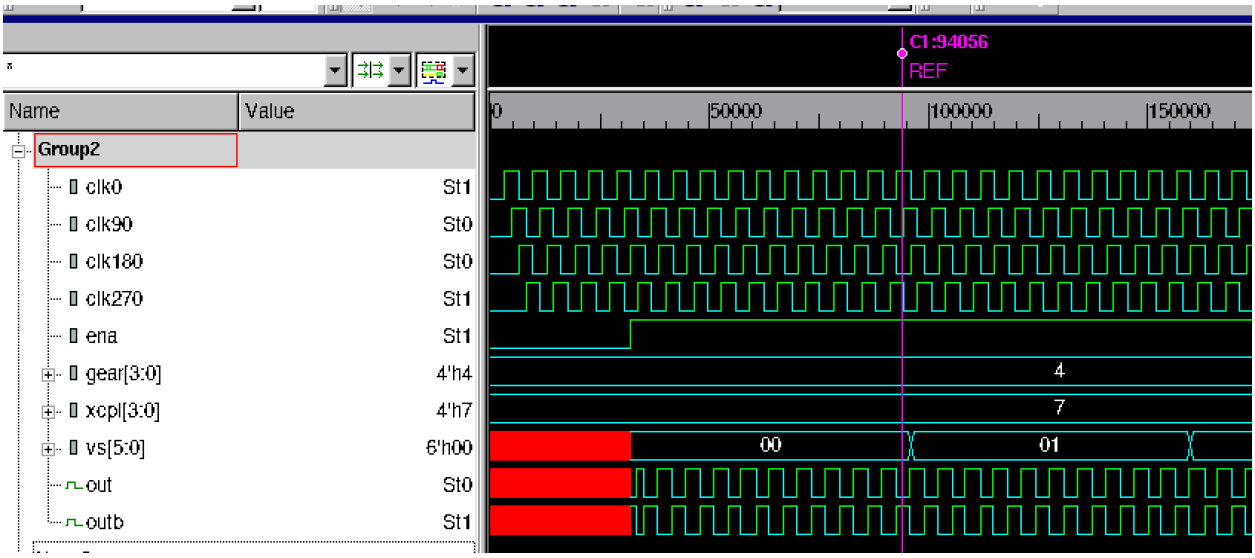


Figure 42. VCS Simulation Waveforms

Then we run the simulation with different configurations. We run 10 clock cycles for each set of configurations. As shown in Figure 42, vs= 00, delay = 119ps and vs=01, delay = 126.8ps. We can see that the real number modeling can reflect the delay of the signal with multiple configurations and matches that of transistor circuit simulation, and while reducing the simulation time by hundreds and thousands of times.

3.7 Open Source on Github

We have the RAIL analog library and analog test cases available on Github, which contain designs in both Tsmc65nm and GlobalFoundries 12nm, including standard cells, IP, and modules. The open-source library provides a set of basic standard analog and mixed-signal cells in different categories and functionalities to enable RAIL, a constraint-driven and technology-independent automatic design flow for AMS circuit design implementation.

The RAIL65 cell library includes the cells under TSMC 65nm technology. It contains 83 cells in total, including switches, delay cell, passive devices, custom defined cells as shown below.

Switch Category: MX2D, MX2U, NSW0, NSW05, NSW1, PSW0, PSW05 PSW1, PXRO, SWFD, TBUF, TGAT
Delay Cell Category: DLY1, PXRO, VCDC, VCDC1, VCDC2, VCDC3, VCDP, VCDP1, VCDP2, VCDP3
Passive Category: MOM1, MOM2, PCAP, PCAP2, NCAP, NCAP2
Custom Defined Cells Category: CDC1
Custom Defined Channel Connect Block: P1N1, P1N2, P1N3, P2N1, P2N2, P2N3, P3N1, P3N2, P3N3, P12N, P22N
Custom Defined Transistor Category: nch_lvt, nch_rvt, pch_lvt, pch_rvt

The RAIL12 cell library includes the cells under GlobalFoundries 12nm and it has 6 circuit examples and 14 cells that are used to compose these circuits.

Switch Category: NSW1, PSW1
Delay Cell Category: DLY, DLY_INV, Pseudo_Differential_DCell
Passive Category: MOM1, MOM2,
Custom Defined Cells Category: CDC1, P1N2SERX1, P1N3SERX1, P2PARN1X1

https://github.com/rail-posh/rail65

https://github.com/rail-posh/rail12

Besides standard cell libraries, we also have 5 complete analog IP and 10 modules released. All designs are fabricated and tested, and each design contains hierarchical schematics, documentation and sanitized CDL netlists. Below is the table of released circuit and modules.

Circuit	Architecture	Design Files
SAR_ADC	12b SAR ADC	SAR_ADC/UW_65nm_SARADC_Apr_2019
DLDO	Digital LDO	DLDO/UW_65nm_DLDO_Apr_2019
MDLL	Multiplying DLL	MDLL/UW_65nm_MDLL_Apr_2019
TX	Transmitter	Transmitter/UW_65nm_Transmitter_Apr_2019
TS	Temperature Sensor	TempSensor/UW_65nm_TS_May_2019

Module	Cell	Function
5b_ADC	5b coarse SAR ADC	Compute the 5-bit MSBs using merged capacitor switching (MCS) method
BTSW	Bootstrapped Switch	Sample input on to CDAC
CDAC_Switch	CDAC Switch for the Coarse SAR ADC	Convert output digit to its analog signal and then subtracted from input
CK_Divider8	Clock Divider by 8	Generate complementary clock to drive DLPF, Comparison enable signal, and reference clock
Coarse_Comp_CK	Coarse Comparator Clock Generator	Generate clock for the comparator
Coarse_SAR_Logic	SAR Logic	Generate control signal for the CDAC and store output digits from comparator
DLPF_RCFilter	CDAC Embedded DLPF	The sensing element with embedded CDAC
EdgeComparator	Edge Comparator	To compare the crossing point with the reference clock
LevelCrossingDetector	Level Crossing Detector	To compare the crossing point with the reference clock
ResTune_Configure	Resistance Tuning Configuration	Logic to tune resistance in the DLPF

https://github.com/rail4open/UW-IDEA_AnalogTestCases

3.8 Summary

This chapter described the RAIL flow and demonstrated the possibility of designing and implementing an analog circuit with the RAIL enhanced digital flow. A SAR-ADC design in TSMC 65nm is used to show the entire design flow from designing and characterizing a RAIL standard analog cell to verifying the performance with post-layout simulation. The simulation result sof 7b SAR ADC in TSMC 65nm and sub-blocks used in SAR ADC designs, including the comparator, S/H switch and CDAC, are described and compared with that o manual implementation. Combined with automated real number model generation, the RAIL flow provides an agile and trustable methodology for open-source mixed-signal circuit implementation.

3.9 References

[1] J. M. Cohn, D. J. Garrod, R. A. Rutenbar and L. R. Carley, "KOAN/ANAGRAM II: New tools for device-level analog placement and routing," *IEEE Journal of Solid-State Circuits,* vol. 26, no. 3, pp. 330-342, March 1991, doi: 10.1109/4.75012.

[2] N. Jangkrajarng, S. Bhattacharya, R. Hartono, and C.-J. R. Shi, "IPRAIL: Intellectual property reuse based analog IC layout automation," *Integration, the VLSI Journal,* vol. 36, no. 4, pp. 237-262, Nov. 2003.

[3] R. Hartono, N. Jangkrajarng, S. Bhattacharya, and C.-J. R. Shi, "Active Device Generation for Automatic Analog Layout Retargeting Tool", 2004.

[4] Bhattacharya, N. Jangkrajarng, and C.-J. R. Shi, "Multi-level symmetry constraint generation for retargeting large analog layouts," *IEEE Trans. on Computer-Aided Design*, vol. 25, no. 6, pp. 945-960, June 2006.

[5] Y. Lei and C.-J. R. Shi, "FROSTY: A program for fast extraction of high-level structural representation from circuit description for industrial CMOS circuits," *INTEGRATION, the VLSI Journal*, vol. 39, pp 311-339, 2006.

[6] L. Zhang, N. Jangkrajarng, S. Bhattacharya, and C.-J. R. Shi, "Parasitic-aware optimization and retargeting of analog layouts: A symbolic template approach," *IEEE Trans. on Computer-Aided Design*, vol. 27, no. 5, pp. 791-802, May 2008.

[7] Zhang, Y. Zhang, Y. Jiang and C.-J. R. Shi "Symmetry-aware placement algorithm using transitive closure graph representation for analog integrated circuits", *International Journal of Circuit Theory and Applications*, vol. 38, no. 3, pp. 221-241, Apr 2010.

[8] E. Chang, et al., "BAG2: A process-portable framework for generator-based AMS circuit design," in *2018 IEEE Custom Integrated Circuits Conference (CICC)*, San Diego, CA, USA, 2018, pp. 1-8, doi: 10.1109/CICC.2018.8357061.

[9] B. Xu, et al., "MAGICAL: Toward fully automated analog IC layout leveraging human and machine intelligence," in 2019 *IEEE/ACM International Conference on Computer-Aided Design (ICCAD),* Nov. 2019, pp. 1–8. doi: 10.1109/ICCAD45719.2019.8942060.

[10] C.-J. R. Shi, "Mixed-signal system-on-chip verification using a recursively-verifying-modeling (RVM) methodology", in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems.*

[11] G. Shi, B. Hu and C.-J. R. Shi, "On symbolic model order reduction," *IEEE Trans. on Computer-Aided Design*, vol. 25, no. 7, pp. 1257-1272, July 2006.

[12] C.-J. R. Shi and X.-D. Tan, "Canonical symbolic analysis of large analog circuits with determinant decision diagrams," *IEEE Trans. On Computer-Aided Design,* vol. 19, no. 1, pp. 1-18, Jan. 2000.

[13] C.-J. R. Shi and X.-D. Tan, "Compact representation and efficient generation of s-expanded symbolic network functions for computer-aided analog circuit design*," IEEE Trans. on Computer-Aided Designs*, vol. 20, no.7, pp. 813-827, July 2001

[14] X.-D. Tan and C.-J. R. Shi, "Efficient approximation of symbolic expressions for analog behavioral modeling and analysis," *IEEE Trans. On Computer-Aided Designs,* vol. 23, no. 6, pp. 907-918, June 2004.

4 DESIGN AND ANALYSIS OF A FULLY-SYNTHESIZABLE BUFFER-FREE SAR-ADC

This chapter presents the analysis of bitwise and sample-wise switched passive charge sharing for successive approximation register (SAR) analog-to-digital conversion (ADC). Closed-form analytic expressions of ADC transfer functions are derived based on charge conservation and validated by behavioral and schematic simulations. This leads to two elegant results for SAR ADCs with bitwise switched reference charge reservoirs (BS-RCRs). First, a binary-weighted SAR ADC implemented with BS-RCRs is transformed into a subradix-2 ADC. Second, the reference error caused by finite reservoir capacitance appears in the form of bit weight error. This error can be corrected digitally or by selecting a sufficiently large bit reference capacitance to bit weight capacitance ratio β. However, the reference error with sample-wise switched reference charge reservoir (SS-RCR) is input dependent. In addition, an equivalent-circuit model-based analysis method is introduced, which shows more circuit intuition why BS-RCRs have better linearity than SS-RCR. A case study of an 11-bit 100-MS/s SAR ADC in 65-nm CMOS is presented.

4.1 Introduction

The core of successive approximation register (SAR) analog-to-digital converters (ADCs) is to convert an analog input voltage to a multibit digital code by sampling the input onto a bit weighted capacitor array and then deciding each bit, starting from the most significant bit (MSB) down to the least significant bit (LSB) by charge redistribution [1]. Using simple architectures and building blocks consisting only of switches, capacitors, comparators, and digital logic with no static power consumption, SAR ADCs are scalable with the digital CMOS technology. With recently introduced switching schemes achieving less switching activity [2]–[4], SAR ADCs have become the most energy-efficient data conversion solutions for 10–100-MHz sampling rates with 10–12-bit resolution [5]–[7]. Furthermore, combined with pipeline [8], [9], time interleaving [9], and noise shaping [10], SAR ADCs are being extended to applications with even higher speed up to gigahertz [11] or higher resolution up to 18 bits [12].

High-speed high-resolution SAR ADCs require accurate voltage references for each bit decision. With bonding wire inductance, bit-switching settling to the needed accuracy often requires a very large decoupling capacitor and/or a large-bandwidth linear buffer. This reference buffer can consume power many times more than the SAR ADC itself and is not scalable with the digital CMOS technology. To remove the power-hungry buffer, passive charge sharing was introduced [13]. One implementation of passive charge sharing [15] is to replace the reference by a sufficiently large capacitor precharged to the reference voltage acting as the reference for all bit switching during the entire analog-to digital conversion (ADC), which is referred to as sample-wise switched reference charge reservoir (SS-RCR). An improvement of SS-RCR for SAR ADCs was introduced in [16]–[18]. For each bit decision, a bit reference charge reservoir (RCR) with the capacitance β times larger than that of the bit weight capacitor is used to replace the reference source. This improvement is referred to as bitwise switched RCRs (BS-RCRs).

This chapter presents the analysis of bitwise and sample wise switched passive charge sharing SAR ADCs. It has two major contributions. First, we show theoretically from the principle of charge conservation that for BS-RCRs, the finite reservoir capacitance error appears in the form of bit weight error. For an N-bit binary-weighted SAR ADC, the ith bit weight error is the bit weight attenuated by $1 + 2^{i+1-N}\beta$ where $i = N - 1, \ldots, 1$. This error can be corrected digitally or by selecting a sufficiently large β. We also show that a binary weighted SAR-ADC with bitwise switched charge reservoirs is essentially being transformed to a subradix-2 ADC. On the contrary, for SS-RCR, we show that the finite reservoir capacitance error appears not in the form of bit weight error and is highly input dependent. An SS-RCR-based SAR-ADC exhibits the subradix-2 property for some inputs and the superradix-2 property for some

other inputs. The missing-level error caused by the super-radix-2 property cannot be corrected. Second, we introduce an equivalent-circuit technique for the analysis of passive charge sharing. Based on the principle of charge conservation, we show that there exist an "equivalent" Thevenin and Norton equivalents and generalized to *Y* transform for a general capacitor network with nonzero initial conditions and controlled by switches. Furthermore, the principle of superposition is applicable. In contrast to the analysis applying the principle of charge conservation, equivalent-circuit analysis helps to provide more circuit intuitions, especially on the better linearity of BS-RCRs over SS-RCR, a seemingly counterintuitive property.

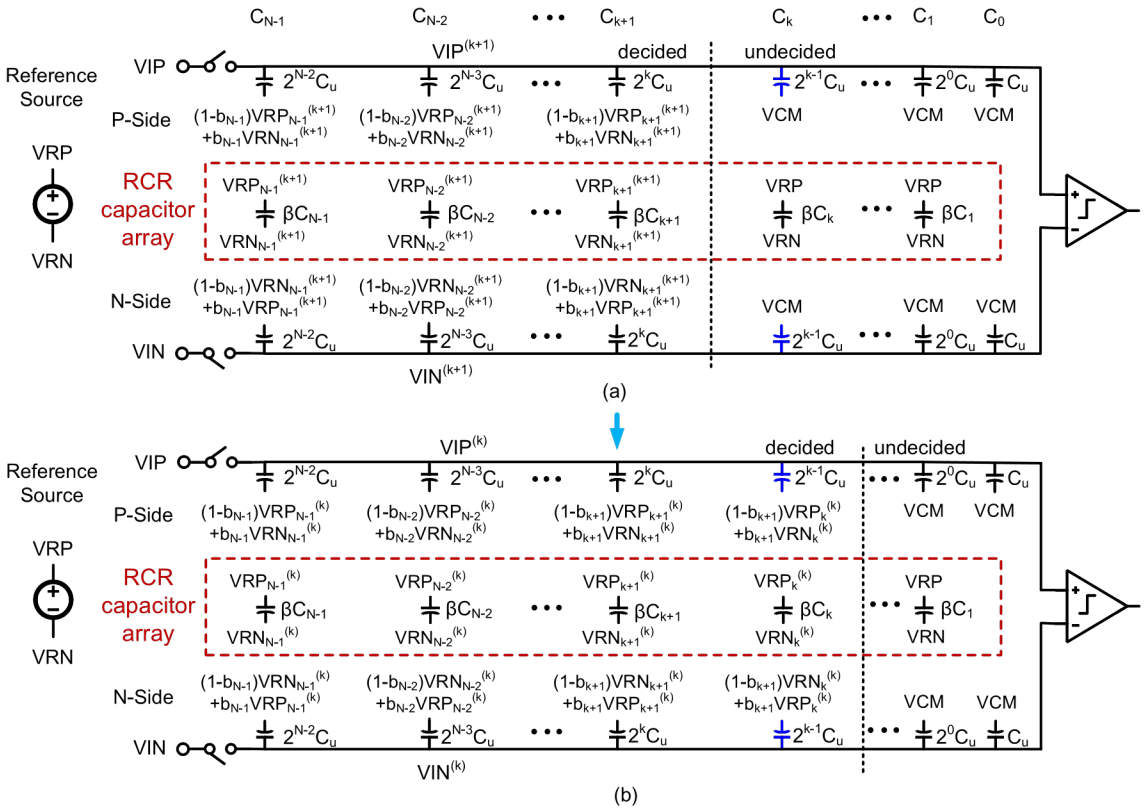


Figure. 1. *N*-bit BS-RCRs-based SAR-ADC with Top-plate Sampling from (a) *(k* + 1*)*th Decision to (b) *k*th Decision

All the theoretical results derived here have been validated by running circuit simulation. An 11-bit 100-MS/s SAR-ADC has been designed in 65-nm CMOS to quantitatively evaluate the effectiveness of bitwise and sample-wise switching RCR techniques, in particular, how *β* affects the linearity, how large *β* causes the settling error, and finally, to validate the theory developed.

This chapter is organized as follows. Section 3.2 introduces the technique of BS-RCR. The SS-RCR technique is presented in Section 3.3. Section 3.4 provides equivalent half-circuit models and simplified derivation of bitwise and sample-wise RCRs. Section 3.5 describes schematic- and behavioral-level simulation validation using an 11-bit SAR-ADC in 65-nm CMOS as a case study. Section 3.6 summarizes this chapter.

4.2 VCM-Based Merged Capacitor Switching

Merged capacitor switching (MCS) [3] is one of the most energy-efficient SAR-ADC bit decision methods. An *N*-bit SAR-ADC converts a pair of differential analog input voltages VIP and VIN with the common mode VCM to an *N*-bit binary codes $b_{N-1}, b_{N-2}, \ldots, b_k, \ldots, b_1, b_0$, where bit b_i can be 0 or 1. Initially, the different input VIP and VIN are sampled on the top plates of two (P-side and N-side) binary-weighted arrays of bit capacitors with capacitances $2^{N-2}C_u$, $2^{N-3}C_u, \ldots, 2^iC_u, \ldots, 2^0C_u$, C_u where C_u is the unit capacitance. All the bottom plates are connected to reference VCM, the common-mode voltage of high reference VRP (often VDD) and low reference VRN (often GND). The comparator first compares and outputs $b_{N-1} = 1$ if VIP ≥ VIN, and 0 otherwise. If $b_{N-1} = 1$, the bottom plate of the P-side (N-side) MSB capacitor is then switched to reference VRN (VRP) and, otherwise, to reference VRP (VRN). As a result of this MSB bit switching, the top-plate voltages change to $\text{VIP}^{(N-1)}$ and $\text{VIN}^{(N-1)}$.

In general, we use bracketed superfix *k* to indicate the *k*th bit switching, $k = N - 1$ down to 1. Given the top-plate voltages $\text{VIP}^{(k+1)}$ and $\text{VIN}^{(k+1)}$, the comparator decides b_k. The bottom plate of the P-side bit capacitor is then switched to reference $(1 - b_k)\text{VRP} + b_k\text{VRN}$, and that of the N-side to $(1 - b_k)\text{VRN} + b_k\text{VRP}$, as marked in Fig. 1. This leads to new top-plate voltages $\text{VIP}^{(k)}$ and $\text{VIN}^{(k)}$.

Based on the principle of charge conservation, the top-plate voltages $\text{VIP}^{(k)}$ and $\text{VIN}^{(k)}$ at the *k*th bit decision can be expressed as follows:

$$\text{VIP}^{(k)} = \text{VIP} + \sum_{i=k}^{N-1} \frac{1}{2^{N-i}} \left(\frac{1}{2} - b_i\right) (\text{VRP} - \text{VRN}) \quad (1)$$

$$\text{VIN}^{(k)} = \text{VIN} + \sum_{i=k}^{N-1} \frac{1}{2^{N-i}} \left(\frac{1}{2} - b_i\right) (\text{VRN} - \text{VRP}). \quad (2)$$

The SAR-ADC based on the above-mentioned operation is referred to as charge redistribution SAR-ADC.

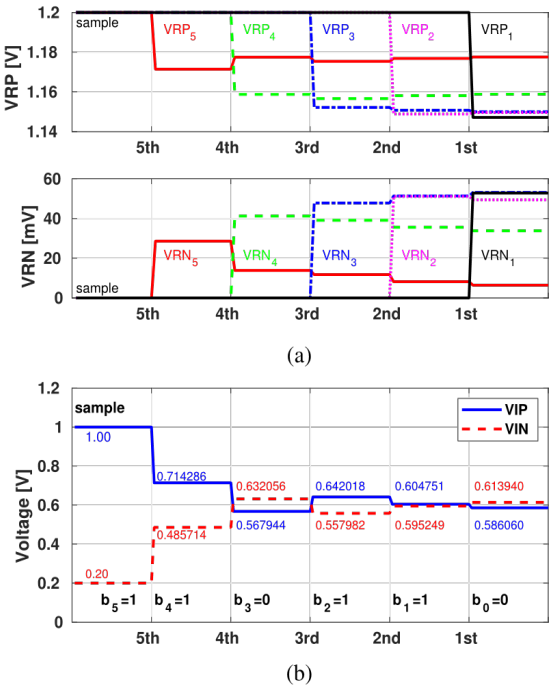


Figure 2. (a) CDAC Reference Voltages and (b) Top-plate Voltage Change at each Bit Decision of a 6-Bit BS-RCR SAR-ADC with β = 5.

4.2.1 Bitwise Switched Reference Charge Reservoir

Now, consider for each bit decision to replace reference sources VRP and VRN with a capacitor whose top plate precharged to VRP and the bottom plate to VRN. This capacitor serves as an RCR. Furthermore, the bit reference capacitance is chosen to be β times larger than the corresponding bit weight capacitance. Fig. 1 shows the middle section of an array of bitwise switched ratioed RCRs (BS-RCRs) with capacitances $\beta C_{N-1}, \beta C_{N-2}, \ldots, \beta C_k, \ldots, \beta C_0$, and C_T is the total bit capacitance of C_{N-1} to C_0. Since no reference sources are used for bit decision, this SAR-ADC is referred to as passive charge sharing SAR-ADC.

Due to charge sharing, the top- and bottom-plate voltages of all bit charge reservoir capacitors are not fixed at VRP and VRN. Let VRP_i and VRN_i denote the top-plate and bottom-plate voltages of the ith bit reservoir capacitor, and $VRP_i^{(k)}$ and $VRN_i^{(k)}$ to denote their values at the kth bit decision step, as shown in Fig. 2(a) for the case of a 6-bit SAR-ADC with β = 5. Fig. 2(b) shows that the capacitive digital-to-analog converter (CDAC) top-plate voltages deviate from their ideal values VRP = 1.2 V and VRN = 0 V.

From the principle of charge sharing, the top-plate voltages $VIP^{(k)}$ and $VIN^{(k)}$ at the kth bit decision can be represented as follows:

$$\mathrm{VIP}^{(k)} = \mathrm{VIP} + \sum_{i=k}^{N-1} \frac{\alpha^{(i)} C_i}{C_T} \left(\frac{1}{2} - b_i\right) (\mathrm{VRP} - \mathrm{VRN}) \quad (3)$$

$$\mathrm{VIN}^{(k)} = \mathrm{VIN} + \sum_{i=k}^{N-1} \frac{\alpha^{(i)} C_i}{C_T} \left(\frac{1}{2} - b_i\right) (\mathrm{VRN} - \mathrm{VRP}) \quad (4)$$

$$\alpha^{(i)} = \frac{2\beta}{2\beta + 1 - \frac{\sum_{j=i}^{N-1} C_j}{C_T}} = \frac{2\beta}{2\beta + (1/2)^{N-i}} \quad (5)$$

where the last equality in (5) holds if bit capacitances are binary weighted.

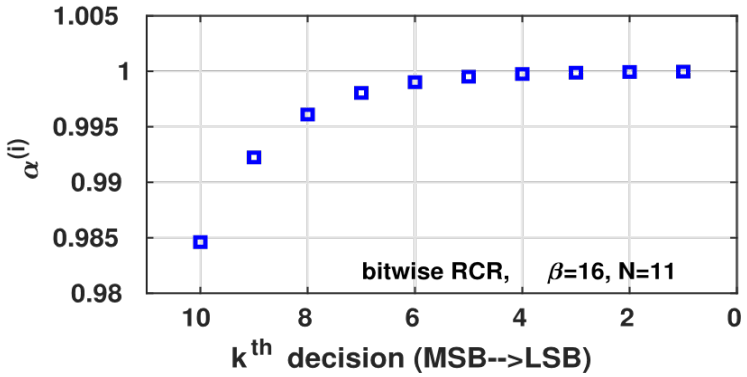


Figure 3. $\alpha^{(i)}$ at Each Bit Decision of an 11-bit BS-RCR SAR-ADC with β = 16

Note when $\beta = \infty$, $\alpha^{(i)} = 1$, (3) and (4) degenerate to (1) and (2). There is no error with an infinite reference charge capacitance. Given finite reservoir capacitance with $\beta \neq \infty$, we have $0 < \alpha^{(i)} < 1$. Furthermore, $\alpha^{(i)}$ increases from the MSB ($i = N - 1$) to the LSB ($i = 1$). Fig. 3 shows $\alpha^{(i)}$ at each bit decision of an 11-bit BS-RCR-based SAR-ADC with β = 16.

4.2.2 Transfer Function and Digital Error Correction

By subtracting (4) from (3) with the left-hand side of the equation truncated to zero, i.e., less than 0.5 LSB, we obtain the transfer function of a BS-RCR-based SAR-ADC as follows:

$$\mathrm{V_{in,diff}} = \sum_{i=1}^{N-1} \alpha^i \frac{C_i}{C_T} (2b_i - 1)(\mathrm{VRP} - \mathrm{VRN}). \quad (6)$$

We observe that the ratio of two consecutive bit weights has the following property:

$$1 < \frac{\alpha^i C_i}{\alpha^{i-1} C_{i-1}} = \frac{2\beta + (1/2)^{N-i+1}}{2\beta + (1/2)^{N-i}} * 2 < 2. \quad (7)$$

With VRP − VRN known as the full-scale (FS) (analog) voltage V_{FS} and binary-weighted bit capacitances, (6) represents a subradix-2 expansion, also known as β-expansion, of a real number [19]. Thus, a binary-weighted SAR-ADC implemented with bitwise charge reservoirs is a subradix-2 ADC [20]. In fact, as long as $\alpha^{(i)}$ increases from the MSB ($i = N - 1$) to the LSB ($i = 1$) then it is a subradix-2 conversion. On the other hand, if $\alpha^{(i)}$ decreases from the MSB ($i = N-1$) to the LSB ($i = 1$), it is a super-radix-2 conversion.

As an example, the transfer functions of a 6-bit BS-RCR based SAR-ADC with $\beta = 0.5$ and $\beta = 5$ are shown in Fig 4. The transfer curve for $\beta = 5$ is almost ideal. The transfer curve for $\beta = 0.5$ deviates from the ideal transfer curve expanding around the middle input (differential input = 0) and reaching the maximum close to the two input ends (±FS). This exhibits the gain error. As a result, the effective analog input range is reduced from the FS. Let the full digital scale to be the range of from code 0000 0000 000 to 1111 1111 111. Around the values of 1/2, 1/4, 3/4 of the full digital scale, where one analog input is mapped into more than one digital output codes. In terms of ADC, some digital codes are missing, exhibiting the missing code error. For a given N, both the possibility of missing code occurrence and the magnitude of the gain error increase with the smaller β.

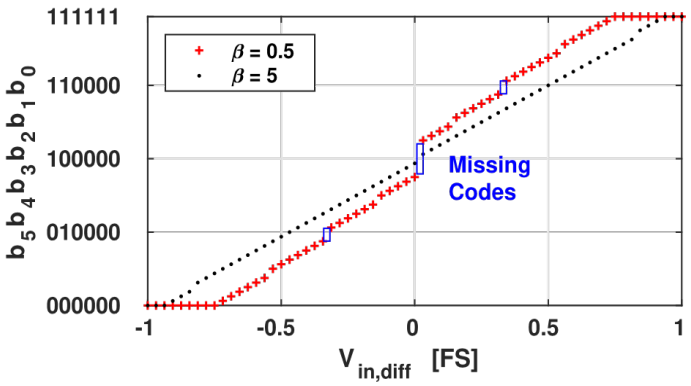


Figure 4. *B*out versus *V*in,diff of a 6-bit BS-RCRs-based SAR-ADC

Note that this BS-RCR induced error appears in the form of bit weight error from (3) and (4). This error can be predicted based on the given value of β and resolution of the ADC N based on (5). Therefore, it can be corrected based on the digital calibration through the following:

$$D_{\text{out}} = \sum_{i=0}^{N-1} b_i \cdot 2^i \cdot \alpha^{(i)}. \qquad (8)$$

Here, $\alpha^{(i)}$ is given in (5). The product of $2^i \cdot \alpha^{(i)}$ represents the actual weights in the comparison for a binary-weighted CDAC.

4.2.3 Error Control by Selecting Sufficiently Large β

Since $1 - \alpha^{(i)}$ is always positive, the maximal linearity error $\varepsilon^{\max}$ for an N-bit SAR-ADC due to finite β can be expressed as

$$|\varepsilon^{\max}| = \sum_{i=1}^{N-1} \frac{C_i}{C_T}(1-\alpha^{(i)})(\mathrm{VRP}-\mathrm{VRN}). \tag{9}$$

With a binary-weighted bit capacitor array, C_i/C_T can be simplified to $2^i/2^N = 1/2^{N-i}$.

To control the linearity error $^{\max}$ to be less than 0.5 LSB requires

$$\sum_{i=1}^{N-1} \frac{1-\alpha^{(i)}}{2^{N-i}} \approx \sum_{i=1}^{N-1} \frac{(1/2)^{N-i}}{2\beta \cdot 2^{N-i}} \le \frac{1}{2^{N-1}}. \tag{10}$$

Solving this, we obtain

$$\beta \ge \frac{2^{N-1}[1-(1/4)^{N-1}]}{6}. \tag{11}$$

Table I shows the minimal βs for an N-bit SAR-ADC with differential nonlinearity (DNL) ≤ 0.5 LSB. It has been verified through simulation. It is interesting to note that for a 2-bit or 3-bit SAR-ADC, a reference reservoir capacitance can be a fraction of the corresponding bit weight capacitance without degrading the linearity.

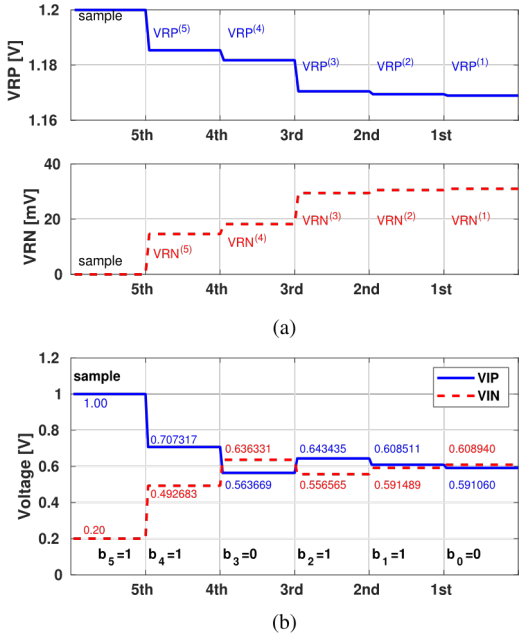


Figure 5. (a) CDAC Reference Voltage and (b) Top-plate Voltage Change at each Bit Decision of a 6-Bit SS-RCR-based SAR-ADC with β = 5

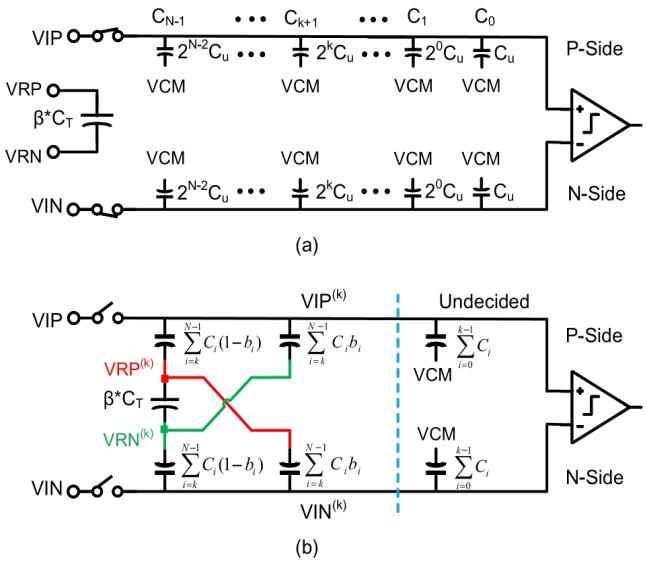


Figure 6. *N*-bit SS-RCR-based SAR-ADC with Top-plate Sampling. (a) Sampling Phase. (b) *k*th Decision

4.3 SAR SWITCHING WITH SS-RCR

Instead of bitwise, if all charge reservoir capacitors are connected in parallel to form a big capacitor with capacitance $\beta(C_{N-1} + C_{N-2} + \cdots + C_1 + C_0) = \beta C_T$ with the top-plate precharged to VRP and the bottom plate to VRN. This is the sample-wise switched charge reservoir introduced in [15]. Fig. 5 illustrates the waveforms of $VRP^{(k)}$, $VRN^{(k)}$, $VIP^{(k)}$, and $VIN^{(k)}$ for the case of a 6-bit SAR-ADC with $\beta = 5$. Compared with a 6-bit BS-RCR-based SAR ADC with $\beta = 5$, both the reference voltages and CDAC top-plate voltages are different from the ones in Fig. 2.

Table I Minimum β OF an N-bit bs-rcrs-based SAR-ADC That Meets DNL ≤ 0.5 LSB without Digital Calibration

N	1	2	3	4	5	6	7	8	9	10	11	12
β	0	1/4	5/8	$1\frac{5}{16}$	$2\frac{21}{32}$	$5\frac{21}{64}$	$10\frac{85}{128}$	$21\frac{85}{256}$	$42\frac{341}{512}$	$85\frac{341}{1024}$	$170\frac{1365}{2048}$	$341\frac{1365}{4096}$

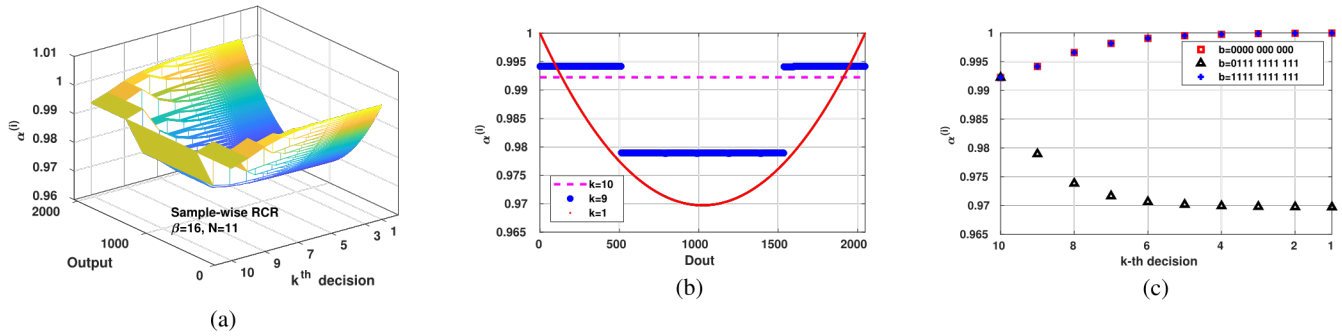


Figure 7. $\alpha^{(i)}$ at each Bit Decision of an 11-bit SS-RCR-based SAR-ADC with β = 16. (a) Three-Dimensional Plot of $\alpha^{(k)}$ Versus Codes and Decision Step *(k)* (b) 2-D Plot of $\alpha^{(k)}$ Versus Codes for Selected Decision Steps *(k)*. (c) Two-dimensional Plot of $\alpha^{(k)}$ Versus Decision Step *(k)* for Selected Codes

In general, consider sample-wise reference switching at the *k*th decision as illustrated in Fig. 6(b). From the principle of charge conservation, the analytic expressions of top-plate voltages for sample-wise switching at the *k*th bit decision can be derived and written in the same form as (3) and (4) but with $\alpha^{(i)}$ redefined by (12)

$$\alpha_{\text{samplewise}}^{(k)} = \frac{2\beta}{2\beta + \frac{\sum_{i=k}^{N-1} C_i}{C_T} - \left[\frac{\sum_{i=k}^{N-1} C_i(2b_i-1)}{C_T}\right]^2}. \qquad (12)$$

We note that when $\beta = \infty$, $\alpha = 1$. For finite β, $0 < \alpha^{(k)} < 1$, and $\alpha^{(k)}$ are functions of β and N.

In contrast to bitwise $\alpha^{(k)}$ as shown in (5), sample-wise $\alpha^{(k)}$ depend on $b_i, i = N-1,\ldots,k$, i.e., the previous more significant bits that have been decided. In fact, the cross $^{N}{}_{-1}$ term $C_i(2b_i - 1)$ in (12) is the total capacitance of those previous bit capacitances connecting to reference plate VRN$_{i=k}$ minus that of those connecting to reference plate VRP as shown in Fig. 6(b). The error caused by the finite-charge reservoir capacitance is, thus, highly nonlinear and no longer in the form of the bit weight error. Furthermore, it does not represent a subradix-2 expansion of a real number and an SSRCR-based binary-weighted SAR-ADC is not a subradix-2 ADC.

Fig. 7 shows α(k) sample-wise of an 11-bit SS-RCR-based SAR-ADC with β = 16. For α(10) sample-wise, there is no previous bit. Therefore, there is only one value of α(10) sample-wise. For α(9) sample-wise, the previous bit b10 has two possible values 0 or 1, so is α(9) sample-wise due to the term [10 i=9 Ci(2bi − 1)] 2 in (12). There are 29 possible values of α(1) sample-wise. The strong input dependence can be clearly seen by comparing Fig. 3 with Fig. 7(a), as further shown in Fig. 7(b).

Fig. 7 shows $\alpha_{samplewise}^{(k)}$ of an 11-bit SS-RCR-based SAR-ADC with β = 16. For $\alpha_{samplewise}^{(10)}$, there is no previous bit. Therefore, there is only one value of $\alpha_{samplewise}^{(10)}$. For $\alpha_{samplewise}^{(9)}$, the previous bit b_{10} has two possible values 0 or 1, so is $\alpha_{samplewise}^{(9)}$ due to the term $[C_i(2b_i - 1)]^2$, i=9 to 10 in (12). There are 2^9 possible values of $\alpha_{samplewise}^{(1)}$. The strong input dependence can be clearly seen by comparing Fig. 3 with Fig. 7(a), as further shown in Fig. 7(b).

One important characteristic of $\alpha_{samplewise}^{(k)}$ is its trend in terms of decision steps kth from the MSB to the LSB. As further shown in Fig. 7(c), at the two ends of the codes (all 0s, all 1s), $\alpha_{samplewise}^{(k)}$ increases from the MSB to the LSB. Thus, sample-wise reference switching behaviors as a subradix-2 conversion like bitwise reference switching. Near the middle code 10000000000, $\alpha_{samplewise}^{(k)}$ decreases from the MSB to the LSB. In this case, sample-wise reference-switching behaviors as a super-radix-2 conversion.

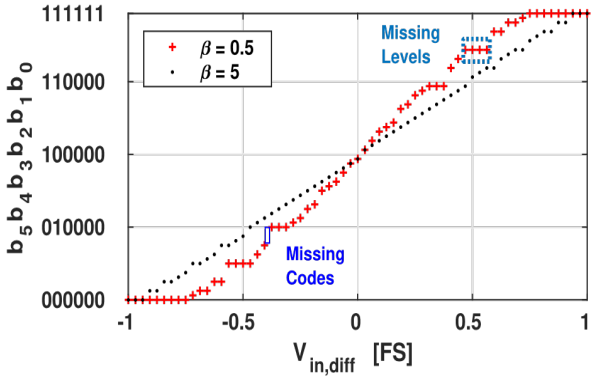


Figure 8. *B*out versus *V*in,diff of a 6-bit SS-RCR-based SAR-ADC with β = 0.5 and β = 5.

As an example, the transfer functions of a 6-bit SS-RCR based SAR-ADC with β = 0.5 and β = 5 are plotted in Fig. 8. The transfer curve for β = 5 is almost ideal. Similar to that of bitwise reference switching, the transfer curve for β = 0.5 deviates from the ideal transfer curve expanding around the middle input point 1/2 of the full digital scale, and reaching the maximum close to the two input ends. This exhibits the gain error. Around the values of 1/8, 1/4, 7/8, and 3/4 of the full digital scale, one analog input is mapped into more than one digital output codes. In terms of ADC, some digital codes are missing, exhibiting the missing code error.

Different from bitwise reference-switching transfer curves shown in Fig. 4, sample-wise reference-switching transfer curves shown in Fig. 8 show missing analog level errors; i.e., there exist multiple analog inputs transferred into the same digital code. Unlike the gain error and missing code error, the missing-level error cannot be corrected.

4.4 Equivalent Half-Circuit Modls: Alternate Derivation of Linearity and Circuit Insight

In this section, we introduce an equivalent-circuit method for analyzing a capacitor network with initial conditions. We first present several general principles for constructing equivalent circuits based on charge conservation, then apply them for the linearity analysis of BS- and SS-RCR-based SAR-ADCs. This equivalent-circuit-based analysis reveals more circuit intuition on why bitwise switching has better linearity than sample wise switching, a seemingly counterintuitive characteristic.

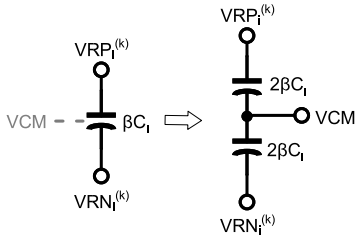


Figure 9. Symmetry Split of a Charge Reservoir Capacitor

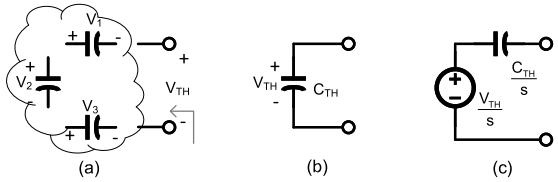


Fig. 10. Thevenin equivalent of a capacitor network with initial conditions

(a) One-port connected capacitor network with initial conditions. (b) Thevenin equivalent in the time domain. (c) Thevenin equivalent in the Laplace domain.

First, we note that all top-plate voltages $VIP^{(k)}$ and $VIN^{(k)}$, reference top-plate and bottom-plate voltages $VRP_{(i}^{k)}$ and $VRN_{i(}^{k)}$, $i = N - 1,...,k + 1$ are symmetrical with respect to the fixed VCM (Symmetry). As a result, the middle point of any capacitor connecting two symmetrical nodes always stays at VCM. We can, thus, split this capacitor in the middle into two capacitors each with double capacitance and the middle being connected to VCM. We refer to this technique as symmetry split. Fig. 9 shows the case of symmetry split of a charge reservoir capacitor.

Second, we observe that any one-port connected capacitor network with nonzero steady-state voltages across capacitors and satisfying the Kirchoff voltage law (called consistent initial conditions) is equivalent to one capacitor, referred to as Thevenin capacitor with a nonzero initial condition, referred to as Thevenin voltage, where the Thevenin voltage is the open-circuit voltage drop across the port, and the Thevenin capacitance is the effective capacitance seen from the port. This observation is referred to as the Thevenin equivalence theorem for a capacitor network, as illustrated in Fig. 10.

Third, any internal node k with node voltage v_k connecting through capacitor C_i, $i = 1,\ldots,k-1$ to $k-1$ neighboring nodes in a capacitor network (known as the star configuration) can be eliminated to yield an equivalent fully connected capacitor network (the clique configuration) among these $k-1$ nodes where the resulting capacitor connecting nodes i and j have the capacitance of C_iC_j/C_T and $C_T = \sum_{i=1}^{k-1} C_i$. This is illustrated in Fig. 11 for the case of $k = 4$. This is the generalized version of Y to transformation for a capacitor network with consistent initial conditions.

This observation can be shown by considering how to eliminate node k and associated charge conservation equation.

$$\sum_{i=1}^{k-1} C_i(v_k - v_i) = \sum_{i=1}^{k-1} C_i(v_k' - v_i') \tag{13}$$

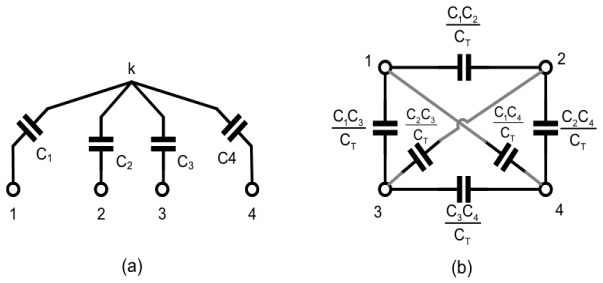


Fig. 11. Generalized Y to transform of a capacitive network.

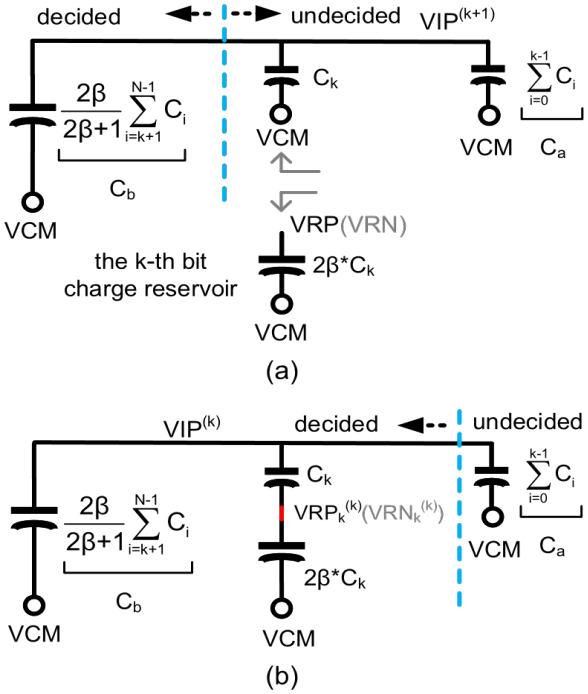


Fig. 12. Half-circuit model and derivation of bitwise reference switching. (a) *(k + 1)*th decision. (b) *(k)*th decision.

where v_i and v_i, $i = 1, \ldots, k$ are voltages before and after charge sharing. Then, the charge contribution of node k to node j through capacitor C_j can be represented as follows:

$$C_j(v'_k - v_k) = \sum_{i=1}^{k-1}\left(C_i C_j / \sum_{l=1}^{k-1} C_l\right)(v'_i - v_i). \qquad (14)$$

This elimination of charge conversation equation (13) and its associated voltage v_k by (14) indicates that the effect of v_k on the solution of the rest of node voltages can be expressed completely in terms of its neighboring node voltages. This gives rise to the principle of superposition. Each time an operation of charge sharing causes a voltage change of a particular node. The final voltage through a sequence of charge sharing operations is the algebraic sum of the voltage changes caused by each charge sharing in sequence. We emphasize that, in general, the final voltage of each node depends on the order of charge sharing operations in each sequence.

Using symmetry split and Thevenin equivalence, the equivalent half-circuit model for bitwise switching at the kth decision step is shown in Fig. 12(a). Here, the bottom plate of the P-side bit weight capacitor C_k can switch from VCM to connect to either VRP or VRN depends on b_k. Let us assume that it is VRP. Here, the Thevenin equivalent initial condition is zero.

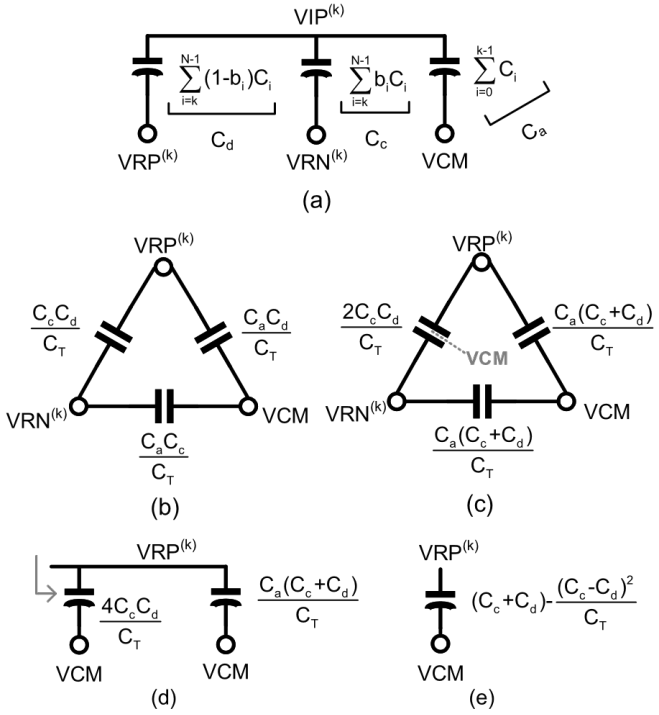


Fig. 13. *Y* to transformation at kth decision of sample-wise reference. (a) CDAC at the P-side. (b) Transfer P-side CDAC from Y-shape to -shape. (c) P-side and N-side networks in parallel. (d) Equivalent capacitance at the reference P-side. (e) Simplify (d) to one capacitor.

This means that when considering the charge sharing of bit k charge reservoir with the CDAC, all the initial voltages in the CDAC do not affect the resulting voltage $VRP^{(k)}$. Voltage $VRP_k^{(k)}$ can be obtained from the following charge sharing equation:

$$(VRP - VCM)(2\beta C_k) = VRP^{(k)}_k - VCM(2\beta C_K + C_{TH})$$

$$C_{TH} = (C_a + C_b)C_k/(C_a + C_b + C_k). \qquad (15)$$

It can be verified that this equation gives rise to the same expression as derived directly from charge conservation.

Then, the P-side top-plate voltage change due to this charge sharing is obtained by capacitive division as

$$\begin{aligned}\mathrm{VIP}^{(k)}-\mathrm{VIP}^{(k+1)} &= \frac{C_k}{C_a+C_b+C_k}\left((1-b_k)\left(\mathrm{VRP}_k^{(k)}-\mathrm{VCM}\right)\right.\\ &\quad \left.+b_k\left(\mathrm{VRN}_k^{(k)}-\mathrm{VCM}\right)\right)\\ &= \frac{C_k}{C_a+C_b+C_k}(1-2b_k)\left(\mathrm{VRP}_k^{(k)}-\mathrm{VCM}\right).\end{aligned} \tag{16}$$

Substituting (15) into (16) with $C_T = C_a + C_b(2\beta + 1)/2\beta + C_k$, we can obtain

$$\mathrm{VIP}^{(k)} - \mathrm{VIP}^{(k+1)} = \alpha^{(k)}\frac{(1-2b_k)C_k}{C_T}(\mathrm{VRP}-\mathrm{VCM}) \tag{17}$$

where $\alpha^{(k)}$ is defined by (5). Clearly, (3) can be obtained by summing (17) from $N-1$ to $k+1$, while (4) is obtained by symmetry.

To analyze sample-wise switching, we first consider charge sharing from the port of the charge reservoir. Y to Δ transformation is used to eliminate nodes $\mathrm{VIP}^{(k)}$ and $\mathrm{VIN}^{(k)}$.

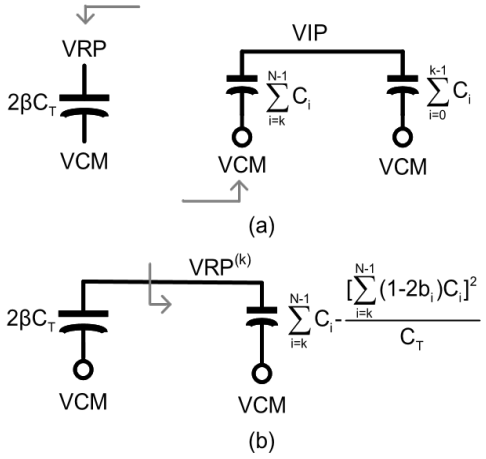


Fig. 14. Half-circuit model and derivation of sample-wise reference switching. (a) Sampling phase. (b) *k*th decision phase

Fig. 13(b) shows the P-side equivalent circuit after Y to Δ from Fig. 13(a). Fig. 13(c) shows the total equivalent circuit transformed from both the P-side and the N-side. Then the symmetry principle is used to split the resulting capacitance between nodes $\mathrm{VRP}^{(k)}$ and $\mathrm{VRN}^{(k)}$. This leads to the equivalent charge sharing circuit shown in Fig. 13(d) for computing $\mathrm{VRP}^{(k)}$, which can be further simplified to Fig. 13(e).

For sample-wise switching, all charge sharing operations from N to $k+1$ are performed on the same port of the charge reservoir capacitor. Fig. 14(a) shows the equivalent circuit at the end of the sampling phase before any charge sharing, and Fig. 14(b) shows the equivalent circuit for computing $\mathrm{VRP}^{(k)}$ after *k*th bit decision. Clearly, the charge is conserved before and after these k bit decision steps, that is

$$(\mathrm{VRP}-\mathrm{VCM})(2\beta C_T)=\left(\mathrm{VRP}^{(k)}-\mathrm{VCM}\right)\times\left[2\beta C_T+(C_c+C_d)-\frac{(C_c-C_d)^2}{C_T}\right]. \quad (18)$$

Then, the P-side top-plate voltage change can be obtained based on capacitive division shown in Fig. 13(a)

$$\begin{aligned}\mathrm{VIP}^{(k)}-\mathrm{VIP}&=\left(\mathrm{VRP}^{(k)}-\mathrm{VCM}\right)\frac{C_d}{C_T}+\left(\mathrm{VRN}^{(k)}-\mathrm{VCM}\right)\frac{C_c}{C_T}\\&=\left(\mathrm{VRP}^{(k)}-\mathrm{VCM}\right)\frac{C_d-C_c}{C_T}. \quad (19)\end{aligned}$$

Substituting (18) into (19), we can obtain the same result as (3) and (4) with $\alpha^{(k)}$ defined by (12).

Comparing Figs. 12(a) and 13(e), we see that sample-wise switching introduces a cross-coupling capacitance that allows the previous bits to affect the current bit decision. This cross-coupling capacitance leads to more finite-reference error, and further, the error is code dependent. Note that the cross-coupling capacitance is 0 for two digital FS codes and the middle code.

4.5 Case Study of an 11-bit SAR-ADC at Behavioral and Schematic Levels

An 11-bit 100-MS/s SAR-ADC has been designed in 65-nm CMOS to quantitatively evaluate the effectiveness of bitwise and sample-wise switching reference charge reservoir techniques, in particular, how β affects the linearity, and to validate the theory developed in this work.

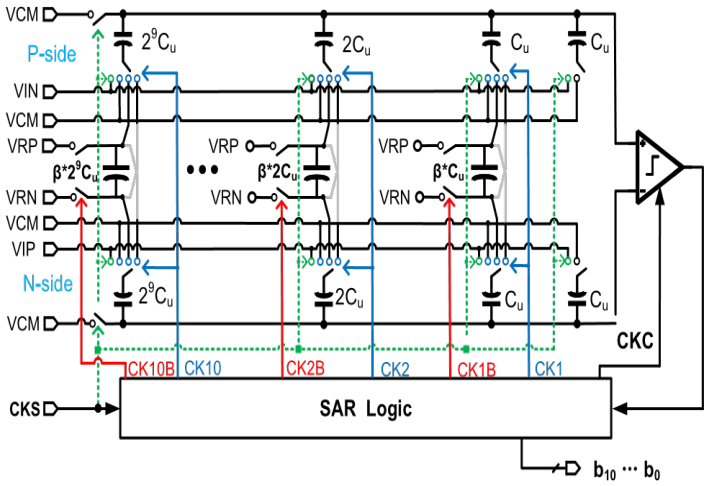


Fig. 15. Eleven-bit SAR-ADC architecture with BS-RCRs.

4.5.1 SAR-ADC Architecture and Schematic Design

The ADC is implemented in the differential architecture shown in Fig. 15. CDACs are binary weighted from C_u to $2^9 * C_u$, where C_u is the unit capacitance. The RCR capacitors are also binary weighted with capacitance β times larger than its bit capacitance. Bottom-plate sampling is used so that the ADC linearity is immune to the top-plate parasitic capacitance.

The ADC then uses VCM-based MCS to decide bits from the MSB to the LSB. Either BS-RCR or SS-RCR technique is employed. The ADC uses a two-stage comparator with a dynamic preamplifier to suppress the kickback noise and latch offset [21]. A 25% duty cycle 100-MHz is used with 25% for sampling and 75% for SAR bit decision. The SAR logic is asynchronous [22] and is the same as [2].

For comparison, an ideal 11-bit SAR-ADC is built out of this schematic ADC with capacitors, switches, comparators, delay line, and digital gates replaced by ideal elements. The linearity of the ideal ADC and schematic-level ADC is simulated using the same test benches with the same settings.

4.5.2 Effect of β on Linearity

A pair of differential input ramp signals with step 0.1 LSB is applied to obtain the static performances. A pair of differential input sinusoidal signals is applied, transient simulation was performed to obtain the dynamic performances. The performances computed include DNL, integral nonlinearity (INL), signal-to-noise distortion ratio (SNDR), spurious-free dynamic range (SFDR), and effective number of bits (ENOB).

Tables II and III summarize, respectively, the simulated static and dynamic performances of ideal 11-bit SAR-ADCs with BS-RCRs and SS-RCR with different βs. These have been verified to be the same as those obtained using analytic expressions (3)–(5), and (12). From Table II, the linearity loss is less than 0.5 LSB when β = 256. When β decreases to 32, the ENOB loss is 1.8 bits. From Tables II and III, we see that all of the performances of bitwise switching are better than that of sample-wise switching.

TABLE II BEHAVIORAL MODEL: STATIC AND DYNAMIC PERFORMANCES OF AN 11-bit BS-RCRS-BASED SAR-ADC WITH DIFFERENT βS

β	32	64	128	256
SNDR(dB)	56.74	61.38	66.70	67.66
ENOB(bits)	9.13	9.90	10.79	10.95
SFDR(dB)	59.26	65.26	72.58	81.32
DNL(LSB)	-1/0.005	-1/0.0026	-0.7/0	-0.4/0
INL(LSB)	-2.7/2.7	-1.4/1.4	-0.7/0.7	-0.3/0.3

TABLE III BEHAVIORAL MODEL: STATIC AND DYNAMIC PERFORMANCES OF AN 11-bit SS-RCR-BASED SAR-ADC WITH DIFFERENT βS

β	32	64	128	256
SNDR(dB)	54.33	59.63	64.62	66.33
ENOB(bits)	8.74	9.61	10.44	10.73
SFDR(dB)	54.54	60.23	67.41	71.86
DNL(LSB)	-1/3.6	-1/1.9	-1/1	-0.5/0.5
INL(LSB)	-3.1/3.1	-1.6/1.6	-0.8/0.8	-0.4/0.4

Tables IV and V summarize the simulated static and dynamic performances of the 11-bit SAR-ADC schematic design with BS-RCRs and SS-RCR with different βs, respectively. In simulation, circuit nonideal effects including clock feedthrough and charge injection introduced by MOS transistor switches as well as kickback noise from comparator and transient effects are included. Comparing Table IV with Table II, the ENOB loss due to circuit nonideality is less than 0.16 bit for BS-RCR-based SAR-ADCs, and the ENOB loss decreases as β increases. Even with circuit nonideality, bitwise switching always has better linearity than sample-wise switching.

Take β = 16 for an 11-bit SAR-ADC with C_u = 1 *f F* as an example, we analyze how the sampled thermal noise on RCR capacitors and the mismatch among RCR capacitors affect the performance in both BS- and SS-based SAR ADCs. The sampled thermal noise on RCR capacitors has the form of $(kT/C)^{1/2}$. Here, k is Boltzmann's constant, T is 300 K at room temperature, and C is the RCR capacitance. The sampled thermal noise on BS-RCR capacitor varies from 22.5 to 509 μV from the MSB to the LSB, while 15.9-μV thermal noise is sampled on the SS-RCR capacitor. The thermal noise is much less than the error due to charge sharing. Thus, it does not influence the performance in both BS- and SS-based SAR ADCs. Now, considering the mismatch among RCR capacitors, the nmoscap is used for the RCR capacitor in the schematic-level implementation. It not only has large capacitance density (14 fF/µm2) at VDD bias but can also be laid out beneath the CDAC to save the area. The 1−σ mismatch of a 16.8- fF nmoscap is 1.875e-17 through Monte Carlo simulation. Apply this 1−σ mismatch to all the RCR capacitors, the dynamic performance does not deteriorate in both BS- and SS-based SAR-ADCs.

4.5.3 Consideration of Settling with Large β

From the quantitative analysis above, to reduce the finite-reference capacitance error to a negligible ENOB loss level without using error correction requires β to be as large as 256 for an 11-bit SAR-ADC. We need to ensure both the charging and discharging of large reference capacitors to settle to 11-bit resolution.

TABLE IV SCHEMATIC MODEL: STATIC AND DYNAMIC PERFORMANCES OF AN 11-bit BS-RCRS-BASED SAR-ADC WITH DIFFERENT βS

β	32	64	128	256
SNDR	56.56	61.16	66.54	67.58
ENOB	9.10	9.87	10.76	10.93
SFDR	58.5	64.74	72.22	81.08
DNL(LSB)	-1/0.008	-1/0.0046	-0.7/0	-0.4/0
INL(LSB)	-2.85/2.85	-1.45/1.45	-0.75/0.75	-0.35/0.35

TABLE V SCHEMATIC MODEL: STATIC AND DYNAMIC PERFORMANCES OF AN 11-bit SS-RCR-BASED SAR-ADC WITH DIFFERENT βS

β	32	64	128	256
SNDR	53.55	59.07	64.24	66.14
ENOB	8.60	9.52	10.38	10.69
SFDR	53.52	59.42	66.82	71.45
DNL(LSB)	-1/3.83	-1/1.91	-1/1	-0.5/0.5
INL(LSB)	-3.28/3.28	-1.66/1.66	-0.85/0.85	-0.43/0.43

First, note that the transfer functions of each bit reference to the ADC outputs are described by the second terms in (3) and (4), i.e., the kth bit reference is attenuated by a factor of $C_T/(\alpha^{(k)}C_K)$, $k = N - 1,...,1$, which is $2^{N-k}/\alpha^{(k)}$ for a binary-weighted capacitor array. To ensure the ADC error caused by the kth bit reference voltage error to be smaller than 0.5 LSB requires that the kth bit reference voltage loss to satisfy

$$\frac{V_k^{\max}}{V_{\mathrm{FS}}} \le \frac{\frac{1}{2^{N+1}}}{\frac{\alpha^{(k)}}{2^{N-k}}} = \frac{2^{-k-1}}{\alpha^{(k)}}. \tag{20}$$

This means that if $\alpha^{(k)}$ is around 1, the kth bit reference needs to have k + 1-bit resolution, i.e., the MSB requires 11-bit resolution and the LSB requires only 2-bit resolution.

Now, consider the bit reference needs to settle to this required bit resolution. There are two situations: at the start of SAR-ADC operation and during SAR-ADC operation. At the start of ADC operation, there is no initial voltages on bit reference capacitors The MSB charging settling dominates and we, thus, only need to consider the MSB reference capacitor charging settling. To charge the MSB capacitor to settle to the needed 11-bit resolution requires the charging settling time to be more than 8.3 times time constant RC where C is the reservoir capacitance and R is the switch ON-resistance. We note that this long settling time is only required one time at the start of SAR-ADC operation. During SAR-ADC operation, since the reference voltage loss is small for each ADC bit decision, the required recharging settling time is short. Quantitatively, the k-bit reference capacitor voltage loss $\mathrm{VR}^{\mathrm{LOSS}}_k$ during kth bit decision can be computed as

$$\frac{\mathrm{VR}_k^{\mathrm{LOSS}}}{V_{\mathrm{FS}}} = \frac{2\beta - 2\beta(C_K/C_T) + (C_K/C_T)}{(2\beta+1)(2\beta + C_K/C_T)}. \tag{21}$$

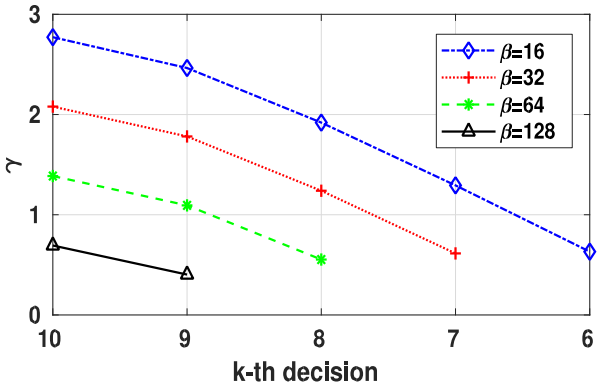


Fig. 16. Settling time factor *γ* versus *k* at different *β*.

We see that the MSB voltage loss is small, but the resolution needed is small, while the LSB voltage loss is large but the resolution required is also large. Using (20) and (21), we can estimate the settling time *t* needed to satisfy

$$\begin{aligned} t &\geq (RC)\ln\frac{V R_k^{\mathrm{LOSS}}}{V_k^{\max}} \\ &= (RC)\ln\left(\frac{2\beta - 2\beta(C_K/C_T) + (C_K/C_T)}{(2\beta+1)(2\beta + C_K/C_T)} 2^k \alpha^{(k)}\right) \\ &= (RC)\ln\left(\frac{\alpha^{(k)}2^k}{\beta}\cdot\frac{1 - 2^{k-N-1} + 2^{k-N}\frac{1}{2\beta}}{(1+\frac{1}{2\beta})(1+2^{k-N}\frac{1}{2\beta})}\right) \\ &= \gamma\,(RC). \end{aligned} \tag{22}$$

Fig. 16 plots the required settling time constant factor *γ* for reference capacitor recharging for an 11-bit SAR-ADC with respect to *k* and *β*. It is interesting to note that the required time constant factor is only around 1 or 3 for recharging (instead of 8.3 for the first-time charging). For a small *k*, the required time constant factor is zero, this is since the bit reference loss at those LSBs does not affect the ADC resolution for a given *β*.

This large reference capacitance is then charge-shared with the corresponding bit capacitance. The time constant for charge sharing is proportional to the effective capacitance of the reservoir capacitance connected in series with the bit capacitance, which is *β/(β* + 1*)* bit capacitance.

4.5.4 Linearity Performance with Small β and Error Correction

Now, we quantitatively show the ADC linearity with relatively small *β* and the effectiveness of error correction. For *β* = 16, the simulated DNL and INL and dynamic spectrum performance are shown in Fig. 17(a) for a BS-RCR-based SAR-ADC. The maximal DNL error of −1 occurs near the 1/2, 1/4, and 3/4 of the full digital scale, which means missing codes at these digits. Dynamic performance shows more than 2.91 bits of resolution loss when *β* = 16 for an 11-bit SAR-ADC.

Fig. 17 (b) shows the simulated static and dynamic performances of an 11-bit SAR-ADC with SS-RCR that has *β* = 16 (the total reservoir capacitance is the same as that of BSRCRs). The maximum DNL is −1*/*7.59 LSB, much worse than that of the bitwise case. It shows both missing codes (DNL = −1 LSB) and missing levels (DNL ≥ 1 LSB) occur at different inputs. We note that sample-wise DNL is 0 near the middle codes. This is because the super-radix-2 operation is near the middle input.

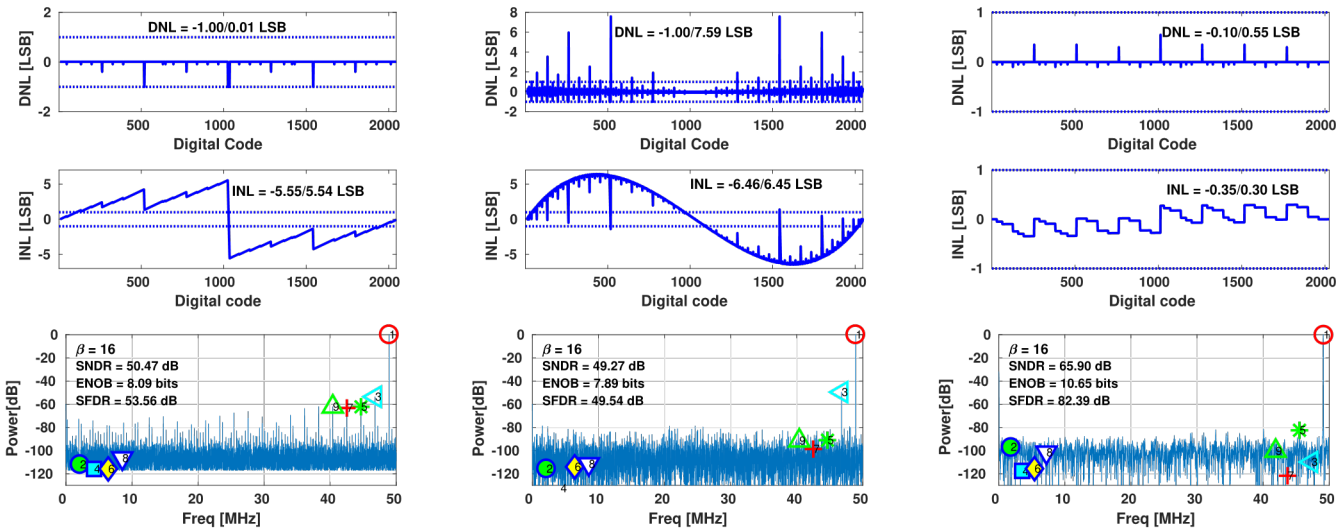


Fig. 17. Simulated performance of an 11-bit SAR-ADC with switched RCRs at β = 16. (a) BS-RCRs. (b) SS-RCR. (c) BS-RCRs after digital calibration.

After digital error correction according to (8), the static and dynamic performances of an 11-bit BS-RCR-based SAR-ADC with β = 16 are shown in Fig. 17(c). We can see that the DNL and INL are both smaller than 0.55 LSB, and the ENOB can reach to 10.65 bit after error correction.

4.6 Summary

This chapter presented the analysis of an energy-efficient BS-RCR technique for reference-buffer-free SAR-ADC design. Complete theoretical analysis and simulation have been performed to analyze the error due to the finite-reference capacitances. It is shown that the reference error in an SAR-ADC with BS-RCRs is in the form of bit weight error, causing the gain and missing code errors, and can be digitally corrected precisely or be eliminated by using a sufficiently large β.

We have shown that an SAR-ADC with bitwise charge reservoirs is essentially a subradix-2 ADC. Future work includes to explore this subradix-2 redundancy for error correction.

The reference error for the previously published sample-wise switched reference charge reservoir (SS-RCR) technique has also been analyzed. It is shown that the SS-RCR error exhibits both the super-radix-2 and subradix-2 properties depends on inputs. In addition to exhibiting the gain and missing code errors, the SS-RCR technique shows the missing-level errors, which cannot be corrected.

All the analyses have been validated and demonstrated using an 11-bit RCR-based SAR-ADC. Quantitative evaluation of β on ADC linearity has been presented. Design consideration of settling error for large β has been discussed. The effectiveness of error correction for small β has also been demonstrated.

4.7 References

1. J. L. McCreary and P. R. Gray, “All-MOS charge redistribution analog-to-digital conversion techniques. I” *IEEE J. Solid-State Circuits*, vol. SSC-10, no. 6, pp. 371–379, Dec. 1975.
2. C.-C. Liu, S.-J. Chang, G.-Y. Huang, and Y.-Z. Lin, “A 10-bit 50-MS/s SAR ADC with a monotonic capacitor switching procedure,” *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 731–740, Apr. 2010.

3. V. Hariprasath, J. Guerber, S.-H. Lee, and U.-K. Moon, "Merged capacitor switching based SAR ADC with highest switching energyefficiency," *Electron. Lett.*, vol. 46, no. 9, pp. 620–621, Apr. 2010.
4. H.-Y. Tai, Y.-S. Hu, H.-W. Chen, and H.-S. Chen, "A 0.85 fJ/conversionstep 10 b 200 kS/s subranging SAR ADC in 40 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 196–197.
5. W. Liu, P. Huang, and Y. Chiu, "A 12-bit, 45-MS/s, 3-mW redundant successive-approximation-register analog-to-digital converter with digital calibration," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2661–2672, Nov. 2011.
6. C.-C. Liu, "A 0.35 mW 12b 100 MS/s SAR-assisted digital slope ADC in 28 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig.*
7. *Tech. Papers*, Jan./Feb. 2016, pp. 462–463.
8. W.-H. Tseng, W.-L. Lee, C.-Y. Huang, and P.-C. Chiu, "A 12-bit 104 MS/s SAR ADC in 28 nm CMOS for digitally-assisted wireless transmitters," *IEEE J. Solid-State Circuits*, vol. 51, no. 10, pp. 2222–2231, Oct. 2016.
9. C. C. Lee and M. P. Flynn, "A SAR-assisted two-stage pipeline ADC," *IEEE J. Solid-State Circuits*, vol. 46, no. 4, pp. 859–869, Apr. 2011.
10. F. van der Goes *et al.*, "A 1.5 mW 68 dB SNDR 80 Ms/s 2× interleaved pipelined SAR ADC in 28 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 49, no. 12, pp. 2835–2845, Dec. 2014.
11. J. A. Fredenburg and M. P. Flynn, "A 90-MS/s 11-MHz-bandwidth 62-dB SNDR noise-shaping SAR ADC," *IEEE J. Solid-State Circuits*, vol. 47, no. 12, pp. 2898–2904, Dec. 2012. doi: 10.1109/JSSC.2012. 2217874.
12. L. Kull *et al.*, "A 3.1 mW 8b 1.2 GS/s single-channel asynchronous SAR ADC with alternate comparators for enhanced speed in 32 nm digital SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3049–3058, Dec. 2013.
13. C. P. Hurrell, C. Lyden, D. Laing, D. Hummerston, and M. Vickery,
14. "An 18 b 12.5 MS/s ADC with 93 dB SNR," *IEEE J. Solid-State*
15. *Circuits*, vol. 45, no. 12, pp. 2647–2654, Dec. 2010.
16. J. Craninckx and G. van der Plas, "A 65 fJ/conversion-step 0-to-50 MS/s 0-to-0.7 mW 9 b charge-sharing SAR ADC in 90 nm digital CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2007, pp. 246–600.
17. B. Malki, T. Yamamoto, B. Verbruggen, P. Wambacq, and J. Craninckx, "A 70 dB DR 10 b 0-to-80 MS/s current-integration SAR ADC with adaptive dynamic range," *IEEE J. Solid-State Circuits*, vol. 49, no. 5, pp. 1173–1183, May 2014.
18. R. Kapusta, J. Shen, S. Decker, H. Li, E. Ibaragi, and H. Zhu, "A 14 b
19. 80 MS/s SAR ADC With 73.6 dB SNDR in 65 nm CMOS," *IEEE*
20. *J. Solid-State Circuits*, vol. 48, no. 12, pp. 3059–3066, Dec. 2013.
21. M. Maddox, B. Chen, M. Coln, R. Kapusta, J. Shen, and L. Fernando, "A 16 bit linear passive-charge-sharing SAR ADC in 55 nm CMOS," in *Proc. IEEE Asian Solid-State Circuits Conf. (A-SSCC)*, Nov. 2016, pp. 153–156.
22. J. Shen *et al.*, "A 16-bit 16-MS/s SAR ADC with on-chip calibration in 55-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 53, no. 4, pp. 1149–1160, 2018.
23. B. Chen, M. Maddox, M. C. W. Coln, Y. Lu, and L. D. Fernando, "Precision passive-charge-sharing SAR ADC: Analysis, design, and measurement results," *IEEE J. Solid-State Circuits*, vol. 53, no. 5, pp. 1481–1492, May 2018.
24. W. Parry, "On the β-expansions of real numbers," *Acta Math. Hungarica*, vol. 11, no. 3, pp. 401–416, Sep. 1960.
25. B. Murmann, "On the use of redundancy in successive approximation A/D converters," in *Proc. IEEE Int. Conf. Sampling Theory Appl.*, Jul. 2013, pp 1–4.
26. M. van Elzakker, E. van Tuijl, P. Geraedts, D. Schinkel, E. Klumperink, and B. Nauta, "A 1.9 μW 4.4 fJ/conversion-step 10 b 1 MS/s charge-redistribution ADC," in *IEEE Int. Solid-State Circuits Conf.*
27. *(ISSCC) Dig. Tech. Papers*, Feb. 2008, pp. 244–610.
28. S.-W. Chen and R. W. Brodersen, "A 6 b 600 MS/s 5.3 mW asynchronous ADC in 0.13 μm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2006, pp. 2350–2359.
29. J.-H. Tsai *et al.*, "A 0.003 mm^2 10 b 240 MS/s 0.7 mW SAR ADC in 28 nm CMOS with digital error correction and correlated-reversed switching," *IEEE J. Solid-State Circuits*, vol. 50, no. 6, pp. 1382–1398, Jun. 2015.

5 ARCHITECTURE AND ANALYSIS OF SYNTHESIZABLE SEGMENTED BUFFER-FREE SAR-ADCS

This chapter presents the theoretical analysis of passive charge sharing-based segmented successive-approximation register (SAR) analog-to-digital converter (ADC), where the precise reference source in a capacitive digital-to-analog converter (CDAC) is replaced by a capacitor that is β times larger than its bit capacitor and precharged to the reference level, known as a reference charge reservoir (RCR); this leads to buffer-free SAR-ADCs, which are readily implantable using the RAIL flow. A segmented SAR-ADC uses a coarse SAR-ADC to compute some most significant bits (MSBs). Four methods, namely aligned switching (AS) with bitwise RCRs, AS with a subsample-wise RCR, detect-and-skip aligned switching (DAS-AS) with bitwise RCRs, and DAS-AS with a subsample-wise RCR are introduced for setting fine MSBs. Closed-form analytic expressions of the reference error due to the finite reference capacitance are derived and validated by behavioral modeling and circuit simulation of an 11-bit 50 MS/s segmented SAR ADC realized in a 65-nm CMOS technology. The error expressions can be used to select one of the four methods for setting the fine MSBs and to determine β for the required linearity or for implementing digital circuitry for precise error correction.

5.1 Introduction

Successive approximation register (SAR) analog-to digital converters (ADCs) are the most energy-efficient data conversion solutions for 10–100-MHz sampling rates with 10–12 bit resolution [1]–[4]. Combined with pipeline [5], [6], time interleaving [6], and noise shaping [7], SAR ADCs are being extended to applications with even higher speed up to GHz [8] or higher resolution up to 18 bits [9].

One major reason why the SAR-ADC architecture is increasingly popular is its scalability with the process technology and its amenability to a digital-centric design methodology. To remove power-hungry buffers associated with precise capacitive digital-to-analog converter (CDAC) references, various forms of passive charge sharing have been introduced [10]-[16]. A Sample-wise Switched Reference Charge Reservoir (SS-RCR) technique was introduced [12], where a sufficiently large capacitor (reservoir), β times larger than the total bit capacitors, precharged to the reference voltage level during the sample phase acts as the reference for all bit switchings during the entire ADC. Very recently, a Bitwise Switched Reference Charge Reservoir (BSRCR) technique was proposed [13], where the charge reservoir capacitor is split into the corresponding bit, and before each bit switching, the corresponding bit charge reservoir capacitor is precharged to the reference levels and used as the reference during each bit decision. With BS-RCRs, a 16-bit 1 MS/s SAR-ADC in 55 nm CMOS with 6.95 mW with a FoM of 738 fJ/conversion-step [14] and a 16-bit 16 MS/s SAR in 55 nm CMOS with FoM 157.4 fJ/conversion-step [15] have been demonstrated. It has been approved theoretically [16] that the BS-RCR technique yields better linearity than the SS-RCR technique, a seemingly counterintuitive fact. Furthermore, the reference error due to finite BS-RCRs appears in the form of digitally correctable *bit weight error*, where the ith bit weight is attenuated by $1 + 2^{i+1-N}\beta$ for an N-bit SAR-ADC.

To reduce the energy associated with most significant bit (MSB) switching, a dominating factor in affecting SAR-ADC switching energy efficiency, a segmented architecture has been developed, where a coarse SAR-ADC computes the MSBs for a fine SAR-ADC [17]. This article extends reference charge reservoirs (RCRs) to segmented SAR-ADC design. Since BS-RCR has better linearity, we use it for the coarse ADC, as well as fine ADC LSB switching. For copying the results from coarse ADC to the MSBs of the fine ADC, there are two switching methods: aligned switching (AS), all bits are switched together; detect-and-skip AS (DAS-AS) [17], where only some bits are switched.

For the fine MSBs reference, it can be bitwise RCRs and subsample-wise RCR. This leads to four switching methods for setting fine MSBs, known as AS with bitwise RCRs, AS with a subsample-wise RCR, DAS-AS with bitwise RCRs, and DAS-AS with a subsample-wise RCR.

Theoretically, we have derived the analytical formula of the reference error for various RCR-based segmented SARADC switching schemes. From these formulae, we have proved that successive decisions using bitwise RCRs and fine MSB switching using DAS-AS subsample-wise RCR yields the near the smallest reference error while saving most switching energy.

This chapter is organized as follows. Section 4.2 describes the operations of charge-redistribution and charge-sharing SAR-ADC. Section 4.3 presents the theoretical analysis of the RCR-based segmented SAR-ADCs. Simulation validation and discussion are described in Section 4.4. Section 4.5 concludes this chapter.

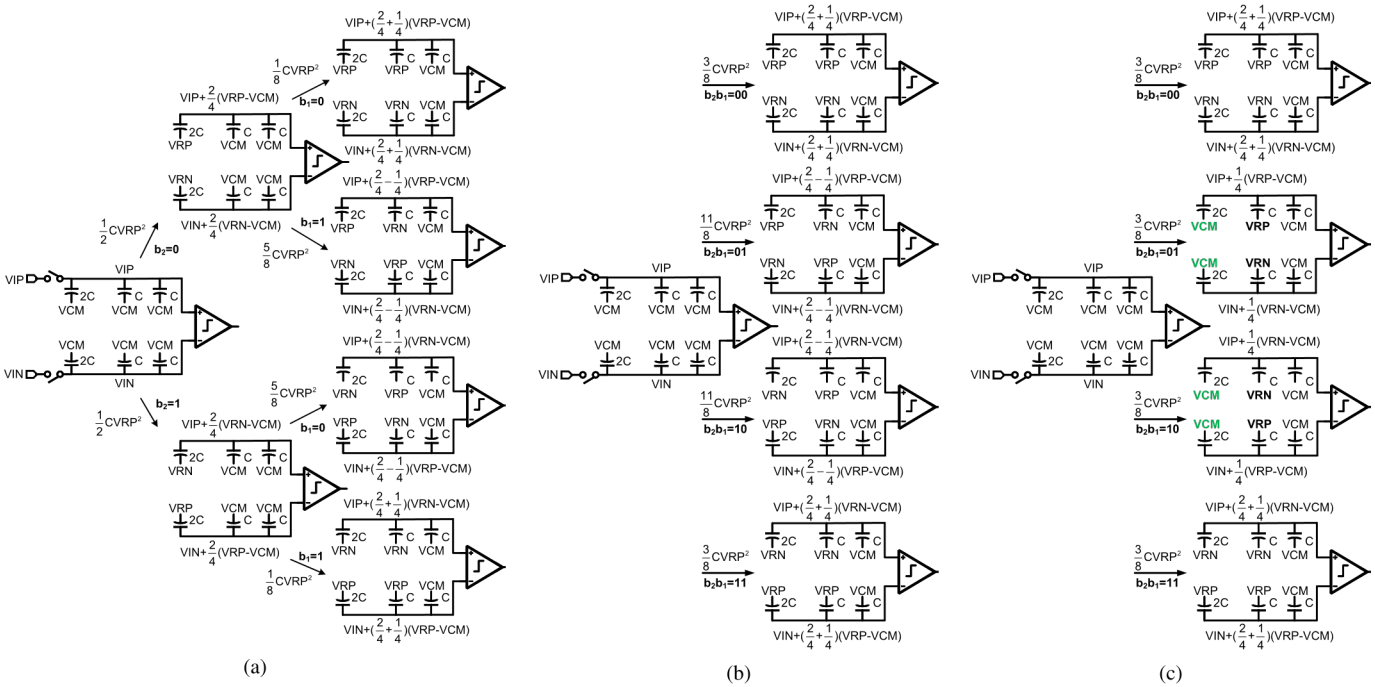


Fig. 1. Three bit charge redistribution CDAC switching examples. (a) MCS. (b) AS. (c) DAS-AS.

5.2 Charge Redistribution and Charge Sharing-based SAR ADCs: A Tutorial Example

In this section, a 3-bit SAR ADC is used to illustrate three switching methods: VCM-based merged capacitor switching (MCS), AS, and DAS-AS, and their switching energy consumption. First, reference sources are used for CDAC; this is the *charge-redistribution* SAR ADC. We then replace reference sources with precharged capacitors, known as RCRs; this leads to the *passive charge-sharing* SAR ADC.

5.2.1 Charge-Redistribution SAR-ADC Switchings

Fig. 1(a) illustrates the operation of a 3-bit binary-weighted differential charge-redistribution SAR ADC using MCS [18]. Two reference levels are VRP (high, often VDD) and VRN (low, often GND) with the common mode voltage VCM = *(*VRP + VRN*)/*2. Bit capacitors are binary weighted with 2*C*, *C*, and *C* where *C* is the unit capacitance.

Initially, differential inputs VIP and VIN are sampled onto the top plates of the CDAC, and all bottom plates are connected to VCM. Then, the comparator compares the two top-plate voltages and determines the MSB b_2. The bottom plate of the p-side CDAC MSB bit capacitor is switched to VRN if $b_2 = 1$ otherwise to VRP, whereas the bottom plate of the N-side MSB bit capacitor switches oppositely to VRP and VRN. The resulting top plate voltages are shown in the figure. This process continues by comparing the two resulting top plate voltages to obtain the second MSB and switches the bottom plates to VRP or VRN accordingly. The LSB is finally obtained by comparing the resulting top-plate voltages.

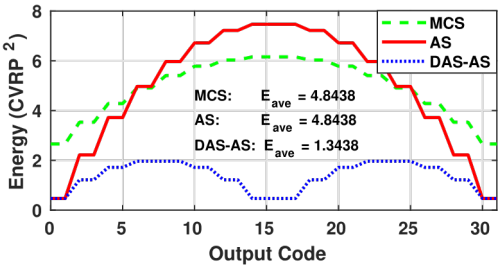


Fig. 2. Calculated switching energy of a 5-bit SAR ADC using MCS, AS, and DAS-AS.

Hence, a total of four cases of MCS are shown in Fig. 1(a), where for each switching, the resulting top-plate voltage and the energy consumed are marked. Now consider that all digital bits are known, and all the bottom plates of the 3-bit CDAC are switched simultaneously in one step. This is referred to as AS [17]. The four cases are illustrated in Fig. 1(b). If we are interested only in the final top-plate voltages with AS, opposite switching can be avoided. This was developed like the DAS-AS operation [17]. Fig. 1(c) shows all DAS-AS four cases.

We observe that all the three switching methods yield the same top-plate voltage but different energy consumption values. The average energy consumed by AS is the same as that of MCS. DAS-AS consumes the least energy. Fig. 2 shows the switching energy for each code of a 5-bit SAR-ADC uses these three different switching methods.

5.2.2 Charge-Sharing SAR ADC Switchings

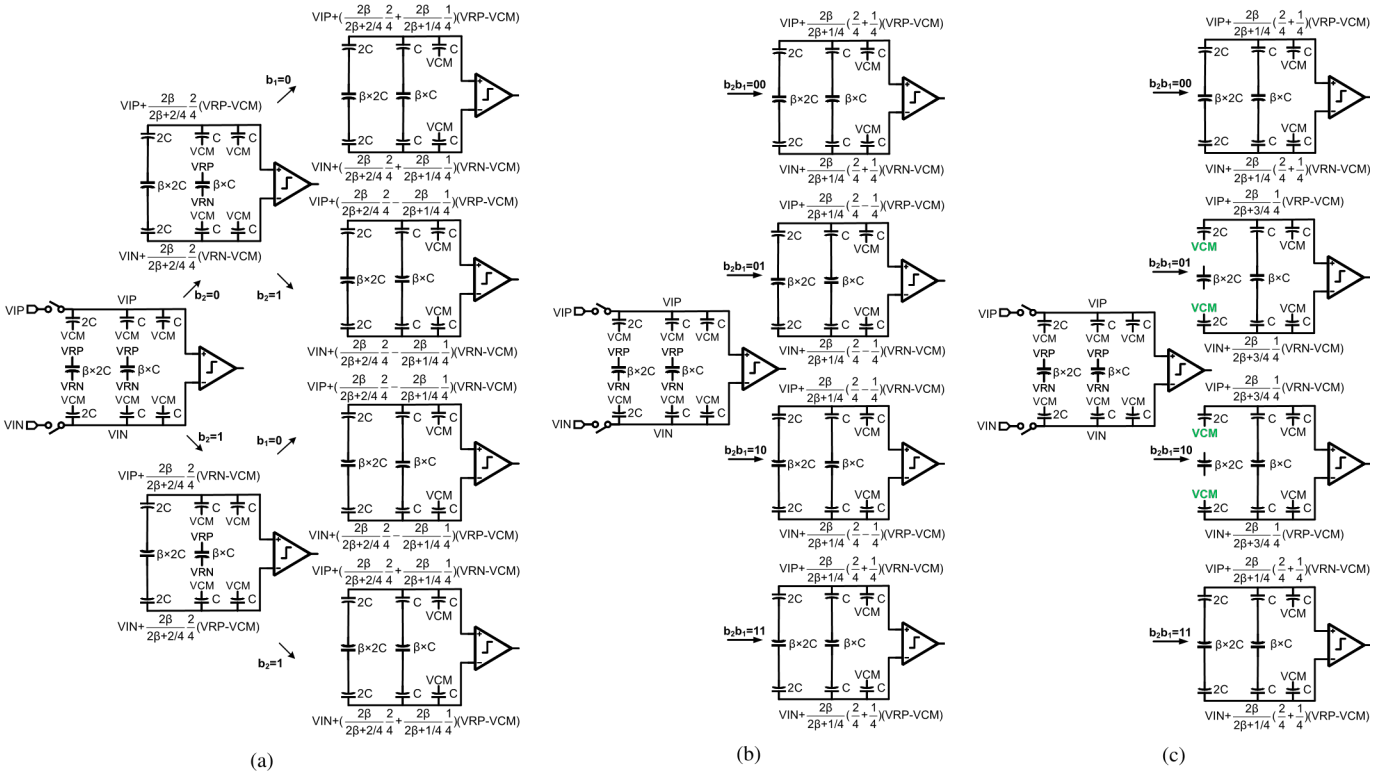


Fig. 3. Three-bit bitwise charge-sharing CDAC switching examples. (a) MCS. (b) AS. (c) DAS-AS.

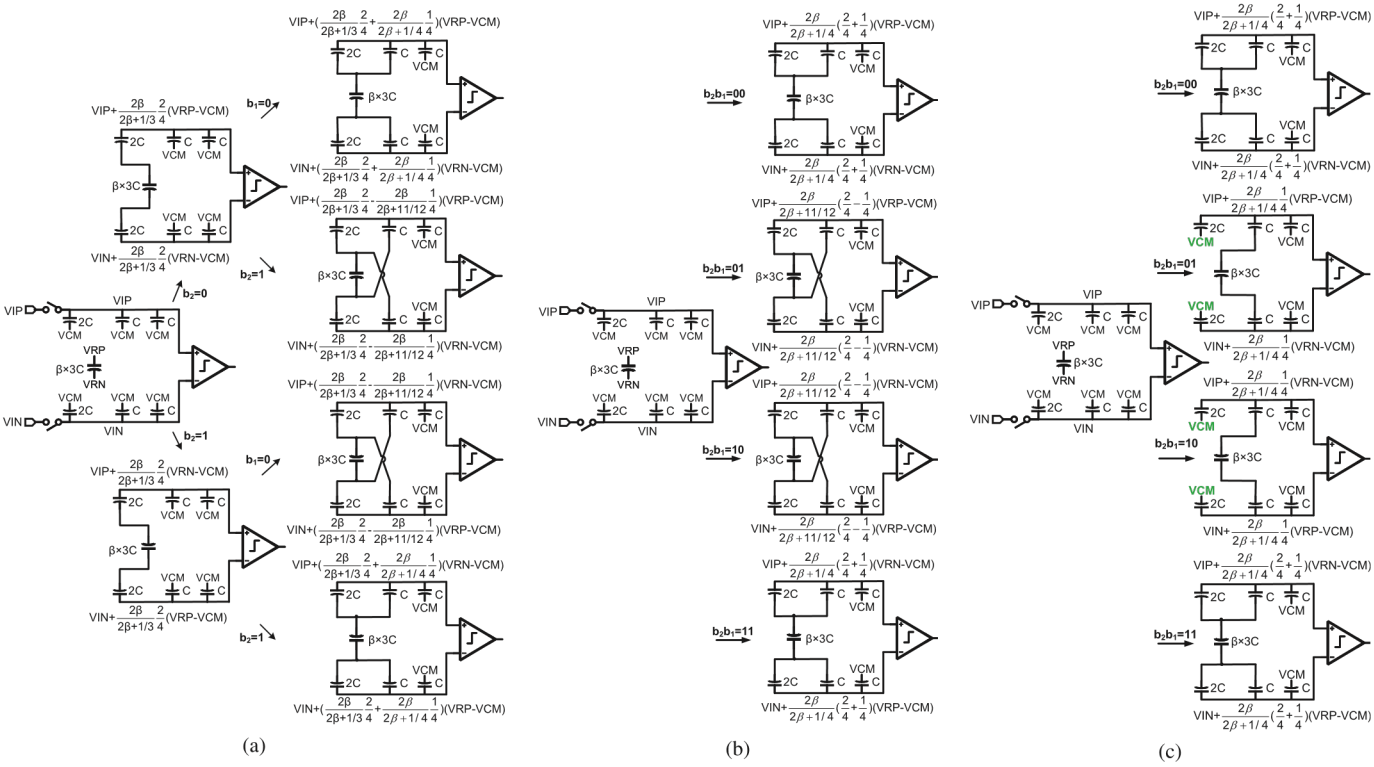


Fig. 4. Three-bit sample-wise charge-sharing CDAC switching examples. (a) MCS. (b) AS. (c) DAS-AS.

Fig. 3(a)–(c) shows, respectively, the decision process of a 3-bit charge-sharing SAR ADC using MCS, AS, and DAS-AS methods. For each bit decision, the reference sources VRP and VRN are replaced by charge reservoir capacitors

β-times larger than the corresponding bit capacitances, precharged to VRP and VRN. This is known as bitwise reference charge reservoir. Figure 4 shows the decision process of another structure of a 3-bit charge sharing SAR-ADC where all reference capacitors connected in parallel to form a sample-wise charge reservoir.

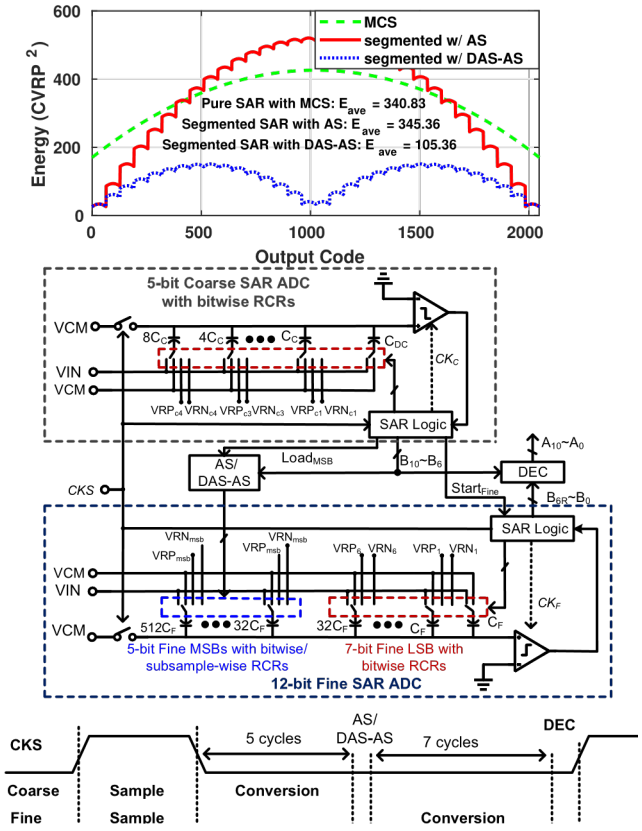


Fig. 5. Prototype segmented SAR-ADC architecture.

For all the three methods, for each bit switching, the resulting top plate voltages are labeled in the figures. We can see that there is a β-dependent term, referred to as α, $0 < \alpha < 1$, reflecting the reference error due to finite β. When $\beta = \infty$ and $\alpha = 1$. We observe that with bitwise RCRs, for term α in AS is not less than that in MCS, which is not less than that in DAS-AS. Thus, for bitwise RCR, AS has better linearity than MCS, and MCS has better linearity than DAS-AS. With sample-wise RCR, the term α in AS is not greater than that in MCS, which is not greater than that in DAS-AS. Hence, for sample-wise RCR, DAS-AS has better linearity than MCS, and MCS has better linearity than AS.

Fig. 6. Calculated switching energy of an 11-bit SAR ADC with MCS, a segmented SAR ADC shown in Fig. 5 with AS and with DAS-AS.

5.3 Analysis of Segmented SAR Switching with RCRs

This section presents the theoretical analysis of the SARADC linearity for various switching methods for a segmented charge-sharing *(N +M)*-bit SAR ADC where a coarse SAR-ADC computes the *N* MSBs, and a fine SAR ADC computes the *M* LSBs. Whenever possible, the analysis is based on the equivalent circuit model introduced in [16] since it provides more circuit intuition. Nevertheless, all the results in this section have been derived directly using the principle of charge conservation following the same procedure as in the Appendices of [16].

5.3.1 Segmented SAR-ADC Architecture

Fig. 5 shows the architecture of the proposed RCR-based segmented SAR ADC. For simplicity, it shows only the single-ended architecture, but the implementation is differential. It consists of a 5-bit coarse SAR ADC, a 12-bit fine SAR ADC including one redundant bit, an AS or DAS-AS logic block for setting fine MSBs, and a digital error correction (DEC) block. The coarse or fine SAR ADC contains a bootstrapped sampling switch, an RCR-based CDAC, a voltage comparator [17], and an SAR logic block.

Differential inputs VIP and VIN are sampled on the CDAC bottom plates in both coarse and fine ADCs, whereas the top plates are connected to VCM during the sampling phase. Simultaneously, the bit reservoir capacitor sample references VRP and VRN. After sampling, the 5-bit coarse SAR ADC computes the 5 MSBs bit by bit using VCM-based MCS [18]. For RCR-based CDAC switching, the reservoir capacitor precharged to voltage level VRP or VRN is used instead of the reference source VRP or VRN. Then, the AS or DAS-AS logic block loads digits from the coarse ADC to the 5-bit fine MSBs in one step. Since the fine 5 MSBs are set at one step, one of the four switching methods: AS with bitwise RCRs, AS with a subsample-wise RCR, DAS-AS with bitwise RCRs, and DAS-AS with a subsample-wise RCR can be used. Finally, the remaining seven fine LSBs are determined successively using MCS with bitwise RCRs. The five MSBs from the coarse ADC and the seven LSBs from the fine ADC are combined into the DEC block to derive the final 11-bit output.

Fig. 6 compares the switching energy of each code of a 11-bit charge-redistribution segmented SAR-ADC and that of a conventional 11-bit charge-redistribution SAR-ADC. The conventional 11-bit SAR-ADC uses MCS. The segmented ADC uses MCS for the coarse ADC, uses AS or DAS-AS to copy the MSBs for the fine ADC, and MCS for resolving the remaining LSBs. We can see that the segmented SAR-ADC with DAS-AS reduces switching energy by 69.1% compared with a conventional SAR with MCS. Here, the coarse unit capacitance is twice the fine unit capacitance in the segmented architecture.

5.3.2 Bitwise RCR-Based Coarse SAR ADC Successive Switching

The *N*-bit coarse SAR ADC decides each bit successively using VCM-based MCS [18] with bitwise RCRs. We use a bracketed superfix $^{(k)}$ to indicate the *k*th *bit decision*, and a subscript $_i$ to represent the *i*th bit. The equivalent half-circuit model [16] for charge-sharing from *(k*+1*)*th to *k*th bit decision is shown in Fig. 7. The bottom plate of the *k*th bit weight capacitor C_k is switched from connecting to VCM to either VRP or VRN dependent upon b_k. Without loss of generality, we assume VRP. Voltage $\mathrm{VRP}_k^{(k)}$ can be obtained from the following charge-sharing equation:

$$(\mathrm{VRP}-\mathrm{VCM})(2\beta C_k) = (\mathrm{VRP}_k^{(k)}-\mathrm{VCM}) \\ [2\beta C_K + (C_a+C_b)C_k/(C_a+C_b+C_k)] \quad (1)$$

(1)

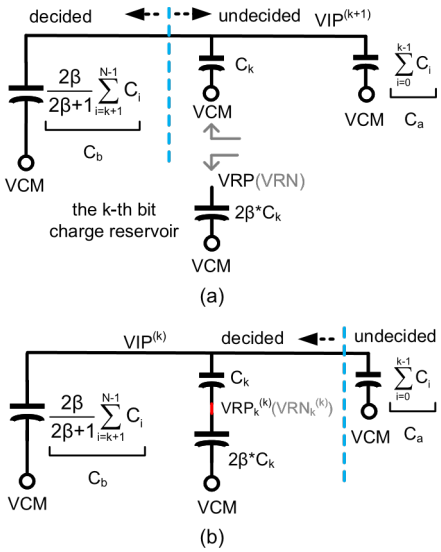


Fig. 7. Half-circuit model and derivation of bitwise reference switching in the coarse SAR ADC. (a) *(k* + 1*)*th decision. (b) *(k)*th decision.

Then the p-side top-plate voltage change due to this charge sharing is obtained by capacitive division as

$$\mathrm{VIP}^{(k)} - \mathrm{VIP}^{(k+1)} = \frac{C_k(1-2b_k)}{C_a + C_b + C_k}(\mathrm{VRP}_k^{(k)} - \mathrm{VCM}) \quad (2)$$

Substituting (1) to (2) with $C_T = C_a + C_b(2\beta + 1)/2\beta + C_k$, we obtain the following:

$$\mathrm{VIP}^{(k)} - \mathrm{VIP}^{(k+1)} = \alpha^{(k)}\frac{(1-2b_k)C_k}{C_T}(\mathrm{VRP} - \mathrm{VCM}) \quad (3)$$

$$\alpha^{(k)} = \frac{2\beta}{2\beta + \sum_{j=0}^{k-1}\frac{C_j}{C_T}} = \frac{2\beta}{2\beta + (1/2)^{N-k}} \quad (4)$$

where N is the number of bits in the coarse ADC, C_T is the sum of the coarse CDAC capacitance from C_0 to C_{N-1}, and b_k is the kth bit value being either 1 or 0. The last equality in (4) holds if bit capacitances are binary weighted.

Note that $0 < \alpha^{(k)} < 1$, $\alpha^{(k)}$ increases from the MSB ($k = N - 1$) to the LSB ($k = 1$) and $\alpha^{(k)} = 1$ when $\beta = \infty$.

Fig. 8 shows $\alpha^{(k)}$ at each bit decision of a 5-bit BS-RCR-based coarse SAR ADC with $\beta = 16$. The term $\alpha^{(k)}$ =1 in (3) is caused by the reference error due to charge-sharing between the reference capacitor and the bit capacitor. The elegance of bitwise RCR-based SAR ADCs is that the reference error is the form of linear bit weight error as shown in (4) and in Fig. 8 [16]. Bit weight error factor $(1-\alpha^{(k)})$ depends only on β, k and N, irrespective of the input.

5.3.3 Fine MSB AS and DAS AS: Bitwise and Subsample Wise

Let C_F denote the fine ADC unit capacitance, C_{TM} the N-bit MSB capacitance, C_{TL} the M-bit LSB capacitance, and C_{TF} the total capacitance in the fine ADC.

Fig. 8. $\alpha^{(k)}$ at each bit decision for the coarse SAR ADC in a segmented architecture.

1) Fine MSB AS: Bitwise and Subsample Wise: When loading the coarse digits directly to the fine MSBs in one step, the fine CDAC top-plate voltages after AS can be expressed

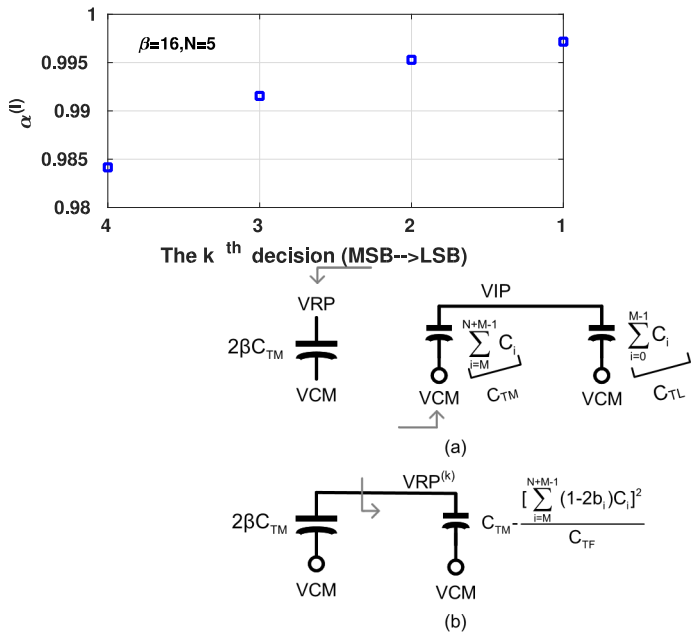


Fig. 9. Half-circuit model and derivation of sample-wise reference switching in fine MSB. (a) Sampling phase. (b) AS decision phase.

in the following equation:

$$\mathrm{VIP}^{(M)} = \mathrm{VIP} + \alpha^{(M)} \sum_{i=M}^{N+M-1} \frac{C_i}{C_{\mathrm{TF}}}(1-2b_i)(\mathrm{VRP}-\mathrm{VCM}) \tag{5}$$

$$\mathrm{VIN}^{(M)} = \mathrm{VIN} + \alpha^{(M)} \sum_{i=M}^{N+M-1} \frac{C_i}{C_{\mathrm{TF}}}(1-2b_i)(\mathrm{VRN}-\mathrm{VCM}) \tag{6}$$

where b_i are the digits of —— the N-bit fine MSBs, which are equal to coarse bits. For AS, the coefficient $\alpha^{(M)}$ is extracted to the outside of the sum symbol. This is because AS is an one-step operation involving N-bits.

For bitwise referenced, $\alpha^{(M)}$ can be expressed as follows:

$$\alpha^{(M)} = \frac{2\beta}{2\beta+1-\sum_{i=M}^{N+M-1}\frac{C_i}{C_{\mathrm{TF}}}} = \frac{2\beta}{2\beta+1-\frac{C_{\mathrm{TM}}}{C_{\mathrm{TF}}}} \tag{7}$$

For subsample-wise referenced, the equivalent half-circuit model from the sampling phase to the AS phase is shown in Fig. 9 [16]. Charge is conserved before and after AS; that is

$$(\mathrm{VRP}-\mathrm{VCM})(2\beta C_{\mathrm{TM}})=(\mathrm{VRP}^{(M)}-\mathrm{VCM})$$
$$[2\beta C_{\mathrm{TM}}+C_{\mathrm{TM}}-\frac{[\sum_{i=M}^{N+M-1}C_i(2b_i-1)]^2}{C_T}] \quad (8)$$

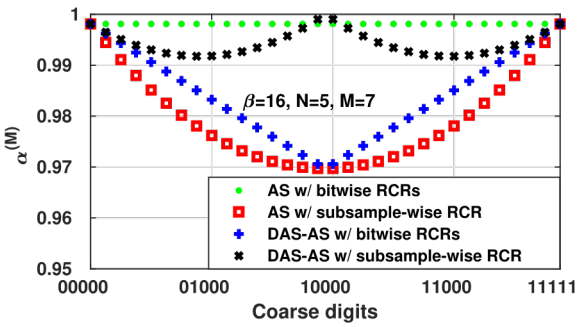


Fig. 10. $\alpha^{(M)}$ versus coarse codes for the fine MSBs with four different switching methods in a segmented architecture.

Then the p-side top-plate voltage change can be obtained based on capacitive division

$$\mathrm{VIP}^{(M)}=\mathrm{VIP}+(\mathrm{VRP}^{(M)}-\mathrm{VCM})\sum_{i=M}^{N+M-1}\frac{C_i}{C_{\mathrm{TF}}}(1-2b_i) \quad (9)$$

Substituting (8) to (9) with $C_{\mathrm{TF}}=C_{\mathrm{TM}}+C_{\mathrm{TL}}$, we obtain the same result as (5) with $\alpha^{(k)}$ defined by the following equation:

$$\alpha^{(M)}=\frac{2\beta}{2\beta+1-\frac{[\sum_{i=M}^{N+M-1}C_i(2b_i-1)]^2}{C_{\mathrm{TF}}C_{\mathrm{TM}}}} \quad (10)$$

We note that coefficient $\alpha^{(M)}$ in (7) is input-independent because both bitwise RCRs and AS are input-independent. $\alpha^{(M)}$ is shown in Fig. 10 by the dot · marker. The last term in the denominator of (7) represents the ratio of the N-bit fine MSB capacitance to the total fine capacitance. The subsample-wise reference coefficient $\alpha^{(M)}$ in (10) is input-dependent because of the subsample-wise RCR. In this case, $\alpha^{(M)}$ has 2^N possible values based on the N-bit MSBs, which is shown in Fig. 10 by the square marker. We further observe that the square term in (10) is less than or equal to C_{TM}. Thus, $\alpha^{(M)}$ in (7) ≥ $\alpha^{(M)}$ in (10). Therefore, AS with bit-wise charge reservoirs always has better linearity than AS with a subsample-wise charge reservoir.

2) Fine MSB DAS AS: Bitwise and Subsample Wise: Now, we consider that the fine MSBs use DAS-AS based on the computed coarse MSBs. If the coarse MSB is 1, then the p-side (n-side) CDAC bottom plates either connect to VCM or switch to VRN (VRP). The switched bit capacitance in the fine MSBs is

$$C_{\mathrm{sw}}=\sum_{i=M}^{N+M-1}C_ib_i \quad (11)$$

where b_i are the digits of the N-bit fine MSBs, which are obtained by left rotating coarse bits. The differential fine CDAC top-plate voltages after DAS-AS operation are given in the following equation:

$$\mathrm{VIP}^{(M)} = \mathrm{VIP} + \alpha^{(M)} \sum_{i=M}^{N+M-1} \frac{C_i b_i}{C_{\mathrm{TF}}} (\mathrm{VRN} - \mathrm{VCM}) \quad (12)$$

$$\mathrm{VIN}^{(M)} = \mathrm{VIN} + \alpha^{(M)} \sum_{i=M}^{N+M-1} \frac{C_i b_i}{C_{\mathrm{TF}}} (\mathrm{VRP} - \mathrm{VCM}) \quad (13)$$

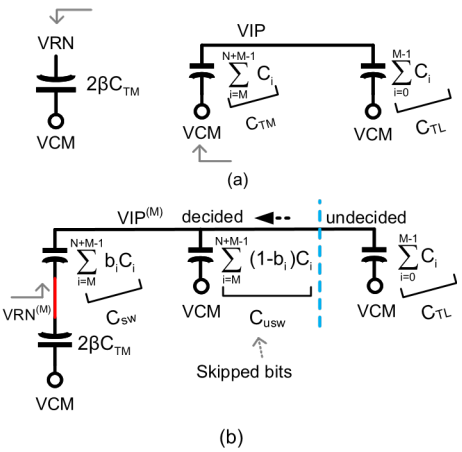


Fig. 11. Half-circuit model and derivation of subsample-wise reference switching when coarse MSB is 1 in fine MSB. (a) Sampling phase. (b) DAS-AS decision phase.

If the coarse MSB is 0, the total switched bit capacitance in the fine MSBs is

$$C_{\mathrm{sw}} = \sum_{i=M}^{N+M-1} C_i (1 - b_i) \quad (14)$$

The CDAC top-plate voltages can be expressed by

$$\mathrm{VIP}^{(M)} = \mathrm{VIP} + \alpha^{(M)} \sum_{i=M}^{N+M-1} \frac{C_i (1 - b_i)}{C_{\mathrm{TF}}} (\mathrm{VRP} - \mathrm{VCM}) \quad (15)$$

$$\mathrm{VIN}^{(M)} = \mathrm{VIN} + \alpha^{(M)} \sum_{i=M}^{N+M-1} \frac{C_i (1 - b_i)}{C_{\mathrm{TF}}} (\mathrm{VRN} - \mathrm{VCM}) \quad (16)$$

For bitwise-referenced DAS-AS, $\alpha^{(M)}$ is given in (17)

$$\alpha^{(M)} = \frac{2\beta}{2\beta + 1 - \frac{C_{\mathrm{sw}}}{C_{\mathrm{TF}}}} \quad (17)$$

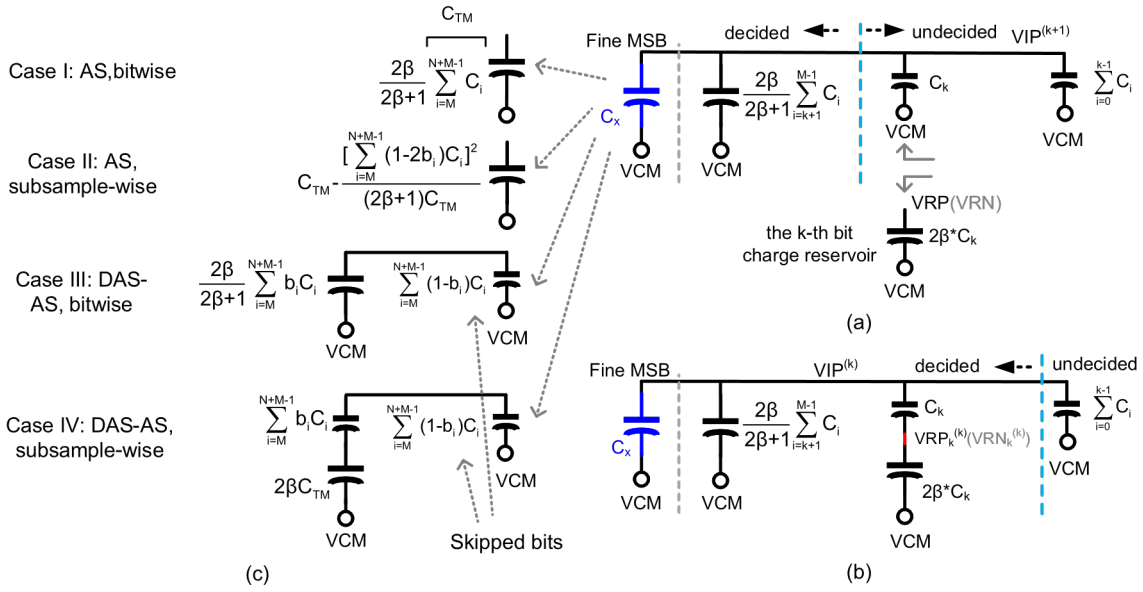


Fig. 12. Half-circuit model and derivation of bitwise reference switching in fine LSBs. (a) *(k* + 1*)*th decision. (b) *(k)*th decision. (c) C_x in four different switching methods

For subsample-wise referenced DAS-AS, there is no cross coupling between the p-side and the n-side CDACs, so the equivalent half-circuit model can be used to derive the results directly. Fig. 11 shows the equivalent half-circuit model from the sampling phase to the DAS-AS phase. The charge is conserved before and after AS; that is

$$(\mathrm{VRN}-\mathrm{VCM})(2\beta C_{\mathrm{TM}})=(\mathrm{VRN}^{(M)}-\mathrm{VCM})$$
$$[2\beta C_{\mathrm{TM}}+(C_{\mathrm{TL}}+C_{\mathrm{usw}})C_{\mathrm{sw}}/(C_{\mathrm{TL}}+C_{\mathrm{usw}}+C_{\mathrm{sw}})] \quad (18)$$

Then the p-side top-plate voltage change can be obtained based on capacitive division

$$\mathrm{VIP}^{(M)}=\mathrm{VIP}+(\mathrm{VRN}^{(M)}-\mathrm{VCM})\sum_{i=M}^{N+M-1}\frac{C_i b_i}{C_{\mathrm{TF}}} \quad (19)$$

Substituting (18) to (19) with $C_{\mathrm{TM}}=C_{\mathrm{sw}}+C_{\mathrm{usw}}$ and $C_{\mathrm{TF}}=C_{\mathrm{TM}}+C_{\mathrm{TL}}$, we can obtain the same result as (12) with $\alpha^{(M)}$ defined by the following equation:

$$\alpha^{(M)}=\frac{2\beta}{2\beta+\frac{C_{\mathrm{sw}}}{C_{\mathrm{TM}}}(1-\frac{C_{\mathrm{sw}}}{C_{\mathrm{TF}}})} \quad (20)$$

We have $\alpha^{(M)}$ in (20) ≥ $\alpha^{(M)}$ in (17) by comparing (20) with (17). This is because, in the DAS-AS operation, the ratio of the reservoir capacitance to the switched bit capacitance is $\beta C_{\mathrm{TM}}/C_{\mathrm{sw}}$ for subsample-wise RCR, greater than or equal to β for bitwise RCRs. Thus, subsample-wise referenced DAS-AS always has better linearity than bitwise referenced DAS-AS. $\alpha^{(M)}$ in (17) and (20) versus coarse digits are shown in Fig. 10 with the plus '+' marker and cross '*x*' marker.

3) Comparison of Four MSB Switching Methods: Now, we compare $\alpha^{(M)}$ in (7), (10), (17), and (20) by examining the last two terms in the denominator of $\alpha^{(M)}$. Consider the DAS-AS method, the total switched capacitance C_{sw}

expressed in (11) and (14) is mathematically equal to the term $|\sum_{i=M}^{N+M-1} C_i(2b_i-1)$ in AS shown in the last term in the denominator of (10). We rewrite (10) as

$$\alpha^{(M)} = \frac{2\beta}{2\beta+1-\frac{C_{\rm sw}^2}{C_{\rm TF}C_{\rm TM}}} \tag{21}$$

Thus, $\alpha^{(M)}$ in (10) ≤ $\alpha^{(M)}$ in (17). Up until now, we have $\alpha^{(M)}$ in (7) ≥ $\alpha^{(M)}$ in (10), and $\alpha^{(M)}$ in (20) ≥ $\alpha^{(M)}$ in (17) ≥ $\alpha^{(M)}$ in (10). Finally, we compare $\alpha^{(M)}$ in (7) with the one in (20) by comparing $1-C_{\rm TM}/C_{\rm TF}$ with $(C_{\rm sw}/C_{\rm TM})/(1-(C_{\rm sw}/C_{\rm TF}))$. We have $\alpha^{(M)}$ in (7) ≥ $\alpha^{(M)}$ in (20) ≥ $\alpha^{(M)}$ in (17) ≥ $\alpha^{(M)}$ in (10) except at the middle codes. Therefore, in terms of linearity, bitwise AS is better than subsample-wise DAS-AS, which is better than bitwise DAS-AS, which is better than subsample-wise AS.

5.3.4 Bitwise RCR-Based Fine LSB Successive Switching

Once the *N* MSBs in the fine SAR-ADC are set based on the coarse ADC bits using one of the four methods, the remaining *M*-bit LSBs in the fine SAR-ADC can be decided by bitwise RCR-based MCS with $k = M-1,\ldots,1$. The top-plate voltages at step *k* can be analyzed using the equivalent half circuit models shown in Fig. 12, where C_x is the equivalent capacitance for the MSBs that have been set using one of these four methods. This is exactly similar to the analysis for the bitwise switched coarse ADC in Section III-B except with an extra C_x. Hence, the top-plate voltages at step *k* can be obtained in the same form as (1), which is rewritten as follows:

$$\mathrm{VIP}^{(k)} = \mathrm{VIP}^{(k+1)} + \frac{\alpha^{(k)}C_k}{C_{\rm TF}}(1-2b_k)(\mathrm{VRP}-\mathrm{VCM}) \tag{22}$$

$$\alpha^{(k)} = \frac{2\beta}{2\beta+1-\sum_{j=i}^{M-1}\frac{C_j}{C_{\rm TF}}-\frac{(C_{\rm TM}-C_x)(2\beta+1)}{C_{\rm TF}}} \tag{23}$$

Substituting C_x in Fig. 12(c) into (23), we have $\alpha^{(k)}$ for the case of *N* MSBs set by AS with bitwise RCRs, aligned switching with a subsample-wise RCR, DAS-AS with bitwise RCRs and DAS-AS with a subsample-wise RCR shown in the following equation, respectively:

$$\alpha^{(k)} = \frac{2\beta}{2\beta+1-\sum_{j=i}^{M-1}\frac{C_j}{C_{\rm TF}}-\frac{C_{\rm TM}}{C_{\rm TF}}} \tag{24}$$

$$\alpha^{(k)} = \frac{2\beta}{2\beta+1-\sum_{j=i}^{M-1}\frac{C_j}{C_{\rm TF}}-\frac{[\sum_{i=M}^{N+M-1}C_i(2b_i-1)]^2}{C_{\rm TF}C_{\rm TM}}} \tag{25}$$

$$\alpha^{(k)} = \frac{2\beta}{2\beta+1-\sum_{j=i}^{M-1}\frac{C_j}{C_{\rm TF}}-\frac{C_{\rm sw}}{C_{\rm TF}}} \tag{26}$$

$$\alpha^{(k)} = \frac{2\beta}{2\beta+1-\sum_{j=i}^{M-1}\frac{C_j}{C_{\rm TF}}-\frac{C_{\rm sw}}{C_{\rm TF}}\frac{2\beta+1}{2\beta\frac{C_{\rm TM}}{C_{\rm sw}}+1}} \tag{27}$$

We see that in all the four cases shown in Fig. 12, $\alpha^{(k)}$ differs only in the last term in the denominator. The third term in the denominator represents the ratio of the capacitance of the switched bits to the total fine capacitance of the successively decided bits in the fine LSBs. It has the same format as in (4). However, the switched bits in the

MSBs are different using different switching methods as reflected in the fourth term in the denominator for each case.

The error increases by C_x. We see that C_x in case II is greater than C_x in case IV, greater than in case III, and greater than in case I. Hence, for fine LSB switching, AS with bitwise for MSB switching yields the minimal error. But since each of LSB switching has an exponentially smaller weight compared to MSB switching, the reference error introduced by $\alpha^{(k)}$ is much smaller than $\alpha^{(M)}$.

Fig. 13(a) plots $\alpha^{(i)}$ versus coarse digits (from 00000 to 11111) at each fine bit decision (from k = 6 to 1) for DASAS. Fig. 13(b) shows $\alpha^{(i)}$ versus coarse digits at k = 6, and (c) shows $\alpha^{(i)}$ at each fine bit decision when coarse digits are 01000 at four difference cases. From Fig. 13(b) we see $\alpha^{(i)}$ is almost the same when 5-bit fine MSBs use AS or DAS-AS with subsample-wise RCR. Observe the last term in (27),

$$(C_{sw}/C_{TF})(2\beta + 1/2\beta(C_{TM}/C_{sw}) + 1) \geq (C_{sw}/C_{TF})(C_{sw}/C_{TM}).$$

It has the same format as the last term in the denominator of (25). C_{sw} is symmetric according to (11). It has the minimum value at coarse digits 01111 and 10000. Thus, C_{sw}^2 is minimum at 01111 and 10000. This explains why $\alpha^{(i)}$ in (27) has a single valley at the middle codes.

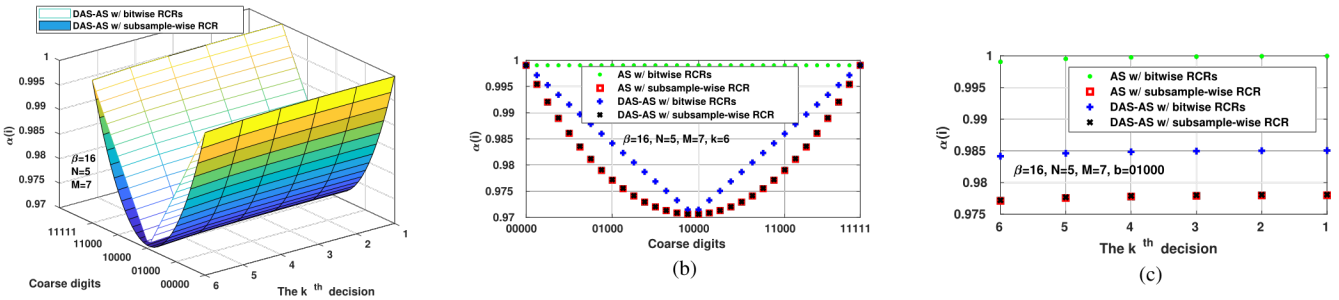


Fig. 13. $\alpha^{(i)}$ versus coarse digits at each bit decision for the fine LSBs in a segmented architecture. (a) 3-D plot of $\alpha^{(i)}$ versus coarse digits and decision step *(k)*. (b) 2-D plot of $\alpha^{(i)}$ versus coarse digits for selected decision step *(k)*= 6. (c) 2-D plot of $\alpha^{(i)}$ versus decision step *(k)* for selected coarse digits = 01000.

5.4 Simulation Validation and Discussion

Behavioral modeling and schematic level simulations were performed to validate the analysis of the segmented SAR ADCs with RCRs. The simulated segmented SAR ADCs with RCRs is shown in Fig. 5 and is implemented in a 65 nm CMOS technology. The coarse unit capacitance C_c is 2 fF whereas the fine unit capacitance C_F is 1 fF.

5.4.1 Static Performance of RCR-Based Segmented SAR ADC

Using analytic formula derived, we can compute, simulate, and compare the static performance differential nonlinearity (DNL) and integral nonlinearity (INL) of RCR-based segmented SAR ADCs. Fig. 14(a)–(d) upper four plots show the respective DNLs and INLs of a segmented 11-bit SAR ADC with a 5-bit coarse and a 12-bit fine SAR ADC using four switching methods for the fine MSBs and β = 5. With β = 5, the 5-bit coarse ADC can compute output digits correctly. These five digits are loaded to the 5-bit fine MSBs using one of four switching methods. As we can see, AS with bitwise RCRs in Fig. 14(a) yields the best static performance (INL/DNL less than 0.5 LSB). In fact from (7), we can compute the reference error for the 5-bit fine MSBs with β = 5 as

$$(1-\alpha^{(M)})(\mathrm{VRP}-\mathrm{VRN}) \approx 0.006(\mathrm{VRP}-\mathrm{VRN}) < \frac{1}{2^7}(\mathrm{VRP}-\mathrm{VRN}) \quad (28)$$

Thus, the 7-bit fine LSBs can compute results correctly. This yields the INL to be in the range of ±0.5 LSB.

The DNLs and INLs in Fig. 14(b)–(d) have some similarities. The segmented misaligned ADC transfer curves lead to different analog input mappings to the same digits, thus causing the DNL at the digits corresponding to misaligned transfer curves to be more than +1 LSB. The INL is the deviation in LSB of the actual transfer function of the ADC from the ideal transfer curve ($\beta = \infty$). The linearity of the segmented ADC is mainly decided by the *reference error introduced in the 5-bit fine MSBs*. This can be seen from $\alpha^{(M)}$, input-normalized ideal and actual transfer curves of 5-bit fine MSBs versus coarse digits plotted in Fig. 14(a)–(d) for four cases. The factor $\alpha^{(M)}$ affects the transfer curve through modulating the reference voltage in a manner of $\alpha^{(M)}$(VRP − VRN). Thus, the larger $\alpha^{(M)}$, the smaller the reference error $1 - \alpha^{(M)}$, the modulated actual transfer curve is closer to the ideal curve. To have a better visualization of reference modulation, $\beta = 1$ is used in plotting $\alpha^{(M)}$ and actual transfer curves. We see that the deviation of the actual transfer curve from the ideal curve in the 5-bit MSBs has the same shape of that of the INL of the 11-bit segmented SAR ADC. To minimize the reference error, $\alpha^{(M)}$ shall be as large as possible. While for minimum switching energy associated with the MSBs switching, DAS-AS with a subsample-wise RCR is a good option.

5.4.2 Comparison of MCS Based Successive Decision Only and Segmented SAR ADCs With RCR

Table I shows the minimum β to achieve nonlinearity ≤ 0.5 LSB of an 11-bit SAR ADC of different architectures with different RCR types. In a segmented architecture, fine MSBs using AS with bitwise RCRs only requires $\beta \geq 5$, whereas $\beta \geq 150$ is needed by using DAS-AS with a subsample-wise RCR. Table II shows the dynamic and static performance, as well as the average switching energy, of an 11-bit SAR ADC with $\beta = 16$ of different architectures using different RCR types. The dynamic performance is simulated using the sinusoidal input with an amplitude of 0.45 V.

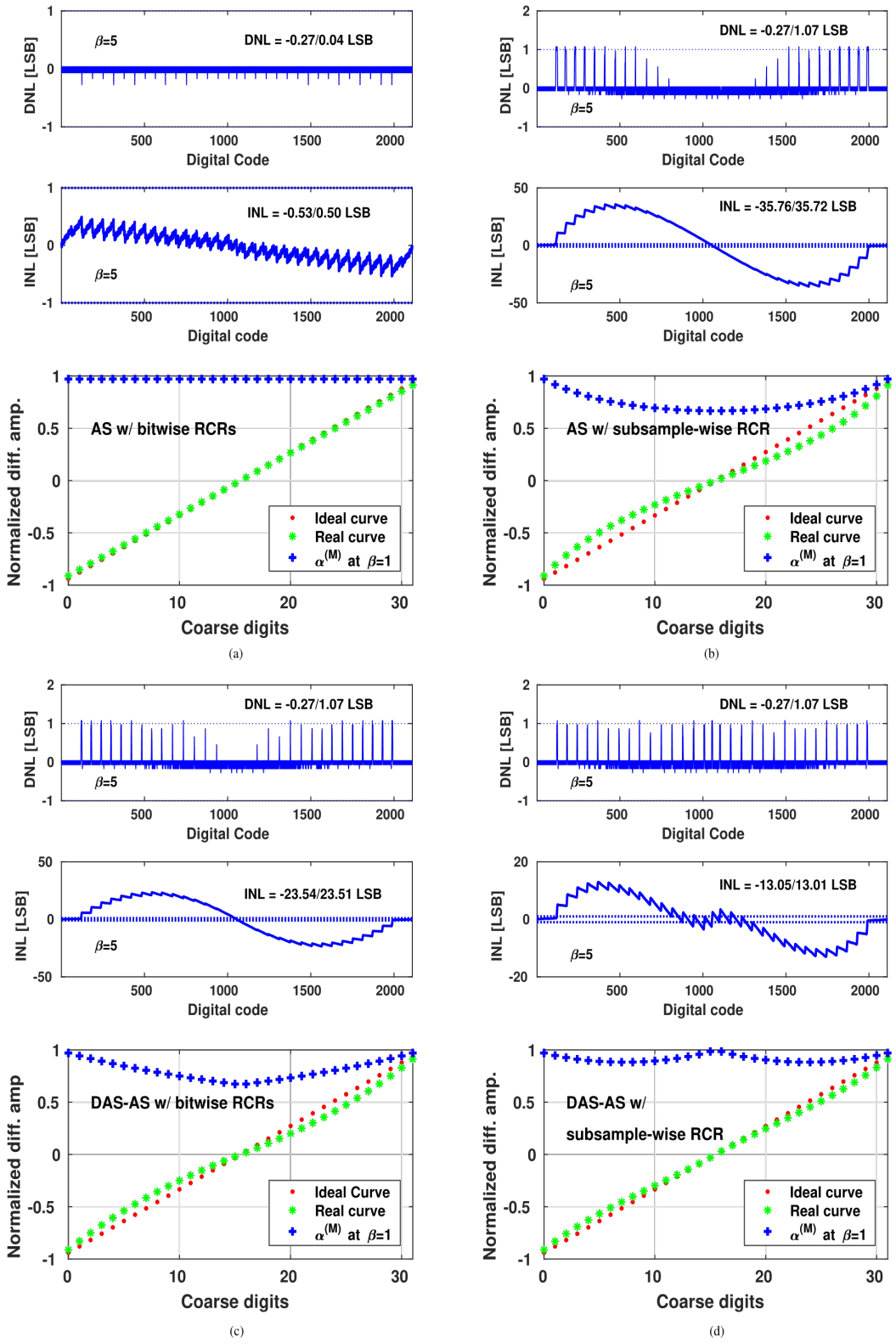


Fig. 14. Static performance and the fine CDAC top-plate voltage versus coarse digits of an 11-bit RCR based charge sharing segmented SAR ADC in which fine MSBs use different switching methods with bitwise or subsample-wise RCRs shown in (a)–(d).

The architecture using DAS-AS with a subsample-wise RCR has segmented architecture using AS with bitwise RCRs has the lowest switching energy at the cost of a slight reduction best linearity than the other cases, whereas the segmented in the SNDR and linearity. We note that with closed-form analytic expressions of the reference error

derived in this article implemented in digital circuitry, precise error correction can be accomplished even if smaller βs are used.

TABLE I

MINIMUM β TO ACHIEVE NONLINEARITY ≤ 0.5 LSB OF AN 11-BIT SAR ADC BASED ON BEHAVIORAL MODEL SIMULATION

Architecture	MCS Decision Only		Segmented Architecture, Coarse: 5-bit, fine:(5+7)-bit *			
RCR Type	bitwise	sample-wise	**AS, bitwise**	AS, subsample-wise	DAS-AS, bitwise	**DAS-AS, subsample-wise**
β	170	256	5	400	280	150

TABLE II

DYNAMIC AND STATIC PERFORMANCE COMPARISON OF AN 11-BIT 50 MS/S SAR ADCS WITH β = 16 BASED ON BEHAVIORAL MODEL SIMULATION

Architecture	MCS Decision Only		Segmented Architecture, Coarse: 5-bit, fine:(5+7)-bit *			
RCR Type	bitwise	sample-wise	**AS, bitwise**	AS, subsample-wise	DAS-AS, bitwise	**DAS-AS, subsample-wise**
SNDR (dB)	50.47	49.27	66.50	46.3	47.69	59.56
SFDR (dB)	53.56	49.54	84.8	46.43	47.92	63.75
ENOB (bits)	8.09	7.89	10.75	7.4	7.63	9.6
DNL (LSB)	-1/0.01	-1/7.59	-0.07/0.03	-0.17/1.07	-0.17/1.07	-0.17/1.07
INL (LSB)	-5.55/5.54	-6.46/6.45	-0.14/0.2	-11.59/11.66	-7.61/7.68	-4.31/4.38
$E^{\dagger}_{ave}$ (CVRP^2)	340.83	340.83	345.36	345.36	105.36	105.36

* The RCR type in segmented architecture only shows the fine MSBs. Both coarse and fine LSB successive decisions use bitwise RCRs.
† Calculated average switching energy. The energy consumed by C_{RCR} is almost the same as the DAC switching energy when the linearity is in the range of LSB.

5.5 Summary

This chapter presented the architecture and analysis of buffer-free segmented SAR-ADCs, which are synthesizable using the RAIL flow. Theoretical analysis and simulation have been performed to analyze the error due to the finite reference capacitance without using power-consuming not-synthesizable reference buffers. Both the analysis and simulation show that in a segmented SAR ADC, successive bit decisions in both coarse and fine SAR ADCs using *bitwise* RCRs and the MSB copy from the coarse ADC to the fine ADC using *AS with bitwise RCRs*, can reduce β significantly compared with other cases without any performance loss. While the fine MSB copy using detect-and-skip, AS with subsample-wise RCR can reduce reference error and switching energy compared with a successive decision only SAR ADCs with bitwise or sample-wise RCR that have the same β.

5.6 References

[1] W. Liu, P. Huang, and Y. Chiu, "A 12-bit, 45-MS/s, 3-mW redundant successive-approximation-register analog-to-digital converter with digital calibration," IEEE J. Solid-State Circuits, vol. 46, no. 11, pp. 2661–2672, Nov. 2011.

[2] H. Chang, H.-S. Lee, and D. Boning, "A 12 b 50 MS/s 2.1 mW SAR ADC with redundancy and digital background calibration," in Proc. ESSCIRC (ESSCIRC), Sep. 2013, pp. 109–112.

[3] C.-C. Liu, "A 0.35 mW 12 b 100 MS/s SAR-assisted digital slope ADC in 28 nm CMOS," in IEEE ISSCC Dig. Tech. Papers, Jan. 2016, pp. 462–463.

[4] W.-H. Tseng, W.-L. Lee, C.-H. Huang, and P.-C. Chiu, "A 12bit 104 MS/s SAR ADC in 28 nm CMOS for digitally-assisted wireless transmitters," IEEE J. Solid-State Circuits, vol. 51, no. 10, pp. 2222–2231, Oct. 2016.

[5] C. Lee and M. P. Flynn, "A SAR-assisted two-stage pipeline ADC," IEEE J. Solid-State Circuits, vol. 46, no. 4, pp. 859–869, Apr. 2011.

[6] F. van der Goes et al., "A 1.5 mW 68 dB SNDR 80 Ms/s 2 × interleaved pipelined SAR ADC in 28 nm CMOS," IEEE J. Solid-State Circuits, vol. 49, no. 12, pp. 2835–2845, Dec. 2014.

[7] J. A. Fredenburg and M. P. Flynn, "A 90-MS/s 11-MHz-bandwidth 62-dB SNDR noise-shaping SAR ADC," IEEE J. Solid-State Circuits, vol. 47, no. 12, pp. 2898–2904, Dec. 2012.

[8] L. Kull et al., "A 3.1 mW 8b 1.2 GS/s single-channel asynchronous SAR ADC with alternate comparators for enhanced speed in 32 nm digital SOI CMOS," IEEE J. Solid-State Circuits, vol. 48, no. 12, pp. 3049–3058, Dec. 2013.

[9] P. Hurrell, C. Lyden, D. Laing, D. Hummerston, and M. Vickery, "An 18 b 12.5 MS/s ADC with 93 dB SNR," IEEE J. Solid-State Circuits, vol. 45, no. 12, pp. 2647–2654, Dec. 2010.

[10] J. Craninckx and G. van der Plas, "A 65fJ/conversion-step 0-to-50 MS/s 0-to-0.7 mW 9b charge-sharing SAR ADC in 90 nm digital CMOS," in IEEE ISSCC Dig. Tech. Papers, Feb. 2007, pp. 246–600.

[11] B. Malki, T. Yamamoto, B. Verbruggen, P. Wambacq, and J. Craninckx, "A 70 dB DR 10 b 0-to-80 MS/s current-integrating SAR ADC with adaptive dynamic range," IEEE J. Solid-State Circuits, vol. 49, no. 5, pp. 1173–1182, May 2014.

[12] R. Kapusta, J. Shen, S. Decker, H. Li, E. Ibaragi, and H. Zhu, "A 14b 80 MS/s SAR ADC with 73.6 dB SNDR in 65 nm CMOS," IEEE J. Solid-State Circuits, vol. 48, no. 12, pp. 3059–3066, Dec. 2013.

[13] B. Chen, M. Maddox, M. C. W. Coln, Y. Lu, and L. D. Fernando, "Precision passive-charge-sharing SAR ADC: Analysis, design, and measurement results," IEEE J. Solid-State Circuits, vol. 53, no. 5, pp. 1481–1492, May 2018.

[14] M. Maddox, B. Chen, M. Coln, R. Kapusta, J. Shen, and L. Fernando, "A 16 bit linear passive-charge-sharing SAR ADC in 55 nm CMOS," in Proc. IEEE Asian Solid-State Circuits Conf. (A-SSCC), Nov. 2016, pp. 153–156.

[15] J. Shen et al., “A 16-bit 16MS/s SAR ADC with on-chip calibration in 55 nm CMOS,” in IEEE Symp. VLSI Circuits Dig. Tech. Papers, Jun. 2017, pp. C282–C283.
[16] C.-J.-R. Shi and A. Wang, “Analysis of bitwise and samplewise switched passive charge sharing SAR ADCs,” IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 27, no. 9, pp. 1977–1989, Sep. 2019.
[17] H.-Y. Tai, Y.-S. Hu, H.-W. Chen, and H.-S. Chen, “A 0.85fJ/conversionstep 10 b 200 kS/s subranging SAR ADC in 40 nm CMOS,” in IEEE ISSCC Dig. Tech. Papers, Feb. 2014, pp. 196–197.
[18] V. Hariprasath, J. Guerber, S.-H. Lee, and U.-K. Moon, “Merged capacitor switching based SAR ADC with highest switching energy efficiency,” Electron. Lett., vol. 46, no. 9, pp. 620–621, 2010.

6 DESIGN AND ANALYSIS OF A SYNTHESIZABLE SAR-BASED TEMPERATURE SENSOR

With the SAR implementation as the basis, this chapter presents a synthesizable resistor-based area-compact temperature sensor for on-chip thermal detection. This sensor has three unique features: 1) the use of a differential low-pass RC filter (DLPF) for thermal sensing, which reduces the area; 2) successive approximation register (SAR)-quantization embedded in the DLPF, which reuses the DLPF capacitor for capacitive digital-to-analog conversion (CDAC), eliminates the CDAC references, and utilizes the full sensing range for quantization; and 3) a highly digital circuit architecture, which can be easily implemented using a standard digital design flow and migrated to different processes. The temperature sensor was fabricated in a 65-nm CMOS technology occupying 8400-μm^2 silicon area. It achieves 0.38 ◦C resolution at room temperature. After a two-point calibration, the sensor achieves a 3σ inaccuracy of ±1.2 ◦C from −30 ◦C to 100 ◦C. It consumes 35.3-µW power from a 1.1-V supply. With a 2.5-µs conversion time, the sensor achieves an 88 pJ/conversion energy efficiency, which yields a 12.7-pJ·K^2 resolution figure-of-merit (FoM).

6.1 Introduction

Integrated temperature sensors are widely used in high-performance integrated circuits (ICs), especially CPUs and GPUs, to provide the thermal information, which can be used to control system clocks or used by circuit designers to develop advanced power and thermal management schemes [1]. The increase in the transistor number, the high core operating frequencies and switching activities, lead to high power density. Further, a single die usually integrates many different functional blocks, where some can be heated rapidly and may create hot spots. Without proper thermal detection, the accumulated heat may shorten the lifetime of a chip and cause unreliable operations or even damage the chip permanently [2]. By monitoring the die temperature with integrated temperature sensors, a microprocessor can reduce the clock frequency, the supply voltage, or both, to reduce the power consumption before the chip reaches the factory-set temperature limit [1]. Thus, integrated on-chip temperature sensors are crucial for high-performance ICs to maintain the proper thermal environment for reliable, long-term system operation.

The multi-core processor architecture prevails because high performance computation is required for many applications. In general, each core works for a particular computation task with its own clock frequency and supply voltage. This leads to a non-uniform thermal distribution. Further, the thermal gradient among multiple cores can change for different instructions and applications [4]. Thus, multiple temperature sensors are required to enable localized measurement. For examples, three [5], ten [3], and forty [7] temperature sensors can be used in a processor. Therefore, it is desirable for integrated temperature sensors to have low power consumption and a small area. To save the power, temperature sensors are in the sleep mode for most of the time. Once activated by an enable signal, they perform several temperature measurements, then return to the sleep mode. Thus, it is important for temperature sensors to have a short conversion time (≤1 ms) [3], [6], [7] and latency so that the thermal management unit (TMU) can perform readout and make decision in time. For general IC thermal monitoring applications, it is sufficient to have a sensor whose accuracy is about ±1◦C at the throttle and about ±3◦C over the rest of the range [3], [6], [7]. It has been pointed out that “it is not necessary to have ultra-high resolution (≤ 0.05◦C) temperature sensors in microprocessors, since the non-linearity associated with high-volume manufacturing will likely be much higher than resolution limits reported” [6] and [7].

Various types of integrated temperature sensors have been developed in CMOS technologies. Parasitic BJT-based bandgap sensors are the most established ones for industry level applications [3], [6], [8]. The linear temperature-to voltage relationship leads to high linearity and accuracy. This physical mechanism is considered

very reliable. However, BJT-based sensors need high supply voltages to bias a BJT in the forward-active region. Further BJT sensors are mostly analog in nature, require power-consuming opamps, which are difficult to scale.

CMOS transistor-based temperature sensors have been developed for compact area and low power consumption. These sensors use temperature-dependent carrier mobility and threshold voltage to generate temperature dependent current to control the frequency of oscillation. A sub-threshold oscillator-based temperature sensor with an exponential temperature-to-frequency conversion was introduced to relax the reference timing constraint, and realized a 0.6 nJ/conversion energy efficiency [9]. A very compact (area is 0.004 mm^2) ring oscillator-based temperature sensor was designed in 65nm CMOS [10]. A fully integrated CMOS temperature sensor based on temperature independent/dependent current sources was developed, which uses oscillators and counters to generate a digital temperature code [11]. However, the temperature dependency of the carrier mobility and threshold voltage is less stable among different processes, which prevents its usage in the volume-production [6], [7]. On the other hand, the leakage current increases at high temperature.

Very recently, resistor-based temperature sensors gained a lot of attention since a modern CMOS technology offers several types of resistors whose temperature coefficients are relatively constant. These temperature sensors can work at sub-1V supply voltages. A resistor-based temperature sensor was presented based on the measurement of an *RC* time constant [7]. A phase-domain ADC quantizes the temperature dependent phase shift generated by an RC Wien-bridge filter [12]. It is known that an *ADC* requires area nonscalable and power consuming opamp circuits. A highly digital resistor-based temperature sensor that uses a frequency locked loop (FLL) as the readout circuit achieved small area and high resolution [13]. The boosted output voltage from the poly-phase filter (PPF) puts extra constraints on the input stage of the following circuitry.

This chapter presents the design and implementation of a resistor-based highly digital temperature sensor [14]. A differential low pass filter (DLPF) based sensing element is proposed to provide a temperature dependent phase shift. The phase shift is quantized by a successive approximation register (SAR) ADC whose CDAC is embedded into the DLPF. The embedded CDAC eliminates the CDAC reference circuits and utilizes the full input range of the ADC. The proposed design was fabricated in a 65nm LP CMOS technology that occupies 8400 μm^2. It achieves 0.38 $^\circ$C resolution at room temperature. After a 2-point calibration with removing the second order systematic error, the sensor achieves a 3σ inaccuracy of ±1.2 $^\circ$C from −30 to 100 $^\circ$C. It consumes 35.3 μW power from a 1.1 V supply. With a 2.5 μs conversion time, the sensor achieves an 88 pJ/Conversion energy efficiency, which yields a 12.7 $pJ \cdot K^2$ resolution FoM.

This chapter is organized as follows. Section 5.2 presents the principle of on-chip temperature sensing using a differential low pass RC filter. The entire sensor architecture embedding SAR-based quantization is described in Section 5.3. Section 5.4 details the circuit implementation. Circuit nonidealities and their impacts on circuit performances are analyzed in Section 5.5. Section 5.6 reports the experimental results. Section 5.7 concludes the chapter.

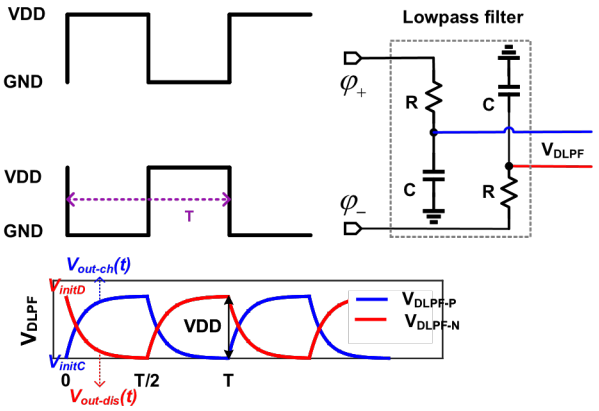


Fig. 1. The transient response of a differential RC LPF to a complementary clock with a 50% duty cycle.

6.2 On-chip Temperature Sensing with a Differential Low-Pass RC Filter

6.2.1 Response of a DLPF to a Complementary Clock

Figure 1 shows a differential RC low-pass filter (LPF) driven by a complementary periodic clock of period T with 50% duty cycle, and its output responses: differential output voltage V_{DLPF-P} and V_{DLPF-N}. One response charge while the other response discharges at the same time, and this behavior repeats periodically. Considering one clock period $t = 0$ *to* T, the charging response can be expressed as

$$V_{out-ch}(t) = V_{DD} - (V_{DD} - V_{initC})e^{\frac{-t}{RC}} \quad (1)$$

where V_{DD} is the high level of the input clock, while the low level is ground, and V_{initC} is the initial voltage at the beginning of charging. The discharging response can be expressed as

$$V_{out-dis}(t) = V_{initD}e^{\frac{-t}{RC}} \quad (2)$$

where V_{init} is the initial voltage at the beginning of discharging.

Solving (1) and (2) with the conditions that V_{initD} is the final voltage at the end of charging and V_{initC} is the final voltage at the end of discharging when $t = T/2$, we have

$$V_{initC} = V_{DD}(e^{\frac{-T}{2RC}} - e^{\frac{-T}{RC}})/(1 - e^{\frac{-T}{RC}}) \quad (3)$$

$$V_{initD} = V_{DD}(1 - e^{\frac{-T}{2RC}})/(1 - e^{\frac{-T}{RC}}) \quad (4)$$

When $T > 10 \cdot RC$, then $V_{initC} \approx 0$ and $V_{initD} \approx V_{DD}$.

The differential output crossing point (DOCP) is defined as the time when V_{DLPF-P} crosses V_{DLPF-N}; i.e., they both reach to $V_{DD}/2$. We can obtain

$$t_{DOCP} = RC \cdot ln(2/(1 + e^{\frac{-T}{2RC}})) \quad (5)$$

When $T/RC \rightarrow \infty$, $t_{DOCP} = RC \cdot ln2$.

TABLE I

CLOCK PERIOD *T* IN TERMS OF *RC* TIME CONSTANT FOR SELECTED DOCP REFERENCE PHASES

t_{DOCP} or Ref. phase	$T/8$	$T/16$	$T/32$
T	$8ln(1.8393)\cdot RC$	$16ln(1.992)\cdot RC$	$32ln(2)\cdot RC$

* Calculation based on 50% duty cycle clocks.

Clearly the t_{DOCP} of a DLPF is determined uniquely by the *RC* time constant and period *T*. Note that we can use digital circuits to generate $T/2^i, i = 1,2,3,...$ as *temperature independent reference phases*. Then the design principle of a DLPF-based temperature sensor is to select *T*, *R*, and *C* for a reference phase so that $t_{DOCP} = T/2^i, i = 1,2,3,...$ for a selected *i*. This is to select *T* to satisfy the following equality

$$T/2^i = RC \cdot ln(2/(1 + e^{\frac{-T}{2RC}})) \qquad (6)$$

Let $x = e$, then (6) is simplified to the following equation

$$x + \frac{1}{x^{2^{i-1}-1}} = 2 \qquad (7)$$

This equation has a valid solution for $i > 2$ as when $i = 1$ *or* 2, $x = 1$, which leads to $T = 0$. Table I shows the required clock period *T* in terms of the *RC* time constant so that the DOCP occurs at a selected reference phase $T/2^i$ for $i = 3,4,5$. Then the temperature-induced phase difference can be measured and quantized.

6.2.2 The Choice of Sensing Resistor and Capacitor in a CMOS Process

In a typical CMOS technology, different types of resistors have different temperature coefficients (TCs), sheet resistances, and supply sensitivities. In general, a resistor that has a large and relatively constant TC, a large sheet resistance, and a low supply sensitivity is preferred for temperature sensing. The relationship between resistance *R* and temperature *temp* can be expressed in (8)

$$R(temp) = R_o \cdot [1 + k_1(temp - 27) + k_2(temp - 27)^2] \qquad (8)$$

where R_o is the resistance at 27◦C, k_1 is the first order TC, and k_2 is the second order TC. To achieve fine resolution and good linearity, k_1 shall be as large as possible, while k_2 should be as small as possible. From Table II, we can see that the Nwell resistor under the oxide layer, the p+ diffusion resistor with salicide, and the p+ poly resistor with salicide have large first order TCs. However, the Nwell resistor has large nonlinear temperature and voltage dependency. The diffusion resistor has large parasitic capacitance and spreads more compared with the poly resistor [12]. Therefore, salicided p+ poly resistors are used in this design for its large first order TC (2360 ppm/◦C) and its small nonlinear temperature dependence (second order TC: −6.8 ppm/(◦C)^2). Note that the salicided p+ poly resistor has small sheet resistance, which can lead to a large area. However, the capacitor array occupies almost half of the area in this design. There is no area penalty to use the salicided p+ poly resistors since they are placed under the capacitor array.

Metal-oxide-metal (MOM) capacitor is used in this design. Compared with metal-insulator-metal (MIM) capacitor, it has following advantages. First, the MOM capacitor has slightly smaller first-order TC (−12 ppm/◦C) compared with the TC (−16 ppm/◦C) of the MIM capacitor. Second, the MIM capacitor requires additional fabrication masks into the process, which potentially increases the cost. Third, the capacitance density of the MOM capacitor (2.4 $fF/\mu m^2$) is slightly larger than that of the MIM capacitor (2 $fF/\mu m^2$) when more metal layers are used, there is no area penalty by using the MOM capacitor. Finally, the routing of the MOM capacitor is more flexible than the routing of the MIM capacitor that is being limited to M8 in 65nm CMOS.

The phase shift of the DOCP within a DLPF is mainly determined by the temperature dependent resistor when it is driven by a fixed cycle clock. The dynamic range of the phase shift (time difference between two extreme DOCPs) within a DLPF can be obtained by combining (5) and (6) and cancelling *C*,

$$t_{dyn} = \frac{T}{2^i} \cdot \frac{R(temp_H) - R(temp_L)}{R_o} \qquad (9)$$

Here i = 3,4,5,..., $temp_H$, and $temp_L$ are the high and low temperatures that the sensor can detect. The dynamic range is proportional to the chosen reference phase $T/2^i$ and the TC of the resistance based on (9). It has no relationship with the absolute value of the resistance *R* and the capacitance *C*.

TABLE II CHARACTERISTICS OF RESISTORS IN A 65NM CMOS TECHNOLOGY

	k_1 †	k_2 †	sheet res.	supply dep.
N+ diff w s*	2.16e-3	-5.77e-6	15.5	7.73e-5
N+ diff wo s*	1.62e-3	-1.36e-6	120	1.63e-4
N+ poly w s*	2.15e-3	-6.08e-6	15	5.61e-5
N+ poly wo s*	1.16e-4	5.35e-7	153	1.98e-4
Nwell under OD	2.45e-3	1.24e-5	327	3.09e-3
Nwell under STI	2.04e-3	1.57e-5	595	4e-3
P+ diff w s*	2.36e-3	-7.1e-6	14.55	8.15e-7
P+ diff wo s*	1.34e-3	-3.01e-7	245	9.22e-5
P+ poly w s*	2.36e-3	-6.8e-6	15	1.41e-4
P+ poly wo s*	-3.2e-4	1.45e-6	690	-5.11e-4

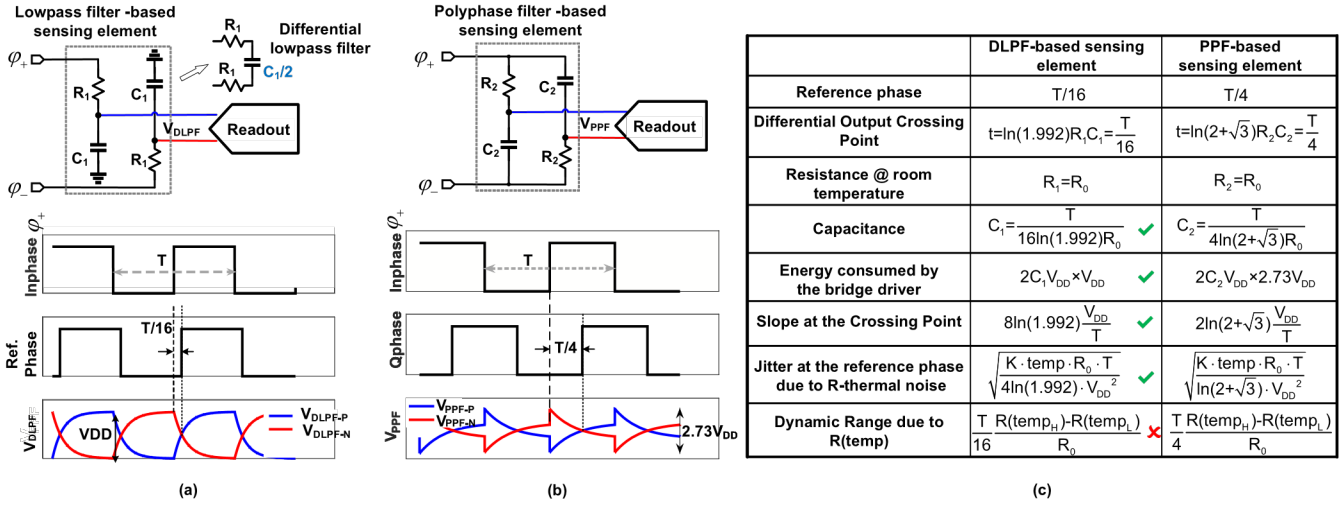
s* stands for salicide.
†: Simulation results based on 10 kΩ resistors with expression in (8)

6.2.3 Comparison of DLPF and PPF -Based Sensing Elements

Figure 2 (a) and (b) show the principle of the proposed DLPF-based and the previous PPF-based [13] temperature sensors. Both sensing elements consist of temperature-dependent salicided p-ploy resistors ($R_{1,2}$) and

temperature-insensitive MOM capacitors ($C_{1,2}$). Driven by the same cycle complementary clock ϕ_+ and ϕ_-, the differential output response of V_{DLPF} and V_{PPF} charges and discharges symmetrically. The phase shift of the differential output is quantized by a readout circuit.

The difference between these two topologies lies in two aspects [14]. First, the connection of the capacitor is different. The connection of the capacitor in the DLPF can maintain its differential output voltage V_{DLPF} in the range of V_{DD}. The two C_1 can be merged into one $C_1/2$ that saves the capacitance value by 4X. However, the connection of the capacitor in the PPF boosts its differential output voltage V_{PPF} to $2.73V_{DD}$. The boosted voltage improves the slope at the crossing point. However, it puts extra constraints on the input stage of the following circuitry [13]. Second, the reference phase in the DLPF is chosen $T/16$ delay of the input clock ϕ_+ to have a large slope of the output voltage at the DOCP without boosting its output voltage. A smaller reference phase increases the slope of the output voltage at the DOCP. However, it also decreases the dynamic range according to (9) and requires a higher order frequency divider to generate a smaller reference phase. Whereas one-quarter delay of ϕ_+ is chosen as the reference phase in the PPF which is referred to quadrature phase, Qphase [13].



	DLPF-based sensing element	PPF-based sensing element
Reference phase	T/16	T/4
Differential Output Crossing Point	$t=\ln(1.992)R_1C_1=\frac{T}{16}$	$t=\ln(2+\sqrt{3})R_2C_2=\frac{T}{4}$
Resistance @ room temperature	$R_1=R_0$	$R_2=R_0$
Capacitance	$C_1=\frac{T}{16\ln(1.992)R_0}$ ✔	$C_2=\frac{T}{4\ln(2+\sqrt{3})R_0}$
Energy consumed by the bridge driver	$2C_1V_{DD}\times V_{DD}$ ✔	$2C_2V_{DD}\times 2.73V_{DD}$
Slope at the Crossing Point	$8\ln(1.992)\frac{V_{DD}}{T}$ ✔	$2\ln(2+\sqrt{3})\frac{V_{DD}}{T}$
Jitter at the reference phase due to R-thermal noise	$\sqrt{\frac{K\cdot temp\cdot R_0\cdot T}{4\ln(1.992)\cdot V_{DD}^2}}$ ✔	$\sqrt{\frac{K\cdot temp\cdot R_0\cdot T}{\ln(2+\sqrt{3})\cdot V_{DD}^2}}$
Dynamic Range due to R(temp)	$\frac{T}{16}\frac{R(temp_H)-R(temp_L)}{R_0}$ ✘	$\frac{T}{4}\frac{R(temp_H)-R(temp_L)}{R_0}$

(c)

Fig. 2. Sensing element comparison: DLPF versus PPF with ideal readout circuits. (a) Proposed DLPF-based temperature sensor. (b) PPF-based temperature sensor. (c) Performance summary and comparison table.

In Fig. 2 (a), the charging and discharging responses can be obtained from (1) and (2) with the initial condition of $V_{initC}=0$ and $V_{initD}=V_{DD}$. They cross each other at the time

$$t_{DOCP}=ln(1.992)R_1C_1=\frac{T}{16} \qquad (10)$$

Similarly, in Fig. 2 (b), the charging response is

$$V_{PPF-ch}(t)=V_{DD}-\frac{1}{2}(2+\sqrt{3})V_{DD}e^{\frac{-t}{R_2C_2}} \qquad (11)$$

and the discharging response is

$$V_{PPF-dis}(t)=\frac{1}{2}(2+\sqrt{3})V_{DD}e^{\frac{-t}{R_2C_2}} \qquad (12)$$

They cross each other at the time

$$t_{DOCP} = ln(2+\sqrt{3})R_2C_2 = \frac{T}{4} \qquad (13)$$

If the resistors in the DLPF and the PPF have the same resistance, i.e., $R_1 = R_2 = R_0$ and driven by the same cycle clock, the required capacitance in the DLPF is 2.1X smaller than that in the PPF based on (10) and (13). The switching energy of the DLPF is 5.75X lower than that of the PPF since the PPF output voltage is boosted to 2.73X V_{DD} and the capacitance is 2.1X greater.

The slope of the output response is its first order derivative. At its reference phase, the slope of DLPF S_{DLPF} is given in (14) based on (1) and (10)

$$\begin{aligned} S_{DLPF} &= \frac{dV_{out-ch}(t)}{dt}|_{t=ln(1.992)R_1C_1} \\ &= 8ln(1.992)\frac{V_{DD}}{T} \end{aligned} \qquad (14)$$

while the slope of PPF S_{PPF} at its reference phase is given by (15) based on (11) and (13)

$$\begin{aligned} S_{PPF} &= \frac{dV_{PPF-ch}(t)}{dt}|_{t=ln(2+\sqrt{3})R_2C_2} \\ &= 2ln(2+\sqrt{3})\frac{V_{DD}}{T} \end{aligned} \qquad (15)$$

Comparing (14) with (15), the slope of the DLPF is 2.1X greater than that of the PPF. The thermal noise of the resistor affects the crossing point inversely proportional to the slope *S* through (16) [15]

$$\sigma_t \propto \frac{\Delta V_n}{S} \qquad (16)$$

Therefore, the jitter in the DLPF due to resistor thermal noise at its reference phase is

$$\sigma_{t_{DLPF}} \propto \frac{\sqrt{K \cdot \text{temp}/C_1}}{S_{DLPF}} = \sqrt{\frac{K \cdot \text{temp} \cdot R_0 \cdot T}{4ln(1.992) \cdot V_{DD}^2}} \qquad (17)$$

whereas, the jitter in the PPF at its reference phase is

$$\sigma_{t_{PPF}} \propto \frac{\sqrt{K \cdot \text{temp}/C_2}}{S_{PPF}} = \sqrt{\frac{K \cdot \text{temp} \cdot R_0 \cdot T}{ln(2+\sqrt{3}) \cdot V_{DD}^2}} \qquad (18)$$

where *K* is the Boltzman constant. From (17) and (18), we can see that with the same resistance, driven clock, and supply voltage, the DLPF has $\sqrt{2.}$ 1X smaller jitter at its reference phase compared with the PPF sensing element.

The dynamic range given in (9) is directly proportional to the reference phase. From this, the dynamic range of the DLPF is reduced by 4X compared with that of the PPF. The jitter in the DLPF is improved by $\sqrt{2.}$ 1X. Thus, the temperature resolution is reduced by 2.76X. The reduction in resolution is acceptable for the targeted IC thermal monitoring applications as 0.5 to 1 $^{\circ}C$ resolution is sufficient [3], [6], [7].

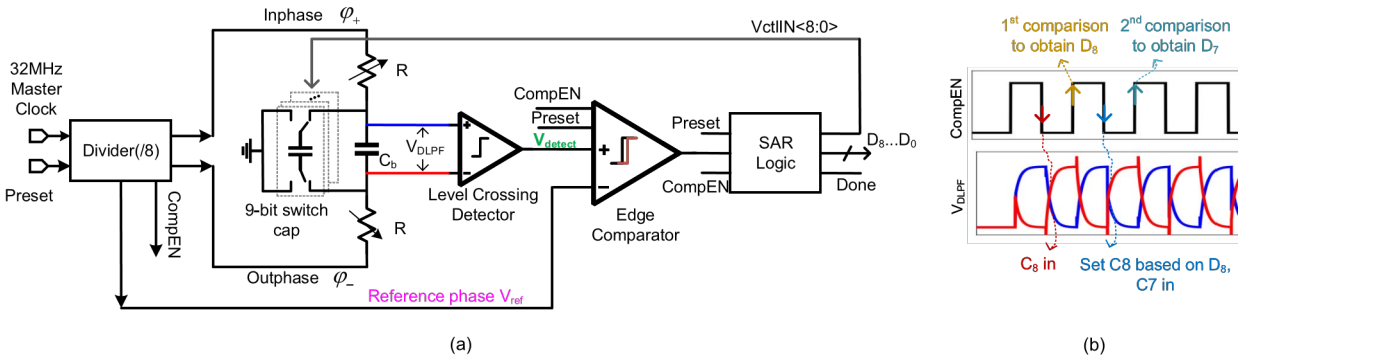


Fig. 3. Block diagram and operation principle of the DLPF-based SAR-quantization embedded sensing. (a) SAR-quantization embedded DLPF temperature sensing architecture. (b) SAR operation.

The two C_1 capacitors in the DLPF can be merged to one $C_1/2$ to form a differential LPF. With this, the total capacitance is reduced by 4X, and there is no need to match the two C_1 capacitors as in the PPF-based design. The total capacitance is reduced by (2.1×4) 8.4X compared with the capacitance in the PPF, this reducing the capacitance area by about 4X. The performance comparison is summarized in Fig. 2 (c). From the comparison table we can see that the dynamic range (resolution) is effectively used to trade off the area and power consumption.

6.3 SAR-Quantized Embedded LPF Temperature Sensing Architecture

To be used for on-chip SoC thermal management, temperature sensors require a resolution of 0.5 to 1 ◦C [3], [6], [7] among the range of −40 to 100 ◦C. This specification corresponds to an 8-to-9-bit quantization, which is the most suitable for the SAR ADCs. Straightforward SAR ADC implementation requires additional area-consuming capacitor arrays, voltage reference generators, and delay-to-voltage converters (such as mixers in [12]). This work eliminates all these additional blocks by re-using the capacitor in the DLPF. We refer to this technique as *SAR-quantization embedded DLPF temperature sensing*. The phase shift in the DLPF due to the temperature-dependent resistance is quantized by successively adjusting the binary weighted capacitance in the DLPF.

Figure 3 shows the block diagram and operation principle of the proposed SAR-quantization embedded DLPF-based temperature sensor. The *Preset* signal and a 32 MHz master clock are derived from an on-chip system clock. The 32 MHz master clock is divided by 8 to generate a 4 MHz complementary clock ϕ_+, ϕ_-, the reference phase clock, which is T/16 delay of ϕ_+, and a *CompEN* that is generated by ϕ_+ and the inverse of *Preset*. These four signals have little temperature dependency. The resistors in the DLPF can be tuned from 9 to 12.5 K with 8-bit non-uniform thermometer weights to compensate process variation of the DLPF and the level crossing detector. The capacitor $C_1/2$ in Fig. 2 (a) was implemented as a fixed base capacitor C_b and a 9-bit binary weighted capacitor array with the unit capacitance C_u of 1 fF. The dynamic range (phase shift) due to the temperature dependent resistance is about 5.2 ns from −40 to 120 ◦C, which can be estimated from (9) when T/16 reference phase is used.

The sensor is initiated by the *Preset* signal followed by the 32 MHz clock. The SAR logic performs the conventional binary searching algorithm; i.e., setting the most significant bit (MSB) capacitor C_8 to the DLPF bridge and the rest of the bits $C_7,...,C_0$ to GND for the first comparison shown in Fig. 3 (b). Then, the level crossing detector detects the differential output crossing point of V_{DLPF-P} and V_{DLPF-N} and generates V_{detect}. Then, the edge comparator compares V_{detect} with the temperature independent reference phase. If V_{detect} lags the reference phase, the edge comparator outputs digital "1"; i.e., $D_8 = 1$, and the MSB $C_8 = 256C_u$ keeps connecting to the DLPF bridge. Otherwise, $D_8 = 0$, and C_8 disconnects from the DLPF but to GND. Then, $C_7 = 128C_u$ connects to the DLPF bridge for the next bit comparison. This process continues till to the least significant bit (LSB) C_0. V_{detect} is close to the reference phase within one LSB resolution after setting the LSB bit.

It is noted that connecting or disconnecting capacitors to the DLPF bridge disturbs the charging and discharging processes with changing the capacitance ($C_b + D \cdot C_u$). To eliminate this disturbance and the error associated with the switch on/off activities such as charge injection and clock feedthrough, the bit capacitors to the DLPF are connected or disconnected at the falling edges of the *CompEN* signal, while comparing V_{detect} with the reference phase at the next rising edges, at which time the capacitance in the DLPF is unchanged and also there is no switching activity. The first two cycles are denoted in Fig. 3 (b). The outcome of embedding SAR-quantization in the sensing circuit is to quantize the RC phase shift due to the temperature dependent resistance through successively adjusting the binary weighted capacitance in the DLPF. The digital code *D* can be expressed as

$$ln(1.992) \cdot R \cdot (C_b + D \cdot C_u) = T/16 \tag{19}$$

with *R* given in (8).

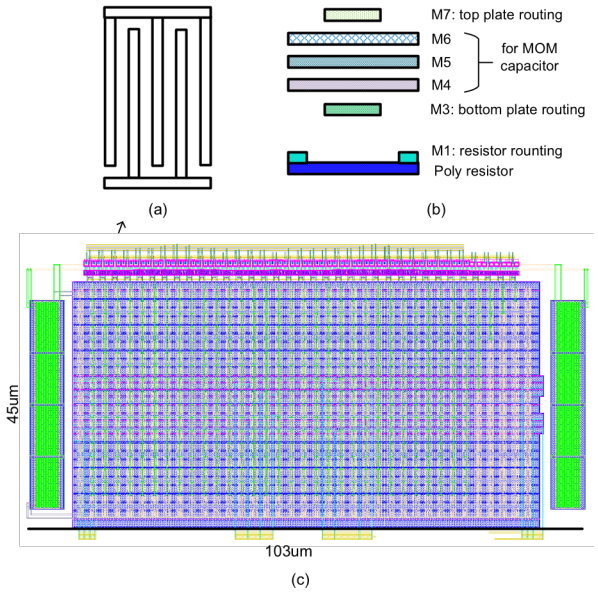


Fig. 4. The custom designed unit capacitor and the DLPF layout. (a) Unit capacitor: top view. (b) Unit capacitor: cross section view. (c) The DLPF layout.

6.4 Circuit Implementation

6.4.1 DLPF-Based Sensing Elements

In the DLPF, the base capacitor is a MOM capacitor from the process library provided by the foundry. The 9-bit capacitor array is a custom designed interdigitated MOM capacitor array shown in Fig. 4 (a) and (b). Metal layers M4, M5, and M6 are used to form unit capacitors utilizing coupling, while M3 is reserved for bottom plate routing and M7 for top plate routing. This minimizes the parasitic routing capacitance. The poly resistors in the DLPF are laid out under the capacitor array. Overlapping the MOM capacitor array and the poly resistors not only saves the area for the poly resistors but also makes it easy to meet the poly and metal density rules. The 9-bit capacitor array occupies 0.004 mm^2, which takes 47.4% of the entire core area. Thus, the width of the poly resistors is enlarged to 1.4 μm to make sure two tunable resistor ladders occupy the same area as the capacitor array. To this extent, the mismatch of the resistance is decreased by 3.44X compared with using the minimum width (0.4 μm) poly resistors. Both the resistor ladders and the capacitor array are surrounded by their dummies to decrease mismatch. The fixed base MOM capacitor is placed symmetrically at the two sides of the capacitor array as shown in Fig. 4 (c).

6.4.2 Level Crossing Detector

The level crossing detector is implemented as an inverter-based threshold comparator shown in Fig. 5 for lower power consumption and good scaling capability [18]. V_{DLPF-P} crosses with V_{DLPF-N} at $V_{DD}/2$. The comparison between V_{DLPF-P} and V_{DLPF-N} is equivalent to the comparison between V_{DLPF-P} and $V_{DD}/2$. In the inverter-based threshold comparator, the switching threshold voltage V_{ST} is internally decided by the sizes of M0 and M1 in Fig. 5, while the comparison voltage is V_{DLPF-P}. V_{ST} is defined as the voltage at which the inverter $V_{in} = V_{out} = V_{DD}/2$.

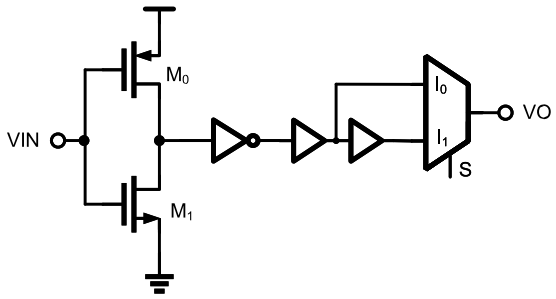


Fig. 5. An inverter based level crossing detector.

With high supply voltages, i.e., $V\,DD > V_{tn} + |V_{tp}|$, both M0 and M1 work in the saturation region with V_{ST} as the input. Then V_{ST} is obtained by equaling the currents through M0 and M1 as shown in (20)

$$V_{ST} = \frac{V_{DD} - |V_{tp}| + V_{tn}\sqrt{\beta_n/\beta_p}}{1 + \sqrt{\beta_n/\beta_p}} \qquad (20)$$

where $\beta_n = \mu_n C_{ox}(W/L)_n$, and $\beta_p = \mu_p C_{ox}(W/L)_p$. $V_{ST} \approx V\,DD/2$ through optimizing the sizes of M0 and M1 in the inverter [19], [20].

In Fig 5, two buffers following the inverter are added to shape the waveform and a 2:1 MUX is used to tune the delay at different supplies. The reference phase also passes through the same level crossing detector to cancel the supply and temperature dependency in the first order. Simulation results show that the delay difference between the reference phase path and the signal path is 16 ps when V_{DD} varies from 1.0 to 1.2 V at room temperature, which corresponds to 0.48 ◦C temperature error. The delay difference changes at a slope of 1.2 to 2.5 ps/◦C from −40 to 120 ◦C at a 1.1 V supply, which corresponds to 0.1 ◦C temperature error. The simulated input referred noise is 60 μV_{rms}, which translates into to 0.16 ◦C temperature error. From the Monte Carlo simulations, 1-σ variation of the offset delay is 172 ps, which is translated into a temperature inaccuracy of ±1.2 ◦C (3σ) after 2-point digital calibration.

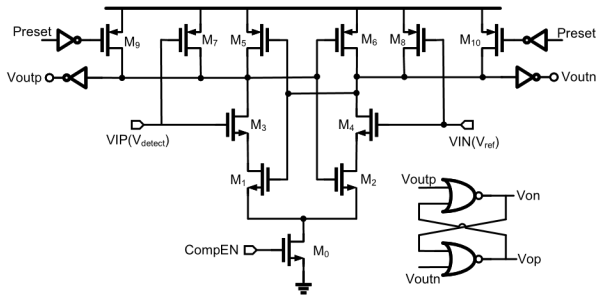


Fig. 6. A dynamic latch based edge comparator.

.

6.4.3 Edge Comparator

The inputs to the edge comparator are two square wave signals: V_{detect} and reference phase V_{ref}. The edge comparator detects whether V_{detect} leads or lags V_{ref}. A dynamic latch-based comparator is adopted shown in Fig. 6. The comparator is enabled through *CompEN*, which is before the rising edge of its two input signals. NMOS $M_3(M_4)$ and PMOS $M_7(M_8)$ are both controlled by the input $V_{detect}(V_{ref})$. When both inputs are low, the outputs are set to low. There is no current flow through the circuit, as M_3 and M_4 are off.

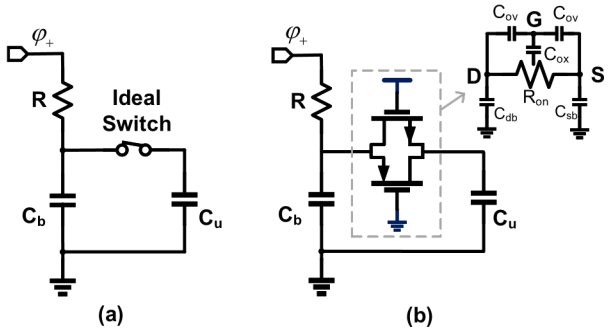


Fig. 7. (a) A bit capacitor with an ideal switch. (b) A bit capacitor with a transmission-gate based switch and its equivalent circuit illustrates parasitics.

If V_{detect} leads V_{ref}, then V_{outp} is set to low, Otherwise, V_{outp} is set to high. An SR latch is used to keep the outputs of the edge comparator for the whole cycle. The size of the transistors is optimized for low offset and small noise [21], [22]. The average current of the edge comparator is 90 nA from a 1.1 V supply at a 4 MHz clock. The 3σ offset and the input referred noise converted to time are 1.8 ps and 0.1 ps, respectively. These are translated into 0.048 and 2.7m ◦C temperature errors.

6.5 Circuit Nonlinearities

6.5.1 Switch Nonideality

The CDAC is embedded into the DLPF. The parasitic capacitance due to the switch, the temperature dependent switch on-resistance, and the non-idealities associated with the switching activity affect the charging and discharging process directly. Thus, these non-idealities shall be analyzed.

Connecting or disconnecting bit capacitors to the DLPF bridge is controlled by the SAR logic. A transmission gate (TG) based switch is used for connecting the bit capacitor to the DLPF, while an NMOS switch is used for connecting it to ground. To analyze the effects of the nonidealities mentioned above, the single-ended circuit is used shown in Fig. 7. With an ideal switch, the bit capacitor is in parallel with the base capacitor shown in Fig. 7 (a). So, the phase shift due to the temperature dependent resistance can be quantized by adjusting the capacitance. As illustrated in Fig. 7(b), the equivalent circuit of a TG-based switch can be modeled as an on-resistance R_{on}, gate-to-drain/source overlap capacitance C_{ov}, gate-to-channel capacitance C_{ox}, and junction capacitance C_{db} and C_{sb} between the drain/source and the body. Since the TG switch is in the triode region, so $C_{gd} = C_{gs} = WC_{ov} + \frac{1}{2}WLC_{ox}$ and yields $C_{gs}/2$ when they are in series. From Fig. 7 (b), the parasitic capacitors affect C_b and C_u directly through $C_b = C_b + 2\times C_{gs}/2 + C_{db}$ and $C_u = C_u + 2 \times C_{gs}/2 + C_{sb}$. The temperature dependent R_{on} is in series with C_b and C_u. The R_{on} is about 2.6 K with the first order TC 1390 ppm/◦C when the gate-source voltage is at half V_{DD}. However, the

temperature-dependent R_{on} does not affect the RC response as it is in series with a fF capacitor whose impedance is thousands of times larger than the switch on-resistance at a 4 MHz clock. However, the parasitic capacitance contributes to the delay directly. Figure 8 shows the time differences of the charging response at the half V_{DD} with a TG switch and an ideal switch. There is an offset at −40 ◦C, which is due to the enlarged C_b and C_u . The positive slope is mainly due to temperature-dependent junction capacitance C_{db} and C_{sb}. Thus, the size of the TG shall be as small as possible to minimize the parasitic capacitance. The temperature-dependent parasitic junction capacitance incurs only 0.25 LSB error in the entire range from −40 to 120 ◦C, which corresponds to 0.078 ◦C temperature error.

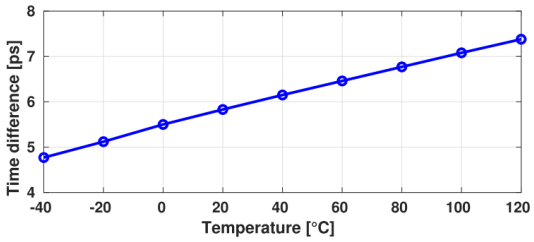


Fig. 8. The time difference of the charging response at half *VDD* of Fig. 7 (b) and (a).

Charge injection and clock feedthrough disturb the charging and discharging response. The incurred error is avoided by selecting different clock edges for switching and comparison as shown in Fig. 3 (b).

6.5.2 Supply Sensitivity

For integrated applications, the supply voltage of the temperature sensor is usually from the voltage regulators. The supply voltage has some variations because of the loading effect. The temperature sensor shall work well in the range of the regulated supply variations. In the proposed temperature sensor, the clock divider and the SAR logic are purely digital circuits. Their performance is robust to the supply variation at a 4 MHz clock. Therefore, supply sensitivity analysis mainly focuses on the DLPF, the level crossing detector, and the edge comparator.

The charging and discharging responses of the DLPF cross each other when they are equal. Based on (1) and (2), we have $t_{DOCP} = ln(1.992)R_1C_1$, which is independent of VDD. Thus, as long as the two resistors in the DLPF match, the DLPF is supply independent. The level crossing detector is an inverter-based threshold comparator. Even the supply dependency of the level crossing detector is partly compensated by passing both the signal path and the reference phase path with the same level crossing detector and using a 2:1 MUX. The lower supply voltage is limited to 1.0 V. When the supply is less than 1.0 V, it is close to or less than the sum of the threshold voltage of M0 and M1 in Fig. 5. It is more challenging to balance the current through these two transistors in the non-saturation region. Thus, the delay of the level crossing detector varies much when the supply is less than 1.0 V.

The edge comparator shown in Fig. 6 contains positive feedback. It has a large voltage gain to obtain the comparison result in a short time with the inputs that have little difference. Therefore, the edge comparator is more sensitive to the input difference than the supply variations. Transient noise simulation shows that the edge comparator can make the correct decision when V_{detect} leads/lags V_{ref} by 1 ps when the supply voltage is from 0.5 to 1.2 V. This indicates that the edge comparator has robust supply immunity.

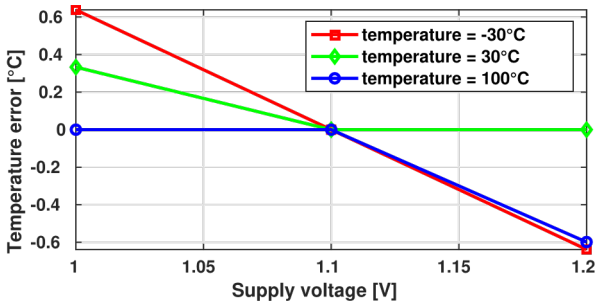


Fig. 9. Simulated digits versus supply voltage at different temperatures.

The simulated temperature sensor error versus supply voltage at three different temperatures is shown in Fig. 9. The supply sensitivity at the nominal supply (1.1 V) is 1.56 ◦C/V when the temperature is 30 ◦C. It increases to 6.3 ◦C/V at −30 ◦C.

6.6 Measurement Results

The temperature sensor prototype was fabricated in a 65nm LP CMOS technology. The die photograph is shown in Fig. 10 with a 0.0084 mm^2 active core area. Replicas of resistors and capacitors are placed beside the core to test the process of the fabrication. The chip has an SPI interface, which is used to connect to a micro-controller/FPGA board (PYNQ-Z1) for testing. The clock is provided by the PLL on the microcontroller, acting like the SoC master clock with noisy jitters in practice. The measurement was done by placing the test board in a temperature-controlled Tenney environmental chamber, and the PYNQ-Z1 board driven by a computer connecting to the test board is placed outside of the chamber. No extra measurement equipment is required.

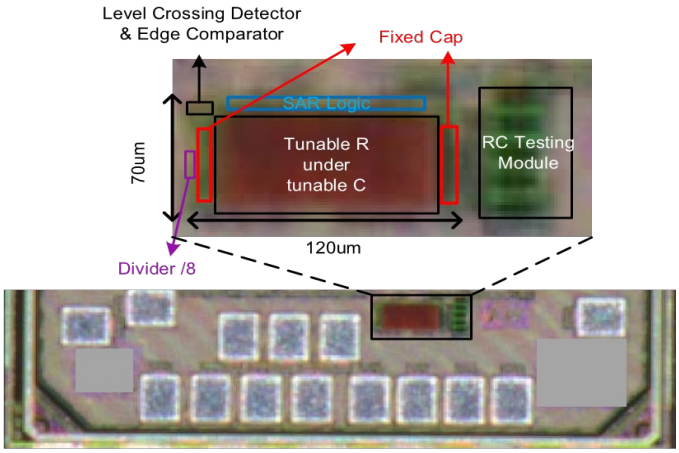


Fig. 10. Die micrograph.

When running at the conversion time of 2.5 μs (one SAR cycle is 250 ns, nine SAR cycles and one done cycle count for one conversion), the entire temperature sensor consumes 35.3 μW under a 1.1 V supply. This yields an energy efficiency of 88 pJ/Conversion. The power breakdown is shown in Fig. 11. The DLPF driver consumes 31 μW, which is 88% of the total power consumption.

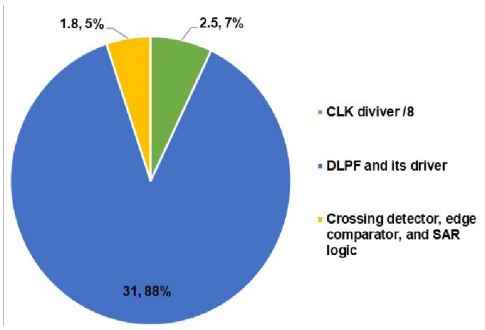


Fig. 11. The power consumption breakdown under a 1.1V supply Voltage.

Twelve samples in QFN40 packages from one MPW shuttle are characterized from −30◦C to 100 ◦C. Figure 12 (a) shows the measured raw digital output versus temperature. The nonlinear digits versus temperature relationship are introduced by the design formula shown in (19). Figure 12 (b) upper plot shows the temperature error after a two-point calibration over −30 to 100 ◦C. It clearly shows the second-order nonlinearity leads to the 3σ temperature error as much as 3.5 ◦C. The second-order nonlinearity is mainly from the second order temperature coefficient of the resistor shown in (8). This nonlinearity is very similar among sensors, therefore can be nulled. After removing the second order systematic nonlinearity, a 3σ inaccuracy of ±1.2 ◦C is achieved using two-point calibration shown in the Fig. 12 (b) lower plot. A two-point calibration is used as the D_{out} to the temperature relationship has at least three variables.

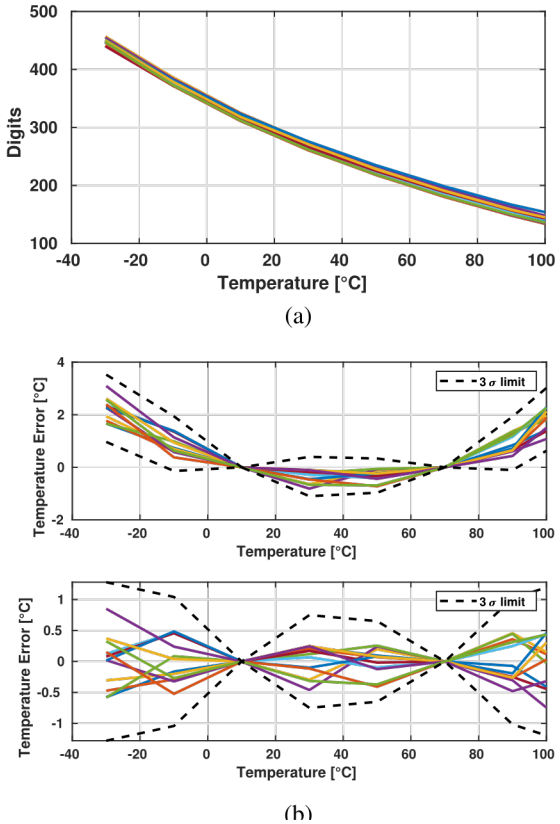


Fig. 12. (a) Measured sensor output digits versus temperature. (b) The temperature error after a 2-point calibration without and with removing the 2nd systematic error.

The effective resolution of the sensor at is obtained by calculating the root mean square (rms) value of 500 consecutive readouts at each temperature. Figure 13 (a) shows the temperature error of 500 consecutive readouts measured at 25◦C. It shows the rms resolution of 0.38 ◦C. The temperature resolution is characterized at the entire temperature range from -30 to 100 ◦C using rms noise calculation. The result is shown in Fig. 13 (b). The resolution decreases with the temperature as more noise presents at higher temperatures. The supply sensitivity is characterized from 1.0 to 1.2 V at three different temperatures shown in Fig. 14. The temperature sensor has 2.8 ◦C/V supply sensitivity at 30 ◦C. While it increases to 7 ◦C/V at −30 ◦C. The measured results match the simulation results.

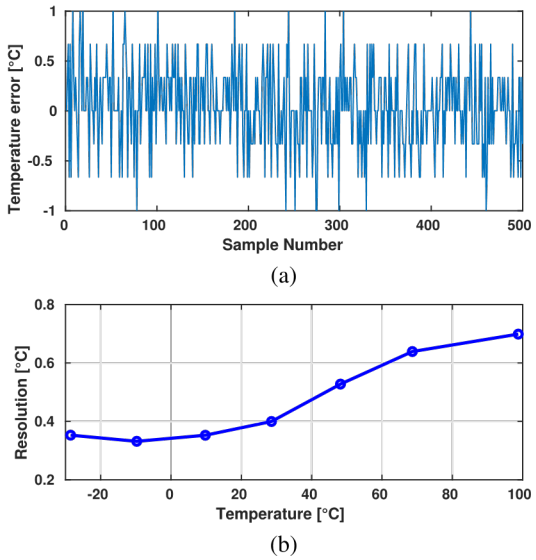


Fig. 13. (a) Measured temperatures error at 25◦C. (b) Measured rms resolution versus temperature.

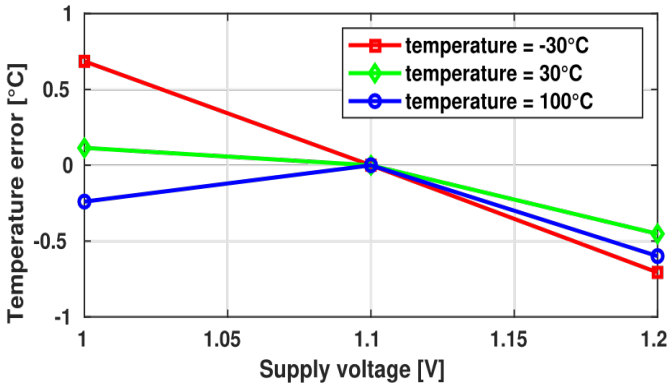


Fig. 14. Measured temperature error versus supply at different temperatures.

Table III summarizes the measured sensor performance and compares it with the state-of-the-arts. This work achieves a FoM of 12.7 pJ·K^2, which is 40X better than that of the BJT-based sensor with the similar resolution [8], The area is 5X smaller than [17], 13X smaller than [24] that use the SAR ADCs as the readout circuits. The sensor has 4X-400X faster conversion time than the state-of-the-arts. Figure 15 shows the resolution FoM benchmark of the proposed work. The energy/conversion of 0.088 nJ is achieved with smaller area and fast conversion time [25].

TABLE III. PERFORMANCE SUMMARY AND COMPARISON WITH THE STATE OF THE ARTS

	This work	JSSC 2018 [7]	JSSC 2018[13]	ISSCC 2018 [8]	JSSC 2018 [17]	SSCL 2018 [24]
Sensor Type	**Resistor**	Resistor	Resistor	PNP/MOS	Resistor	Resistor
Readout Type	**SAR**	TDC	FLL	SAR	Flash-SAR	SAR
Technology	**65nm**	65nm	65nm	22nm	65nm	65nm
Supply [V]	**1.0∼1.2**	1.1∼1.4	0.85∼1.05	0.97∼1.3	0.6∼1.2	0.65, 1
Supply sensitivity [°C/V]	2.8†	3	0.5	1.75	0.28	N/A
Power [μW]	**35.2**	12.8	68 *	50	47.2	0.488
Conversion Time [μs]	**2.5**	80	1000	32	10	10
Energy/Conversion [nJ/Conv.]	**0.088**	0.9	68	1.6	0.472	0.00488
Area [mm^2]	**0.0084**	0.01	0.007*	0.0043	0.044	0.11
Temperature Range [°C]	**-30∼100**	-40∼110	-40∼85	-30∼120	-45∼85	0∼100
Resolution [°C]	**0.38**	0.15	0.0025	0.58	0.121	0.61
Calibration Type	**2-point**	2-point	2-point	1-point	2-point	1-point
3σ Inaccuracy [°C]	**±1.2**	± 1.4	±0.12	±1.07	+1.6/-1	-1.1/1.5
FoM [pJ·K^2]	**12.7**	20	0.43*	540	6.9∼13.8	1.82

*: Does not include an on-chip frequency to digital converter.
†: Supply sensitivity at 30 °C

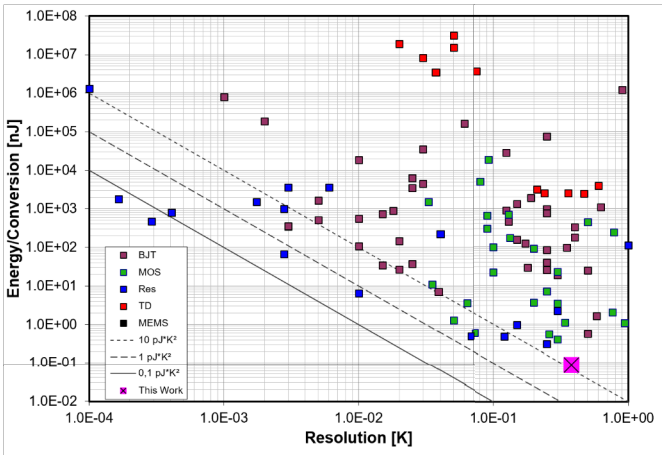


Fig. 15. Resolution FoM benchmark of this work compared with the published temperature sensors [25].

6.7 Summary

A synthesizable resistor-based, area compact (0.0084 mm^2 implemented in 65nm CMOS) temperature sensor was presented. It uses a salicided p-poly resistor to sense the temperature, a differential low-pass filter (DLPF) to generate a temperature-dependent phase, which is then quantized by adjusting the MOM capacitor in the DLPF via successive approximation registration (SAR). This SAR-quantization embedded DLPF sensing architecture utilizes the full temperature-sensing range for digitization, reuses SAR capacitors, eliminates CDAC reference circuits, is much more area efficient than previous PPF and WB followed by SAR quantization architectures. Measurement results show that it achieves a 12.7 $pJ{\cdot}K^2$ resolution FoM. With a 2.5 μs conversion time, it achieves an 88 pJ/Conversion energy efficiency, which is the fastest sample rate with good energy efficiency comparing to the state-of-the arts. Furthermore, the highly digital architecture enables easy process migration and the use of a standard digital design flow.

6.8 References

[1] M. Berktold and T. Tian, "CPU monitoring with DTS/PECI," Intel, Santa Clara, CA, USA, Tech. Rep. 322683, Sep. 2010. [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/whitepapers/cpu-monitoring-dts-peci-paper.pdf

[2] *Thermal Management for Intel Xeon Processors*, Intel, Santa Clara, CA, USA, Aug. 2018, Art. no. 000006710. [Online]. Available:

https://www.intel.com/content/www/us/en/support/articles/000006710/ processors.html

[3] J. S. Shor and K. Luria, "Miniaturized BJT-based thermal sensor for microprocessors in 32- and 22-nm technologies," *IEEE J. Solid-State Circuits*, vol. 48, no. 11, pp. 2860–2867, Nov. 2013.

[4] K. Luria and J. Shor, "Miniaturized CMOS thermal sensor array for temperature gradient measurement in microprocessors," in *Proc. IEEE Int. Symp. Circuits Syst.*, May/Jun. 2010, pp. 1855–1858.

[5] D. E. Duarte, G. Geannopoulos, U. Mughal, K. L. Wong, and G. Taylor, "Temperature sensor design in a high volume manufacturing 65nm CMOS digital process," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Sep. 2007, pp. 221–224.

[6] T. Oshita, J. Shor, D. E. Duarte, A. Kornfeld, and D. Zilberman, "Compact BJT-based thermal sensor for processor applications in a 14 nm tri-gate CMOS process," *IEEE J. Solid-State Circuits*, vol. 50, no. 3, pp. 799–807, Mar. 2015.

[7] A. Mordakhay and J. Shor, "Miniatured, 0.01 mm^2, resistor-based thermal sensor with an energy consumption of 0.9 nJ and a conversion time of 80 μs for processor applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 10, pp. 2958–2969, Oct. 2018.

[8] C.-Y. Lu, S. Ravikumar, A. D. Sali, M. Eberlein, and H.-J. Lee, "An 8b subthreshold hybrid thermal sensor with ±1.07◦C inaccuracy and single element remote-sensing technique in 22nm FinFET," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2018, pp. 318–320.

[9] K. Yang *et al.*, "A 0.6nJ -0.22/+0.19◦C inaccuracy temperature sensor using exponential subthreshold oscillation dependence," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2017, pp. 160–161.

[10] T. Anand, K. A. A. Makinwa, and P. K. Hanumolu, "A VCO based highly digital temperature sensor with 0.034 ◦C/mV supply sensitivity," *IEEE J. Solid-State Circuits*, vol. 51, no. 11, pp. 2651–2663, Nov. 2016.

[11] S. Jeong, Z. Foo, Y. Lee, J.-Y. Sim, D. Blaauw, and D. Sylvester, "A fully-integrated 71 nW CMOS temperature sensor for low power wireless sensor nodes," *IEEE J. Solid-State Circuits*, vol. 49, no. 8, pp. 1682–1693, Aug. 2014.

[12] S. Pan, Y. Luo, S. H. Shalmany, and K. A. A. Makinwa, "A resistor-based temperature sensor with a 0.13 pJ·K^2 resolution FoM," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 164–173, Jan. 2018.

[13] W. Choi *et al.*, "A compact resistor-based CMOS temperature sensor with an inaccuracy of 0.12 ◦C (3σ) and a resolution FoM of 0.43 pJ·K^2 in 65-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 53, no. 12, pp. 3356–3367, Dec. 2018.

[14] A. Wang, C. Chen, and C.-J. R. Shi, "A 9-bit resistor-based all-digital temperature sensor with a SAR-quantization embedded differential lowpass filter in 65nm CMOS consuming 57pJ with a 2.5us conversion time," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2019, pp. 1–4.

[15] A. A. Abidi and R. G. Meyer, "Noise in relaxation oscillators," *IEEE J. Solid-State Circuits*, vol. 18, no. 6, pp. 794–802, Dec. 1983.

[16] J. Lee, A. George, and M. He, "A 1.4V 10.5MHz swing-boosted differential relaxation oscillator with 162.1dBc/Hz FOM and 9.86 PSRMS period jitter in 0.18μm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Jan./Feb. 2016, pp. 106–108.

[17] H. Park and J. Kim, "A 0.8-V resistor-based temperature sensor in 65-nm CMOS with supply sensitivity of 0.28◦C/V," *IEEE J. Solid-State Circuits*, vol. 53, no. 3, pp. 906–912, Mar. 2018.

[18] A. Tangel and K. Choi, "'The CMOS inverter' as a comparator in ADC designs," *Analog Integr. Circuits Signal Process.*, vol. 39, no. 2, 147-155, May 2004.

[19] P. Mroszczyk and V. F. Pavlidis, "Mismatch compensation technique for inverter-based CMOS circuits," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[20] J. Segura, J. L. Rossello, J. Morra, and H. Sigg, "A variable threshold voltage inverter for CMOS programmable logic circuits," *IEEE J. SolidState Circuits*, vol. 33, no. 8, pp. 1262–1265, Aug. 1998.

[21] J. He, S. Zhan, D. Chen, and R. L. Geiger, "Analyses of static and dynamic random offset voltages in dynamic comparators," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 5, pp. 911–919, May 2009.

[22] P. Nuzzo, F. De Bernardinis, P. Terreni, and G. Van der Plas, "Noise analysis of regenerative comparators for reconfigurable ADC architectures," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 6, pp. 1441– 1454, Jul. 2008.

[23] M.-T. Tan, J. S. Chang, and Y.-C. Tong, "A process-independent threshold voltage inverter-comparator for pulse width modulation applications," in *Proc. IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, vol. 3, Sep. 1999, pp. 1201–1204.

[24] H. Xin, M. Andraud, P. Baltus, E. Cantatore, and P. Harpe, "A 174 pW–488.3 nW 1 S/s–100 kS/s all-dynamic resistive temperature sensor with speed/resolution/resistance adaptability," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 3, pp. 70–73, Mar. 2018.

[25] K. A. A. Makinwa. (Sep. 2018). *Temperature Sensor Performance Survey*. [Online]. Available: http://ei.ewi.tudelft.nl/docs/TSensor_survey.xls

7 DESIGN AND ANALYSIS OF A FULLY SYNTHESIZABLE INJECTION-LOCKED MULTIPLYING DELAY-LOCKED LOOP

This chapter presents a low-reference spur ring oscillator (RO) based injection-locked clock multiplier (ILCM). Unlike conventional calibrators that deal with error sources separately, an injection pulse shaping (IPS) technology is proposed to accurately characterize the output pulse and eliminate phase error caused by non-ideal injection. An injection shaping cell (ISC) is proposed to be used as the injection cell in the RO, allowing to reform the output by the injection pulse. The phase error is extracted by the proposed sampling error detector (SED) that compares the captured edges from the high-frequency output. The SED enables two phases every injection cycle to identify the phase errors from frequency mismatch, pulse distortion and delay offset. Based on this, a digital calibrator is proposed to track the injection caused phase error. To save power and area, digitally controlled delay lines with switched-capacitor integrator are used to compensate errors in time domain. The prototype of the IPS-ILCM is fabricated in a 65 nm CMOS technology and the active area is 0.021 mm^2. The measured reference spur is -79 dBc at 2.5 GHz output with 50 MHz reference. The RMS jitter is 819 fs and the power consumption is 1.23 mW, achieving the figure of merit (FoM) of -240.8 dB.

7.1 Introduction

The increasing demand for highly integrated systems has motivated the development of frequency synthesizers with low jitter, compact area, wide tuning range, and the immunity of electromagnetic coupling. Phase-locked loops (PLLs) using an LC-based voltage-controlled oscillator (VCO) are extensively used for better noise/power compromise. However, a few passive devices in the LC VCO dominate the silicon area, which is not scaling-friendly for the advanced CMOS technology.

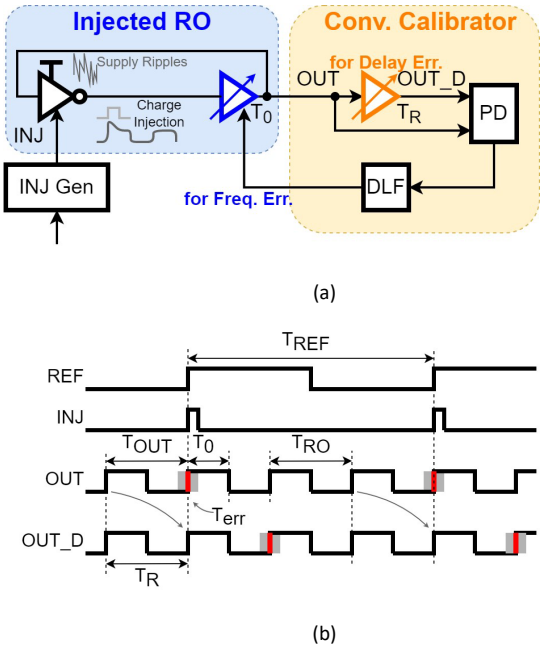


Figure 1. ILCM with a conventional calibrator: (a) block diagram and (b) timing diagram.

As an attractive alternative, an injection locked clock multiplier (ILCM) employing a ring oscillator (RO) can provide a low-jitter clock with a limited budget in terms of area and power consumption [1]–[8]. The injection generated from a clean reference clock at the frequency F_{ref} can reset the noisy edges every N cycle, and aggressively filter out the RO noise with a much wider bandwidth compared to that of a conventional type-II PLL [9]–[11].

However, the phase realignment makes ILCMs susceptible to imperfect injection that causes instantaneous disturbance. This periodic timing variation causes the deterministic jitter in the time domain, which appears as the reference spur in the spectrum. The timing certainty and the frequency purity are degraded, resulting in the risk of timing and spectral violations. Prior works have proposed several calibrations to eliminate the injection-caused phase error [5]–[8], [12]. They share a similar principle of first detecting the non-ideal sources and then enabling specific calibrators to correct them. One main cause discussed comprehensively is the frequency drift [1]– [7], [12], [13]. When the free-frequency $F_{RO} = 1/T_{RO}$ of the RO deviates away from the target frequency $N \cdot F_{REF}$, the injection clear the accumulated timing error T_{err} between the output and the injection from the frequency error F_{err}. The resulting reference spur can be estimated as [2]

$$\text{ref spur} \approx 20\log(|T_{err}|/T_{RO}) = 20\log(|F_{err}|/F_{REF} \qquad (1)$$

which suggests that even a tiny frequency drift in the RO, e.g. $F_{err} = 500$ KHz (0.02% away from the target), can lead to a spur of -40 dBc at the reference of 50 MHz. To address this problem, a frequency-tracking loop (FTL) is usually employed to track the target frequency against the process, voltage, and temperature (PVT) variations. Unfortunately, since the RO frequency is adjusted based on the injection-caused error extracted by a phase detector (PD), the delay offset T_{off} between the injection and the error detection can generate to severe reference spur due to the incorrect frequency tuning, which is required to be counteracted by additional calibrators [6], [7], [12], [13]. Moreover, the slope modulation has been considered for the reference spur [8].

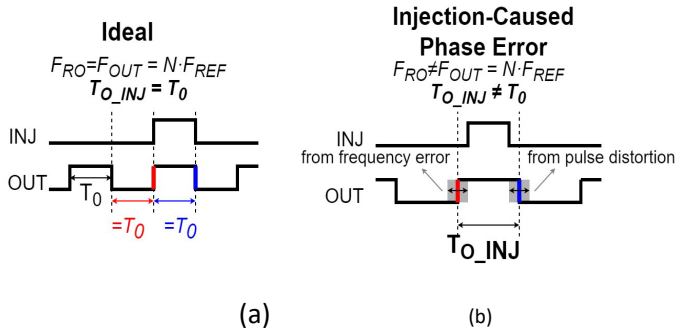


Figure 2. (a) Ideal output. (b) Phase error caused by injection.

To achieve decent spur performance, calibrations are supposed to be sufficiently fine and offset-free to capture the exact errors and eliminate them completely, where the high resolution is usually realized with a larger area and higher power for a bulk of unit devices. Nevertheless, besides the errors mentioned above, there are lots of other non-idealities that can affect the injection accuracy, such as supply ripples, charge injection of the switches, and clock feedthrough. They not only disturb the edge reset by the injection but also affect the following edges. In practice, it is not a trial to employ multiple calibration loops to simultaneously mitigate each of them. The residual phase error always limits the

capability of conventional calibrations. On the other hand, conventional phase-error-detection schemes [5], [7], [8], [13] suffer from the timing race between the calibration loop and the injection, hence elaborate timing control is required to extract errors.

In this work, we present an RO-based ILCM that can achieve a low reference spur using an injection pulse shaping (IPS) technique to suppress all kinds of phase errors in the distorted RO output [14]. The ideal output can be recovered by the injection shaping cell in the RO. In addition to the frequency and delay calibration, the pulse distortion is tackled as the injected output is accurately shaped. Thanks to the multi-source calibration implemented by the circuits mostly operating in discrete time, the reference spur is minimized with low power and compact area.

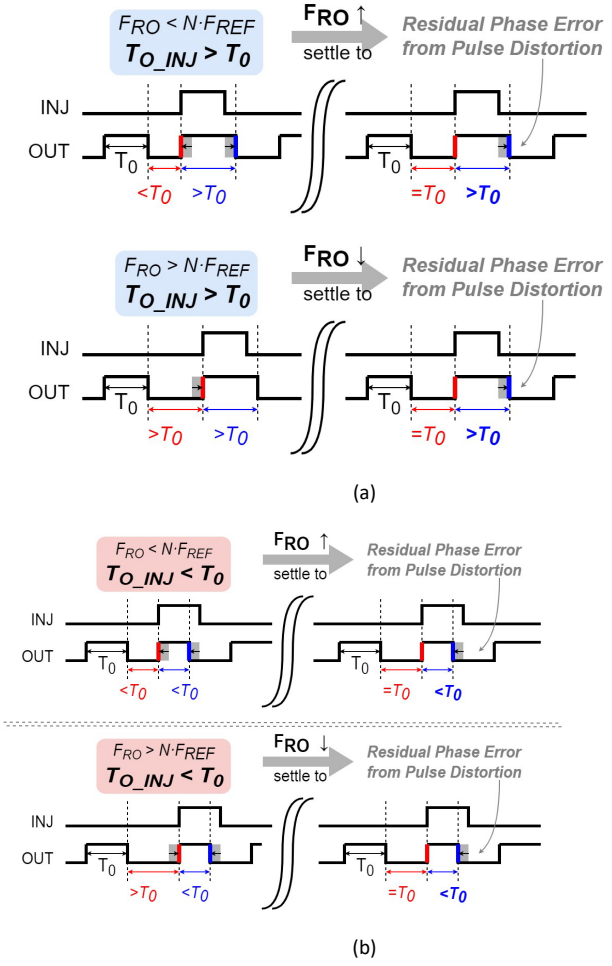


Fig. 3.ILCM with conventional calibrator settles with frequency error and (a) $T_O INJ > T0$ and (b) $T_O INJ > T0$.

The rest of this chapter is organized as follows. In Section 6.2 , the injection-caused phase error and the limits of conventional calibrators are discussed. Section 6.3 presents the principle of the proposed injection pulse shaping technique. Section 6.4 describes the overall implementation and the key blocks, while Section 6.5 presents measurement results. Conclusions are drawn in Section 6.6.

7.2 Injection-Caused Phase Error

In a typical ILCM, the free-running RO output OUT is reset by a pulse INJ generated from the reference REF. The injected RO experiences periodic phase disturbance due to imperfect injection. Prior works usually leverage this issue by a calibrator that deals with the phase error from frequency mismatch and delay offset. Fig. 1(a) shows the operation principle of the conventional calibrator, where a PD is introduced to monitor the phase error in the injection moment. Fig. 1(b) describes the effects of the frequency and delay errors and explains the calibration principle. OUT_D is a replica of the RO output OUT and the delay T_R from OUT to $OUT\ D$ satisfies $T_R = \cdot T_{OUT}$, where $T_{OUT} = 1/F_{OUT}$ is the average output period. Thus, OUT_D can be used to indicate the ideal position. When the RO runs at the target frequency F_{OUT} that satisfies $F_{RO} = F_{OUT} = N \cdot F_{REF}$, OUT and OUT_D are aligned. Otherwise, the frequency mismatch is accumulated as the timing error T_{err} between OUT and $OUT\ D$, which can be detected to adjust the RO frequency F_{RO}. However, if $T_R \neq T_{OUT}$, the delay offset is involved in T_{err} as well. Since the injection resets the RO phase and forces the average output period to be $T_{OUT} = T_{REF}/N$, F_{RO} must settle with error, which results in a periodic phase error happens for every injection cycle. In Fig. 1(a), the unit delay T_0 in the RO and T_R in the delayed replica are calibrated based on the phase error T_{err} detected by a PD. Thus, both frequency error and delay offset are minimized, and the ideal injection is achieved when OUT and $OUT\ D$ are aligned.

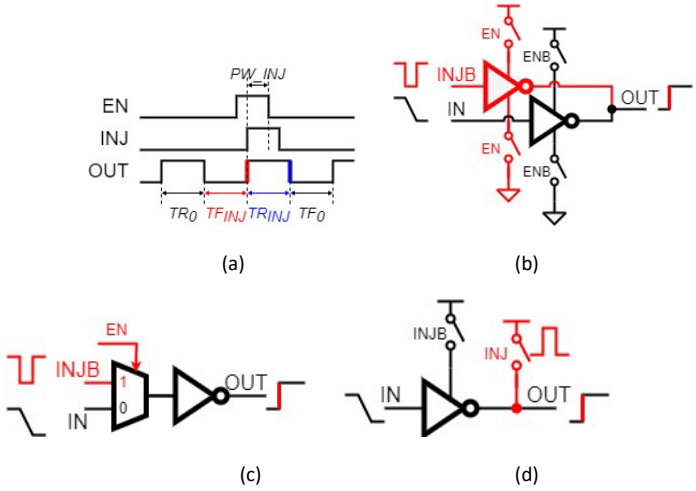


Fig. 4. Implementation of the injection: (a) timing diagram; (b) tri-state inverters; (c) multiplexer; (d) direct injection.

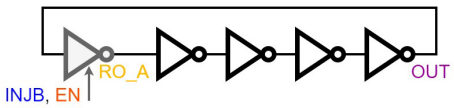
However, this calibrator is designed under the assumption that the injection only affects one single rising edge. Fig. 2 explains the possible phase distortion in a practical situation. When the injection is ideal, OUT keeps constant edges with the injection and all pulses have the same width T_0 as shown in Fig. 2(a). Considering the RO settles with frequency error as Fig. 2(b) shows, the injected rising edge varies due to the phase error accumulated from the frequency error. Besides, the following falling edges can be affected by non-ideal injection as well. The injected output pulse is distorted to $T_{O\ IN} \neq T_0$. A slope modulation in [8] was proposed to compensate for this pulse distortion. The regulated injection slope can improve the pulse timing, but it is limited when there are various error sources that also introduce errors in the falling edge.

Fig. 3 describes how the distorted pulses (DP) affect the calibration, where the delay offset is neglected to simplify the analysis. In the initial condition, the frequency error in F_{RO} manifests as the shifted rising edge, and the calibrator can detect the phase error and correct it by tuning F_{RO}. Due to the pulse distortion caused by the injection, the RO always settles with residual phase error. Fig. 3(a) shows the case that the injection causes a pulse width $T_{O\,INJ} > T_0$. The too large $T_{O\,INJ}$ remains after the calibration settles. To keep the average frequency, the RO must run at $F_{RO} > N \cdot F_{REF}$.

Similarly, if $T_{O\,INJ} < T_0$, the RO settles to $F_{RO} > N \cdot F_{REF}$. In both cases, the calibrator is incapable of detecting and correcting the residual error in the varied pulse width related to the falling edge. The distorted output pulse occurs every injection cycle and gives rise to the reference spur.

Fig. 5. A Simplified diagram of the injected RO.

The pulse distortion should be considered in the calibration, which is highly dependent on the



implementation of the injection. Fig. 4 shows different injection cells in conventional works that realize the timing in Fig. 4(a). Here, an enable signal EN can be introduced to avoid the race between the injection INJ and the intrinsic phase recirculation. The effective injection pulse width PW_{INJ} is determined by the common activated region of EN and INJ. Since the injection reshapes the output, the duration TF_{INJ} and TR_{INJ} deviate from their value TF_0 and TR_0 in the free-running cycle. In Fig. 4(b), two tri-state inverters are used as the main delay cell and the injection cell, respectively. When EN is high, the inverted $INJB$ makes the rising edge in OUT instead of the original input IN, while a multiplexer in Fig. 4(c) realizes this path selection. In Fig. 4(d), the rising edge is created by INJ that pulls up OUT directly.

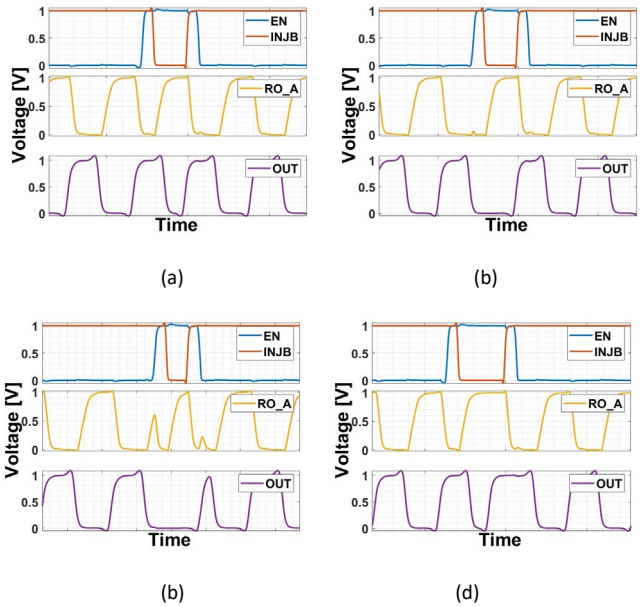
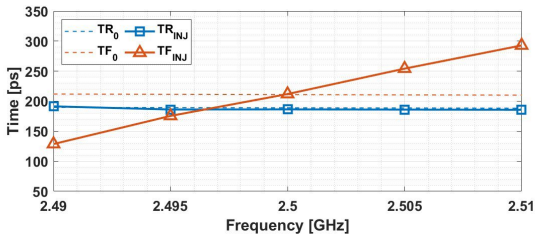


Fig. 6. Injection-locked RO: (a) $F_{RO} > NF_{REF}$; (b) $F_{RO} > NF_{REF}$; (c) $PW_{INJ} \ll T_0$; (d) $PW_{INJ} \gg T_0$.

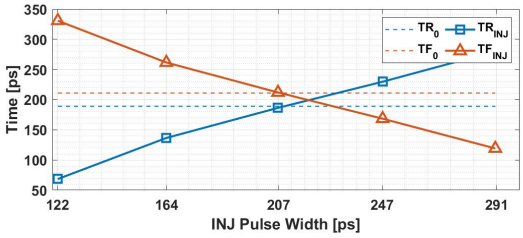
The injection cell can be employed to form an injected RO that is simplistically depicted in Fig. 5. To analyze the output pulse distortion, we simulate with a five-stage RO using tristate inverters with 4-finger

transistors. F_{RO} = 2.5 GHz and PW_{INJ} = 208 ps are adopted as the default value while one of them is varied to observe its effect on the phase distortion.

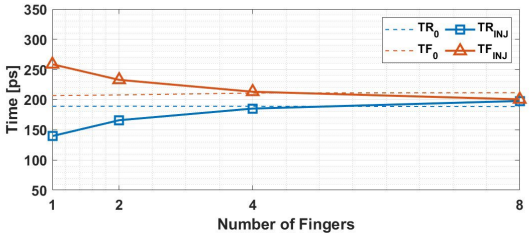
Fig. 6(a) and Fig. 6(b) show that the injected rising edge varies as the free-running frequency F_{RO} differs from the target F_{OUT} = 2.5 GHz, where TF_{INJ} is smaller when F_{RO} = 2.49 GHz and larger with F_{RO} =2.51 GHz. This scheme is on the premise that other cycles of OUT are constant and TF_{INJ} reaches the ideal value as long as F_{RO} is correct. However, this condition is not valid with non-ideal injection pulse width PW_{INJ}. In Fig. 6(c) and Fig. 6(d), TF_{INJ} and TR_{INJ} are both affected by PW_{INJ} and F_{RO} settles with residual error as what discussed above. Fig. 7 explores how TF_{INJ} and TR_{INJ} are affected in an injected RO without calibration, where tri-state inverters are used. In Fig. 7(a), TF_{INJ} changes as F_{RO}, and TR_{INJ} is constant with the fixed



(a)



(b)

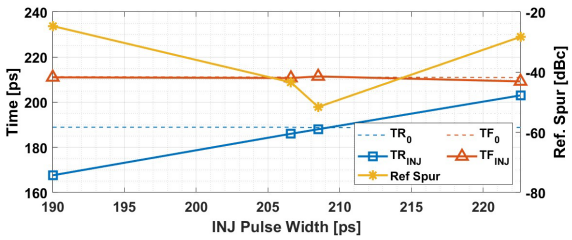


(c)

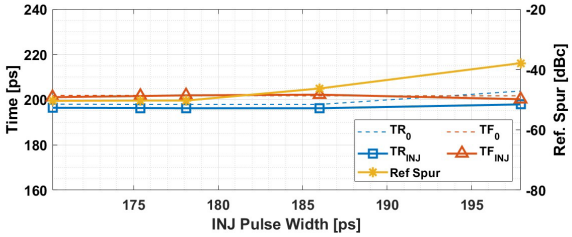
Fig. 7. Output pulse in the injected RO varies with (a) free-running frequency F_{RO}, (b) injection pulse width PW_{INJ}, and (c) the number of transistor fingers.

PW_{INJ}. However, when PW_{INJ} varies, both TF_{INJ} and TR_{INJ} deviate away from their ideal value, which suggests that TF_{INJ} cannot indicate correct frequency information. Further, TF_{INJ} and TR_{INJ} vary as different transistor sizes as shown in Fig. 7(c). With the same F_{RO} and PW_{INJ}, using fewer fingers in the injection cell results in a slower transition in OUT and a wider output pulse, while the output pulse gets narrower with the fast transition when increasing the number of fingers.

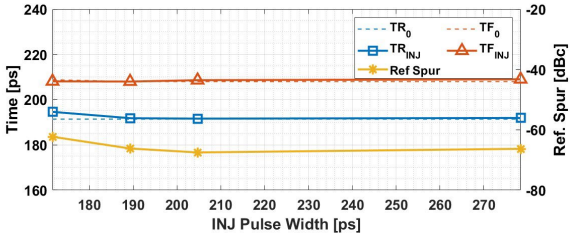
The discussion above explains that pulse distortion is sensitive to various parameters. Even if the calibrator can correct TF_{INJ} by tuning F_{RO}, the residual error from other variations remains. Simulations are performed to compare different calibrated ILCMs. As Fig. 8 shows, TF_{INJ} can be calibrated to TF_0 in all cases, while TR_{INJ} is highly dependent on the exact implementation. In Fig. 8(a), tri-state inverters are used as the injection cell and EN covers the duration of $INJB$, which allows $INJB$ to make a new pulse in OUT.



(a)



(b)



(c)

Fig. 8. Output pulse of the calibrated ILCM: the conventional calibrator with (a) tri-state inverters, (b) multiplexer, and (c) the proposed IPS-ILCM.

In this way, TR_{INJ} follows the variation in the pulse width of $INJB$, so the reference spur is limited by the residual phase error. Further, the duration of EN also introduces uncertainty in the output. The reference spur reaches the minimum value only when TR_{INJ} approaches TR_0. On the other hand, EN ends before $INJB$ in Fig. 8(b) with the multiplexer-based cell. Instead of the whole output pulse, only the rising edge is reshaped by the injection. Here, TR_{INJ} is relatively constant, especially when PW_{INJ} is small. As PW_{INJ} increases, the reference spur gets slightly higher due to the modulation effect from $INJB$ in the multiplexer input. In both cases, the minimum reference spur is limited to about -50 dBc, which results from unexpected pulse distortion.

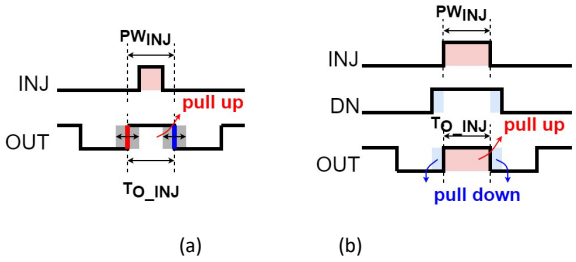


Fig. 9. Injection realized by: (a) pulling up only in a conventional implementation and (b) the proposed injection pulse shaping.

Conventional calibrators cannot deal with the pulse distortion caused by the injection. The key restriction is the incapacity of shaping the output pulse. Fig. 9(a) shows the injection in conventional implementation. To reset the output phase, the injection is supposed to form a rising edge by pulling up the output. However, $T_{O\,INJ}$ is not always equal to T_0, where both rising and falling edges can be distorted by non-ideal injection timing and circuit nonidealities. To recover the desired output waveform, we use injection pulse INJ to shape the output accurately. As shown in Fig. 9(b), besides pulse INJ pulling up OUT, a slightly wider pulse DN is introduced. It provides down pulling only when INJ is deactivated. Thanks to the explicit falling edge, the output pulse is well characterized by INJ. Fig. 8(c) presents the simulation results with different initial injection pulse widths. TF_{INJ} and TR_{INJ} maintain constant after the PW_{INJ} is calibrated, and the reference spur can reach below -65 dBc across a wide-range variation.

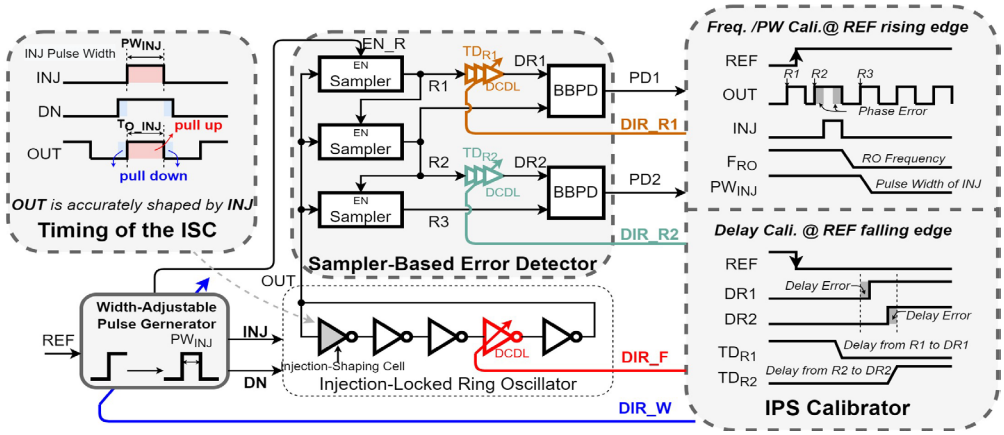


Fig. 10. Overall architecture of the proposed IPS-ILCM.

7.3 Injection Pulse Shaping Technique

7.3.1 Proposed Injection Pulse Shaping

Fig. 10 shows the overall architecture of the proposed IPSILCM. The RO frequency error, PD delay offset, and pulse distortion are calibrated by an IPS calibrator. In the RO, an injection shaping cell (ISC) is employed to reshape the output with two pulses INJ and DN, which are generated by a width-adjustable pulse generator (WAPG). The output pulse width $T_{O\,INJ}$ is characterized by the width PW_{INJ} of INJ. We propose a sampler-based error detector (SED) to detect output phase disturbance in the injection. Three high-frequency sequential rising edges $R1$, $R2$, and $R3$ in the RO output around INJ are captured by samplers, which allows the calibration to work at the rate of F_{REF}. Two bangbang phase detectors

(BBPDs) are adopted to identify the edge misalignments, which can be resulted from the frequency drift in F_{RO}, the delay errors in TD_{R1} and TD_{R2}, and the injection-caused pulse distortion in PW_{INJ}. The SED is shared by all calibrations, which prevents extra offset. The IPS calibrator adjusts F_{RO} and PW_{INJ} at the rising edge of REF, while at the falling edge of REF, TD_{R1} and TD_{R2} are updated. These corrections are realized in the time domain by digitally controlled delay lines (DCDLs). Neglected in Fig. 10, a counter based FTL is used for coarse frequency tracking.

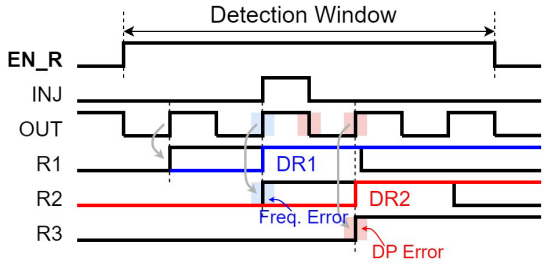
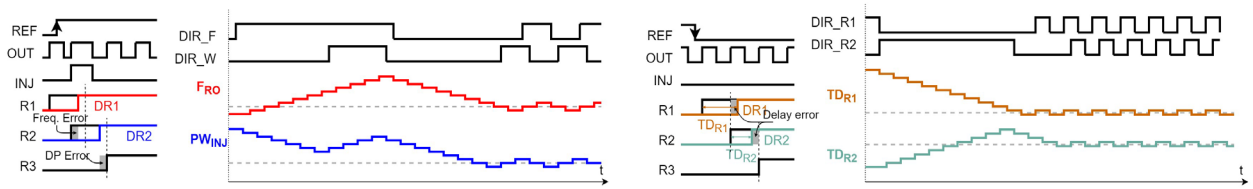


Fig. 11. Timing diagram of the sampling error detector.

7.3.2 IPS Calibrator

To identify frequency error, delay error, and pulse distortion separately, the IPS calibrator is activated twice for every injection cycle as shown on the right side of Fig. 10. This dual phase operation is realized with an enabling signal EN_R that combines two pulses triggered by both transition edges of REF and provides two detection windows. Fig. 11 describes the timing in the phase that REF rising edge arrives and generates the injection pulse. EN_R triggers the snapshot of the high-frequency edges in OUT, which are captured as three low-frequency signals $R1$, $R2$ and $R3$. Ideally, the delay between two adjacent rising edges in OUT is equal to T_{RO}, which holds true for the captured edges as well. Based on this, we delay $R1$ and $R2$ by T_{RO} to get $DR1$ and $DR2$, and adopt them as the reference to extract the phase error in $R2$ and $R3$ by the SED. In this phase, the injection can cause a phase error in output edges. The timing variations in $R2$ and $R3$ result from the frequency error and the pulse distortion, respectively. Thus, the calibrator adjusts F_{RO} and PW_{INJ} to minimize the phase error.



(a) (b)

Fig. 12. Settling behaviors of the IPS calibrators. (a) Frequency and pulse width are calibrated at the reference rising edge. (b) Delays are calibrated at the falling edge.

EN_R has another detection window at the REF falling edge, which triggers the edge capture similar to what is in Fig. 11, while the RO output cycles are supposed to be even in the absence of the injection. Here, the timing variations come from the delay errors in $DR1$ and $DR2$, which are calibrated by updating TD_{R1} and TD_{R2}.

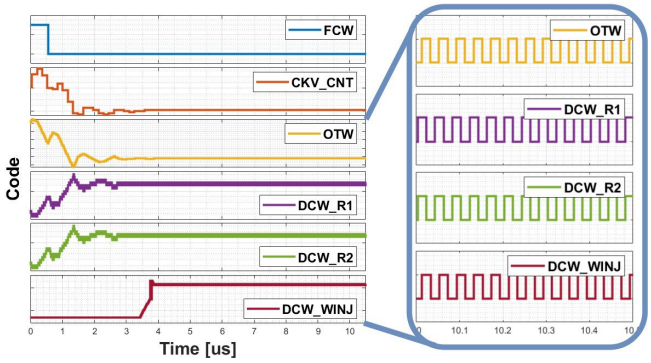


Fig. 13. Simulated settling process of the IPS-ILCM and the steady state.

Fig. 12 illustrates the settling behavior of the proposed IPS calibrator. The REF rising phase shown in Fig. 12(a) is for calibrating the frequency error and pulse distortion. $R2$ and $R3$ deviate away from their ideal positions indicating as $DR1$ and $DR2$ due to the non-ideal injection. The tuning directions $DIR\ F$ and $DIR\ W$ for correcting F_{RO} and PW_{INJ} are based on the BBPD outputs $PD1$ and $PD2$. At REF falling phase in Fig. 12(b), TD_{R1} and TD_{R2} are updated, where $PD1$ and $PD2$ give the tuning directions DIR_R1 and DIR_R2. Aligning $DR1$ to $R2$ and $DR2$ to $R3$ in both phases eliminates phase errors and achieves the target state where $F_{RO} = F_{OUT}$ and $TD_{R1} = TD_{R2} = T_{RO}$.

Noted that the ideal values of PW_{INJ}, TD_{R1}, and TD_{R2} depend on the output frequency F_{OUT}, and they are compensated in the time domain by the IPS calibrator. Fig. 13 presents the simulated loop settling of the proposed IPS-ILCM. As the RO frequency and delay offset are calibrated in different phases, they are operating without interrupting each other. The counter based FTL works first when the frequency control word FCW changes. After the coarse locking completes, the IPS calibrator starts to adjust the RO frequency and delay offset. The injection pulse width tuning is turned on only when the RO oscillation control word OTW and delay control words $DCW\ R1$ and DCW_R2 vary in a narrow range. Finally, the IPS-ILCM enters the steady state, where each word toggles in adjacent steps if the random noise is not considered. The residual error is determined by the maximum step size of T_{RO}, TD_{R1}, TD_{R2}, and PW_{INJ}. To achieve a low reference spur, the step size should be minimized, while more units are required to cover the frequency range and make a slower settling. In this design, the frequency tuning gain $K_F = 10$ KHz/bit is equivalent to $K_T = N \cdot K_F/F_{RO}^2 = 100$ fs/bit, which is comparable to the step of PW_{INJ}, TD_{R1} and TD_{R2}. In practice, the random noise can break the limit cycle and allows the reference spur lower than that (1) predicts.

7.3.3 Noise Analysis

In an ILCM, the injection introduces periodic phase realignment in the oscillator and feeds the reference noise to the output directly, while the noise is also filtered by the FTL. The proposed IPS-ILCM has a similar noise mechanism to this model, while it includes four calibration paths to update the RO and DCDLs. Specifically, TD_{R1} and TD_{R2} are used for calibrating F_{RO} and PW_{INJ} separately. In the IPS calibrator, the FTL only processes the frequency error, while PW_{INJ} and TD_{R1} only affect the direct forwarding reference noise. To simplify the noise analysis, the noise of delay tuning is included in the reference jitter, which is simulated by treating them as the high-pass filtering output noise of a delay-

locked loop. Based on this, the multi-rate operation in the IPS-ILCM can be modeled in the F_{REF} domain. The cross-frequency conversion can be avoided, which gives an explicit prediction of the output jitter and phase noise [15].

Fig. 14(a) describes the time-domain model of the IPSILCM. $tr[k]$ and $td[k]$ describe the time stamps of the reference and the feedback, where the latter is down converted from the output $tv[h]$ in F_{OUT} domain. The timing error $\Delta t[k]$ is detected by the BBPD and quantized as $dt[k] = \pm 1$. This quantization can be represented as a block with the linear gain K_{bpd} and the quantization error $q[k] = K_{bpd} \cdot \Delta t[k] - dt[k]$. As evident in previous works [16]–[18], K_{bpd} is dependent on the input-referred jitter $\sigma_{\Delta t}$ and expressed as

$$K_{bpd} = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sigma_{\Delta t}}. \tag{2}$$

The variance of $q[k]$ is $\sigma_q^2 = 1 - 2/\pi$. An accumulator is used as the digital loop filter (DLF) and its gain is chosen as $\alpha = 1$ to minimize the quantization error in the oscillation

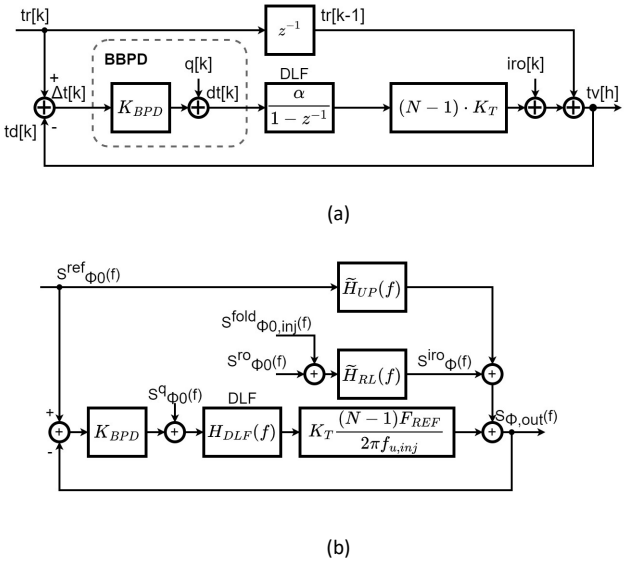


Fig. 14. Phase noise models of the IPS-ILCM: (a) time-domain model and (b) frequency-domain model.

controlling word $otw[k]$. $iro[k]$ is the down-sampled output from the injected RO. Using this closed-loop model, the input referred jitter $\sigma_{\Delta t}$ of the BBPD can be solved as

$$\sigma_{\Delta t} = \frac{\eta}{2} + \sqrt{\left(\frac{\eta}{2}\right)^2 + 2\sigma_{tr}^2 + (N-1)\sigma_{Tro}^2}, \tag{3}$$

where $\eta = {}^{\mathrm{p}}\pi/8(N-1)\alpha K_T$, and σ_{tr} and σ_{Tro} are the derivations of the reference jitter and the cycle jitter of the free-running RO, respectively. Consider the ILCM output $tv[h]$ updates at $F_{OUT} = NF_{REF}$, it has the variance $\sigma_{tv}^2 = \sigma_{\Delta t}^2/2$ due to the periodic accumulation and reset operation in the injected RO. $\partial\sigma_{\Delta t}/\partial(K_T) > 0$ suggests that $\sigma_{\Delta t}$ is positively related to the RO tuning step K_T, and so as σ_{tv}. However, as discussed in [15], they maintain relatively constant when the random noise is dominant. Thus, K_T has a small impact on the noise performance, and it should be minimized for low reference spur.

The phase noise spectrum of the IPS-ILCM can be predicted using the frequency-domain model in Fig. 14(b). The injection resets the RO phase every *REF* cycle and introduces the zeroorder-hold in the RO phase, which is approximated as

$$\widetilde{H}_{up}(f) = \frac{N}{1 + jf/f_{u,inj}} \tag{4}$$

and

$$\widetilde{H}_{rl}(f) = \frac{1}{1 + f_{u,inj}/jf} \tag{5}$$

It is obvious that $H_{eup}(f)$ and $H_{erl}(f)$ have low-pass and highpass shapes, respectively, where the noise filtering bandwidth can be estimated as $f_{u,inj} \approx 0.28F_{REF}$ [15].

The noise sources $S^{ref}_{\Phi 0}(f)$ from the reference, $S^{q}_{\Phi 0}(f)$ from the quantization in the BBPD and S^{iro} from the injected RO are considered in Fig. 14(b). $S^{ref}_{\Phi 0}(f)$ includes the noise from the off-chip reference and the DCDLs in the IPS calibrator. It contributes to the output noise $S_{\Phi,out}(f)$ through the forwarding path $H_{eup}(f)$ and the FTL. $S^{q}_{\Phi 0}(f)$ is filtered by the DLF $H_{dlf}(f)$. The phase noise S^{iro} from the injected RO is considered as

$$S^{iro}_{\Phi}(f) = \left[S^{ro}_{\Phi 0}(f) + S^{fold}_{\Phi 0,inj}(f)\right] \cdot |\widetilde{H}_{rl}(f)|^2, \tag{6}$$

which is the combined result of the free-running noise $S^{ro}_{\Phi 0}(f)$ and its folding harmonics $S^{fold}_{\Phi 0,inj}(f)$. The latter term is introduced to count for the far-out 3-dB increase in the injected RO and it has a high-pass profile as

$$S^{fold}_{\Phi 0,inj}(f) \approx S^{ro}_{\Phi 0}(f) \cdot |\widetilde{H}_{rl}(f)|^2. \tag{7}$$

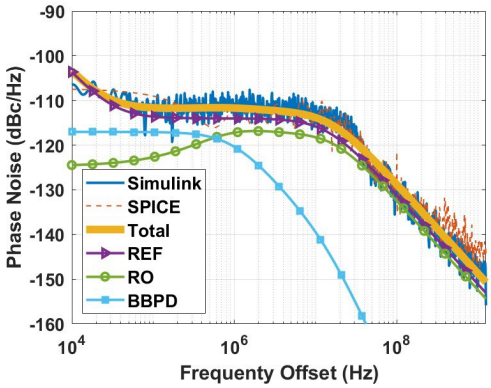


Fig. 15. Simulated output phase noise of the proposed IPS-ILCM.

The FTL has a different response from that of a PLL. Here, the autonomous phase accumulation in the RO is reset periodically and degraded to a linear gain $(N\text{-}1)K_TF_{REF}/(2\pi f_{u,inj})$. The loop gain of the FTL can be expressed as

$$\widetilde{G}_{FTL}(f) = K_{bpd} H_{dlf}(f) \cdot \frac{(N-1) F_{REF} K_T}{2\pi f_{u,inj}}, \qquad (8)$$

where $H_{dlf}(f) = \alpha F_{REF}/(j2\pi f)$. The FTL bandwidth can be estimated by its unity-gain frequency [15]

$$f_{u,ftl} \approx \frac{N-1}{10} \alpha K_T K_{bpd} F_{REF}, \qquad (9)$$

which increases as K_T. Thus, the noise transfer functions of the reference, BBPD, and the injected RO can be derived as

$$\begin{aligned} NTF_{REF}(f) &= \frac{\widetilde{H}_{up} + N\widetilde{G}_{FTL}(f)}{1 + \widetilde{G}_{FTL}(f)} \\ &\approx \frac{N}{1 + jf/f_{u,inj}} \end{aligned} \qquad (10)$$

$$\begin{aligned} NTF_{BPD}(f) &= \frac{1}{K_{BPD}} \cdot \frac{\widetilde{H}_{up} + \widetilde{G}_{FTL}(f)}{1 + \widetilde{G}_{FTL}(f)} \\ &\approx \frac{1}{K_{BPD}} \frac{1}{1 + jf/f_{u,inj}} \end{aligned} \qquad (11)$$

$$\begin{aligned} NTF_{IRO}(f) &= \frac{\widetilde{H}_{rl} + \widetilde{G}_{FTL}(f)}{1 + \widetilde{G}_{FTL}(f)} \\ &\approx \frac{1}{(1 + f_{u,inj})/jf) \cdot (1 + f_{u,ftl}/jf)} \end{aligned} \qquad (12)$$

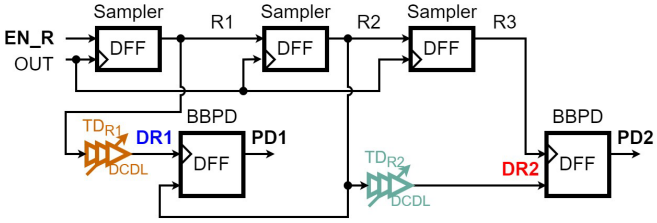


Fig. 16. Schematic of the SED.

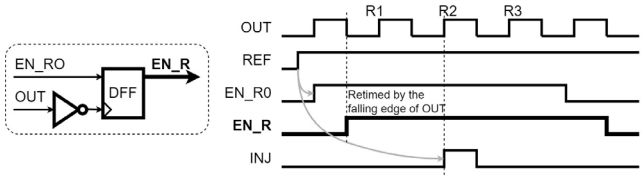


Fig. 17. Retiming for $EN\ R$.

Using the model in Fig. 14(b), the overall output phase noise of the proposed IPS-ILCM can be estimated as shown in Fig. 15. The frequency-domain estimation coincides well with the simulated results in Simulink and SPICE. Since the noise from the reference and DCDLs contributes to the output mainly through the direct forwarding path, the FTL in the IPS calibrator has limited filtering for them. To optimize the noise performance, using wide bandwidth is preferable for the suppression of the RO phase noise, which has little effect on the noise from the reference DCDLs.

7.4 Circuit Implementation

7.4.1 SED

Fig. 16 shows the schematic of the SED. The high-frequency edges in OUT are down-converted to the injection rate. The error detection is performed at low frequency, which relieves the timing requirement of the IPS calibrator and saves power. Three D-flip-flops (DFFs) are utilized as samplers. They are triggered by the RO output within the enabling window EN_R, which is sampled by the first DFF and the other two DFFs sample their previous DFF output. In this way, three adjacent rising edges in OUT are captured as $R1$, $R2$, and $R3$ after EN_R is activated in both transitions of REF.

Assuming all samplers have the same latency, $R1$, $R2$, and $R3$ are supposed to have the same timing relationship as the cycles in OUT.

EN_R and INJ are both generated from REF and the delay between them is constant. However, the timing relationship between REF and OUT is unknown in different frequencies and varies during the settling. This entails possible timing violation in DFFs when EN_R gets close to the edges to be captured. To avoid this risk, EN_R is retimed as shown in Fig. 17. The primary pulse EN_R0 is re-sampled by the falling edge of OUT. Thus, the first rising edge in OUT always has a well-defined delay from EN_R for all operating states. With the clear timing provided by the SED, the IPS calibrator works more robustly than conventional calibrators that suffer from complicated timing requirements in the high frequency domain.

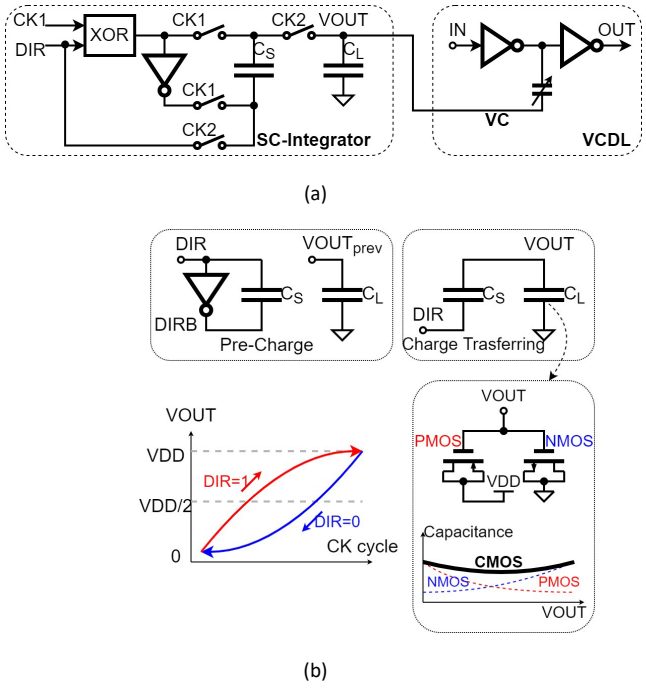


Fig. 18. Implementation of the DCDL: (a) Schematic of the DCDL and (b) Principle of the switched-capacitor integrator.

7.4.2 DCDL

The proposed IPS-ILCM uses DCDLs presented in Fig. 18 to realize error calibration. As Fig. 18(a) shows The DCDL is based on a voltage-controlled delay line, where the controlling voltage VC is discrete-time

updating with a switched capacitor (SC) integrator. For each DCDL, the charge transfer is controlled by the direction signal DIR with two nonoverlapped clocks $CK1$ and $CK2$. Fig. 18(b) shows the principle of the SC integrator. In the pre-charge phase, C_S is charged according to DIR, then $V\,OUT$ is updated by transferring the charge to C_L.

To increase $V\,OUT$ when DIR is high, the top plate C_S is connected to $V\,DD$ and its bottom plate is forced to the ground. Then, the bottom plate is inverted to $V\,DD$ and more charge is transferred from C_S to C_L. Based on the charge conservation, the increased output voltage can be calculated as

$$\Delta VOUT = (-VOUT_{prev} + 2VDD) \cdot \frac{C_S}{C_S + C_L}, \quad (13)$$

where $V\,OUT_{prev}$ is the voltage in the previous state. Similarly, the decreased output voltage when DIR is

$$\Delta VOUT = (-VOUT_{prev} - VDD) \cdot \frac{C_S}{C_S + C_L}. \quad (14)$$

In the SC integrator, small C_S and large C_L are preferable to achieve a fine step $\Delta V\,OUT$. To save area, MOS-caps are adopted for C_L. The intrinsic voltage-dependent capacitance is compensated using a complementary configuration. Even though the charge integration is nonlinear, it has a relatively constant step around $V\,DD/2$. Using this discrete-time SC integrator renders low-power and high-resolution delay tuning without massive passive devices.

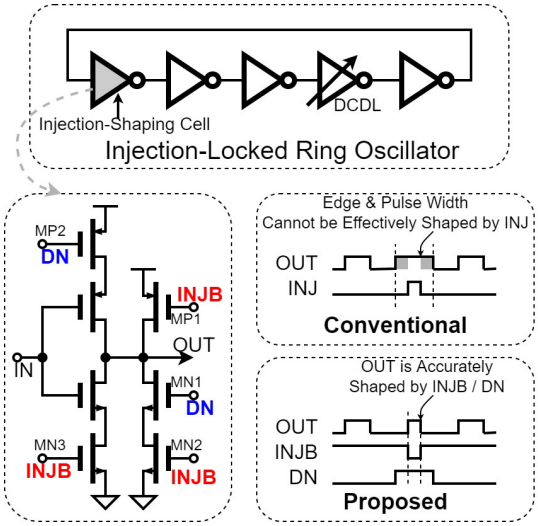


Fig. 19. Schematic of the injection-locked RO with the injection shaping cell (ISC).

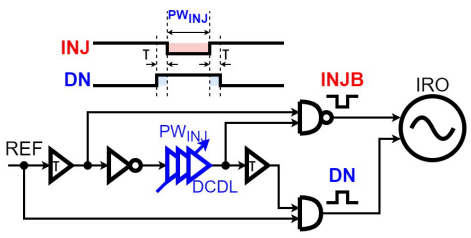


Fig. 20. Schematic of the width-adjustable pulse generator.

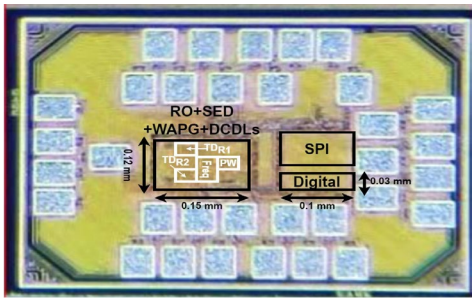
7.4.3 Injection-Locked RO

Fig. 19 illustrates the schematic of the injection-locked RO with the injection shaping cell (ISC). In the proposed ISC, the inverted $INJB$ is used to pull up OUT by MP1, where MN1 and MN2 give $INJB$ a higher priority than the down-pulling of DN. Disabling IN by MP2 and MN3 avoids the racing between the injection and the autonomous phase incoming in the RO. Unlike in a conventional ILCM that resets the output phase only by pulling up, the proposed ISC recovers the ideal output waveform by forcing the pulse width TO_{INJ} to PW_INJ.

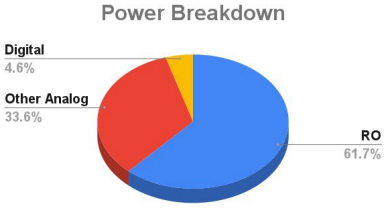
Fig. 20 shows the schematic of the WAPG that generates the injection pulses $INJB$ and DN for the ISC. The pulse width PW_INJ is adjusted by a DCDL. DN is supposed to be a little wider than $INJB$, but extra error can be caused by a too wide DN. This can be realized by inserting two fixed delay cells of τ between DN and $INJB$, which is implemented with two fixed delay cells at the input and the output of the DCDL.

7.5 Measurement Results

A prototype chip has been fabricated in a 65 nm CMOS technology with an active area of 0.021 mm^2. The chip micrograph is shown in Fig. 21(a). It is characterized using Rohde & Schwarz FSW26 signal analyzer and Keysight E5052B signal source analyzer. The proposed IPS-ILCM covers output frequency 2.1-2.7 GHz with 50 MHz reference, the proposed IPS-ILCM consumes 1.23 mW at 2.5 GHz output. As the power breakdown presented in Fig. 21(b), the RO dissipates more than 60% power. The majority area is occupied by the DCDLs with the SC integrator, where a large load capacitance is employed to minimize the residual error. Thanks to the controlling logic and the timing that have been significantly simplified with the SED and the discrete-time SC integrator, the digital block is very compact and power-efficient.



(a)

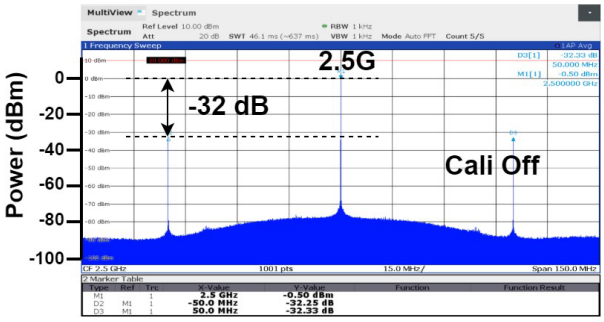


(b)

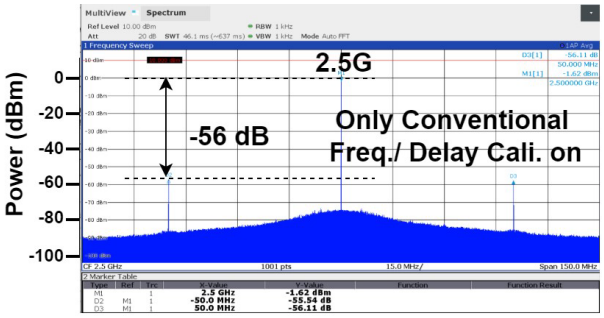
Fig. 21. (a) Chip micrograph (active area=0.021mm^2). (b) Power breakdown.

Fig. 22 shows the measured output spectrum of the ILCM. In Fig. 22(a), all calibrations are off after the coarse FTL brings the RO near the target frequency. The RO is locked to 2.5 GHz by injection. The free-running frequency of the RO drifts slowly away, resulting in poor reference spurs of -32 dBc. In Fig. 22(b), the RO frequency and delay offset of the PD path are tracked and corrected in the background as what has been implemented in conventional calibrators. The reference spur can be reduced to -56 dBc. As Fig. 22(b) shows, the measured reference spur is sensitive to system variations such as technological deviation, supply voltage, and the initial state of the ILCM. In Fig. 22(c), all features of the proposed IPS calibrator are enabled. As all phase error caused by imperfect injection is eliminated by the accurate pulse shaping, the level of reference spurs is further decreased to -79 dBc.

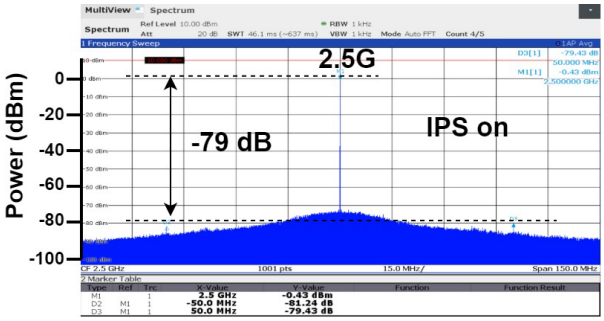
Fig. 23 shows the measured output spectrum at 2.1 GHz and 2.7 GHz. With the IPS calibrator, the reference spur in both cases is significantly reduced to lower than -80 dBc. The proposed IPS-ILCM can maintain low reference spur over output frequency and supply variations. As Fig. 24(a) shows, for three measured samples, the reference spur is all below -75 dBc from 2.1 GHz to 2.7 GHz. Moreover, the measured reference spurs of the IPS-ILCM are stable for the supplies ranging 1.0-1.1 V as shown in Fig. 24(b). On the other hand, when only the frequency and delay calibrations are enabled, the reference spur varies with voltage significantly.



(a)



(b)

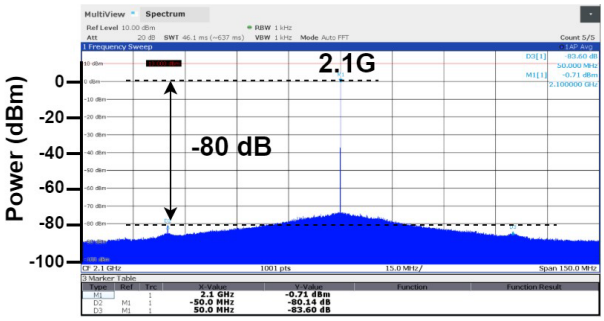


(c)

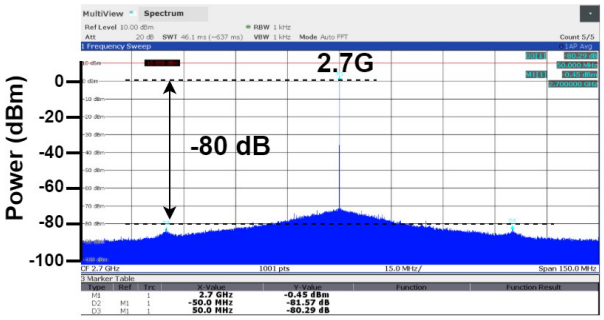
Fig. 22. Measured output spectrum at 2.5 GHz. (a) All calibrations are off; (b) Only frequency and delay calibrations are enabled as in conventional calibrators. (c) Proposed IPS calibrator.

Fig. 25 illustrates the measured output phase noise of the IPS-ILCM using 50 MHz reference, as well as the free-running RO profile around 2.5 GHz. Since the injection provides high-pass filtering with wide bandwidth, the RO phase noise is effectively suppressed. At 100 KHz and 1 MHz offset frequency, the spot noise noises are about -111.5 dBc/Hz. The RMS jitter integrated from 10KHz to 40 MHz is about 819 fs. Consuming 1.23 mW power to generate the output of 2.5 GHz, the proposed IPS-ILCM achieves an FoM of 240.8 dB and power efficiency of 0.49 mW/GHz. Table I summarizes the comparison with the state-of-the-art RO-based ILCMs, where prior works rarely achieve low reference spur because they usually focus on frequency and delay error. On the other hand, a high-frequency reference is preferable for reducing jitter and reference spur, which increases power consumption. Since the proposed IPS-ILCM can eliminate all kinds of phase error without complicated calibration logic, it reaches the lowest

reference spur with great power efficiency as shown in Fig. 26. This work provides a promising solution for a clock synthesizer with low jitter and compact area.

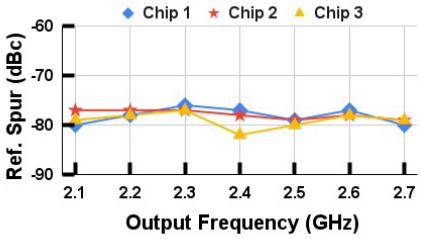


(a)

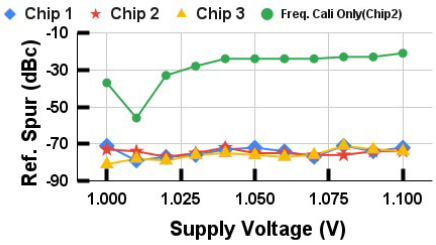


(b)

Fig. 23. Measured output spectrum of the IPS calibrator at (a) 2.1 GHz and (b) 2.7 GHz.



(a)



(b)

Fig. 24. Measured reference spur of the IPS-ILCM over (a) output frequency and (b) supply voltages.

TABLE I COMPARISON WITH RO-BASED ILCMS.

Reference	This Work	[19] Zhang JSSC'21	[12] Kundu JSSC'20	[8] Yoo JSSC'19	[20] Megawer ISSCC'18	[13] Kim ISSCC'17
Technology	65 nm CMOS	65 nm CMOS	22 nm FinFET	65 nm CMOS	65 nm CMOS	65 nm CMOS
Architecture	Int.-N ILCM	Frac.-N MDLL	Frac.-N MDLL	Int.-N ILCM	Int.-N ILCM	Int.-N ILPLL
FTL Type	IPS	Freq./Delay	Time/Period	Freq./Delay/Slope	Freq./Delay	Freq./Delay
Reference (MHz)	50	50	80	100	54	156.25
Output (GHz)	2.1-2.7	0.8-2	1.2-3.8	2.2-2.5	2.6-5.2	2.5
RMS Jitter (fs)	819	1670	2550	140	2550	198
Int. Range	10K-40MHz	10K-10MHz	10K-100MHz	10K-30MHz	10K-30MHz	10K-100MHz
Ref. Spur (dB)	-79	-44	-56	-72	-53	-65
Power (mW)	1.23	11.95	3	11.0	6.5	13.5
FoM (dB)	-240.8	-224.8	-226.3	-246.7	-240.5	-242.8
Area (mm^2)	0.021	0.18	0.00525	0.055	0.16	0.064

$$\text{FoM} = 10\log\left[\left(\frac{}{\text{RMS jitter}_{1\text{s}}}\right)^2 \cdot \frac{\text{Power}}{1\text{mW}}\right]$$

7.6 Summary

This chapter presented an RO-based ILCM in 65-nm CMOS with the IPS technique. The phase error caused by the injection is detected by a SED and calibrated by an IPS calibrator through discrete-time DCDLs. The SED detects the phase error with captured edges captured from the high-frequency output and provides clear timing of dual-phase operation for the IPS calibrator. Besides calibrations for the RO frequency and delay offset in the PD, the ideal output is recovered by accurately shaping the output pulse in the injection. Using the discrete time SC integrator in DCDLs realizes compact area and high power-efficiency. The FoM is -240.8 dB measured at 2.5 GHz. The proposed IPS-ILCM achieves -79 dBc reference spur with 50 MHz reference over variations.

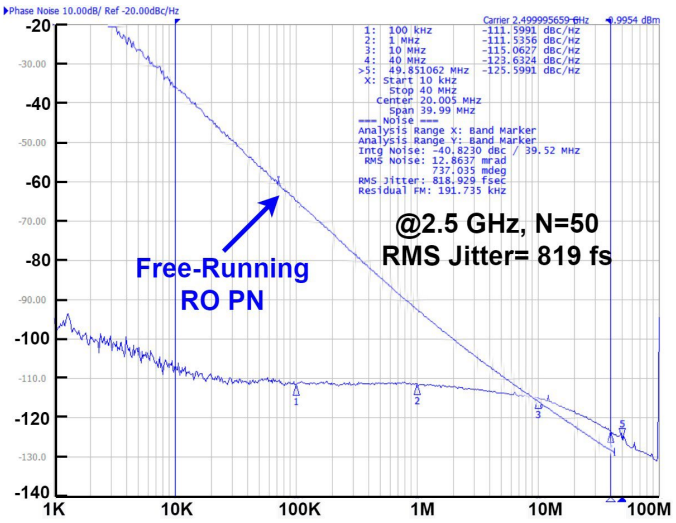


Fig. 25. Measured phase noise of the IPS-ILCM.

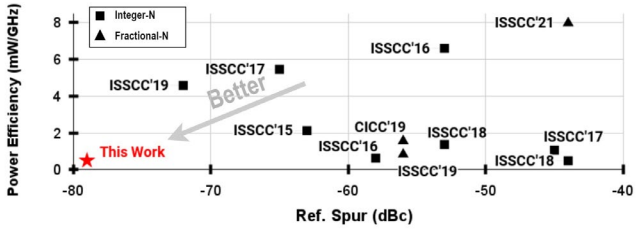


Fig. 26. Comparison with state-of-the-art RO-based ILCMs in terms of power efficiency and reference spur.

7.7 References

[1] S. L. J. Gierkink, "Low-spur, low-phase-noise clock multiplier based on a combination of PLL and recirculating DLL with dual-pulse ring oscillator and self-correcting charge pump," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2967–2976, 2008.

[2] B. M. Helal, C.-M. Hsu, K. Johnson, and M. H. Perrott, "A low jitter programmable clock multiplier based on a pulse injection-locked oscillator with a highly-digital tuning loop," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1391–1400, 2009.

[3] A. Elshazly, R. Inti, B. Young, and P. K. Hanumolu, "Clock multiplication techniques using digital multiplying delay-locked loops," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 6, pp. 1416–1428, 2013.

[4] A. Elkholy, M. Talegaonkar, T. Anand, and P. K. Hanumolu, "Design and analysis of low-power high-frequency robust sub-harmonic injectionlocked clock multipliers," *IEEE J. Solid-State Circuits*, vol. 50, no. 12, pp. 3160–3174, Dec. 2015.

[5] S. Choi, S. Yoo, Y. Lim, and J. Choi, "A PVT-robust and low-jitter ring-VCO-based injection-locked clock multiplier with a continuous frequency-tracking loop using a replica-delay cell and a dual-edge phase detector," *IEEE J. Solid-State Circuits*, vol. 51, no. 8, pp. 1878–1889, Aug. 2016.

[6] H. Kim, Y. Kim, T. Kim, H.-J. Ko, and S. Cho, "A 2.4-GHz 1.5-mW digital multiplying delay-locked loop using pulsewidth comparator and double injection technique," *IEEE J. Solid-State Circuits*, vol. 52, no. 11, pp. 2934–2946, 2017.

[7] S. Yang, J. Yin, P. Mak, and R. P. Martins, "A 0.0056-mm² -249-dB-FoM all-digital MDLL using a block-sharing offset-free frequency-tracking loop and dual multiplexed-ring VCOs," *IEEE J. Solid-State Circuits*, pp. 1–11, 2018.

[8] S. Yoo, S. Choi, Y. Lee, T. Seong, Y. Lim, and J. Choi, "A low-jitter and low-reference-spur ring-VCO-based injection-locked clock multiplier using a triple-point background calibrator," *IEEE J. Solid-State Circuits*, vol. 56, no. 1, pp. 298–309, 2021.

[9] S. Ye, L. Jansson, and I. Galton, "A multiple-crystal interface PLL with VCO realignment to reduce phase noise," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1795–1803, 2002.

[10] N. D. Dalt, "An analysis of phase noise in realigned vcos," *IEEE Trans. Circuits Syst. II*, vol. 41, no. 1, pp. 143–147, Mar. 2014.

[11] A. Santiccioli, C. Samori, A. L. Lacaita, and S. Levantino, "Time-variant modeling and analysis of multiplying delay-locked loops," *IEEE Trans. Circuits Syst. I*, vol. 66, no. 10, pp. 3775–3785, 2019.

[12] S. Kundu, L. Chai, K. Chandrashekar, S. Pellerano, and B. R. Carlton, "A self-calibrated 2-bit time-period comparator-based synthesized fractional-N MDLL in 22-nm FinFET CMOS," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 1, pp. 43–54, 2021.

[13] S. Kim, H.-G. Ko, S.-Y. Cho, J. Lee, S. Shin, M.-S. Choo, H. Chi, and D.-K. Jeong, "29.7 a 2.5GHz injection-locked ADPLL with 197fs rms integrated jitter and -65dBc reference spur using time-division dual calibration," in *Proc. IEEE Solid-State Circuits Conf. Dig. Tech. Papers*, 2017, pp. 494–495.

[14] R. Xu, D. Ye, and C.-J. Shi, "13.3 a 0.021mm^2 65nm CMOS 2.5GHz digital injection-locked clock multiplier with injection pulse shaping achieving -79dBc reference spur and 0.496mW/GHz power efficiency," in *Proc. IEEE Solid-State Circuits Conf. Dig. Tech. Papers*, 2022, p. ababa.

[15] R. Xu, D. Ye, and C. J. R. Shi, "Analysis and design of digital injectionlocked clock multipliers using bang-bang phase detectors," *IEEE Trans. Circuits Syst. I*, vol. 69, no. 7, pp. 2832–2844, Jul. 2022.

[16] N. D. Dalt, "A design-oriented study of the nonlinear dynamics of digital bang-bang PLLs," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 1, pp. 21–31, Jan. 2005.

[17] ——, "Linearized analysis of a digital bang-bang PLL and its validity limits applied to jitter transfer and jitter generation," *IEEE Trans. Circuits Syst. I*, vol. 55, no. 11, pp. 3663–3675, Dec. 2008.

[18] H. Xu and A. A. Abidi, "Design methodology for phase-locked loops using binary (bang-bang) phase detectors," *IEEE Trans. Circuits Syst. I*, vol. 64, no. 7, pp. 1637–1650, Jul. 2017.

[19] Q. Zhang, S. Su, C.-R. Ho, and M. S.-W. Chen, "A fractional-N digital MDLL with background two-point DTC calibration," *IEEE J. Solid-State Circuits*, vol. 57, no. 1, pp. 80–89, 2022.

[20] K. M. Megawer, A. Elkholy, D. Coombs, M. G. Ahmed, A. Elmallah, and P. K. Hanumolu, "A 5GHz 370fs rms 6.5mW clock multiplier using a crystal-oscillator frequency quadrupler in 65nm CMOS," in *Proc. IEEE Solid-State Circuits Conf. Dig. Tech. Papers*, 2018, pp. 392–394.

8 DESIGN AND ANALYSIS OF A FULLY SYNTHESIZABLE DIGITAL LOW DROPOUT VOLTAGE REGULATOR

This chapter describes a fully synthesizable digital low dropout (DLDO) regulator using an automatic offset control and reuse technique implemented in a 65-nm CMOS technology implemented using the RAIL flow. To realize the fully synthesizable DLDO design, all components of core blocks are made with standard logic cells. The proposed offset control and reuse technique is adopted to cancel the offset voltage from the logic cells automatically, to provide the adaptive equivalent thresholds for the voltage comparison window, and to speed up the dropout voltage response. Besides, the modified comparator-triggered oscillator introduces the output of the comparator to the oscillation loop to choose the optimum clock edge in the delay line to feedback, which ensures both of the fast response and enough time margin for the comparator delay at the same time. The core area of the output-capacitance-less DLDO is 0.027 mm^2. The simulated results show that a 100 mA step in load current produces a voltage drop of 140 mV with the response time of 1.2 ns. The steady-state error is less than 4 mV. The peak current efficiency is 99.9%.

8.1 Introduction

Low dropout regulator (LDO) is an important power management block needed for energy-efficient system-on-chip (SOC) design. Traditional analog LDO regulators can achieve a fine output voltage, a fast-transient response and good immunity to droop or overshoot [1], [2]. But their performance degrades in lower operating voltages, and they are not easy to scale down or up with different processes. Therefore, digital LDO regulators are explored due to their process scalability, compactness, process-voltage-temperature (PVT) immunity and easy programmability [3].

However, the current highly synthesizable core logic design of DLDO cannot completely avoid using analog techniques or precise voltage references to help the voltage comparison process [3]–[13], which hinders further integration. In [7], an autonomous gain tracking technique is introduced to the digital LDO loop for fast and stable transient response across PVT. A flash analog-to-digital converter (ADC) is required for the voltage comparison, which is a complete analog design and not friendly for process scalability. The non-linear variation of the power-gate (PG) resistance and loop output pole are discussed in [8] and a PG number computation technique is proposed to find the required number of PGs under different operating conditions quickly, while several digital-to-analog converters (DACs) are used in [8] to generate the reference voltages. Jonghyun *et al.* [9] presents a highly synthesizable digital LDO, which realizes almost all the core logic blocks by standard cells, but two reference voltages V_H and V_L are required from the outside of the chip. Therefore, how to generate the reference voltages based on all standard digital cells design is a critical problem that DLDO needs to solve for further integration improvement.

Besides, as an essential block of a LDO, the comparator provides the comparison result between the output voltage and the reference voltage to the feedback loop for the output current adjustment. Many DLDOs adopt the analog comparator to achieve higher sensitivity, lower offset voltage and more accurate output [11]–[13]. But this is not good for the process scalability and higher integration. As such, more research focus moves towards the pure digital comparator made up by standard digital cells [9],

[10]. However, due to the minimum length used in all transistors of the standard digital cells, the mismatch problem becomes more severe. In [9], the design provides a NAND-gate varactor to tune the input-referred offset manually. The manual offset tuning for each fabricated chip is time consuming and not suitable for practical applications. Hence, the automatic offset cancellation is much necessary for the digital standard-cell based comparator in DLDOs to ensure the output accuracy and wide acceptance.

This chapter presents a DLDO that is fully synthesizable in the digital flow without any analog blocks (ADC or DAC) and extra threshold voltages, capable of automatic offset cancellation and fast droop response without the use of output capacitance. This paper is organized as follows. The architecture and blocks are described in Section 7.2. Section 7.3 presents the simulation results in a 65-nm CMOS technology. Section **Error! Reference source not found.** summarizes this chapter.

8.2 Architecture and Block Design

Fig. 1 shows the architecture of the proposed DLDO. It is based on the extension of a fine- and coarse-control loop structure for efficient output voltage regulation. The finecontrol loop is composed of a comparator to sense the error between V_{ref} and V_{out}, a modified CMP triggered oscillator providing the optimum operation clock for the fine loop, fine-control bidirectional shift registers (Bi-SRs) and the fine power gates. The fine loop minimizes the difference between V_{ref} and V_{out}. The coarse loop employs two comparators to generate a voltage window for the fast response when there is

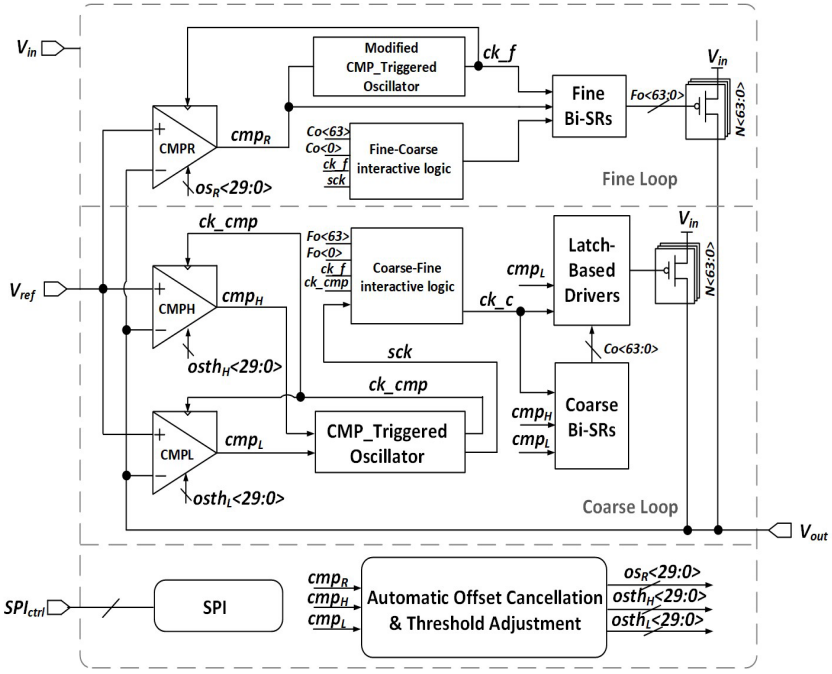


Fig. 1. The architecture of the synthesizable DLDO.

a large regulation error, a traditional CMP triggered oscillator [9] adopted for the fast logic operation and voltage recovery control, the coarse-control Bi-SRs and the latch-based drivers for fast response, and the coarse power gates that are around four times of the fine power gates. No real edge voltages V_H and are required to input to the regulator, the equivalent window is generated by reusing the offset of comparators. The desired offsets are controlled by $osth_H\langle 29:0\rangle$ and $osth_L\langle 29:0\rangle$ from the block of

automatic offset cancellation and threshold adjustment (AOSTH). The AOSTH shares the same logic to remove the dynamic offset voltage automatically for each comparator at the chip start-up time, and then sets up the desired window voltages by combining the threshold code with its corresponding offset code. The offset codes given to the symmetrical paths in the comparator generate the expected offset, which is equivalent to the off-chip analog threshold voltage for comparators in the coarse loop. Unlike the previous design [9], the output voltage switching between the fine and coarse loops is smooth and continuous by introducing the interactive logic between two loops.

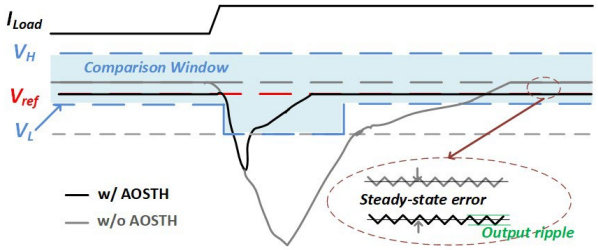


Fig. 2. The load transient response to current step.

Fig. 2 illustrates V_{out} resulting from the load-current step in the two cases of the automatic technique block AOSTH turned on or off. V_H and V_L are two equivalent voltages generated by the comparators in the coarse loop. In the traditional voltage window design, V_H and V_L are fixed, and for the stability, the window is relatively large. The large fixed window results in a slow response when there is a voltage droop. We introduce a variable equivalent low threshold scheme in the loop to ensure the stability during the steady state and improve the voltage sensitivity at the same time. As shown in Fig. 2, V_L will keep a small deviation with V_{ref} when the DLDO is at the steady state. This scheme guarantees the DLDO to have a high sensitivity with V_{out} droop and no need for extra droop detector like [9]. When a sudden V_{out} droop comes, the comparator responses quickly and cmp_L will be given to the latch-based drivers to control the power gates switching. At the same time, the equivalent V_L becomes large when the coarse loop is active. After the tracking process, the equivalent V_L will return to the small value automatically for the next droop detection.

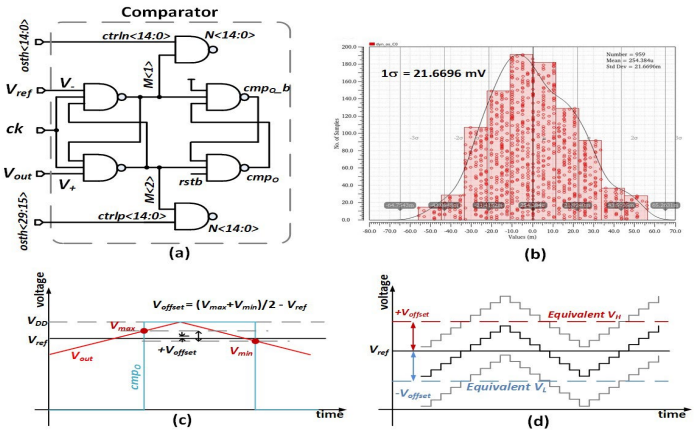


Fig. 3. (a) The comparator; (b) the simulated offset of the comparator by Monte-Carlo simulation; (c) the diagram of the dynamic offset; (d) The diagram of equivalent V_H and V_L.

8.2.1 Dynamic offset cancellation and equivalent threshold voltage adjustment

Comparators are the most sensitive blocks in various LDO designs and affect the accuracy of the output voltage directly. Therefore, it is required to realize the comparators with very low offset voltage. Compared with the traditional analog comparators, the offset becomes more sever for the digital comparators using the standard digital cells due to the poor matching performance caused by the small-length transistors. The offset includes the static offset and the dynamic offset. For a DLDO, the dynamic offset affects the loop mainly and will be discussed in this subsection.

Fig. 3(a) shows the comparator circuit, which employs 3input NAND gates and a NAND-gate array to adjust the offset voltage [10]. The 1-sigma dynamic offset of the comparator is about 21.7 mV which is much larger than that of a typical analog comparator. The diagram of the dynamic offset is given in Fig. 3(c). When V_{out} increases gradually from the lower voltage than V_{ref}, there is a voltage making the output of the comparator cmp_O changing from 0 to 1, and we refer to this voltage as V_{max}. When V_{out} decreases gradually from the voltage higher than V_{ref}, it will achieve some voltage to make cmp_O flipping from 1 to 0, and we refer to this voltage point as V_{min}. And then, the dynamic offset can be calculated by the following equation:

$$V_{offset} = (V_{max} + V_{min})/2 - V_{ref} \quad (1)$$

Once the offset is known, it can be cancelled in the reference comparator and the threshold codes will be combined with the offset in the coarse loop comparators and generate the equivalent threshold levels V_H and V_L, as shown in Fig. 3(d). This proposed method will turn different numbers of loading NAND gates on at the symmetric paths of the comparator, M⟨1⟩and M⟨2⟩, to realize the desired equivalent thresholds.

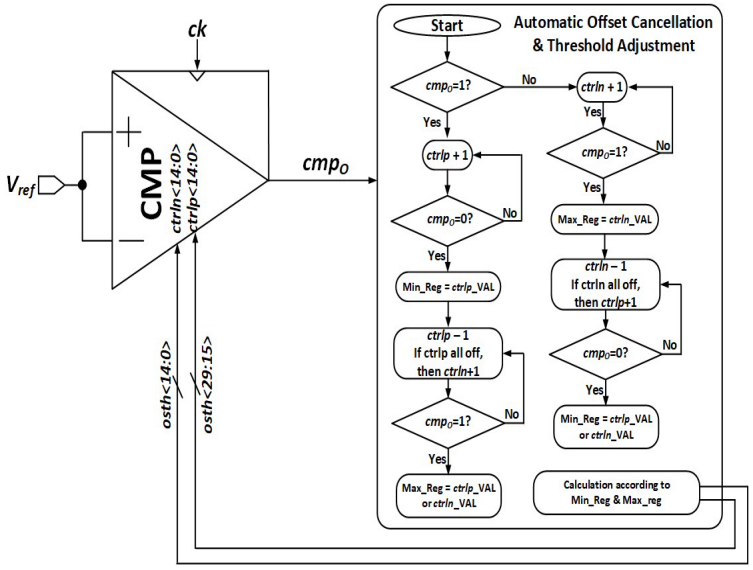


Fig. 4. The simplified flow chart of the automatic offset cancellation and threshold adjustment block with its feedback loop.

The simplified flow chart of the AOSTH block with its feedback loop is shown in Fig. 4. When the DLDO runs the offset detection and cancellation, the differential inputs of the comparator will be connected together to V_{ref}, the clock of the comparator is one sixteenth of the clock generated by the comparator triggered oscillator, and the two tuning NAND gate varactors are all off (*ctrlp*⟨14 : 0⟩ and

ctrln⟨14 : 0⟩ are all ”0”). At the beginning, the automatic process will detect cmp_O. If cmp_O is ”1”, it means that the parasitic capacitor is larger in the *ctrln* path than in the *ctrlp* path in the comparator. The algorithm will start turning on the NAND gates in the *ctrlp* path until cmp_O becomes ”0”, and the register will store the digits as the minimum dynamic offset control digits. This value is equivalent to V_{min} as shown in Fig. 3(c). Then, the NAND gates in the *ctrlp* path will be turned off one by one until cmp_O becomes ”1”. If cmp_O keep ”0” until all the *ctrlp* NAND gates turn off, the NAND gates in the *ctrln* path will turn on one by one until cmp_O becomes ”1”. So we achieve the maximum dynamic offset control digits, which is equivalent to V_{max} as shown in Fig. 3(c). Due to the opposite direction control of *ctrlp* and *ctrln*, we can assume *ctrlp* is a positive value and *ctrln* is a negative value for the convenient calculation. Then, the dynamic offset is the average of the maximum digits and the minimum digits. The offset cancellation digits are finally generated and sent back to the comparator. If cmp_O is ”0” at the beginning, the process is similar as the above description, as shown in Fig. 4.

After the inherent offset cancellation, the offset can be rearranged by *ctrlp*⟨14 : 0⟩ and *ctrln*⟨14 : 0⟩ as the equivalent threshold. The initial threshold setting codes will be given to the AOSTH by SPI block and combined with the calculated inherent offset codes, and then the final offset codes are generated to *ctrlp*⟨14 : 0⟩ and *ctrln*⟨14 : 0⟩. To speed up the sensing time of V_{out} droop, the equivalent threshold V_L will keep a small value when I_{load} has no change. And it will turn to be the largest value as soon as V_{out} droop is coming for better loop stability. This equivalent V_L will go back to the small value automatically after the adjustment of V_{out} has completed and the DLDO loop is back to be steady again. This adaptation is realized by changing the control codes back to the small offset after waiting 4 rising edges of cmp_f signal.

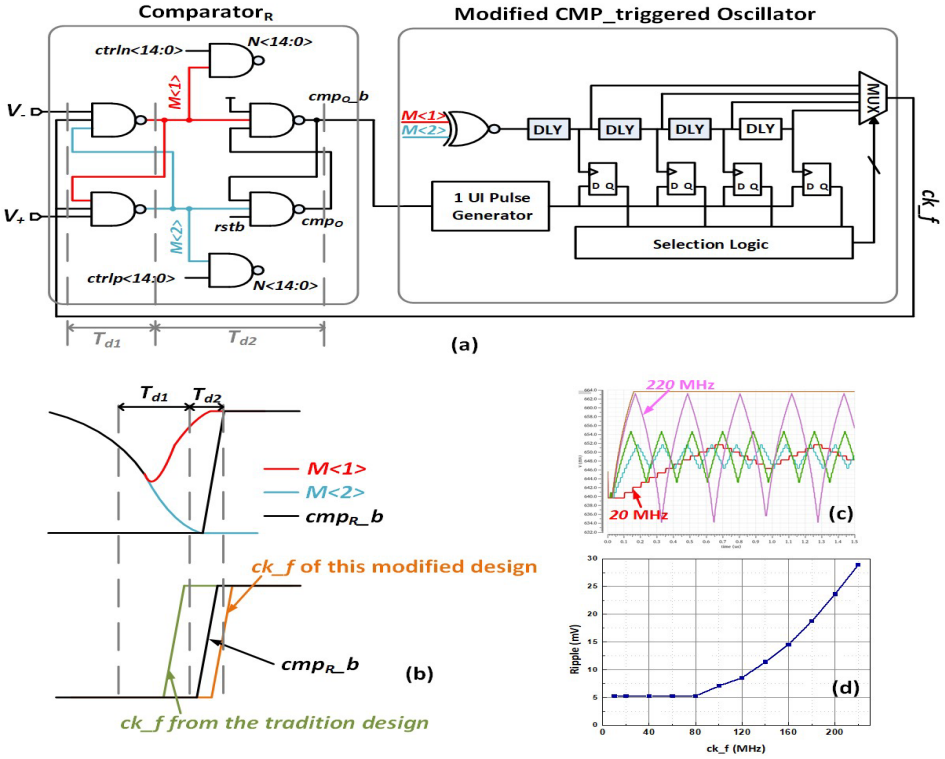


Fig. 5. (a)Implementation of the comparator and the modified CMP triggered oscillator; (b)transient response of the key comparator nodes; (c)simulated steady waveform of DLDO under different fine clock frequency; (d)steadystate ripple vs fine clock frequency.

8.2.2 Modified comparator triggered oscillator

In the output-capacitor-free DLDO, to improve the recovery time of the load current, the clock frequency of the comparator will be increased within the range where the comparators operate normally at the clock frequency, so the comparator triggered oscillator was proposed in [9]. However, only the variable input-stage delay of the comparators (T_{d1}) shown in Fig. 5(a) was considered in [9], which may make the loop fail when V_{out} is close to V_{ref} since the clock is too fast and no time margin for cmp_O to respond. Therefore, in [9], a slow fixed 10 MHz clock from off-chip was adopted for the fine loop, which scarifies the tracking speed to ensure the enough margin for T_{d2}. In the simulated steady case under different comparator clocks as shown in Fig. 5(c), the ripple of V_{out} will increase with the clock frequency increasing in the higher frequency range, because the time margin is not enough for cmp_O under the small difference between V_{ref} and V_{out} and it needs to decrease cmp_O generation time by increasing the difference between V_{ref} and V_{out} to get the steady status. In the lower clock frequency range, the ripple will keep the same value whatever the clock frequency is, since the time margin is enough for cmp_O, as shown in Fig. 5(d). Hence, the best clock frequency for the fine loop is the highest frequency that ensures the loop has minimum V_{out} ripple. In our modified oscillator, to avoid long-period oscillation and extra clock input, the delay time T_{d2} from the delay of the output stage is introduced to the oscillator directly. It will select the proper delay time right after cmp_O is built and ensure to generate the optimum clock frequency for the fastest fine loop response and minimum steady ripple.

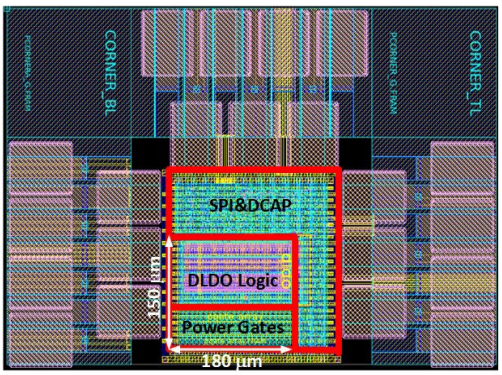


Fig. 6. The layout of the proposed DLDO regulator.

8.3 Simulation Results

The proposed DLDO has been implemented and simulated in the 65-nm CMOS process with the RAIL design flow. Fig. 6 shows the layout of the DLDO regulator. The core area, including the power gates and the DLDO logic, occupies 0.027 mm^2.

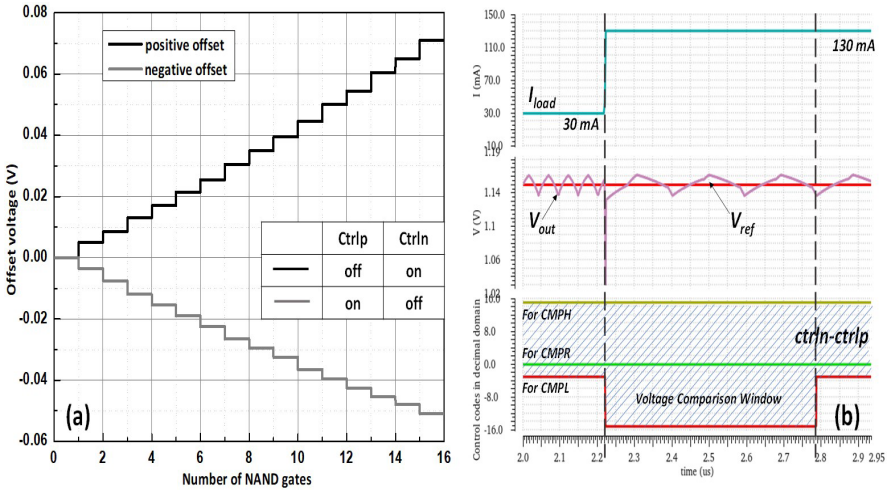


Fig. 7. (a) The controlled range of the comparator offset; (b) load transient response and its corresponding adaptive comparison window.

Fig. 7(a) shows the controllable offset range of the comparator. This comparator can provide the maximum positive offset of 71 *mV* while all *ctrln* bits are ”1” and all *ctrlp* bits are ”0”; and the maximum negative offset of -51 *mV* while all *ctrln* bits are ”0” and all *ctrlp* bits are ”1”. The average step is 4 *mV* , which means the maximum steady-state error is 4 *mV* . According to the Monte-Carlo simulation result shown in Fig. 3(b), this controlled range can cover around 3-sigma inherent input-referred offset of the comparator and provide the maximum equivalent threshold voltage of 71 mV . Fig. 7(b) shows the load transient response with V_{in} of 1.2 *V* while load current step is 100 *mA* with 3 *ns* edge rise-time. The decimal domain numbers of the adaptive offset control bits for the three comparators are also given in Fig. 7(b). We can see the control bits for the comparator with equivalent low threshold keeps a small value while the loop is steady and becomes the largest value as soon as I_{load} is changing for the best stability during the voltage adjustment period. And then, the offset control codes go back to the small value at the new steady status.

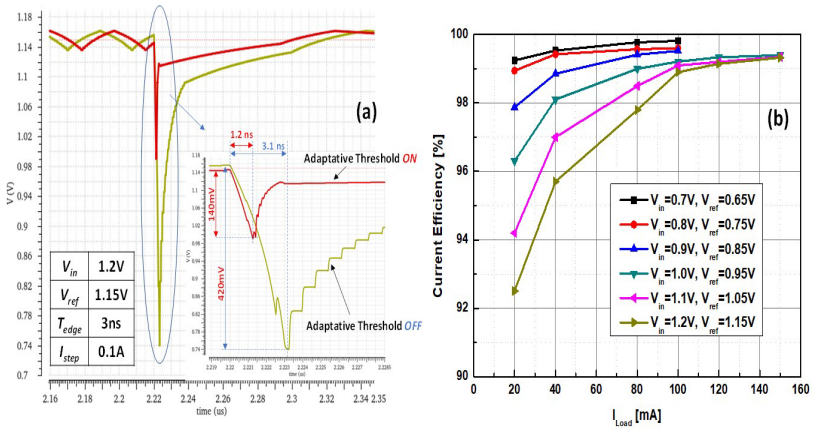


Fig. 8. (a) Simulated load transient response with V_{in} of 1.2V while load current step is 100mA with 3ns edge rise-time; (b) current efficiency while V_{in} is between 0.7 to 1.2V.

The load transient response is shown in Fig. 8(a). When the current load is increased by 100mA with a 3ns edge rise-time, the DLDO exhibits the voltage droop of 140mV and 420mV with the threshold adaptive scheme "ON" and "OFF", respectively. As shown in Fig. 8(b), the peak current efficiency is 99.9% with V_{in} of 0.7V. Table I is the comparison of our work and the state-of-the-art DLDO regulators. Comparing with the previous work in that targets the similar load current range in the mA, the proposed DLDO achieves the steady state error of 4 *mV*, the fastest response time of 1.2ns and the minimum chip area of 0.027mm^2.

8.4 Summary

This chapter presented a fully synthesizable DLDO regulator that achieves the least requirements for analog voltages and blocks for further integration improvement. The automatic offset control and reuse technique ensures a small steady-state error and adaptive comparison window threshold adjustment for the faster time response without extra droop detector (saving area). The simulation results prove that the proposed DLDO regulator saves chip area and improve the response time of the voltage droop, and the fully synthesizable feature makes this design to have better process scalability and higher degree integration.

8.5 References

[1] S. T. Kim *et al.*, "Enabling wide autonomous DVFS in a 22nm graphics execution core using a digitally controlled hybrid LDO/switched capacitor VR with fast droop mitigation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, 2015, pp. 1-3.

[2] Y. Lu, Y. Wang, Q. Pan, W. Ki and C. P. Yue, "A fully-integrated low-dropout regulator with full-spectrum power supply rejection," *IEEE Trans. on Circuits and Systems I: Reg. Papers*, vol. 62, no. 3, pp. 707-716, March 2015.

[3] S. Kundu, M. Liu, S. Wen, R. Wong and C. H. Kim, "A fully integrated digital LDO with built-in adaptive sampling and active voltage positioning using a beat-frequency quantizer," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 1, pp. 109-120, Jan. 2019.

[4] Y. Lu, F. Yang, F. Chen and P. K. T. Mok, "A 500mA analog assisted digital-LDO-based on-chip distributed power delivery grid with cooperative regulation and IR-drop reduction in 65nm CMOS," in *IEEE International Solid - State Circuits Conference (ISSCC)*, San Francisco, CA, 2018, pp. 310-312.

[5] X. Liu et al., "A modular hybrid LDO with fast load-transient response and programmable PSRR in 14nm CMOS featuring dynamic clamp tuning and time-constant compensation," in *IEEE International Solid- State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2019, pp. 234-236.

[6] M. A. Akram, W. Hong and I. Hwang, "Capacitorless self-clocked all digital low-dropout regulator," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 1, pp. 266-276, Jan. 2019.

[7] X. Sun, A. Boora, W. Zhang, V. R. Pamula and V. Sathe, "A 0.6-to-1.1V computationally regulated digital LDO with 2.79-cycle mean settling time and autonomous runtime gain tracking in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2019, pp. 230-232.

[8] K. Z. Ahmed et al., "A variation-adaptive integrated computational digital LDO in 22-nm CMOS with fast transient response," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 4, pp. 977-987, April 2020.

[9] J. Oh, J. Park, Y. Hwang and D. Jeong, "A 480mA output-capacitor-free synthesizable digital LDO using CMP-triggered oscillator and droop detector with 99.99% current efficiency, 1.3ns response time, and 9.8A/mm^2

current density," in *IEEE International Solid- State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2020, pp. 382384.

[10] M. Choi, C. Kye, J. Oh, M. Choo and D. Jeong, "A synthesizable digital AOT 4-phase buck voltage regulator for digital systems with 0.0054mm^2 controller and 80ns recovery time," in *IEEE International Solid- State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2019, pp. 432-434.

[11] S. Li and B. H. Calhoun, "A 745pA hybrid asynchronous binary searching and synchronous linear-searching digital LDO with 3.8×105 dynamic load range, 99.99% current efficiency, and 2mV output voltage ripple," *IEEE International Solid- State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2019, pp. 232-234.

[12] M. Huang, Y. Lu, S. U and R. P. Martins, "An analog-assisted tri-loop digital low-dropout regulator," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 20-34, Jan. 2018.

9 DESIGN AND IMPLEMENTATION OF LPDDR4X, UCIE, AND NOC PHY AND MEMORY CONTROLLER

LPDDR4x DRAM [1] has been widely used in modern electronic systems, especially portable devices such as Apples' M1 chip due to its higher speed and lower energy consumption than the previous generations of DRAM chips. It is often a must-have block required for open-source hardware innovation. Therefore, the basic program objective, as set at the start of this project, is to develop a agile synthesis, implementation and verification methodology and RAIL flows for LPDDR4X PHY design. To enable the demonstration and verification of LPDDR4X PHY, LPDDR4X memory controller has been developed, based on open source platforms [3,4]. We note that a complete LPDDR4 design requires the integration of SAR-ADC, PVT sensors, PLL/DLL and LDO IP blocks developed in this project, also exercises the SPICE-to-SystemVerilog abstraction flow for system-level design verification, a key innovation in enabling an agile mixed-signal design flow.

During the execution of this project, in February 2022, the standard on Universal Chiplet Interconnect express (UCIe) [2] was established for high-speed die-to-die short-reach interconnect. Steaming protocols like PCIe 6.0 and CXL2.0 are defined on top of UCIe's electrical and die-level specification.

Once we studied the UCIe standard, we observe that at the electrical level, it shares a lot of similarity with LPDDR4X. We therefore developed a **software-programable PHY** to support both LPDDR4X and UCIe 1.0 standards. Furthermore, the similarity and difference allow us to propose the Turbo modes to both UCIe and LPDDR4X standards and thus greatly enhances their performances beyond what are in the standards.

A LPDDR4x and UCIe combo PHY has been designed and simulated in the TSMC 28nm Plus process. In addition to standard NRZ signaling in both the LPDDR4x and UCIE standards, PAM4 signaling is also supported to extend its data width, from 16 bit to maximum of 96 bit, with proper pin assignment. 4.5Gbps data rate can be achieved with 0.86pJ/pin power efficiency.

Network-on-Chip (NoC) is also an emerging technology to improve the performance of conventional bus architecture in large systems. A typical NoC fabric can not only achieve higher scalability but also ease electrical constraints (timing closure, wire routing congestion, etc.) during designing procedures. In this work, we implement the NoC over the UCIe interface.

9.1 Introduction

The combo PHY transceiver, designed in the TSMC 28nm HPC Plus 1P10M process, can support multiple standards, including LPDDR4x [1], UCIE v1.0 [2], and the proprietary turbo-UCIE mode. To save power consumption, the voltage-mode instead of current-mode Line Drivers (LD) is adopted. Both Off-Chip-Driver (OCD) and On-Die-Termination (ODT) [5] are implemented and hence no external resistors are needed. To cope with Process-Voltage-Process (PVT) variation, the ZQ calibration circuit is used in the design [6]. With the help of one external precise 240Ω resistor, the on-chip termination resistance can be guaranteed. Signal modulation scheme can be programmable to be Non-Return-to-Zero (NRZ) or Pulse-Amplitude-Modulation 4 (PAM4) [7][8]. Termination characteristic impedance (Zo) is also programmable,

ranging from 30Ω to 240Ω [9], or even can be set to be un-termination. By the training procedure defined by the LPDDR4x and UCIe standards, optimal data sampling phase and comparison reference voltage can be decided [10]. A 6-bit complementary Phase Interpolator (PI) is designed [11] for sampling phase searching. The duty cycle of 50% can be assured for the interpolated output clock. Besides, a reference generator circuit, equipped with a sub-1V bandgap reference circuit and an R-2R DAC, can provide the reference voltage to the data slicer.

This chapter describes key technologies and details of the proposed mixed controller and PHY design logistics, considerations, constraints, and dependencies. This chapter is organized as follows. Section 8.2 describes the system architecture. The memory controller and UCIe-NoC adaptor are described in Section 8.3. Section 8.4 presents the overall specification and architecture of the combo PHY. Then, the circuit implementation is presented in Section 8.5. The link simulation results are described in Section 8.6. Finally, the conclusion is drawn in Section 8.7.

9.2 System Architecture

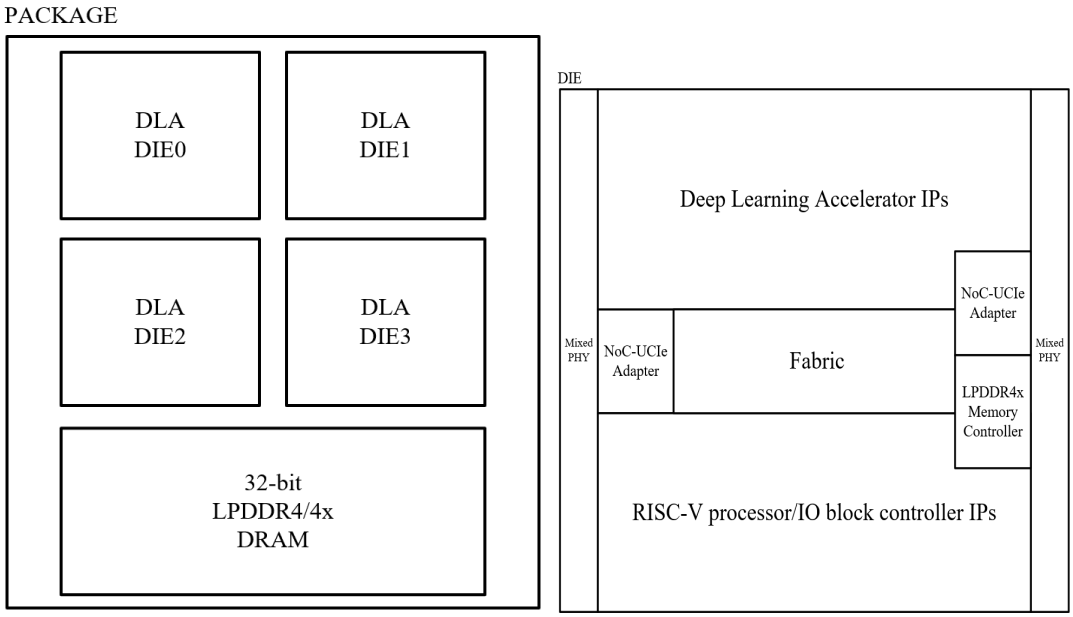


Figure 1. An example of package (left) & die (right) architecture illustration.

Figure 1 shows an example architecture of the package-level and the die-level system applications of LPDDR4x/UCIe. Inside one package, four (4) Deep Learning Accelerator (DLA) dies communicate with each other through UCIe and access to one LPDDR4/4x DRAM die using the advanced fan-out packaging technology. To support the qualification of the UCIe performances and link adaptation, the DLA die bonding microbumps are carefully designed to support three packaging styles (wire bonding, fan-out and flip-chips).

Inside a single die, the DLA, the RISC-V processor, and other IO IPs share the same LPDDR4X DRAM memory space through the Fabric. The mixed-PHY on the west side of the die is configured as the UCIe mode thus transferring data between packages. The one on the east side of the die is connected to both

the NoC-UCIe adapter and LPDDR4x Memory Controller. It can be software-configured to the UCIe mode to realize Network-in-Package or the LPDDR mode to achieve high-speed memory access.

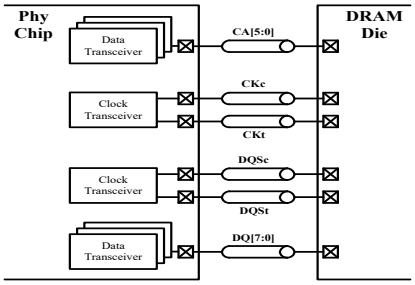


Figure 2. Simplified channel links between the PHY die and the DRAM die.

Simplified channel connection between the PHY die and DRAM die is shown in Figure 2. Command data and address are transmitted via CA channels, with its associated differential clock channels, CKc and CKt. Real data is transmitted via DQ channels, with its associated differential clock channels, DQSc and DQSt. Our SOC chip contains two Phy circuits, one can be programmed to be LPDDR4x link, and the other to be UCIE link. Four SOC chips and a DRAM die are packaged into one Unit, as illustrated in Figure 3.

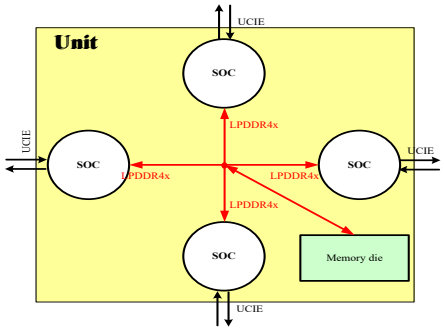


Figure 3. Dies packaged into one unit.

The combo PHY can be programmed in the LPDDR4x hal- duplex mode as in Figure 4(a), and in the UCIE full-duplex mode as shown in Figure 4(b).

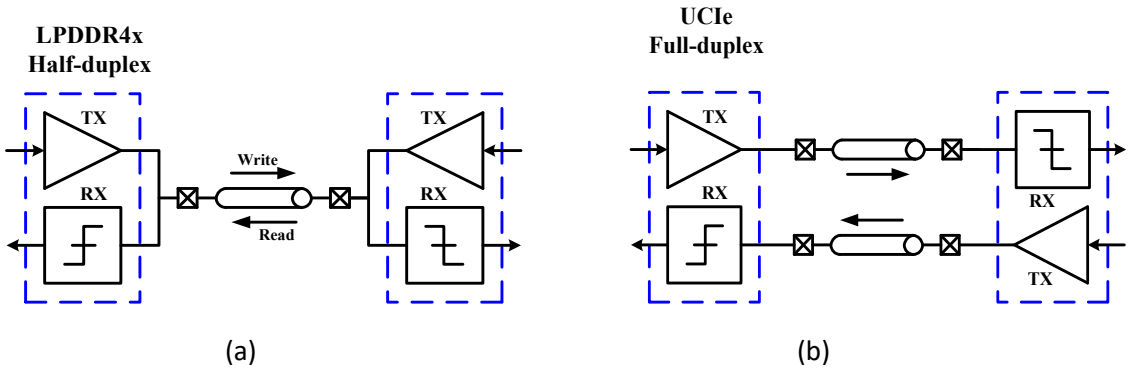


Figure 4. Transceiver (a) in the LPDDR4x mode, and (b) in the UCIE mode.

9.2.1 LPDDR4X Mode

As an example, four ScC dies can have a shared access to one LPDDR4x die, where the mixed PHY can be configured in the LPDDR4x mode.

a. Write : The SOC chip who wants to write data to DRAM operate in standard OCD mode, while the other three chips stay in ***"high-Z"*** mode to maintain correct bus termination impedance.

b. Read : Support four SOC chips ***"simultaneous"*** read. Program all four SOC chips ODT to 240Ω, and DRAM chip ODT to 60Ω.

c. Idle : Program (any) one SOC chips in ***"freeze"*** mode, i.e., 240Ω weakly pull down, and keep the other three chips in high-Z mode. Thus to prevent bus floating.

Table 1. LPDDR4X/UCIE Interface Pin Table.

Pin Name	UCIe	UCIe-turbo	LPDDR4	Width	Description
CA[11:0]		tx/rxdata[21:16]	CA	12	Command/Address Bus, tx/rxdata in UCIe-turbo mode
CS[1:0]	txdatasb/rxdatasb	txdatasb/rxdatasb	CS	2	Chip-select, tx/rx sideband data in UCIe mode
CKE[1:0]	txcksb/rxcksb	txcksb/rxcksb	CKE	2	DRAM clock enable, tx/rxdata sideband clock in UCIe mode
CK_p[1:0]	txvld/rxvld	txvld/rxvld	CK_p	2	Clock_p, txvld/rxvld in UCIe mode
CK_n[1:0]			CK_n	2	Clock_n
DQ[15:0]	txdata[15:0]	txdata[15:0]	DQ[15:0]	16	Data inout in LPDDR4 mode, txdata in UCIe mode
DQ[31:16]	rxdata[15:0]	rxdata[15:0]	DQ[31:16]	16	Data inout in LPDDR4 mode, rxdata in UCIe mode
DMI[0],DMI[2]		tx/rxdata[22]	DMI[0],DMI[2]	2	DMI inout in LPDDR4 mode, tx/rxdata in UCIe-turbo mode
DMI[1],DMI[3]		tx/rxdata[23]	DMI[1],DMI[3]	2	DMI inout in LPDDR4 mode, tx/rxdata in UCIe-turbo mode
DQS0_n	txckn	txckn	ch0_dqs0_n	1	DQS_n in LPDDR4 mode, txckn in UCIe mode
DQS0_p	txckp	txckp	ch0_dqs0_p	1	DQS_p in LPDDR4 mode, txckp in UCIe mode
DQS2_n	rxckn	rxckn	ch1_dqs0_n	1	DQS_n in LPDDR4 mode, rxckn in UCIe mode
DQS2_p	rxckp	rxckp	ch1_dqs0_p	1	DQS_p in LPDDR4 mode, rxckp in UCIe mode
DQS1_p	txtrk	txtrk	ch0_dqs1_p	1	DQS_p in LPDDR4 mode, TX track in UCIe mode
DQS1_n			ch0_dqs1_n	1	DQS_n for LPDDR4
DQS3_p	rxtrk	rxtrk	ch1_dqs1_p	1	DQS_p in LPDDR4 mode, TX track in UCIe mode
DQS3_n			ch1_dqs1_n	1	DQS_n for LPDDR4
ZQ			ZQ	1	ZQ for LPDDR4
RESET_n			RESET_n	1	RESET_n for LPDDR4

9.2.2 UCIE Mode

a. Standard UCIE mode : 16 bit data width.

b. Extended UCIE mode : 32 bit data width.

- program transceiver from full-duplex mode to half-duplex mode.

c. Turbo UCIE mode : 48 bit data width.

- Reuse CA & DMI pins as UCIE data pins. Pin table summary is illustrated in Table 1.

d. PAM4 UCIE mode: 96 bit data width.

- PAM4 instead of NRZ LD, transmitted 2b per pin.

9.3 LPDDR4X Memory Controller and UCIE-NoC Adaptor

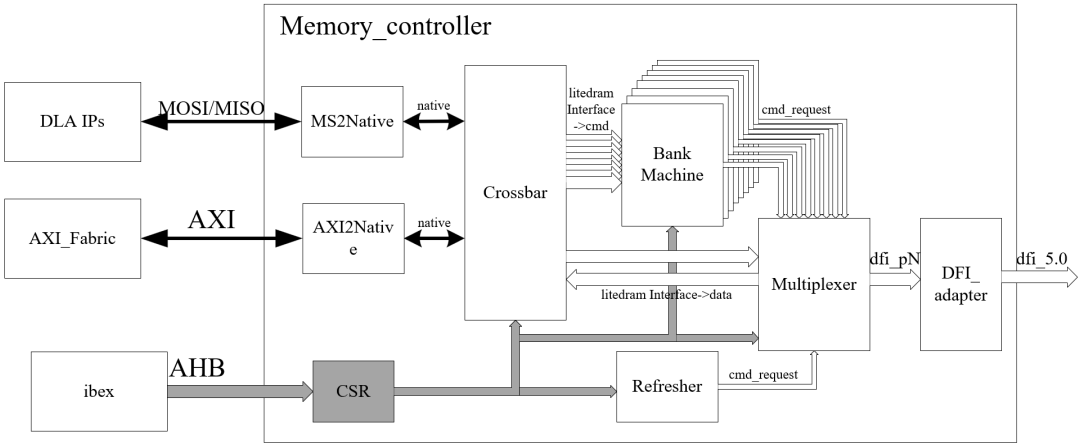


Figure 5. LPDDR4/4x Memory Controller Diagram.

Figure 5 shows the top-level diagram of the designed memory controller. This memory controller originated from the LiteDRAM [3] python source code, which aims to run on FPGAs. The very first version of Verilog code was generated by MiGen generator inside the LiteDRAM project. Based on the logistics provided by the original project, we further developed this memory controller for ASIC design flow and parameterized several parameters by adding software-controlled control-status registers (CSRs) related to latency, timing constraints, and different configurations.

There are two master ports provided. The LA uses 256-bit MOSI/MISO protocol to access the LPDDR DRAM. The ideal maximum throughput of LPDDR DRAM is 256 bit/cycle. The other AXI slave port is provided to connect with the AXI bus. All other IPs connected to the AXI bus can access the LPDDR4X

DRAM through this port. This memory controller has been verified comprehensively by the UVM testbenches.

As illustrated in Figure 6, the NoC-UCIe adapter transfers the original NoC protocol to the standard Raw Die-to-Die Interface (RDI). It applies the credit system to realize data flow control between dies. The nearside credit information is transferred along with data to the far side to inform the FIFO status. The effective maximum throughput is also 256 bit/cycle if the flit header bits are ignored.

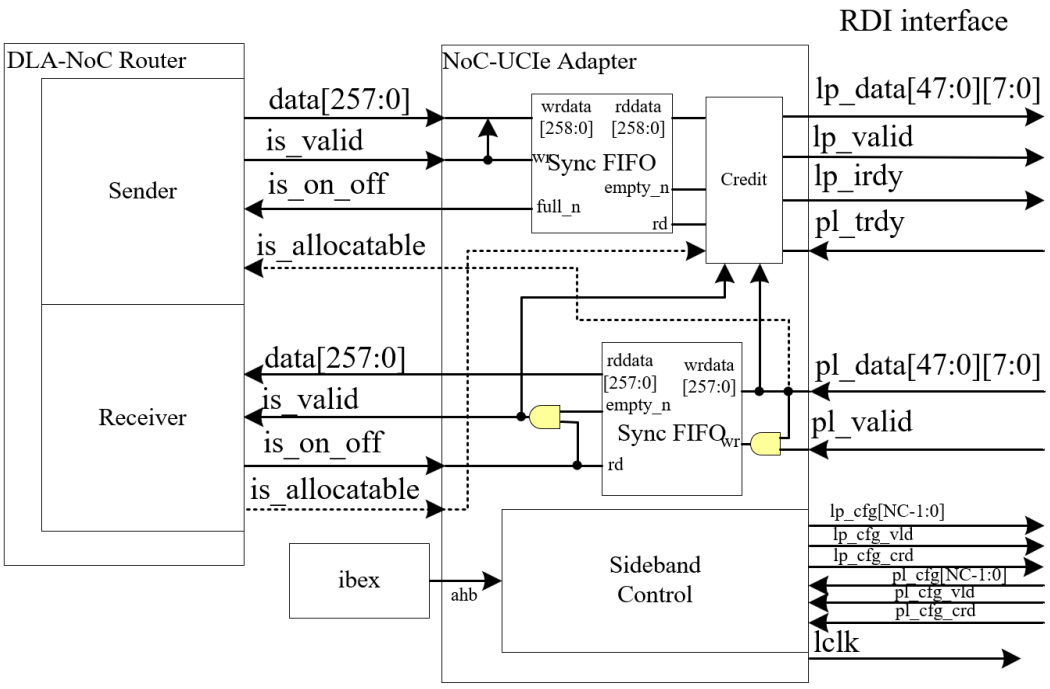


Figure 6. The NoC-UCIe adapter diagram.

9.4 PHY Architecture

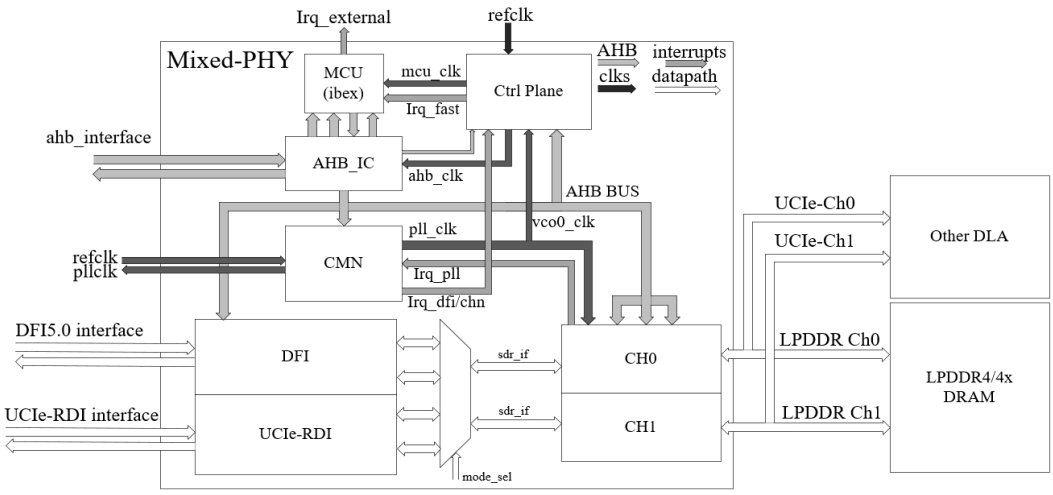


Figure 7. The mixed LPDDR4X/UCIe PHY architecture.

The PHY, as shown in Figure 7, is modified from the Wavious open-source WDDR PHY[4]. All digital and analog hardware configurations are controlled through AHB bus by a small open-source RSIC-V ibex MCU. We further extended the capability of this PHY to support UCIe electrical signals and PAM4 transmission. The memory controller, along with the UCIe-RDI adapter can be connected to the corresponding interface outside the DFI or UCIe-RDI block. A mode-select signal controls which part of data is transferred through channels.

9.4.1 PHY IO Blocks

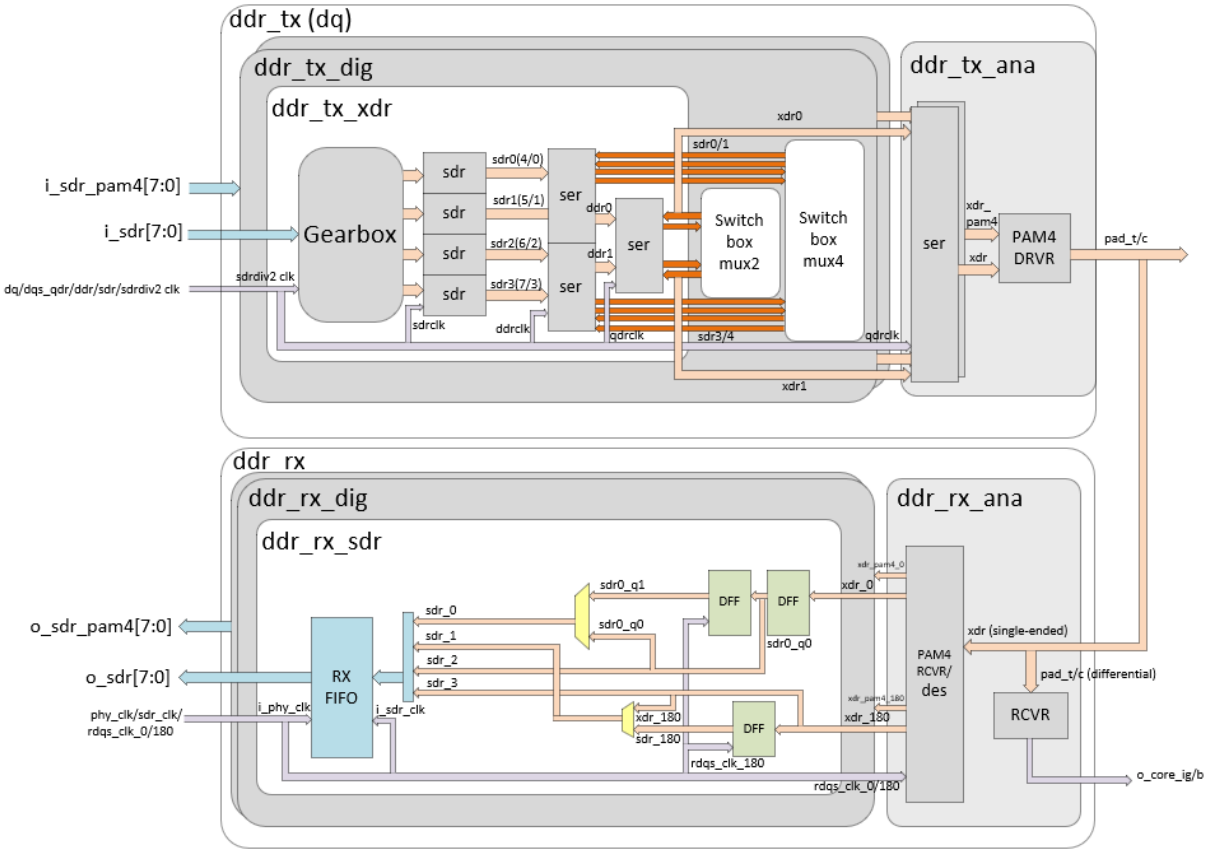


Figure 8. PHY IO block diagram.

In Figure 8, the basic transceiver and receiver diagram is shown. In the LPDDR mode, PAM4 mode is disabled, the IO block implements a 1:4 frequency ratio system which means the clock frequency on the pad side is 4 times that of the system side. The target system frequency is 533MHz. So, the LPDDR DRAM is running at 2133MHz. The maximum data bandwidth of 32-bit LPDDR is $2133MHz \times 32\ bit \times 2(DDR) = 133\ Gbps$.

The UCIe standard mode utilizes LPDDR DQ pins as data pins so that it has the same throughput with the LPDDR mode. However, the UCIe turbo mode tries to utilizes remaining 12 CA pins as data lanes to further boost the bandwidth. Moreover, the PAM4 mode can also be enabled to double the bandwidth. Under the UCIe turbo mode, the maximum bandwidth is $2133MHz \times 48\ bit \times 2(DDR) \times 2(PAM4) = 400\ Gbps$.

9.4.2 Channel Blocks and Pin Maps

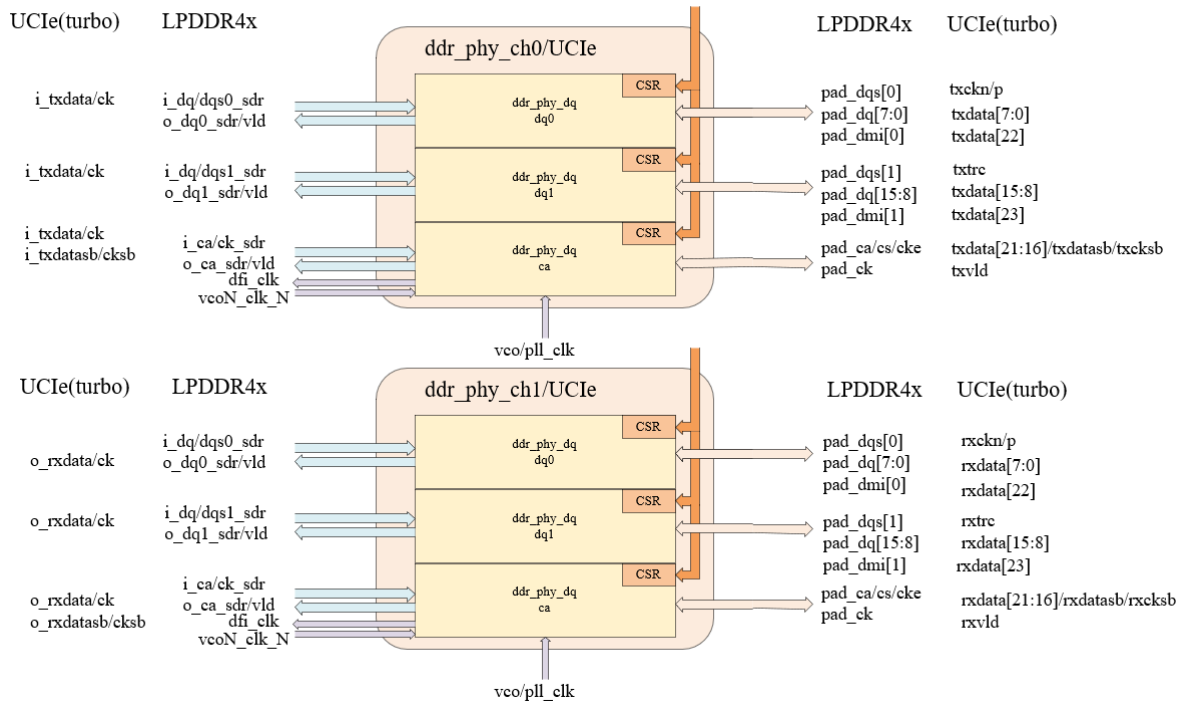


Figure 9. PHY channel block diagram.

Figure 9 shows the channel arrangement for both LPDDR mode and UCIe mode. In UCIe mode, each channel can be configured either as UCIe-TX mode or UCIe-RX mode. In total, the two channels can be configured as 3 modes–24 lanes TX+24 lanes RX, 48 lanes TX, and 48 lanes RX, which improves flexibility. The complete pin map is shown in Table 2.

Table 2. UCIe/LPDDR Pin Map.

Pin Name	UCIe	UCIe-turbo	LPDDR4	Width	Description
CA[11:0]		tx/rxdata[21:16]	CA	12	Command/Address Bus, tx/rxdata in UCIe-turbo mode
CS[1:0]	txdatasb/rxdatasb	txdatasb/rxdatasb	CS	2	Chip-select, tx/rx sideband data in UCIe mode
CKE[1:0]	txcksb/rxcksb	txcksb/rxcksb	CKE	2	DRAM clock enable, tx/rxdata sideband clock in UCIe mode
CK_p[1:0]	txvld/rxvld	txvld/rxvld	CK_p	2	Clock_p,tx/rxvld signal in UCIe mode

CK_n[1:0]			CK_n	2	Clock_n
DQ[15:0]	txdata[15:0]	txdata[15:0]	DQ[15:0]	16	Data inout in LPDDR4 mode, txdata in UCIe mode
DQ[31:16]	rxdata[15:0]	rxdata[15:0]	DQ[31:16]	16	Data inout in LPDDR4 mode, rxdata in UCIe mode
DMI[0],DMI[2]		tx/rxdata[22]	DMI[0],DMI[2]	2	DMI inout for LPDDR4, tx/rxdata[22] in UCIe-turbo mode
DMI[1],DMI[3]		tx/rxdata[23]	DMI[1],DMI[3]	2	DMI inout for LPDDR4, tx/rxdata[23] in UCIe-turbo mode
DQS0_n	txckn	txckn	ch0_dqs0_n	1	DQS_n in LPDDR4 mode, txckn in UCIe mode
DQS0_p	txckp	txckp	ch0_dqs0_p	1	DQS_p in LPDDR4 mode, txckp in UCIe mode
DQS2_n	rxckn	rxckn	ch1_dqs0_n	1	DQS_n in LPDDR4 mode, rxckn in UCIe mode
DQS2_p	rxckp	rxckp	ch1_dqs0_p	1	DQS_p in LPDDR4 mode, rxckp in UCIe mode
DQS1_p	txtrk	txtrk	ch0_dqs1_p	1	DQS_p in LPDDR4 mode, TX track in UCIe mode
DQS1_n			ch0_dqs1_n	1	DQS_n for LPDDR4
DQS3_p	rxtrk	rxtrk	ch1_dqs1_p	1	DQS_p in LPDDR4 mode, TX track in UCIe mode
DQS3_n			ch1_dqs1_n	1	DQS_n for LPDDR4
ZQ			ZQ	1	ZQ for LPDDR4
RESET_n			RESET_n	1	RESET_n for LPDDR4

9.4.3 DFI/RDI Blocks and Firmware

Previous paragraphs demonstrate the basic data path of this PHY. In this paragraph, higher-level abstractions are illustrated as well as basic operations of PHY initialization, training and calibration

through firmware. Figure 10 shows the hardware/software interaction inside DFI/RDI blocks. During training, the MCU can take full control of the channel, sending and receiving data to the far side. Initialization and training are conducted within this period. The ibex will automatically perform training algorithms by controlling Control Registers and receiving feedback data from Status Registers over blocks.

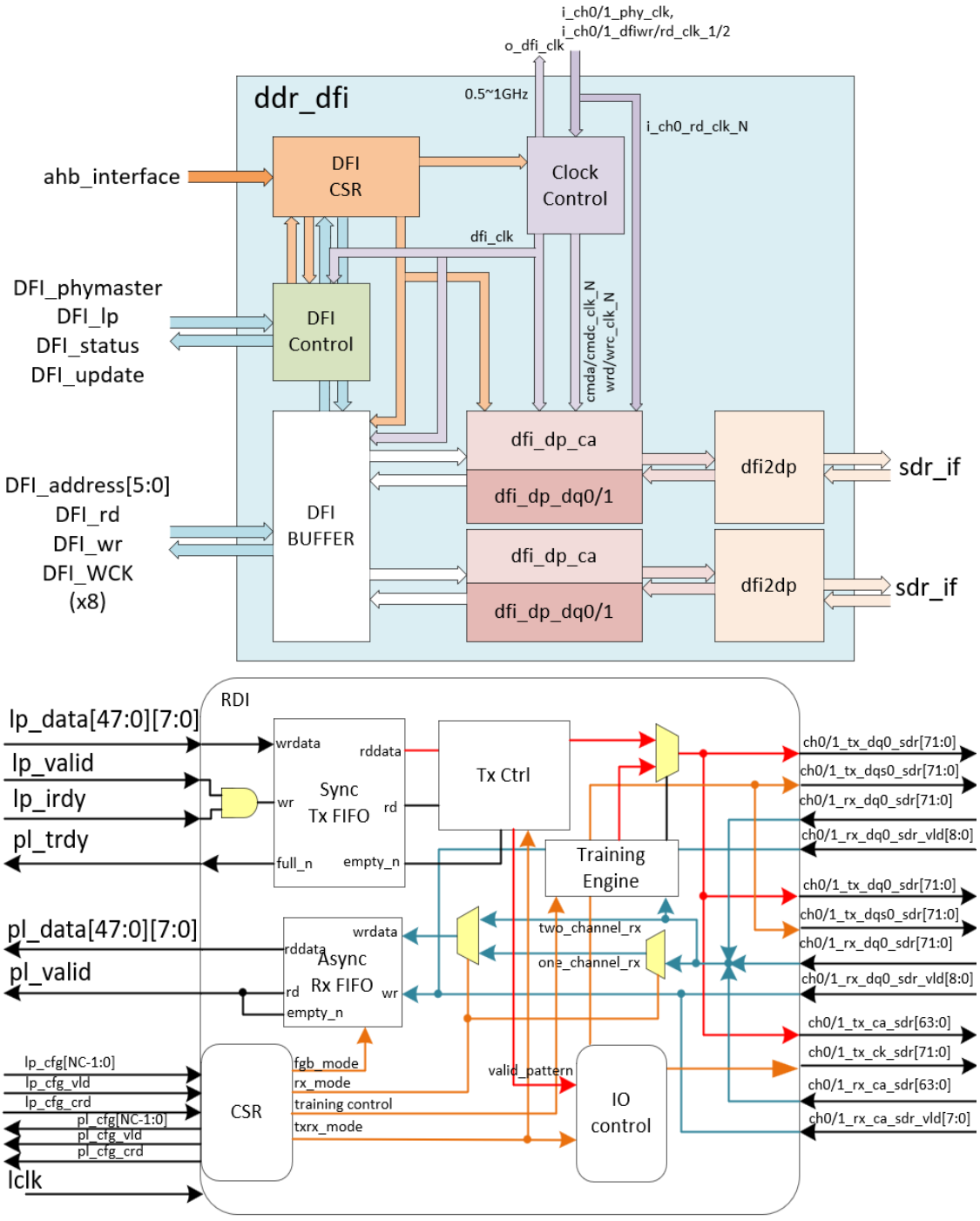


Figure 10. DFI (1) & RDI (2) blocks diagr

9.5 PHY Circuit Implementation

Figure 11 illustrates the simplified block diagram of both data (DQ) and clock (DQS) transceiver. Besides line driver and receiver frontend, phase interpolator, aims for global deskew, and programmable delay, aiming for per-bit deskew, are also included.

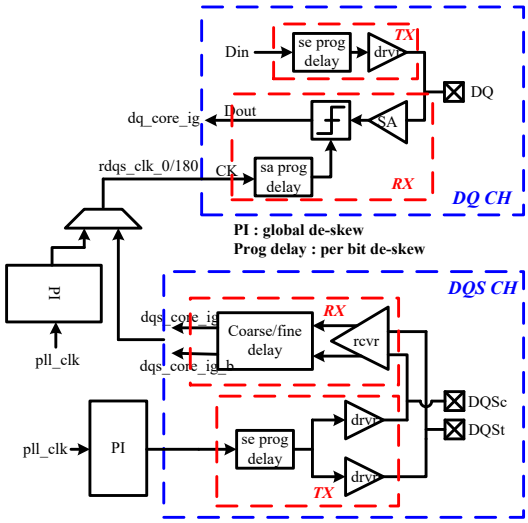


Figure 11. Block diagram of the data and clock transceiver.

9.5.1 OCD/ODT Leg

Termination impedance can be implemented with the help of a resistor and a MOSFET, as shown in Figure 12, called a “slice”. The MOSFET should operate in the triode region to keep its rds (on) to be linear. An OCD comprises both pull-up leg and pull-down leg, while pull-down leg also serves as ODT (with pull-up leg disable). To cover induced resistance variation, number of legs used should be programmable.

A ZQ calibration circuit discussed in next section will decide how many legs to be used to keep both pull-up and pull-down legs to be 240Ω.

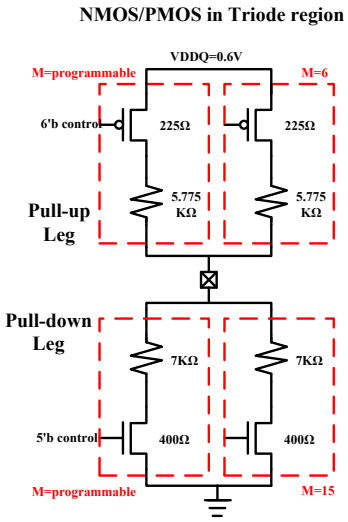


Figure 12. Pull-up and pull-down leg.

9.5.2 ZQ Calibration

Figure 13 shows the ZQ calibration circuit. A pull-down leg is connected to an external resistor with precise 240Ω. The pad voltage is compared with VDDQ/2, and tune 5b control code ncal[4:0] accordingly where ncal[4:0] code controls turn-on pull-down leg number. Final ncal[4:0] code will be set such that pull-down leg resistance equals 240Ω. Then the same ncal[4:0] code applies to a replica pull-down leg, co-work with a pull-up leg, to decide pull-up control code pcal[5:0], and hence pull-up leg resistance (240Ω). Simulation result shows that, in the TT corner, 17 pull-down legs and 19 pull-up legs should be turned on to get 240Ω. Figure 14 proves the effectiveness of the ZQ calibration process. The final codes can be set to 17 (ncal) and 44 pcal).

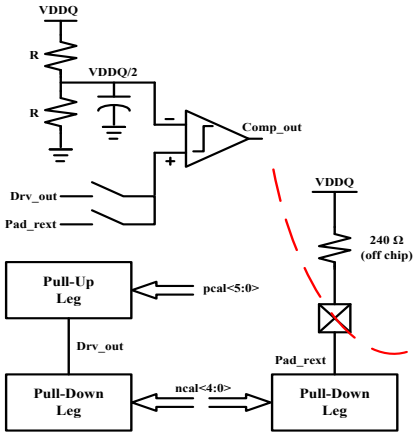


Figure 13. The ZQ calibration circuit.

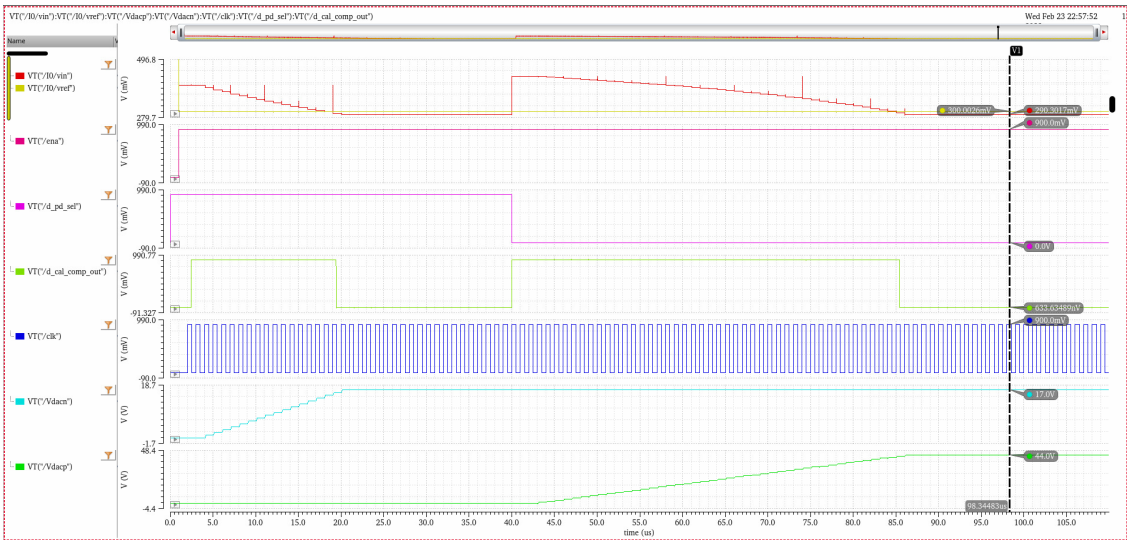


Figure 14. ZQ calibration circuit simulation result.

9.5.3 De-Emphasis & PAM4 LD

A line driver consists of eight slices. One to six slices can be turned on according to the LPDDR4x specification, to achieve 240/120/80/60/48/40Ω link impedance. However, all eight slices should be turned on for the UCIE mode, to meet the 30Ω impedance specification. In addition to that, by manipulating the turn-on slice number of the ODT, both de-emphasis (-2.5dB) and PAM4 features, depicted in Figure 15 and Figure 16 separately, can be achieved.

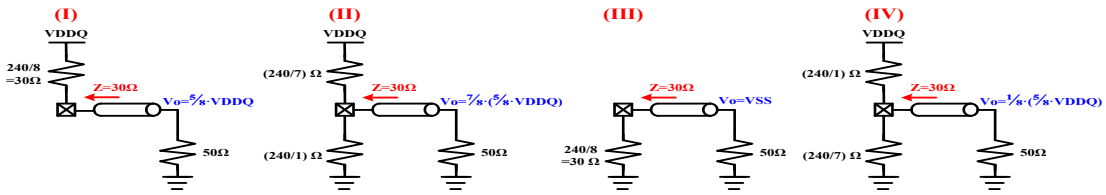


Figure 77. Transmitter de-emphasis waveform

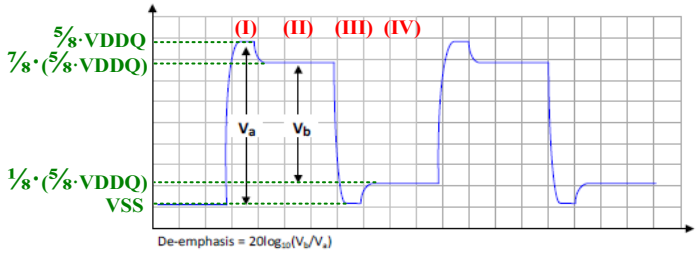


Table 31. Transmitter de-emphasis values

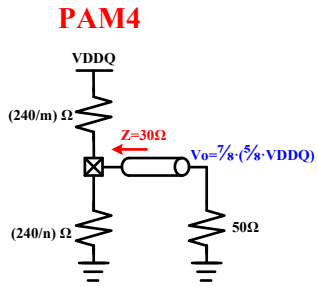
Setting	De-emphasis	Accuracy	C_{+1}	V_b/V_a
1	0.0 dB	-		
2	-2.2 dB	+/- 0.5 dB		

$V_a = \frac{5}{8} \cdot VDDQ$

$V_b = \frac{3}{4} \cdot (\frac{5}{8} \cdot VDDQ)$

De-emphasis amount $= 20\log(\frac{3}{4}) = -2.5$dB

Figure 15. Preemphasis ODT.



1: Vo=VDDQ/[1+(n+24/5)/m] = k*VDDQ
2: m+n=8 (to keep Z=30Ω)

m	n	k	Vo(mV)
7	1	35/64	437.5
5	3	25/64	312.5
3	5	15/64	187.5
1	7	5/64	62.5
Note :	VDDQ=0.8		
	ΔVo=0.125		

Figure 16. PAM4 ODT.

9.5.4 DQS RX

Both the ac path and the dc path are implemented in the DQS receiver frontend, as illustrated in Figure 17. An ac couple capacitance, ~0.5pF, followed by a self-biased inverter based slicer, can convert the LVSTL signal into the CMOS level clock, while preserving a nearly 50% duty cycle. However, a passive high pass filter is formed by this C, thus, this ac path should be enabled only for the high data rate case, e.g., >1Gbps. For the low data rate, the dc path can be adopted with the drawback of PVT-dependent clock duty cycle distortion.

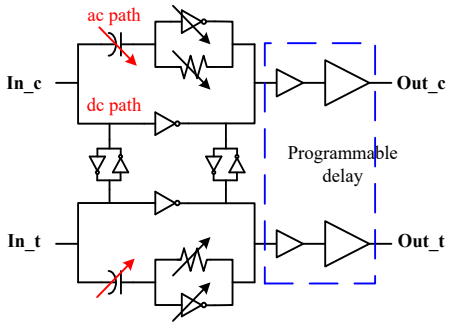


Figure 17. DQS clock channel receiver frontend.

9.5.5 DQ RX

The analog frontend of data path (DQ) receiver is shown in Figure 18. To support PAM4 signaling, three data paths are included in the DQ receiver. A data path comprises a Continuous Time Linear Equalizer (CTLE) [10][11][12], followed by a dynamic comparator serving as the data slicer. To obey the UCIe specification, four equalized levels, i.e., 0/-1/-2/-3dB at Nyquist frequency (2133MHz), can be chosen. Three reference voltages, for example, 125/250/375mV, are provided by the voltage reference circuit. Input offset of comparators should be calibrated before the link starts. A ±56mV calibration range is provided to cover the 3σ variation. A programmable delay cell provides the sampling clock, in which

the optimal sampling phase should be determined in the training sequence. Note that Dout[1] is the default output if NRZ instead of PAM4 signaling is adopted.

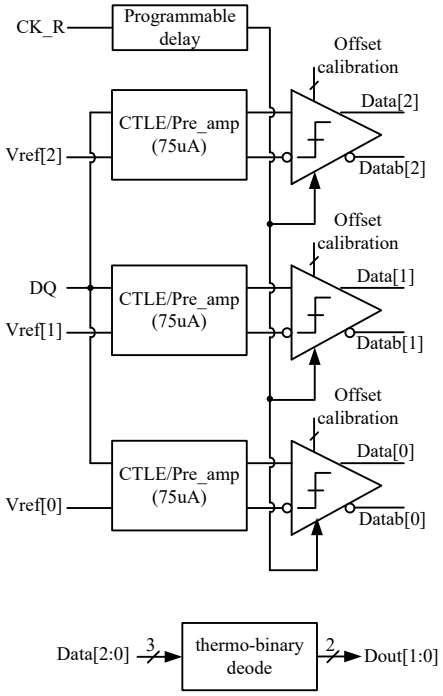


Figure 18. DQ data channel receiver frontend.

9.5.6 Vref Generation

Since both LPDDR4x and UCIE are based on the single-ended transmission scheme, a reference voltage is required for the data slicer. As depicted in Figure 19, the vref generator contains three main blocks. The first one is the sub-1V bandgap reference circuit, which outputs a ~512mV voltage. Followed by two 8b R-2R DACs and a non-inverting amplifier, a reference voltage, ranging from 50mV to 560mV with a step of 2mV, can be provided. Then, a resistor divider and three Unity-Gain-Buffers (UGB) pass Vref[2:0] to the DQ receiver. Ideally Vref should be set to be equal to the highest data signal level in PAM4, and be equal to $V_{OH}/2$ in NRZ signaling. Practically optimal Vref will be determined by training procedure. Note that the input offset of UGBs should be calibrated before the link starts. The ±25mV of calibration range is provided to cover the 3σ variation. However, if NRZ instead of PAM4 is adopted, UGB offset calibration is unnecessary since only one UGB is enabled and thus optimal vref in conjunction with offset will be decided during the training procedure.

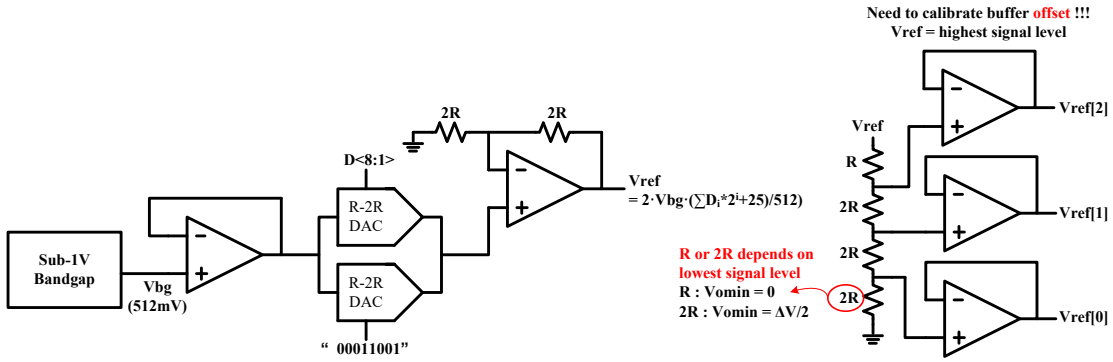
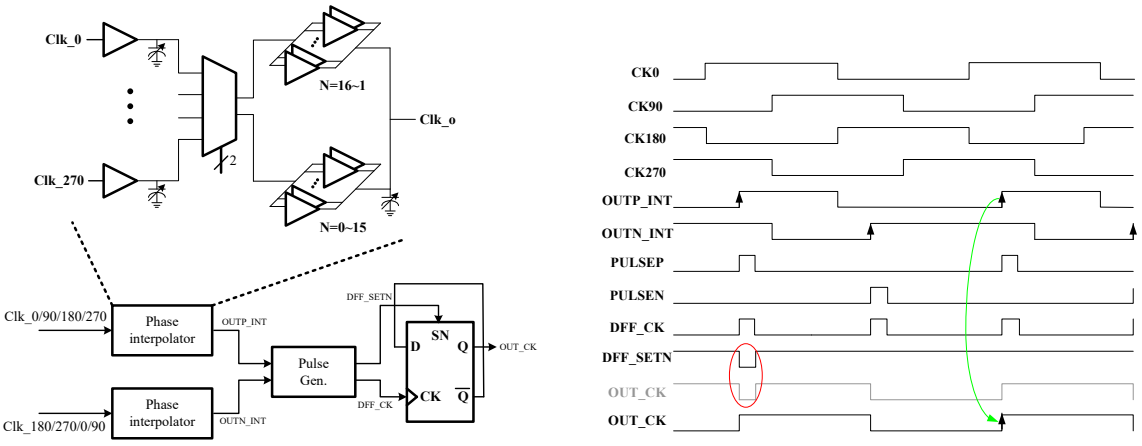


Figure 19. Reference voltage generator.

9.5.7 Phase Interpolator

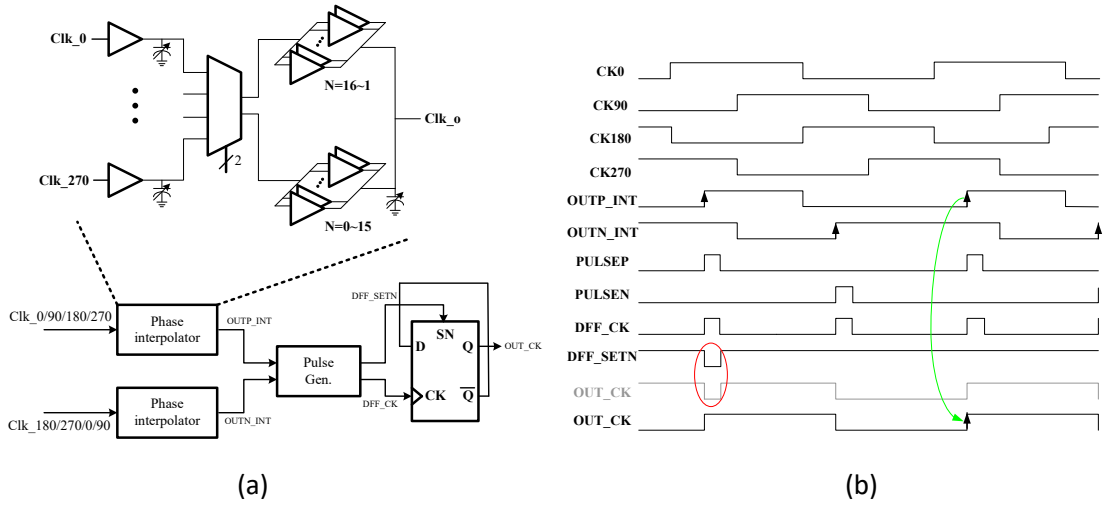
Complementary Phase Interpolation cell, as drawn in Figure 20 (a), is designed to provide global de-skew feature. A 6b PI cell can output a clock whose phase is programmable with 1/64 clock period resolution. Figure 20 (b) show its timing diagram to guarantee correct output phase after divided by 2. Variable capacitance at both input clock buffer output node and interpolation buffer output node are set to shape proper clock waveforms for linearity concern.



(a) (b)

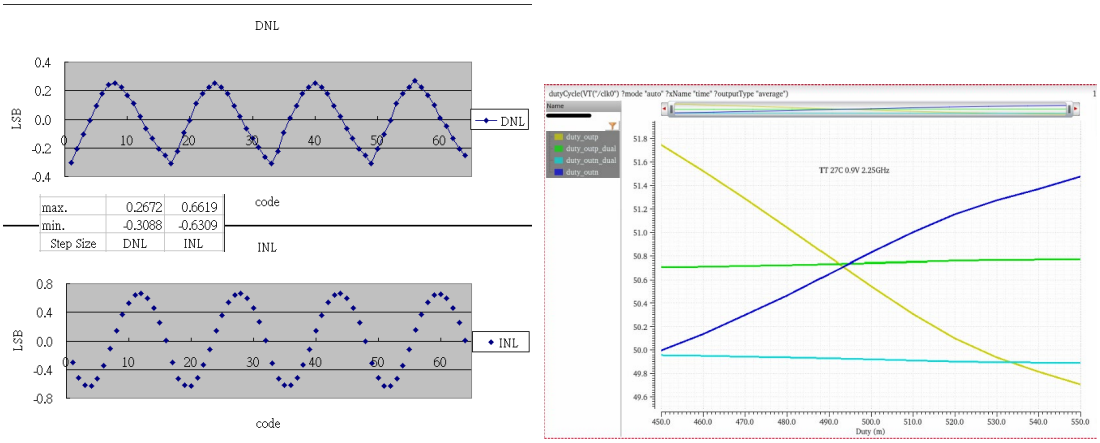
Figure 20 (a) complementary phase interpolation cell . and (b) timing diagram.

At the TT corner case, Figure 21 (a) shows the linearity (both INL and DNL) simulation results of the complementary PI cell. DNL of ±0.3LSB and INL of ±0.66LSB can be achieved. Besides, two PI cells with inverted input clocks, hence named complementary, are used to guarantee an output clock with 50% duty cycle. Figure 21 (b) shows duty cycle simulation result under various input clock duty. The duty of a single PI cell is also illustrated to make a fair comparison. It is clear that the complementary PI cell keeps the constant output clock duty in comparison with a single PI cell, by incurring the cost of slightly more power consumption (~2mA @ 2.133GHz frequency).



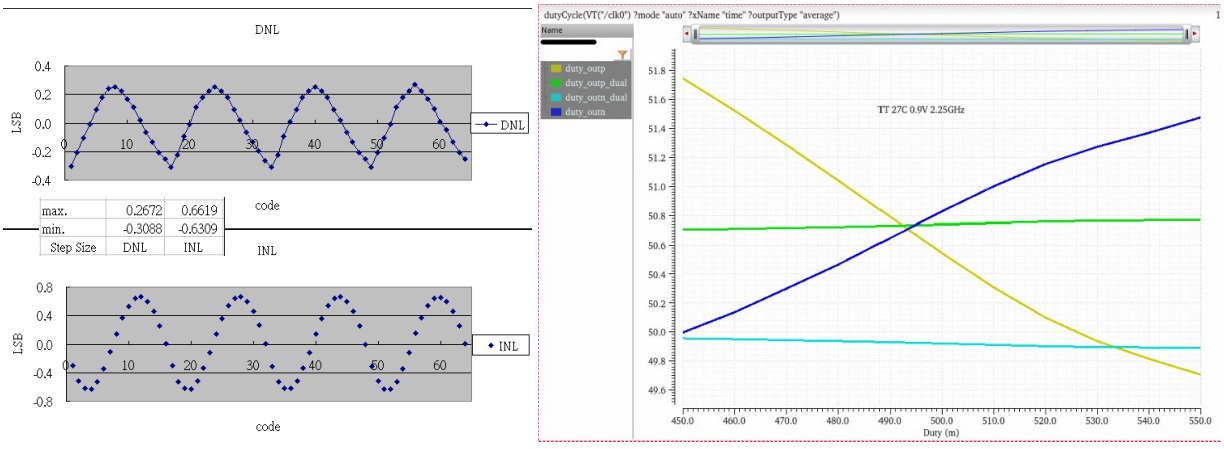
(a) (b)

Figure 21 (a) complementary phase interpolation cell . and (b) timing diagram.



(a) (b)

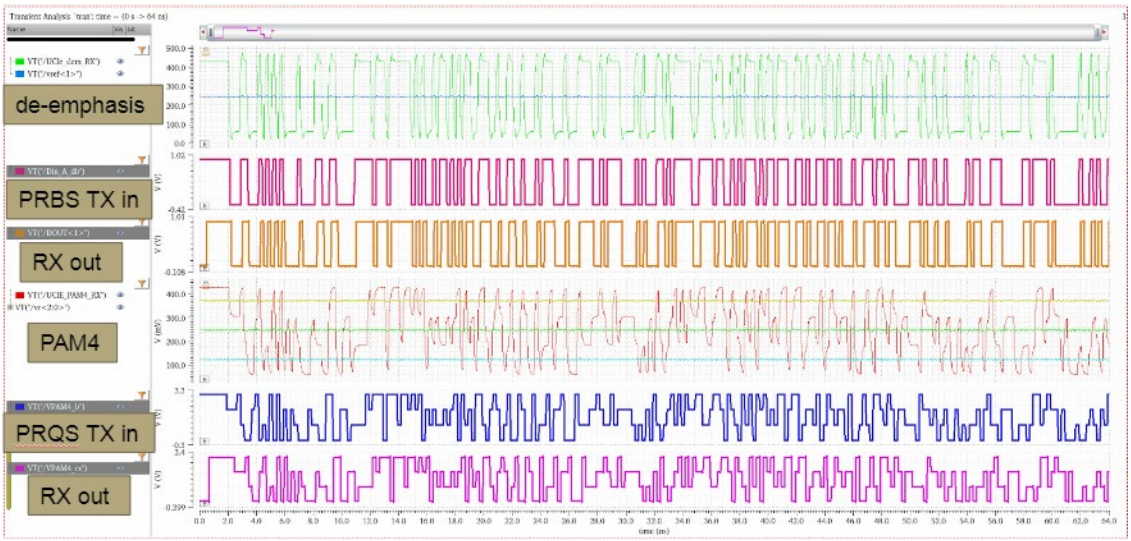
Figure 22. Simulation result of complementary PI cell (a) DNL/INL. and (b) output duty v.s. input duty



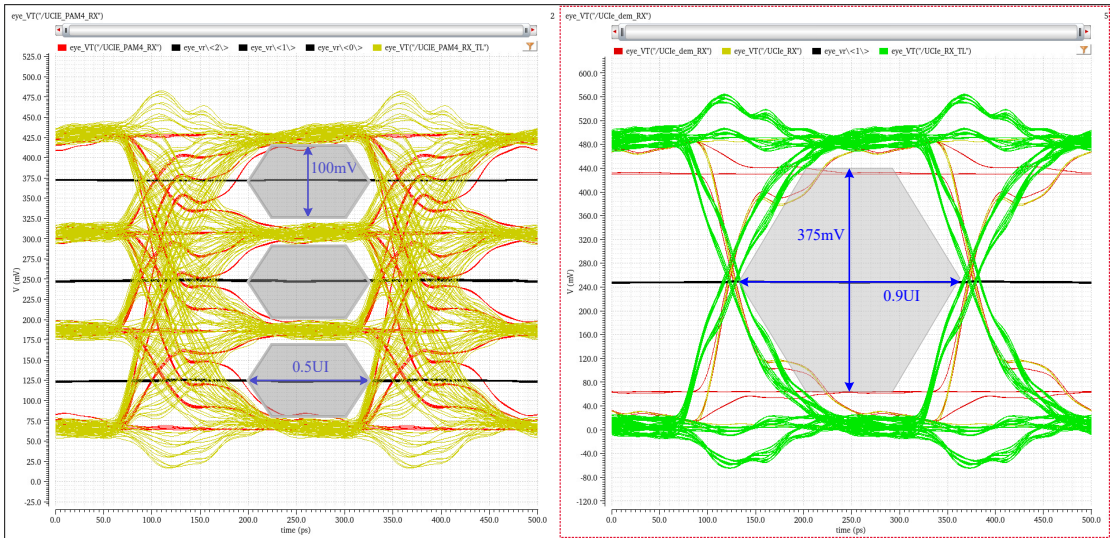
(a) (b)

9.6 Link Simulation Result

While the transceiver functions are simulated and checked. Both ideal transmission line (TLine) model and RLC package model are included in the simulation test bench. Here $R_{package}$=10mΩ, $L_{package}$=150pH, $C_{package}$=100fF, and C_{PCB_load}=300fF. Pre-layout simulation results of UCIE de-emphasis case and PAM4 case are shown in Figure 23 (a), where 2^8-1 PRBS and 2^8-1 PRQS input patterns are fed into transmitters, and same data pattern of receiver output are checked and confirmed. It proves the correct function of the designed combo Phy. Also eye diagrams of both cases are illustrated in Figure 23 (b). Even with the heavy IO pad capacitance loading that provided by TSMC, for the PAM4 case, the eye width can still reach 0.5UI with the eye height of ~100mV. Table II summarizes the PHY power consumption in the LPDDR4x mode. At the 40Ω characteristic impedance condition, the link can operate at ~4.5Gbps with ~0.86pJ/pin.



(a)



(b)

Figure 23. Simulation result of transceiver (a) transient waveform. and (b) eye diagram plot.

Table 3 power consumption of the link.

TT 27C 0.9V PRBS9

VDDQ	0.6
VDDA	0.9
drv_impd (ohm)	40
Dara Rate (Gbps)	4.5

DQ	ivddq(mA)	ivdda(mA)	Power(mW)
TX	5.013	0.1433	3.13677
RX	0.0194	0.8054	0.7365
DQS			
TX	12.37	0.6378	7.99602
RX	0.06541	1.503	1.391946

For DQ : 0.86073 pJ/pin

VDDQ	0.6
VDDA	0.9
drv_impd (ohm)	120
Dara Rate (Gbps)	4.5

DQ	ivddq(mA)	ivdda(mA)	Power(mW)
TX	1.878	0.1374	1.25046
RX	0.01883	0.8033	0.734268
DQS			
TX	4.728	0.5766	3.35574
RX	0.05912	1.579	1.456572

For DQ : 0.441051 pJ/pin

Note : total power depends on following settings
1: ***Delay Line***
2: ***drv_impd<2:0>***

drv_impd<2:0>	RZ(ohm)
1	240
2	120
3	80
4	60
5	48
6	40

9.7 Summary

A combo PHY transceiver is designed and simulated using the TSMC 28nm HPC Plus process models. The PHY can meet several standard interface circuit specifications, including LPDDR4x, UCIe, and a proprietary turbo UCIe specification. Both NRZ and PAM4 signaling are supported. Operated under ~4.5Gbps data rate, the energy efficiency of ~0.86pJ/pin can be achieved. An SoC equipped with two PHY interfaces, one programmed in the LPDDR4x mode and the other in the UCIE mode, can be used to realize the NOC features.

The Mixed-PHY system is designed to be a Memory Controller, an NoC-UCIe Controller, and an LPDDR/UCIe PHY IP that meets high performance, high compatibility, high flexibility, low area, and low power requirements for data communication between deep learning accelerator systems and DRAM chips. The controller can be connected to both memory mapping interfaces like AXI and NoC streaming protocols. The PHY can be configured as LPDDR4 mode, UCIe 48 TX mode, UCIe 48 RX mode, and UCIe 24 TX+24 RX mode. The maximum LPDDR mode data rate is 133 Gbps and the UCIe mode data rate is up to 400 Gbps.

9.8 References

[1] Low Power Double Data Rate 4 (LPDDR4) Specification, *https://www.jedec.org.sites.default/files/docs/JESD209-4.pdf*

[2] UCIe Specification 1.0, *https://uciexpress.org/team-3*.

[3] Enjoy-digital (2021) *litedram [Source code]*. *https://github.com/enjoy-digital/litedram*

[4] Wavious LLC (2021) *wav-lpddr-hw [Source code]*. *https://github.com/waviousllc/wav-lpddr-hw*

[5] Yongsuk Chio, Gyunam Jeon, and Yong-Bi Kim, "Transceiver design for LVSTL signal interface with a low power on-chip self calibration scheme," *INTEGRATION, the VLSI journal*, vol. 63, pp. 148-159, Sep. 2018.

[6] Chang-Kyo Lee, *et al*., "Dual-loop two-step ZQ calibration for dynamic voltage-frequency scaling in LPDDR4 SDRAM," *IEEE J. Solid-State Circuits*, vol. 53, pp. 2906-2916, Oct. 2018.

[7] Hyunsu Park, *et al*., "30-Gb/s 1.11-pJ/bit single-ended PAM-3 transceiver for high-speed memory links," *IEEE J. Solid-State Circuits*, vol. 56, pp. 581-590, Feb. 2021.

[8] Timothy M. Hollis, *et al*., "An 8-Gb GDDR6X DRAM achieving 22Gb/s/pin with single-ended PAM-4 Signaling," *IEEE J. Solid-State Circuits*, vol. 57, pp. 224-235, Jan. 2022.

[9] Hae-Kang Jung, *et al*., "A 4.35Gb/s/pin LPDDR4 I/O interface with multi-VOH level, equalization scheme, and duty-training circuit for mobile applications" in *Symp. VLSI Circuits Dig.*, Jun. 2015, pp. 184-185.

[10] Keunsoo Song, *et al*., "A 1.2V 2y-nm 4.35 Gb/s/pin 8 Gb LPDDR4 mobile device with bandwidth improvement techniques," *IEEE J. Solid-State Circuits*, vol. 50, pp. 1945-1959, Aug. 2015.

[11] J.-H. Chae, *et al*., "266-2133 MHz phase shifter using all-digital delay-locked loop and triangular-modulated phase interpolator for LPDDR4X interface," *Electronics Letters,* vol. 53, no.12, pp. 766-768, Jun. 2017.

[12] Chang-Kyo Lee, *et al*., A 5Gb/s/pin 8Gb LPDDR4X SDRAM with power-isolated LVSTL and split-die architecture with 2-Die ZQ calibration scheme" *IEEE International Solid-State Circuits Conference*, pp. 390-391, Feb. 2017.

[13] Mino Kim, *et al*., "A 4266Mb/s/pin LPDDR4 interface with an asynchronous feedback CTLE and an adaptive 3-step eye detection algorithm for memory controller," *IEEE Transactions on Circuits and Systems II-Express Briefs,* vol. 65, pp. 1894-1898, Dec. 2018.

[14] Ji-Hyo Kang, *et al*., "A 24-Gb/s/Pin 8-Gb GDDR6 with a half-rate daisy-chain-based clocking architecture and I/O circuitry for low-noise operation," *IEEE J. Solid-State Circuits*, vol. 57, pp. 212-223, Jan. 2022.