CIRCUIT TECHNIQUES FOR POWER MANAGEMENT UNIT AND SWITCHED CAPACITOR DC-DC CONVERTER

by

Suyoung Bang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2016

Doctoral Committee:

Professor Dennis M. Sylvester, Chair
Professor David Blaauw
Professor Katsuo Kurabayashi
Associate Professor Zhengya Zhang

TABLE OF CONTENTS

LIST OF FIGURES

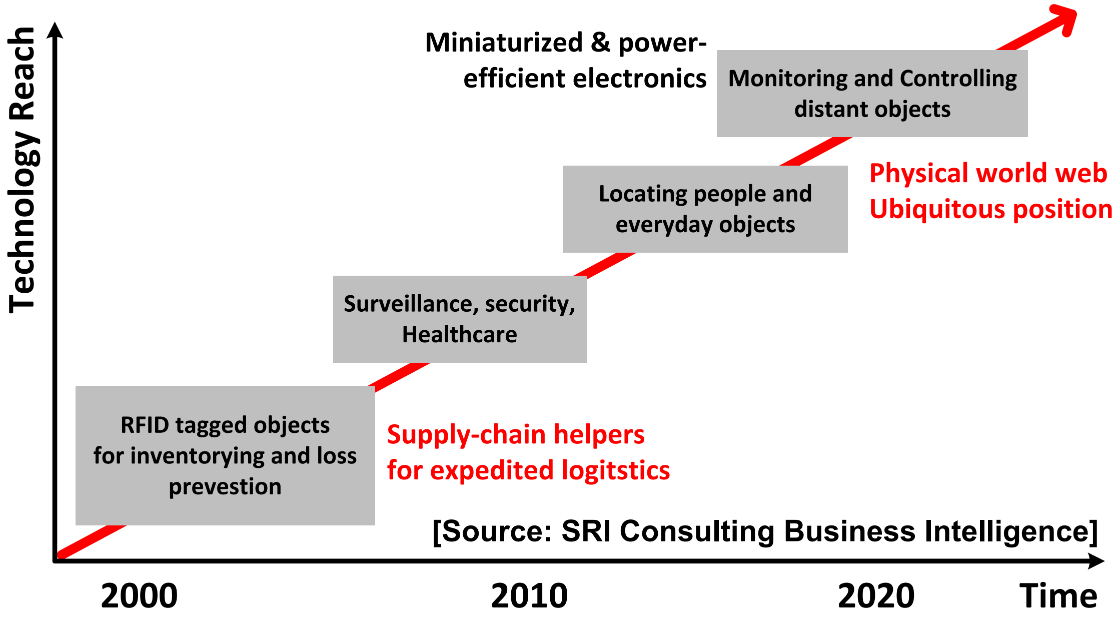
LIST OF TABLES

CHAPTER 1

Introduction

1.1 Power Conversion and Power Management in Internet of Things (IoT) Devices

Efficient power conversion and power management have been important in the era of the Internet of Things (IoT), which was first defined in 1999 by a technologist, Kevin Ashton. When he coined the terminology, IoT was referred to networks and devices collecting data about RFID-tagged objects. As the technologies that he stated have matured, IoT is not constrained to merely RFID technology. IoT is today built with wireless technology-enabled electronics, software and sensors to collect and exchange data with the manufacturer, operator and other connected devices [1].


Technology Reach
Miniaturized & power-efficient electronics
Monitoring and Controlling distant objects
Physical world web
Ubiquitous positioning
Locating people and everyday objects
Surveillance, security, Healthcare
RFID tagged objects for inventorying and loss prevestion
Supply-chain helpers for expedited logitstics
2000
2010
2020
Time

[Source: SRI Consulting Business Intelligence]

Figure 1.1. Advancement in IoT Technology (SRI consulting business intelligence).

Figure 1.1 shows an expectation of Strategic Business Insights (SBI), formerly SRI Consulting-Business Intelligence, for regarding advancement of the technologies, evolving from an era of supply-chain helpers for expedited logistics with RFID-powered devices in 2000s toward an era of physical world web and ubiquitous positioning with miniaturized and power-efficient electronics to monitor/control distant objects in 2020s. According to [2], IoT devices ranging from sub-mm to cm in their size will be demanded to be unplugged and untethered, thereby operating on a battery and harvesting ambient energy from environment. IoT devices will be placed where computing system did not exist before, and they will sense and actuate physical world around us by communicating with mobile and cloud computers in wired/wireless interface. Recently, modern IoT devices has started to be available in market in forms of handheld electronic devices such as Apple Watch [3] and Fitbit [4] that offer simple healthcare features, to start-up companies' challenges such as IoT-enabled farm land monitoring system [5], and object tracker with GPS-embedded tags designed to track pets and everyday objects around neighborhood [6].

The latest advancement of IoT devices toward miniaturized and power-efficient electronics can be understood similarly to evolution of computing platforms. The trend of evolution and miniaturization of computing platforms is shown in Figure 1.2. As found in 1972

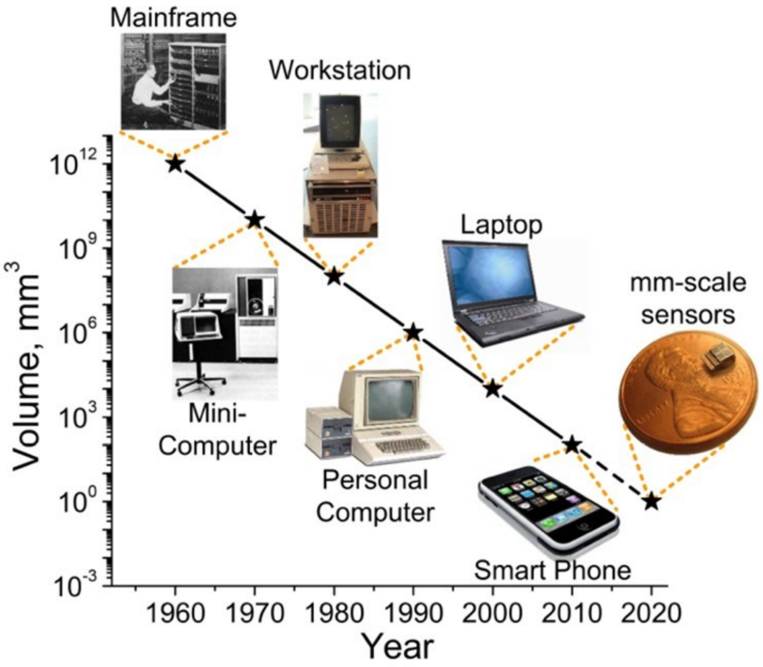


Figure 1.2. Evolution of computing platforms.

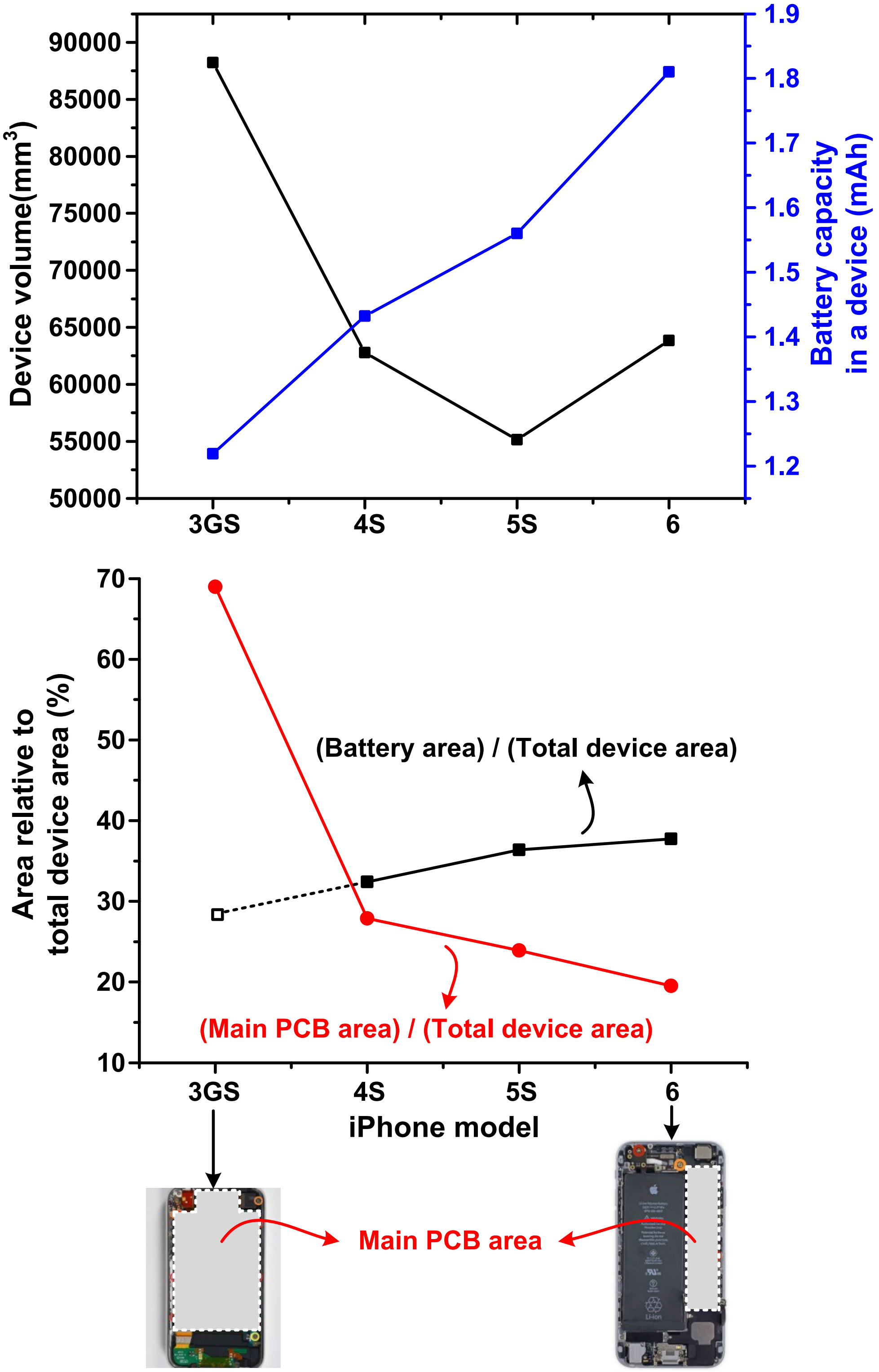


Figure 1.3. Changes in volume of a device, and battery capacity of a device with iPhone models (top), and change in battery area and PCB area with iPhone models (bottom).

by Gordon Bell, and named Bell’s Law [64], every ten years, computing platforms have shrunk in volume by 100 times from the era of mainframe computing to handheld smart-phone. Over time the computing systems not only has scaled in size, but it has also improved computing capabilities and power consumption, enabling for portable electronic devices with remarkable computing performance and extended battery lifetime. Looking ahead and further reducing form factor, today’s researchers are working to realize mm-scale sensor platforms [9-12].

It is helpful in understanding direction of advancement of IoT devices to see how a modern electronic device has evolved. Let us take a look at iPhone, a forerunner of IoT devices, for example. Figure 1.3 (top) shows change in volume of a device, change in battery capacity in a device with iPhone models, and Figure 1.3 (bottom) shows change in battery area and PCB area out of total device area with iPhone models. It can be observed that battery capacity and battery area has increased whereas device volume has decreased. More importantly, it is noted that PCB area has significantly decreased from iPhone 3GS to iPhone 6. This reflects trends of electronic markets that light-weight and long-last battery-built devices are preferred.

In this trend, power converter design and power management have become vital for low-power and energy-efficient operation of electronics. In an attempt to implement energy efficient circuits, optimal voltage has been investigated in prior works [7-8]. Also, as presented in paper [9-10], supply voltages for analog circuits and digital circuits have different levels, both of which can be significantly lower than single input battery voltage. Also, battery voltage decreases as its energy is drained, and current drawn by sub ICs affects intrinsic voltage drop across battery internal resistance. Thus, power converter is required to generate and regulate lower voltages than input. Since its power conversion efficiency (PCE) affects system power consumption such that system power is load circuit power divided by PCE, it is essential to design efficient power converter for extended battery life and low power operation of electronics.

Miniaturized IoT devices proposed in [9-12] are rarely active and spend most of their time in sleep mode, and their average power is largely determined by the standby leakage [24]. As shown in Figure 1.3, these devices require power management circuits that enables to (1) perform efficient power conversion, (2) reduce standby leakage current for prolonged battery life of the devices, (3) design ultra-low power controller that governs system wake-up and operating mode, and (4) enable harvesting for autonomous operation with harvesting of ambient energy in volume-constrained IoT devices.

1.2 Switched-Capacitor DC-DC Converter vs. Buck Converter and Linear Regulator

Switched-capacitor (SC) DC-DC converter is favorable as voltage regulator module (VRM) in the era of IoT, over other regulator candidates such as buck converter and linear regulator. The simplified diagrams for the three voltage regulators are illustrated in Figure 1.4. The advantages of SC converter over buck converter are well explained in [13-14]. First, considering PCB-level discrete components, capacitor has substantially large energy and power density than inductor. It is true that inductive converter can fully energize and de-energize an inductor every period without incurring loss, and SC converter need to store fairly large amount of energy on capacitors and utilize only a small ripple voltage. Regardless, significantly superior energy density of capacitor overcome the operational limiting property of SC converters, enabling for accomplishment of higher efficiency at equal power density, or higher power density at equal efficiency.

Second, SC converters are compatible with CMOS technology, whereas on-chip buck converters need magnetic and thick metal layers for on-chip inductors that have inferior quality factor than off-chip inductors, degrading power conversion efficiency (PCE) of buck converter. A packaged bondwire-based buck converter can be a solution for good quality factor, but this approach needs dedicated bond-wires, which requires post-fabrication efforts [15].

Third, industry trends toward system-on-chip (SoC), which is likely preferred in small form-factor IoT devices, have multiple number of circuit loads, each of them requires decoupling-capacitors. SC converters inherently use capacitors for power conversion, which

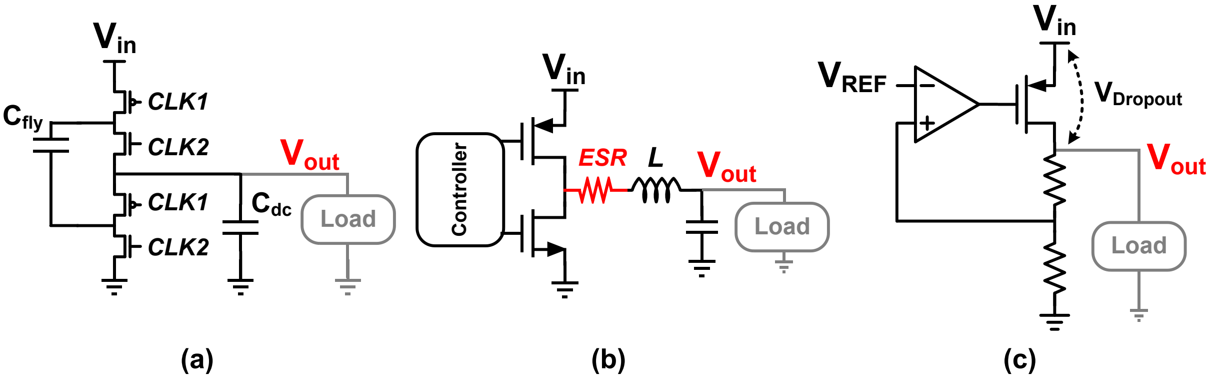


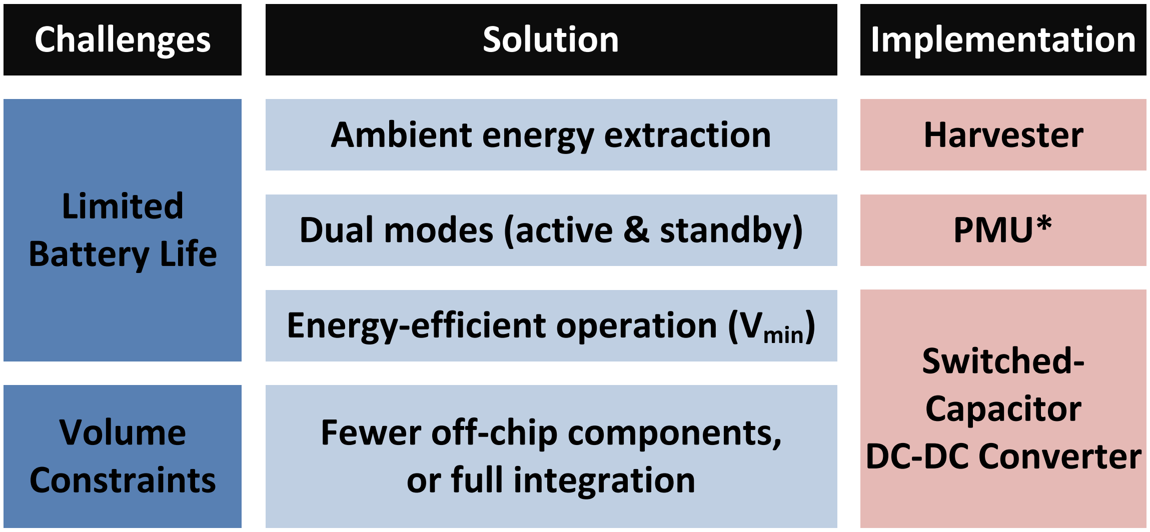
Figure 1.4. Three candidates for voltage regulators: (a) switched-capacitor DC-DC converter, (b) inductive buck converter, and (c) linear regulator.

behaves as decoupling-capacitors in multiple-phase SC converters.

Fourth, low-power application favors SC converter even more. In SC converter, current output level is scalable with clock frequency without any complicated control or operation for efficiency [16]. On the other hand, inductive buck converter need to operate in deep discontinuous conduction mode (DCM) or pulsed frequency mode, and require complicated control circuit [17]. Also, in inductive converter for low-power load, large inductance is not inevitable [18], so there have been no fully-integrated buck converter reported for nW range load, to date.

Finally, linear regulators as an alternate approach for on-chip voltage regulation can be implemented on chip with superior power density. Nonetheless, it has an efficiency bounded by the ratio of output voltage to input voltage (V_{out}/V_{in}), and hence suffers from poor efficiency when difference between input and output voltages is large. Furthermore, in such case, excessive power is dissipated in a small volume and planar area, resulting in over heat issue [19].

SC converters have been known to have (1) a limited number of conversion ratio, restricting achievement of good PCE over wide range of input and output voltages, and (2) poor output voltage regulation capability, showing large output noise. We tackled those issues for step-down converter.



Challenges	Solution	Implementation
Limited Battery Life	Ambient energy extraction	Harvester
	Dual modes (active & standby)	PMU*
	Energy-efficient operation (V_{min})	Switched-Capacitor DC-DC Converter
Volume Constraints	Fewer off-chip components, or full integration	

PMU* = Power Management Unit

Figure 1.5. Challenges and solutions of IoT devices.

As solutions to challenges in IoT devices summarized in Figure 1.5, I developed circuit techniques for ambient energy harvester and power management unit for ultra-low power application. These include (1) adaptive and multi-modal harvester based-on SC converter, (2) ultra-low power management unit, and (3) reconfigurable sleep transistor for reduction of gate-induced drain leakage (GIDL). Furthermore, I developed (1) highly reconfigurable SC converter topology with cascaded form and (2) ripple minimized technique for closed-loop multi-phase SC converter.

1.3 Deep Learning Co-processor for Internet of Things (IoT) Devices

Recently, deep learning algorithm has gained attention from many fields. It can be used in a variety of learning and inference such as keyword speech recognition, object recognition, natural language processing [67]. In computer architecture, energy efficiency has been constrained by power wall and dark silicon, and there has been growing attention and active researches on hardware accelerators design that offers better energy-efficient operation in performing a few digital signal processing algorithms than CPU and GPU [65-66]. The latest advent deep learning has made computer architecture and circuit designers explore deep learning processor.

A prior work investigated deep learning algorithm operating at mobile platform, and highlighted that existing mobile sensing platforms can utilize scalable and configurable properties of deep learning for a wide range of application [67]. Also, several earlier works proposed instruction-based and reconfigurable processor for deep neural network (DNN), or convolutional neural network (CNN) [68-70]. However, the proposed processors are not suitable for low-power mobile application.

We proposes a low-power configurable deep learning co-processor for IoT devices which performs fully-connected layer operation of neural network and FFT, aiming for keyword speech recognition, handwritten digits detection, and other generic digital signal processing tasks.

1.4 Contributions and Organization

In chapter 2, fully-integrated switched-capacitor-based power management unit (PMU) with self-adaptive conversion ratio is proposed for ultra-low power sensor platform, including adaptive step-up and step-down SC converter and battery supervisory finite state machine (FSM). The proposed PMU has two ladder-type SC converters with reconfigurable conversion ratio, using 1.03nF of on-chip MIM capacitance. It automatically adapts to different battery voltages for down-conversion and different harvesting sources/harvesting conditions for up-conversion, and achieves 63.8% / 60.7% down-conversion efficiency at 17.9µW active mode / 12.8nW sleep mode power loading. With the adaptive down-conversion ratio, load power range is improved by 3.76× and 5.48× in sleep and active mode, respectively. It is shown how the proposed adaptation method enables harvesting with solar, microbial fuel cell, and thermal energy sources, increases harvesting efficiency by 1.92× and achieves the peak extraction efficiency of 99.8% for solar cell.

In chapter 3, reconfigurable sleep transistors for GIDL reduction is proposed for ultra-low power sensor platform with volume constrained, where lithium-ion battery is only available input option, generating 3.6-4V. Standby power reduction is critical to battery life and volume reduction in mm-scale sensor nodes, and power gating is extensively adopted to reduce leakage. However, the inserted sleep transistors can suffer from other leakage mechanisms, namely GIDL, which become dominant at battery voltages of 3 V or higher. The reconfigurable sleep transistors have two different topologies are used in active versus sleep mode. In active mode, transistors are stacked as in traditional power gating schemes. In sleep mode, sleep transistors are reconfigured to reduce GIDL current, in addition to subthreshold leakage. Measurements on a 180nm CMOS test chip shows 12.6× standby leakage reduction at *VDD*=4.0 V and T=25°C. This improvement comes with acceptable area penalty due to additional small reconfiguration transistors and separate body contacts, and no impact on active mode operation.

In chapter 4, in an attempt to improve number of conversion ratio of SC converter, successive approximation (SAR) SC converter is demonstrated. The proposed SAR SC converter allows for fine output voltage control with conversion ratio resolution of $V_{IN}/2^{\text{Number of Stages}}$ to enable effective load and line regulation in ultra-low power applications. This chapter also investigates and provides comparisons of slow-switching limit impedance loss (P_{SSL}) and bottom-plate parasitic capacitor switching loss (P_{BOT}) as a function of conversion ratio for series-parallel and conventional ladder topology in addition to SAR SC converter. It is shown that SAR SC converter outperforms conventional topology SC converters in terms of P_{SSL} and P_{BOT}.

In chapter 5, flying-capacitance-dithered SC DC-DC converter is proposed for output voltage ripple minimization. On-chip voltage regulators have been favored in that it enables for

(1) reduction of input current to on-chip system, IR drops, and Ldi/dt drops, (2) fast load response, and (3) efficient per-block power management [38, 55]. SC converter as on-chip regulator has become a popular choice in that it has power conversion efficiency and compatibility to CMOS process in contrast to linear regulator and on-chip buck converter. SC converter is likely to suffer from ripple issue due to its inherent switching property enabling charge transfer for voltage conversion and power generation. In this chapter, effective power conversion efficiency (PCE_{eff}) is defined as an indicator of both voltage regulation capability and power conversion efficiency, and relation of PCE_{eff} and output voltage ripple magnitude is investigated. Proposed dithered capacitance modulation (DCM) implements closed-loop multi-phase SC converter for ripple minimization, by operating at the fastest available clock and modulates amount of flying capacitance (C_{FLY}). A testchip for DCM was fabricated in 65nm CMOS process, and it achieved ripple ranging from 6−16mV for load current of 11−142mA, with peak efficiency of 70.8% at a power density of 187mW/mm^2.

In chapter 6, a low-power configurable deep learning co-processor is proposed, which performs fully-connected layer operation of neural network and FFT. In order to perform energy efficient and low power operation, we propose deep learning processor utilizing the following techniques (1) non-uniform memory architecture (NUMA), (2) temporal and spatial locality, (3) weight matrix tiling, (4) variable data precision, variable multiplier width, and long one line of memory. The expected benefit from the proposed techniques are up to 43%. For implementation, we have developed custom SRAM memory with four hierarchies, and designed optimized data flow which can save unnecessary memory accesses in operations of the fully-connected layer and FFT. The proposed co-processor has low-power digital signal processor (DSP) with four processing elements (PEs) and hierarchical memory, and ARM-Cortex M4. The four PEs of the co-processor operate with instruction, ensuring great configurability, and they are expected to perform a wide range of application within budgets of power and memory capacity.

In chapter 7, conclusions for the completed work as part of this thesis are provided. Finally, this work includes one appendix. Appendix A demonstrates mathematical derivation steps of slow-switching impedance (R_{SSL}) for series-parallel, conventional ladder, variant ladder, and SAR topologies. This appendix derives the exact equations for R_{SSL}, which can be used to predict R_{SSL} loss without simulation, providing insights on R_{SSL} of a variety of SC converter topologies as function of conversion ratios.

CHAPTER 2
Power Management Unit (PMU) for mm3-scale Sensor Node

The thin-film Li-ion battery in the proposed sensor platform outputs voltages as high as 4.1V while low power electronics often operate below 0.5V, resulting in a challenging DC-DC conversion ratio. The PMU must enable both up-conversion during harvesting and down-conversion in the absence of harvesting conditions. Further, it must accommodate a >1000× spread in current draw between sleep/active modes and low/high harvesting conditions. Having switches that can accommodate 10s of μA in active mode, while at the same time maintaining high efficiency conversion with load currents as low as single digit nAs in sleep mode, is extremely challenging. Finally, different harvesting sources have varying optimal operating points that change with harvesting conditions. For instance, among the harvesting sources used in this chapter, a solar cell has an open circuit voltage as low as 350mV in indoor conditions, increasing to 530mV in sunlight, while a microbial fuel cell (MFC) has an open circuit voltage that tends to fluctuate over time from 500 to 800mV. These stringent requirements call for the PMU in the proposed sensor platform to be highly adaptive to load current and harvesting source conditions.

Recently [20] proposed an energy harvesting architecture for multiple energy sources. However, the use of a large off-chip inductor makes it unsuitable for miniature sensor nodes. In [9-10], authors propose the use of fully integrated SCN converters for both down-conversion and harvesting purposes. However, these SCNs are static and cannot adjust to different harvesting sources and conditions, restricting their use to specific scenarios [9]. Furthermore, they use an integrated canary harvester to monitor the harvesting condition, which restricts them to a single type of harvesting.

2.1 Overview of PMU operation

A new fully integrated SCN is proposed with a chip area of 0.95mm^2. As shown in Figure 2.1, active region is 0.24mm^2 and MIM-capacitance area is 0.71mm^2. It automatically adapts to different harvesting conditions and can be configured to operate with different harvesting sources.

The SCN consists of two ladder-type SCNs, as shown in Figure 2.2; one for converting between battery and processor voltages (*battery SCN*) and one for converting between processor and harvester voltages (*harvester SCN*). During down conversion, when the processor draws

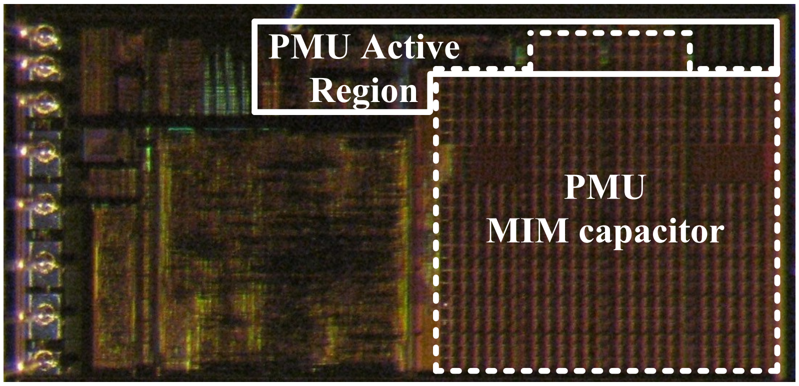


Figure 2.1. Die photo: The proposed power management unit was fabricated in180nm CMOS technology.

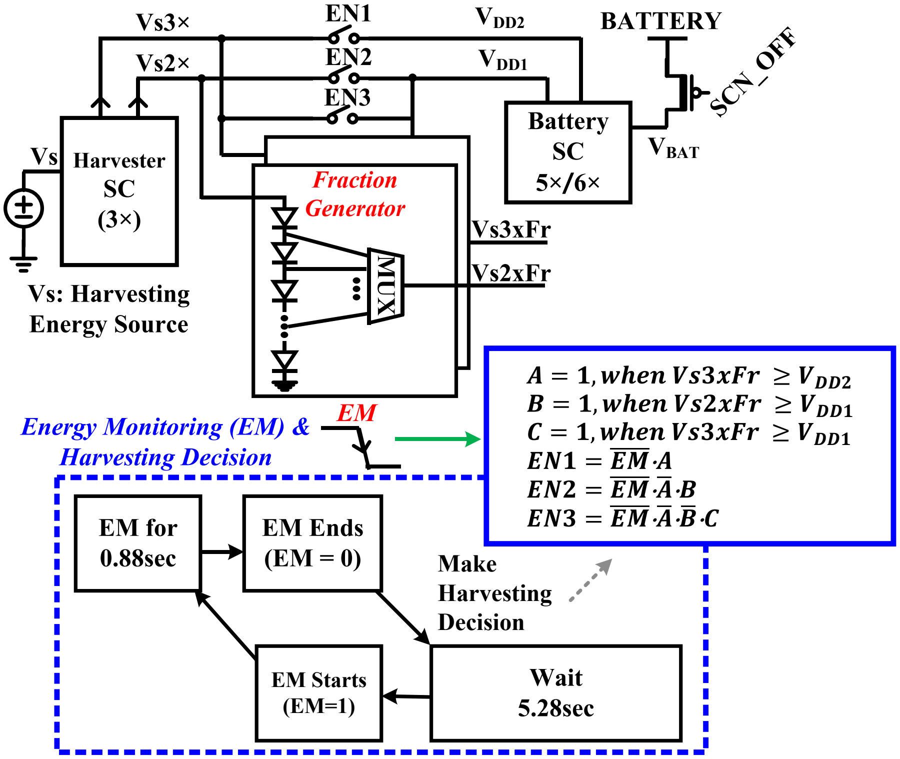


Figure 2.2. Proposed overall PMU diagram and adaptive harvesting technique.

current from the battery, the battery SCN is automatically reconfigured between 5× and 6× modes, providing the ability to adapt to different load currents and battery condition from 3 – 4.1V.

During harvesting, both SCNs up convert the harvested voltage and provide current to the processor and battery. The harvester SCN has 2× and 3× modes and is connected to the battery SCN in one of three possible configurations to automatically adapt to different harvesting sources and harvesting conditions. The key challenge in efficient harvesting is to determine the configuration that forces a voltage at the harvester leading to extraction of maximum energy from the harvester. The solar power profile in Figure 2.2 shows how the extracted power peaks when the extraction voltage is approximately ¾ of the open circuit voltage (V_{OC}). However, the V_{OC} varies between 350 and 500mV depending on light conditions, and hence the harvest extraction voltage must be adjusted accordingly. In addition, different energy sources have different power profiles, leading to the need to adjust the fraction of V_{OC} at which the harvested power is extracted.

The proposed PMU adapts to different harvesting conditions using a two phase process. First, the battery and harvester SCNs are disconnected for a short monitoring time period. During this time, the harvester develops its open-circuit voltage, which is up converted by harvester SCN by 2× and 3×, and then divided using reconfigurable, high-impedance fractional voltage dividers, after which the resulting voltage levels are compared to battery SCN voltages to find the optimal harvesting configuration. During the subsequent harvesting time period, the battery and harvester SCNs are connected according to the optimal configuration determined during the monitoring period, and both processor and battery are powered by the harvesting unit. The monitoring / harvesting cycle is repeated every 6 seconds to allow the PMU to adapt to changing harvesting conditions. If there is insufficient harvested energy, the PMU automatically disconnects the harvester SCN from the battery SCN. The voltage dividers can be configured by the processor to adjust to harvesting sources with different power profiles. There is a trade-off between loss in harvesting time and response time to harvesting condition change with adjusting monitoring / harvesting cycle time. Applications with slow harvesting condition change do not require fast response to harvesting condition change hence can benefit from longer cycle time which minimizes loss in harvesting time.

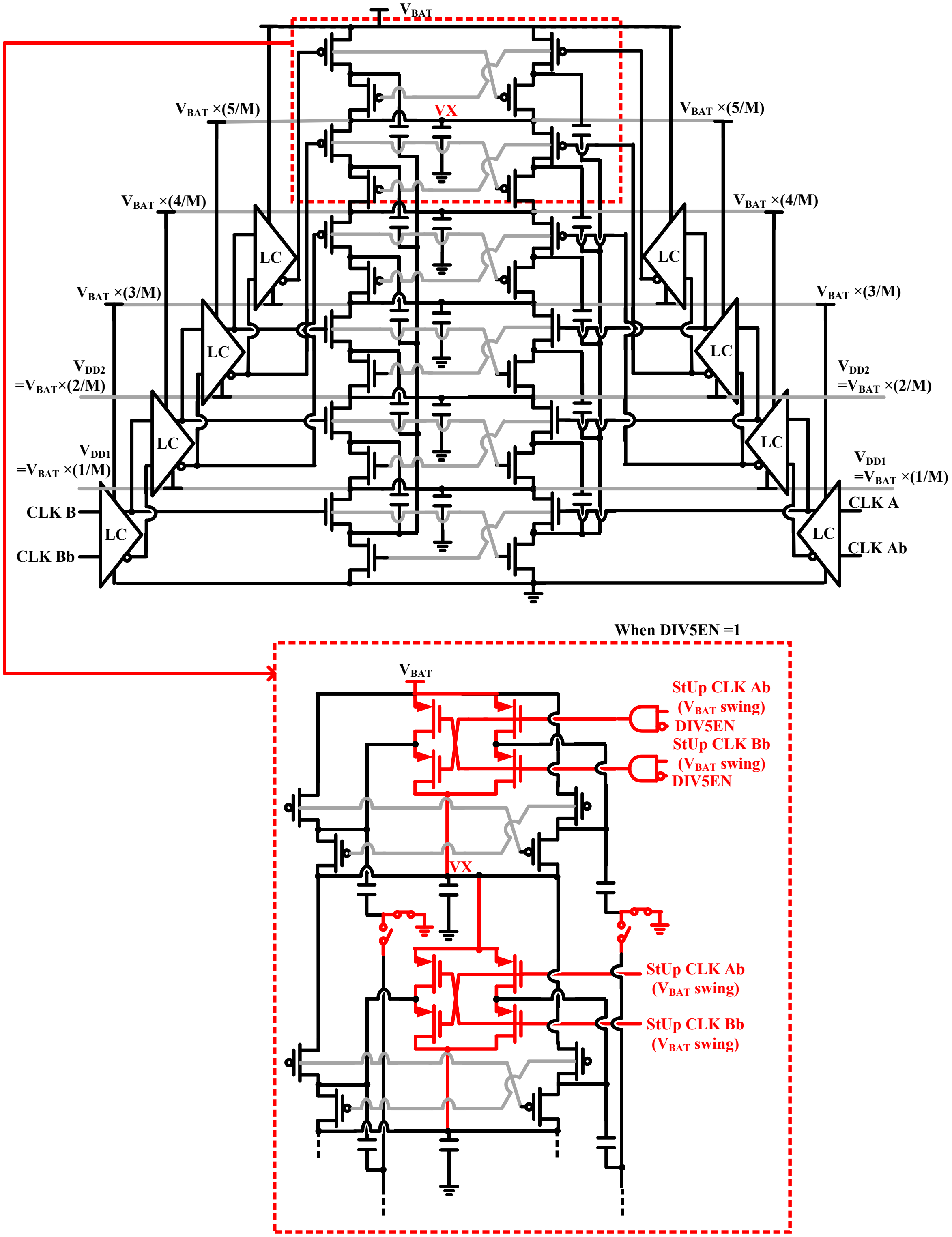

V_{BAT}
VX
V_{BAT} ×(5/M)
V_{BAT} ×(5/M)
V_{BAT} ×(4/M)
V_{BAT} ×(4/M)
V_{BAT} ×(3/M)
V_{BAT} ×(3/M)
LC
V_{DD2} =V_{BAT}×(2/M)
V_{DD2} =V_{BAT} ×(2/M)
V_{DD1} =V_{BAT} ×(1/M)
V_{DD1} =V_{BAT} ×(1/M)
CLK B
CLK Bb
CLK A
CLK Ab
When DIV5EN =1
V_{BAT}
StUp CLK Ab (V_{BAT} swing)
DIV5EN
StUp CLK Bb (V_{BAT} swing)
DIV5EN
VX
StUp CLK Ab (V_{BAT} swing)
StUp CLK Bb (V_{BAT} swing)

Figure 2.3. Two-way phase-interleaved ladder-type SCN is used for conversion of 5× or 6× (M: conversion ratio). CLK A and CLK B are nonoverlapping clocks with the same frequency. Thick-gate switches for start-up sequence are indicated using arrow FET notation. By setting DIV5EN, VX is simply shorted with VBAT and the converter operates in division-5 mode.

2.2 DC-DC down-conversion operation

During down conversion, the PMU converts the battery voltage to two voltage domains (VDD1 = 0.6V and VDD2 =1.2V, nominally) to power the processor and an array of peripherals. The proposed PMU has two operating modes with vastly different power budgets: 1) sleep mode when the processor is inactive with a current draw of 1 – 10nA and 2) active mode when the processor is running with a current draw of 1 – 10uA. To efficiently perform DC-DC down-conversion, the PMU uses four different oscillators: a start-up sequence clock (5kHz, simulated), sleep mode clock (340Hz, measured), active mode clock (335kHz, measured), and an additional SCN clock (3.125kHz, measured) for harvesting in sleep mode. To reduce ripple magnitude on the output voltage levels at a given SCN frequency and capacitance, a two-way phase interleaved

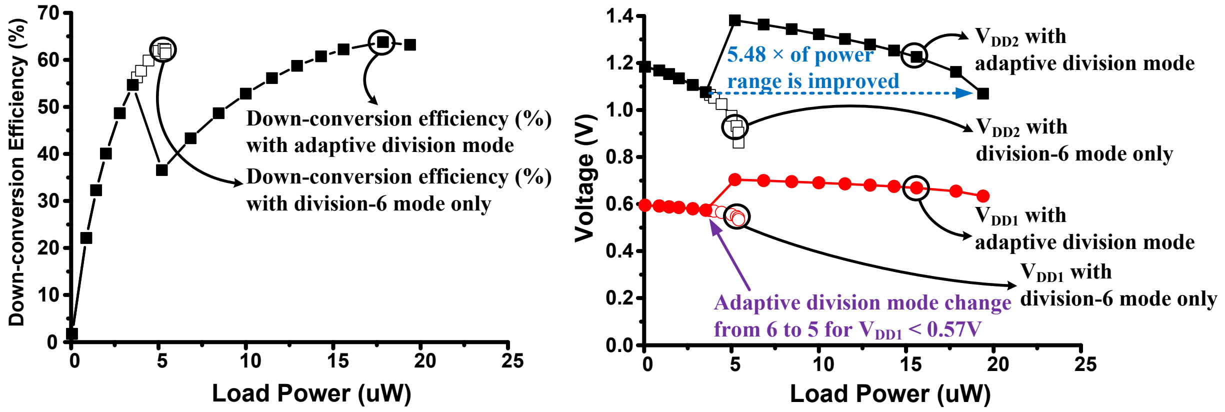


Figure 2.4. Measured down-conversion efficiency in sleep mode with VBAT = 3.6V and 340 Hz sleep mode clock (left).Measured VDD2 and VDD1 with adaptive division mode, and with division 6 mode only (right).

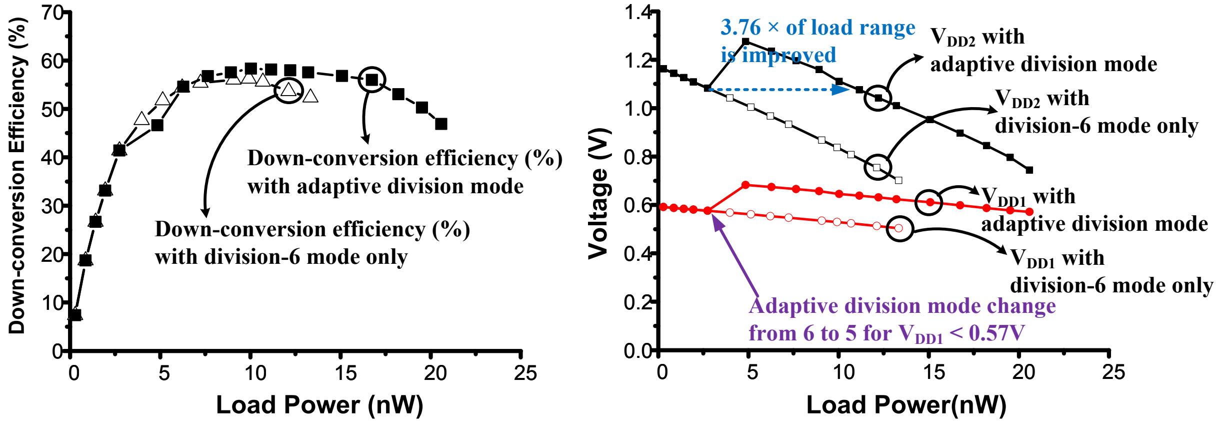


Figure 2.5. Measured down-conversion efficiency in active mode with V and 335 kHz active mode clock (left). Measured VDD2 and VDD1 with with VBAT = 3.6V adaptive division mode, and with division 6 mode only (right).

ladder is used as shown in Figure2.3.

When the battery SCN starts up for the first time, no charge is stored on internal nodes. Hence, the start-up clock operates at the battery voltage with full amplitude signals and thick-gate oxide switches. Once start-up is completed, the PMU enters operational mode, and the clocks operate with VDD1 amplitude in active mode and VDD2 in sleep mode, driving thin-oxide switches. Using lower clock voltage swing and smaller thin-oxide switches significantly reduces PMU power consumption and increases conversion efficiency. Level conversion is required for driving the different switches in the ladder and is implemented using a chain of level converters as seen in Figure 2.4.

The PMU switches between active and sleep mode based on control signals from the processor and timers. The battery has an adaptive conversion ratio of 6× or 5× to address battery voltage variation and a wide range of load power. When 5× is enabled, switches reconfigure the top stage of the ladder such that VBAT and VX are shorted and the bottom plate of the flying capacitor is grounded. This reconfiguration is triggered by two comparators that compare fractions of VDD1 to a voltage reference. Hysteresis is introduced between the transition to prevent oscillation between the 6× and 5× modes.

The measured standby power consumption of the sensor platform ranges from <3nW in sleep mode up to ~20μW in active mode, resulting in peak efficiency of 63.8% and 60.7% in the two modes, respectively. (Figure 2.5). Note that nearly equal efficiency is achieved in sleep mode compared to active mode, due to the use of the VDD1 voltage domain for the SCN clocks in sleep mode. Figure 2.4 and 2.5 show how the configurability of the battery SCN increases the load power range by 3.76×/5.48× for sleep/active modes over the case without this capability (VBAT=3.6V). The battery SCN conversion ratio changes from 6 to 5 when VDD1 drops below 0.57V, and reverts back to 6× when VDD1 exceeds 0.71V.

2.3 PMU up-conversion for adaptive harvesting

During harvesting, the harvester SCN up-converts the harvesting source voltage by 2× and 3×, and connects either of these outputs to the VDD1 or VDD2 ports of the battery SCN (Figure 2.2). The two SCNs are connected in three different ways to accommodate harvesting sources with output voltages from 360 – 800mV. The open-circuit voltage develops during the monitoring phase, which lasts 0.88sec. During this phase a highly resistive diode-connected fraction generator (with configurable output voltage from 0.2× to 0.95×) is applied to both the 2× and 3× outputs of the harvester SCN. These two factional OCVs are then compared to VDD1 and VDD2 from the battery SCN according to three possible configurations (Config #1: 3× = VDD1, Config #2: 2× = VDD1, Config #3: 3× = VDD2). The three configurations are rank ordered according to the harvesting voltage that they induce and a configuration in which the fractional voltage minimally exceeds the voltage from the battery SCN is automatically selected to ensure near-optimal energy extraction from the harvesting source.

The proposed method automatically adjusts the harvesting setting in response to the harvester's VOC (addressing different harvesting conditions), the battery voltage (addressing different battery charge states), and the setting of the fractional generator (addressing different harvesting sources). For example, when battery voltage is 3.0V and the OCV fraction is 0.65, configuration 1 corresponds to an VOC of 307 − 462mV, configuration 2 to 462 − 615mV, and configuration 3 to VOC > 615mV. In a situation with VOC < 307mV, the harvester is automatically disabled to prevent possible drain through the harvesting unit. Finally, after the monitoring phase, the chosen configuration is established during the harvesting phase, which lasts 5.28s.

The proposed approach was tested with three harvesting sources: solar cell, microbial

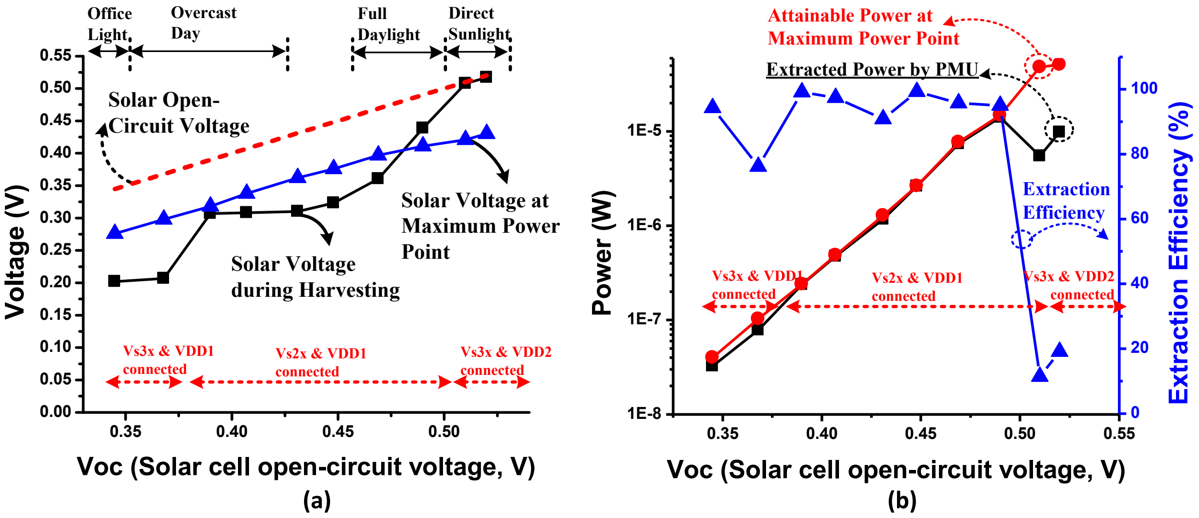


Figure 2.6. (a) Maximum power point tracking, and (b) attainable/harvested power and extraction efficiency of the proposed PMU for energy harvesting with solar cell at VBAT=3.6V and fraction = 0.75.

fuel cell (MFC), and thermoelectric generator (TEG). The proposed PMU achieves an overall energy harvesting efficiency of 39.8% with an MFC, and 28.1% with a 2.56cm2 TEG and 26.9% with a 1.62mm^2 solar cell. To demonstrate the proposed method for maximum power point tracking (MPPT), we show in Figure 2.6 (a) how the harvest voltage forced by the PMU tracks the optimum harvesting voltage. As light intensity increases from office light to outdoor lighting, the PMU automatically switches from the first configuration (with force harvest voltage of 200mV) to the second configuration (300mV). Figure 2.6 (b) compares the power extracted from the solar cell by the PMU with the maximum attainable power and the extraction efficiency defined as following.

$$Extraction\ efficiency = \frac{Extracted\ power\ from\ solar\ cell\ by\ PMU}{Attainable\ power\ at\ maximum\ power\ point\ (MPP)} \tag{2.1}$$

For light intensities from office light to full daylight, the average extraction efficiency is 93.5%, with a maximum efficiency of 99.2% and minimum efficiency of 76.1%. When the solar cell is exposed to direct sunlight, Voc increases to 520mV and the PMU switches to the third configuration (400mV). However, the current at this light condition reaches up to 100μA (> 2 orders of magnitude higher than in office lighting), exceeding the conversion capability of the PMU. Hence the PMU can no longer force the intended 400mV harvest voltage, resulting in a drop in extraction efficiency.

2.4 Battery Voltage Monitoring

A battery voltage monitoring unit [21] is included to detect changes in battery voltage and to allow the PMU to take appropriate action. Figure 2.7 shows PMU state transitions in response to battery voltage fluctuations. Once the battery voltage stabilizes above 3.4V, the PMU enters *Operational* mode and the system is activated. Discharging the battery below 3.0V due to heavy use or wear-out can incur permanent damage to the Li battery. To prevent this, the monitoring unit detects voltage levels below 3.1V and the system enters a 185pW *Deep Sleep* mode where all supplies are turned off. The availability of harvested energy is monitored in Deep Sleep mode and when sufficient light is detected, the system enters *Recovery* mode, recharging the battery from is detected by the POR circuit in each IC layer, which forces local reset for proper initialization.

the harvested energy. After the battery has reached a sufficiently high voltage (>3.4V), the system returns to its normal Operational mode and distributes a power on reset (POR) signal to all IC layers by sequentially releasing the 1.2V and 0.6V supplies. This sequence

2.5 Summary

The proposed PMU for ultra-low power sensing application accomplishes 63.8% and 60.7% peak down-conversion efficiency at 17.9µW load power in active mode and 12.8nW load power in sleep mode, respectively. The adaptive down-conversion ratio of 6×/5× increases load power range by 3.76× and 5.48× in sleep and active mode, respectively. With the adaptive harvesting technique, it improves end-to-end harvesting efficiency by up to 1.92×, achieves the peak extraction efficiency of 99.2% for solar cell and the peak harvesting efficiency of 39.8%, and allows energy-efficient harvesting from solar cell, microbial fuel cell, and TEG. This work was presented and published in [11-12, 60].

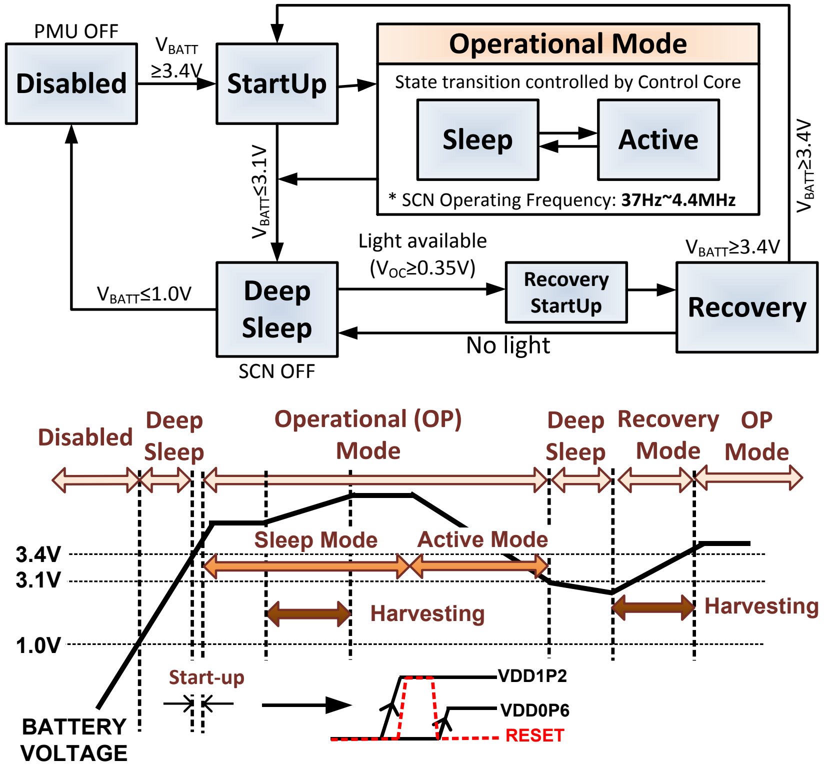


Figure 2.7. PMU FSM state diagram and PMU state change along with battery voltage variations.

CHAPTER 3

Reconfigurable Sleep Transistors for GIDL Reduction

A wide range of techniques have been proposed to reduce subthreshold leakage during sleep mode in ultra-low power systems [24], including aggressive voltage scaling [25], dynamic body biasing [28], and sleep transistors [26-29]. However, such techniques are not necessarily effective in suppressing other leakage contributions such as gate-induced drain leakage (GIDL) of sleep transistors, which dominates at battery voltages (V_{DD}) of 3V or higher. Such battery voltages are commonly observed in highly integrated mm-sized sensor nodes, as they typically rely on batteries with high energy density. For example, Cymbet printed batteries with mm^2 size have a typical operating voltage range of 3.6-4.0V [22-23].

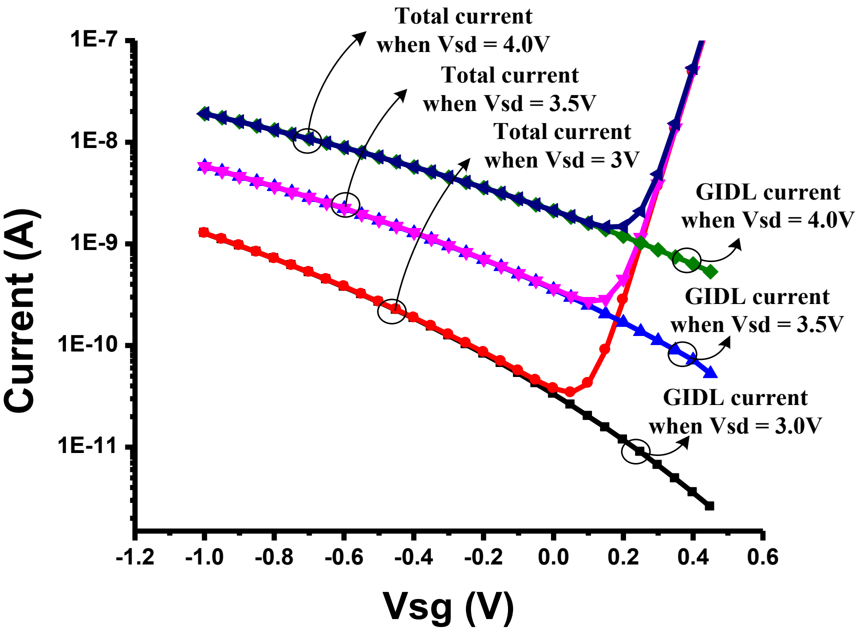


Figure 3.1. Simulated GIDL and overall leakage of p-MOSFET transistor (W/L=1012) versus V_{SG}.

The importance of GIDL at such voltages is shown in Figure 3.1, which plots GIDL and overall leakage of a PMOS transistor versus V_{SG} for V_{DD} of 3, 3.5V, and 4V in 180nm CMOS (note that older technologies are often used in sensor nodes due to low leakage, and the sensor's relatively low performance requirements). Figure 3.1 shows that GIDL becomes dominant in the overall leakage budget for higher V_{SD}, even at V_{SG}=0 V. Hence, new techniques to suppress GIDL in sleep transistors are needed to prolong battery lifetime in mm-size sensor nodes using ultra-compact batteries.

This chapter introduces the concept of reconfigurable sleep transistors to suppress GIDL, in addition to subthreshold leakage. Various topologies based on this concept are proposed and compared to existing solutions through measurements of a 180nm test chip at various temperatures. Reconfigurable sleep transistors are demonstrated to enable 12.6× standby leakage reduction at V_{DD} = 4.0V and T = 25^{o}C, at acceptable area overhead and no active mode performance penalty.

3.1 Sleep Transistor Leakage and Proposed Reconfigurable Topology

3.1.1 Sleep transistor leakage

We begin by considering the typical stacked sleep transistors as shown in Figure 2b, which are preferred over the single sleep transistor in Figure 3.2(a) for extremely low subthreshold leakage [26,29]. Transistors in Figure 3.2 are conventionally sized to set the on-resistance in active mode (and hence the voltage drop) lower than a given specification. An example constraint would be that the virtual supply VV_{DD} remains close to the supply voltage V_{DD} (e.g., within 5-10%). In sleep mode, the sleep transistor conducts leakage current that is dominated by either GIDL or subthreshold leakage, depending on the value of V_{DD}. These leakage mechanisms differ significantly since the former increases with reverse body biasing and negative gate biasing (i.e., super cut-off), in contrast to the latter. GIDL current of a p-MOSFET can be expressed according to the BSIM4 model [30] as

$$I_{GIDL} = AGIDL \cdot W_{eff} \cdot N_f \cdot \frac{V_{gd} - EGIDL}{3 \cdot T_{oxe}} \cdot \exp\left(-\frac{3 \cdot T_{oxe} \cdot BGIDL}{V_{gd} - EGIDL}\right) \cdot \frac{V_{bd}^3}{CGIDL + V_{bd}^3} \tag{3.1}$$

where $AGID$, $BGIDL$, $CGIDL$, and $EGIDL$ are technology-dependent GIDL coefficients, V_{sd} and V_{sge} are the source-drain and effective source-gate voltages, respectively, T_{oxe} is the electrical gate equivalent oxide thickness, and V_{bd} is the bulk-drain voltage. Here, (3.1) was obtained by substituting $V_{sd} - V_{sge} = V_{gd}$ into BSIM4 standard equations. From (3.1), GIDL current is reduced by minimizing voltages V_{gd} and V_{bd}, while channel width W_{eff} cannot be used as a knob for GIDL since it is sized according to active mode performance requirements as discussed above.

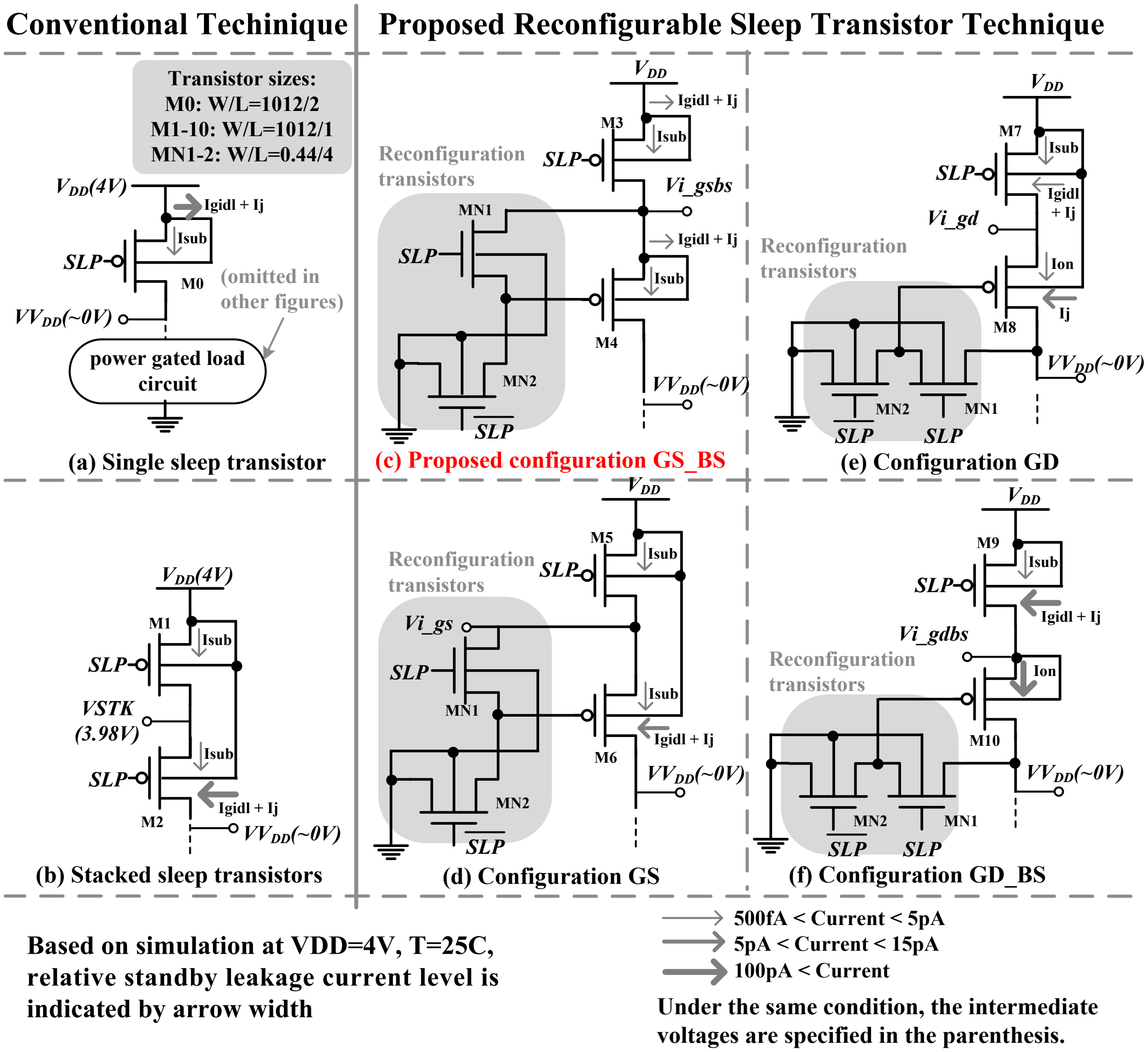

Conventional Techinique
Proposed Reconfigurable Sleep Transistor Technique
Transistor sizes:
M0: W/L=1012/2
M1-10: W/L=1012/1
MN1-2: W/L=0.44/4
VDD(4V)
Igidl + Ij
Isub
SLP
M0
(omitted in other figures)
VVDD(~0V)
power gated load circuit
(a) Single sleep transistor
VDD
M3
Isub
Reconfiguration transistors
SLP
Vi_gsbs
MN1
M4
MN2
VVDD(~0V)
(c) Proposed configuration GS_BS
M7
Vi_gd
Ion
Ij
M8
(e) Configuration GD
VDD(4V)
M1
VSTK (3.98V)
M2
(b) Stacked sleep transistors
M5
Vi_gs
M6
(d) Configuration GS
M9
Vi_gdbs
M10
(f) Configuration GD_BS
Based on simulation at VDD=4V, T=25C, relative standby leakage current level is indicated by arrow width
500fA < Current < 5pA
5pA < Current < 15pA
100pA < Current
Under the same condition, the intermediate voltages are specified in the parenthesis.

Figure 3.2. Six different sleep transistor configurations: configuration GS_BS is proposed for best GIDL reduction. Transistor sizes are indicated. Load is connected to drain of the bottom transistor in each configuration (M0, M2, M4, M6, M8, M10). GS stands for gate and source connected in sleep mode in the bottom transistor. GD stands for gate and drain connected in sleep mode. BS means body and source connected in sleep mode. BS means body and source of the bottom transistors are tied. I_{gidl}, I_j, I_{sub} and Ion refer to GIDL, junction leakage, subthreshold leakage and on-current, respectively.

From the exponential dependency of GIDL on V_{DD} (through V_{gd} and V_{bd}) in (3.1), GIDL becomes negligible and subthreshold leakage dominates at low supply voltages. The stack configuration in Figure 3.2(b) is an effective way to reduce subthreshold leakage due to the stack effect in M1-M2, and is commonly used to reduce standby power in low voltage systems where subthreshold leakage dominates.

3.1.2 Proposed reconfigurable sleep transistors

While the configuration of Figure 3.2(b) achieves ultra-low subthreshold leakage, GIDL becomes a serious issue at larger supply voltages since both V_{gd} and V_{bd} in (3.1) tend to be high (i.e., close to V_{DD}). Specifically, the gate and body voltage of M2 are at V_{DD} and its drain voltage is very close to ground, due to the large voltage drop across well-designed sleep transistors [10]. Ideally GIDL would be significantly reduced by setting the gate and/or body voltage of M2 such that V_{gd} and V_{bd} in (3.1) are minimized. The main idea of this chapter is to minimize GIDL by reconfiguring the topology in Figure 3.2(b) to set the voltages across transistor M2 as follows:

- In active mode (SLP is low), the gate voltages of M1 and M2 (i.e., the sleep signal) are equal to 0, so that active mode operation is the same as in Figure 3.2(b);
- In sleep mode (SLP is high), voltages V_{gd} and V_{bd} of off-transistors are reduced compared to Figure 3.2(b) by setting their gate and/or body voltages lower than V_{DD}.

Such reconfiguration can be conceptually obtained by connecting the gate terminal of the bottom stacked transistor (M2 in Figure 3.2(b)) to ground only in active mode, while connecting it to an existing node with potential lower than V_{DD} in sleep mode (e.g., the intermediate node between the two sleep transistors or VV_{DD}). This permits GIDL reduction without the need for generating an *ad hoc* voltage. To this aim, two reconfiguration transistors (MN1-MN2) are added in Figures 3.2(c)-3.2(f), which show four options found to be promising through preliminary simulations. The detailed operation of these topologies and their potential leakage advantages are detailed in Section 3.2.

3.2 Simulation & Measurement Results of Proposed Reconfigurable Sleep Transistor Topologies

The proposed sleep transistors in Figures 3.2(c)-3.2(f) are sized to have the same on-resistance as the reference stacked sleep transistors in Figure 3.2(b). The area penalty of topologies in Figures 3.2(c)-3.2(f) compared to Figure 3.2(b) is due to the additional reconfiguration transistors MN1-MN2, as well as additional area for a separate body contact for M4 (M10) in Figure 3.2(c) (Figure 3.2(f)). The area penalty due to reconfiguration transistors is low since these transistors can be small (e.g., minimum size), considering that they do not affect the active mode on-resistance. The size of MN1-MN2 is constrained only by the required time to switch from/into sleep mode. The additional area of separate body contacts is also small compared to the large sleep transistors. Typical sleep transistor sizes for the above topologies are reported at the bottom of Figure 3.2, where 10% virtual V_{DD} degradation at V_{DD}=3.25 V and T=25°C was targeted, assuming a 9.3-mA load in active mode.

3.2.1 Active mode

In active mode, configuration GD_BS (Figure 3.2(f)) and GS_BS (Figure 3.2(c)) were

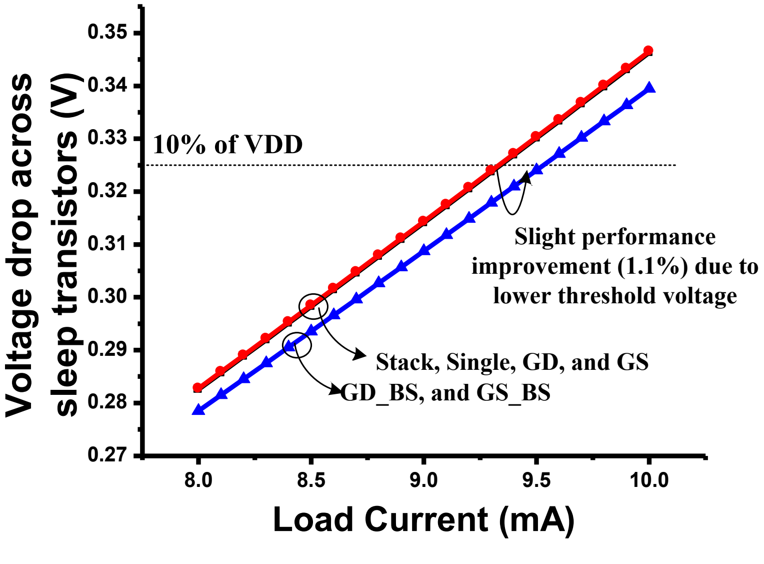


Figure 3.3 Simulated voltage drop across sleep transistors in active mode (VDD=3.25 V, T=25 °C).

found to have the lowest virtual V_{DD} voltage degradation among the considered sleep transistor configurations, although this advantage was found to be limited (about 1%). This small advantage arises since the bottom transistor in configurations GD_BS and GS_BS (M4 and M10 in Figures 3.2(c) and 3.2(f), respectively) has its body connected to its source, thereby canceling body effect and reducing $|V_{TH,p}|$ compared to the other topologies. This is shown in Figure 3.3, which plots the simulated voltage drop across the sleep transistors in active mode versus the current drawn by the power gated load (ranging from 8 to 10 mA) under the above conditions.

3.2.2 Sleep mode

All circuits in Figure 3.2 were fabricated using thick-gate MOSFET in 180nm CMOS to comparatively evaluate the sleep mode leakage reduction over the traditional topologies in Figures 3.2(a)-3.2(b). The adopted sleep transistors sizes are shown in Figure 3.2. The overall leakage current in each configuration was measured by sweeping the battery voltage V_{BAT} from 1.25V to 4V, at 25, 50, and 75°C. Measured results are reported in Figure 3.4. Figure 3.4 shows that the proposed configurations GS_BS, GS, and GD significantly reduce leakage for voltages higher than 2.75V at T=25°C, mainly due to the reduction of the dominant GIDL leakage component. In particular, GS_BS has the lowest leakage at V_{DD}>2.75 V, thanks to the symmetry of transistors M3-M4 in Figure 3.2(c), setting the intermediate voltage *Vi_gsbs* close to $V_{DD}/2$, which in turn reduces V_{bd}, V_{gd} and GIDL in both M3 and M4 (see (1)). Figure 3.5 shows the measured value of the intermediate voltage between the sleep transistors in the various configurations; it can be seen that GS_BS tracks $V_{DD}/2$ fairly well, leading to small V_{gd}values. Similarly, GS has lower GIDL than traditional stacked transistors since the intermediate voltage *Vi_gs* in Figure 3.2(d) is lower than *VSTK* in Figure 3.2(b) (as shown in Figure 3.5, *VSTK* is very close to V_{DD} because of reverse gate biasing of M2 in Figure 3.2(b)). Since gate and source are short-circuited by MN1 in sleep mode, this results in a V_{gd} reduction in M6 (see Figure 3.2(d)), thereby reducing GIDL leakage from (1).

Analogously, GD has lower GIDL than traditional stacked transistors and GD_BS since gate and drain of M8 are short-circuited by MN1 in sleep mode. In GD and GD_BS, M8 and M10 are diode-connected and are still on in sleep mode, so GIDL from these devices is negligible and GIDL from only M7 and M9 are of interest. The intermediate voltage of GD (*Vi_gd* in Figure 3.2(e)) is determined by negative feedback between GIDL current and body effect in M8.

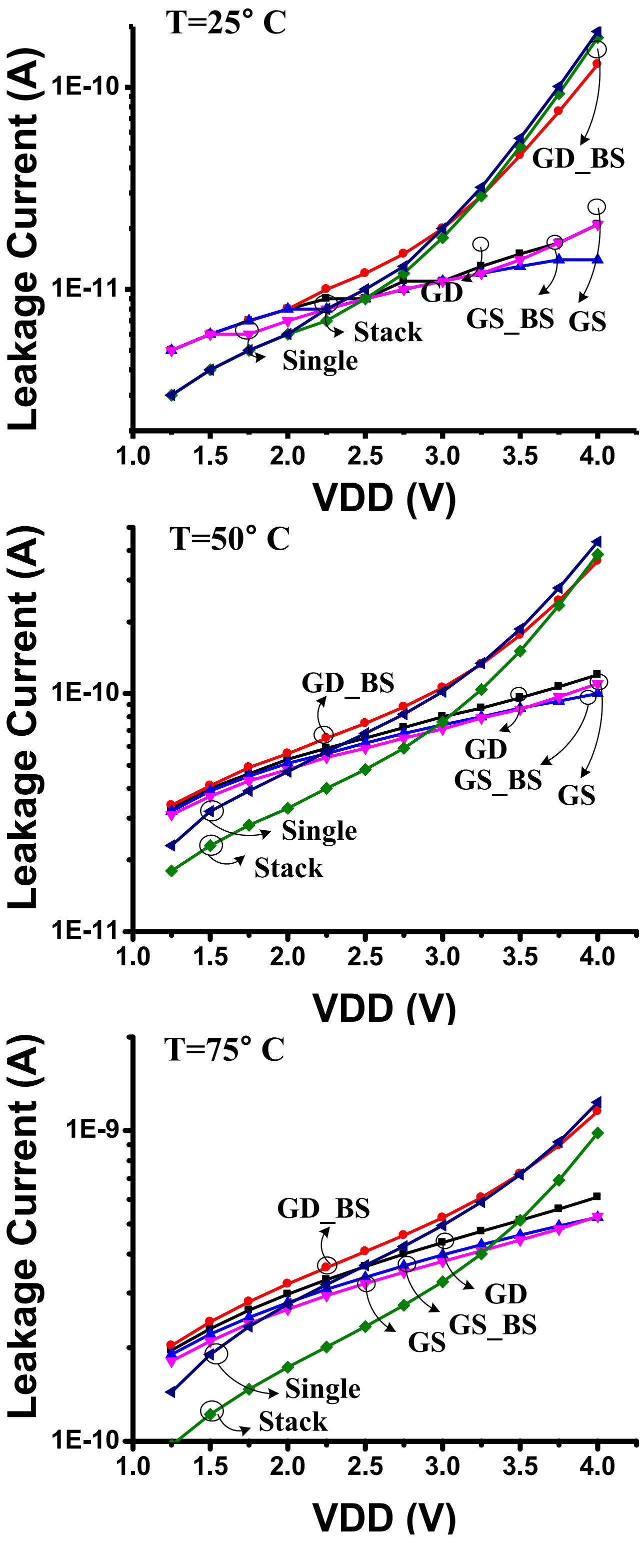


Figure 3.4. Measurement results of overall leakage for when sleep transistor is in sleep mode at 25°C (left), 50°C (center), and 75°C (right).

M7 GIDL current in Figure 3.2(f), but reduces M8 on-current in Figure 3.2(f) due to body effect. As a result Vi_gd is higher and closer to $V_{DD}/2$ than Vi_gdbs. On the other hand, GD_BS does not provide any GIDL reduction since intermediate voltage Vi_gdbs is close to 0 V (see Figure 3.5) due to the zero body-source and gate-drain voltage of M10 in sleep mode. This increases V_{gd} and V_{bd} (and hence GIDL) for M9, compared to M1 in Figure 3.2(b). For these reasons, GD_BS provides no leakage reduction at either high or low V_{DD}, and its leakage profile resembles that of a single sleep transistor as Vi_gdbs is clamped to approximately the threshold voltage of M10.

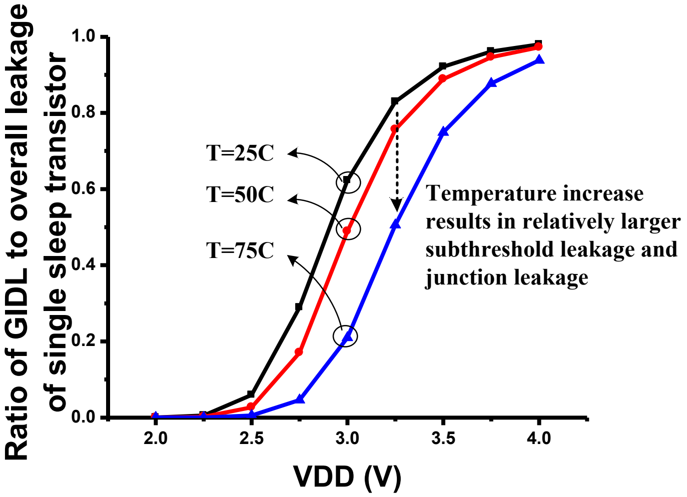


Figure 3.5. Simulated ratio of GIDL to overall leakage of a single sleep transistor (W/L=1012/2) in sleep mode vs. VDD at different temperatures.

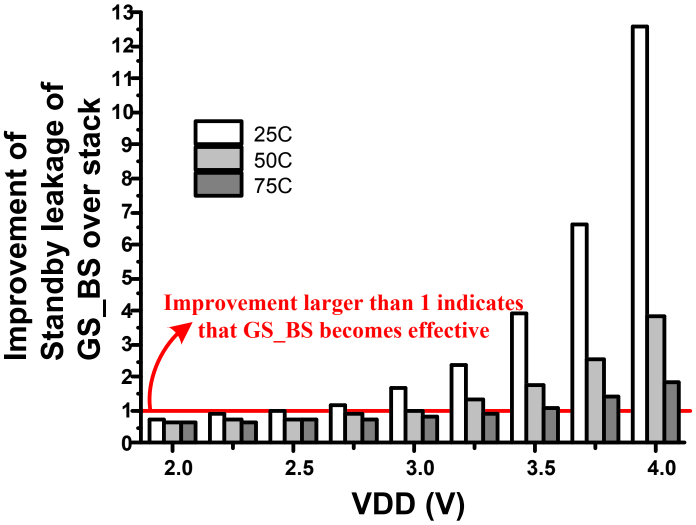


Figure 3.6. Standby leakage improvement(I_stack/I_(GS_BS)) of proposed configuration GS_BS over stack.

The above considerations are summarized in Figure 3.2, where the relative leakage current magnitude is qualitatively indicated by the arrow width, based on simulation results at V_{DD}=4V and T=25°C. Among the proposed topologies, GS_BS has the lowest GIDL and overall leakage. Regarding GS, it has higher V_{bd} in M6 (its body is tied to V_{DD}), which in turn leads to higher GIDL and junction leakage compared to GS_BS. GD exhibits higher junction leakage than GS_BS due to a higher V_{bd} in M8 compared to M4.

As expected from qualitative considerations in Section 3.1, the proposed reconfigurable sleep transistor GS_BS, GS, and GD offer a substantial leakage reduction at moderately high V_{DD} through GIDL mitigation, whereas they provide no benefit at low voltages at which subthreshold leakage dominates. This is shown in Figure 3.6, which depicts the simulation results of the total and non-GIDL leakage for configuration GS_BS and the traditional stacked version (Figure 3.2(b)) at 25°C. Regarding the temperature dependence, higher temperatures exponentially increase subthreshold leakage, while weakly affecting GIDL leakage, as shown in Figure 3.7. GIDL therefore becomes a smaller fraction of total leakage at higher temperatures, hence the GIDL reduction offered by configuration GS_BS has a smaller impact. In particular, our measurements show that configuration GS_BS has the lowest leakage for V_{DD}>2.5 V at T=25°C (representative of most applications in ultra-low power sensors since no self-heating does not occur), for V_{DD}>3

Temp.	Config.	single	stack	GS_BS	GS	GD	GD_BS
25 C	Mean	269.60	269.40	**2.25**	16.86	12.73	146.70
	S.D.	63.90	63.83	**0.65**	4.82	2.03	39.81
50 C	Mean	359.50	358.60	**9.60**	28.95	22.82	235.50
	S.D.	80.82	80.69	**3.89**	8.96	5.58	60.45
75 C	Mean	482.70	478.80	**46.19**	68.00	63.35	379.30
	S.D.	107.00	106.80	**18.62**	22.98	20.67	96.38

Table 3.1. Statistical Process and Mismatch Variation Simulation Results of Total Standby Leakage Current (pA) at Different Temperatures and VDD =4V (500 runs for each).

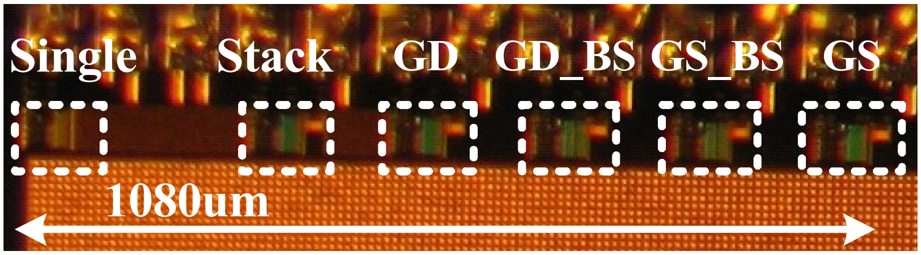


Figure 3.7. Die photo: Configuration single, stack, GD, GD_BS, GS_BS, GS from left.

V at T=50°C (possible in very high ambient temperatures), and for V_{DD}>3.25~3.5 V at T=75°C. As shown in Figure 3.8, configuration GS_BS can reduce leakage by 2.41× at V_{DD}=3.25 V and by 12.6× at V_{DD}=4 V (T=25°C), compared to stacked transistors.

To consider the impact of process and mismatch variation, statistical simulation results (Monte Carlo) at different temperatures (T=25/50/75 °C) and V_{DD} =4V are summarized in Table 3.1. The results show the consistent effectiveness of GS_BS even under process and mismatch

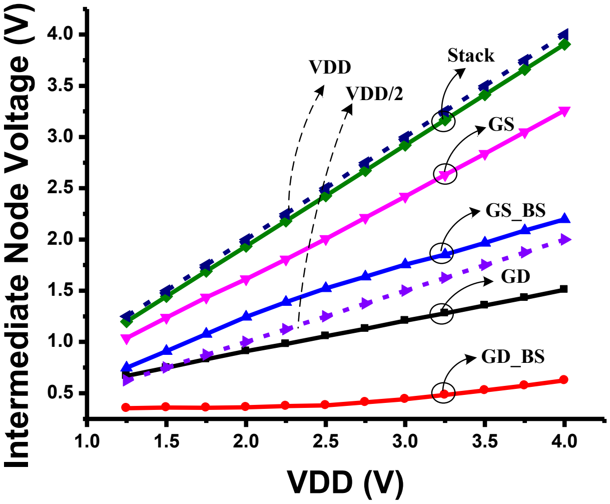


Figure 3.8. Measured intermediate voltages of proposed sleep transistor configurations and traditional stack sleep transistor in sleep mode T=25°C.

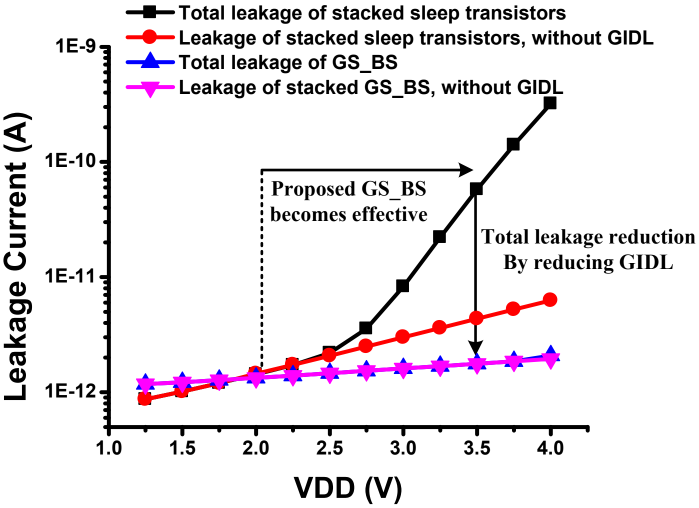


Figure 3.9. Simulated total and non-GIDL leakage vs. VDD in configuration GS_BS (Figure 3.2(c)) and traditional stack (Figure 3.2(b)) at 25°C.

variation. The die photo is shown in Figure 3.7.

3.3 Summary

We have proposed novel reconfigurable sleep transistors to reduce GIDL current in battery-driven ultra-low power sensors. Experimental comparisons with traditional configurations in 180nm CMOS were made to evaluate the efficacy of the new techniques. The proposed configuration GS_BS reduces leakage by up to 12.6× at VDD=4 V and room temperature. Such an advantage comes with acceptable area penalty and no degradation in active mode performance. This work was presented in [61].

CHAPTER 4

Switched-Capacitor DC-DC Converter with Fine-Grained Conversion Ratios in Successive Approximation Approach

Efficient power conversion of SC converters is performed around available discrete conversion ratios, and only one pair of frequency and output voltage level offer peak efficiency at a given conversion ratio and load current. Operating points that deviate from these optimal ones result in efficiency degradation. Adding more conversion ratios in conventional SC converters can help provide greater design flexibility, but this increases implementation complexity, often leading to increased losses and degraded efficiency. For instance, series-parallel SC converters that are reconfigurable with several conversion ratios have irregular structures with additional switches, which limit their reconfigurability while also harming efficiency [32- 38].

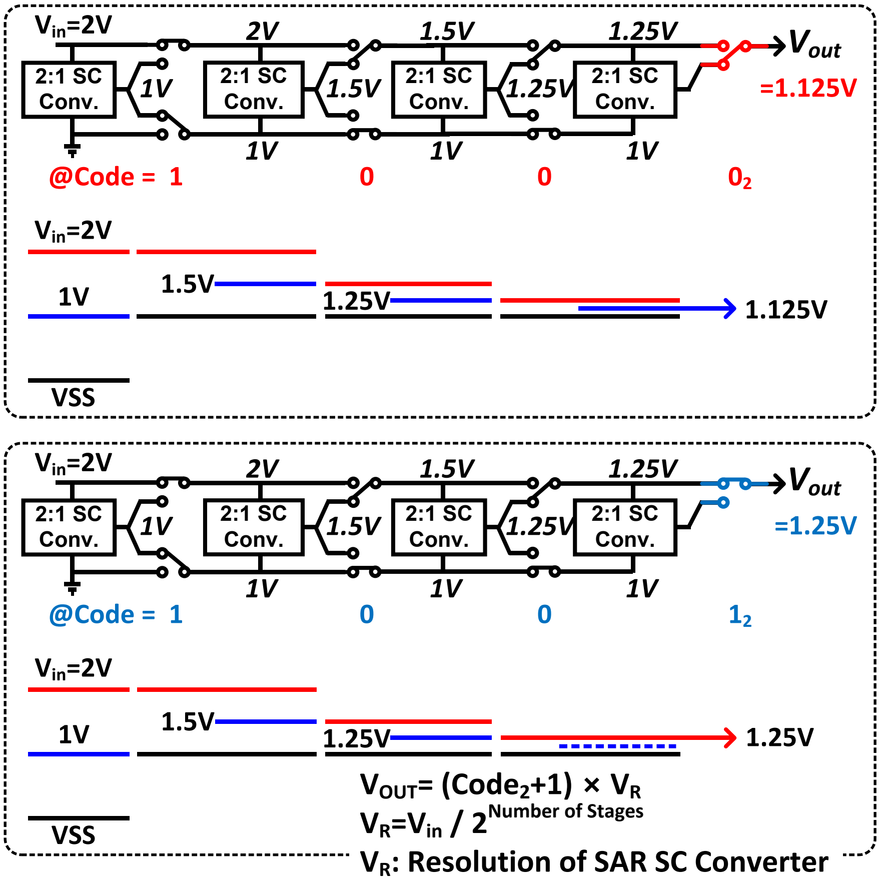
This chapter presents a successive-approximation switched-capacitor (SAR SC) DC-DC converter that provides a conversion ratio resolution of $V_{IN}/2^{\text{Number of Stages}}$ [40]. A SAR SC converter cascades multiple stages of 2:1 SC converters, The proposed SAR SC converter offers a large number of conversion ratios, and has smaller slow-switching limit impedance, or charge-sharing loss [40], and smaller switching loss due to bottom-plate parasitic capacitor than other conventional SC converters with the same number of conversion ratios. In the SAR SC converter, the 2:1 SC converter cell in each stage generates the voltage resolution set by the number of stages between the 1st stage and the corresponding stage regardless of conversion ratio, and a conversion ratio is successively obtained between neighboring stages. Thus, minimal change of configuration is required in case of conversion ratio adjustment.

4.1 Successive-Approximation Switched-Capacitor Converter

4.1.1 Concept of Successive-Approximation Switched-Capacitor Converter

Figure 4.1 describes the operating principle of the SAR SC DC-DC converter. The main idea is to cascade multiple 2:1 SC-stages using configuration switches to obtain a fine grain output voltage (V_{OUT}). Each SC-stage takes two inputs (V_{high}, V_{low}) and produces an output $V_{mid} = (V_{high}+V_{low})/2$. The high voltage of the next stage is connected to either the high or mid voltage of the previous stage. The low voltage of the next stage is connected to either the mid or low voltage of the previous stage. In the 4-stage example of Figure 4.1, V_{in} = 2V is converted to V_{OUT} = 1.125V with configuration code = 1000_2 and to V_{OUT} = 1.250 with code = 1001_2, providing a 125mV step under no load condition.

With configuration code = 1000_2, the 1st stage converter output is 1V, representing the average of 2V and 0V. Thus, the second stage takes 2V and 1V as high and low voltage. The



V_{OUT}= ($Code_2$+1) × V_R

V_R=V_{in} / $2^{\text{Number of Stages}}$

V_R: Resolution of SAR SC Converter

Figure 4.1. A conceptual example of 4-b SAR SC DC-DC converter operation for code = 1000_2 (top) and 1001_2 (bottom).

MSB is one and controls the switch configuration between the 1st and 2nd stages. Hence, the mid voltage of the 2nd stage becomes 1.5V. The high voltage of the 3rd stage is connected to the mid voltage of the 2nd stage, and the low voltage of the 3rd stage is connected to the low voltage of the 2nd stage. In other words, the 3rd stage takes 1.5V and 1V as its high and low voltages, since the 2nd bit is 0. As a result, the mid voltage of the 3rd stage becomes 1.25V. Since the 3rd bit is also 0, the mid voltage of the final stage becomes 1.125V. To generate an output of 1.125V, the mid voltage of the final stage is connected to the final output, and the LSB is set to 0. On the other hand, in order to make the output 1.25V, the high voltage of the final stage is connected to the final output, and the LSB is set high, resulting in code = 1001_2. Note that SAR converter can generate an output voltage level that is successively approximated, so each stage of 2:1 converter is designed to generate a fixed level of voltage. When V_{IN}, target voltage, or load current changes, only a successive (i.e, minimal) change of configuration is required for regulation.

In summary, output voltage is given as Equation (4.1).

$$\mathrm{V_{out}} = (\mathrm{Code_2} + 1)\times \mathrm{V_R} \tag{4.1}$$

where $\mathrm{V_R}$ is SAR SC converter resolution is given as Equation (4.2).

$$\mathrm{V_R} = \mathrm{V_{in}}/2^{\mathrm{number\ of\ stages}} \tag{4.2}$$

For a 4b design with 2V input, resolution becomes 125mV. Hence, the key benefit of the proposed converter is the possibility of very fine output voltage resolution over a wide output voltage range.

4.1.2 Current Flow and Sizing of Successive-Approximation Switched-Capacitor Converter

Depending on conversion ratio and the corresponding code, the SAR SC converter is reconfigured and the current flow through the SC converter is changed. We explain the current flow with an example using a 4-b SAR SC converter, as illustrated in Figure 4.2. With code 1000_2, a conversion ratio of 9/16 is achieved, and each 2:1 SC converter should output current of 7/8 × I_o, ¼ × I_o, ½ × I_o, and 1 × I_o on average, where I_o is the output load current. On the other hand, with code 1010_2, a conversion ratio of 11/16 is achieved, and each 2:1 SC converter outputs currents of 5/8 × I_o, ¾ × I_o, ½ × I_o, and 1 × I_o on average. In addition, the current output of the LSB 2:1 SC converter stage is either zero or I_o according to the LSB code. As such, current flow in the SAR SC converter can be generalized as follows:

$$I_{H(0)} = \frac{1}{2}\overline{C_0}I_o \tag{4.3}$$

$$I_{L(0)} = \frac{1}{2}\overline{C_0}I_o \tag{4.4}$$

$$I_{o(1)} = \overline{C_1}C_0I_o + \left(\overline{C_1}I_{H(0)} + C_1I_{L(0)}\right) \tag{4.5}$$

$$I_{H(1)} = I_{L(1)} = \frac{1}{2}I_{o(1)} \tag{4.6}$$

Similarly,

$$\begin{aligned} I_{o(k)} = \overline{C_k} \cdot \textstyle\sum_{i=0}^{k-2}\left(\prod_{j=i+1}^{k-1} C_j\right)I_{H(i)} + C_k \cdot \sum_{i=0}^{k-2}\left(\prod_{j=i+1}^{k-1} \overline{C_j}\right)I_{L(i)} \\ +\left(\overline{C_k}I_{H(k-1)} + C_kI_{L(k-1)}\right) \end{aligned} \tag{4.7}$$

$$I_{H(k)} = I_{L(k)} = \frac{1}{2}I_{o(k)} \tag{4.8}$$

where code is given as $\overline{C_{n-1}C_{n-2}\cdots C_1C_0}_{(2)}$ and $I_{o(k)}, I_{H(k)}, I_{L(k)}$ are the average output current, the current flow from V_{high}, and the average current flow from V_{low} of (k+1)-th stage,

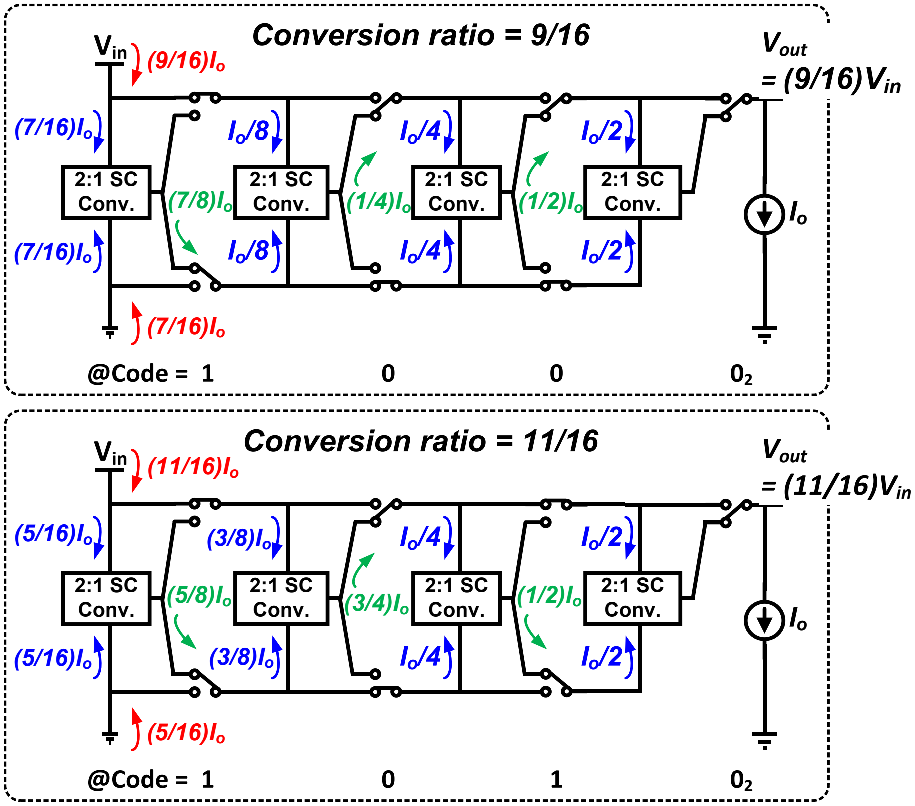


Figure 4.2. Current flow of 4-b SAR SC converter for code = 1000_2 (top) and 1010_2 (bottom).

respectively. Due to varying levels of current flow in each stage of 2:1 SC converter depending on configuration, 2:1 SC converters in all stages are sized identically in our implementation.

4.1.3 Loss Analysis and Comparison against Switched-Capacitor Converters

It is widely known that switched-capacitor (SC) converter has two main loss factors: output impedance and switching loss [19, 54]. In the following sub-sections, slow-switching limit impedance, which represents output impedance under assumption of ideal switches and interconnects, and switching loss due to bottom-plate parasitic capacitor, which mainly contributes to total switching loss, will be investigated in generalized forms for successive-approximation (SAR), series-parallel and ladder topologies of SC converters.

4.1.3.1 Analysis of Slow Switching Limit Impedance for Non-constant Output Voltage

Slow switching limit impedance (R_{SSL}) is the low-frequency output impedance (R_o) of SC converter under the assumption that switches and all other interconnects are ideal [40]. R_{SSL} arises from charge-sharing mechanism of SC converters. R_o determines the maximum available output of a SC converter, and smaller R_o and R_{SSL} indicate higher driving capability of output power. In prior works [40, 44], R_{SSL} is derived for various topologies of SC converters under the assumption that output voltage is constant. However, in practice, only finite decoupling output capacitance is available, and hence the output voltage cannot be a constant level and instead

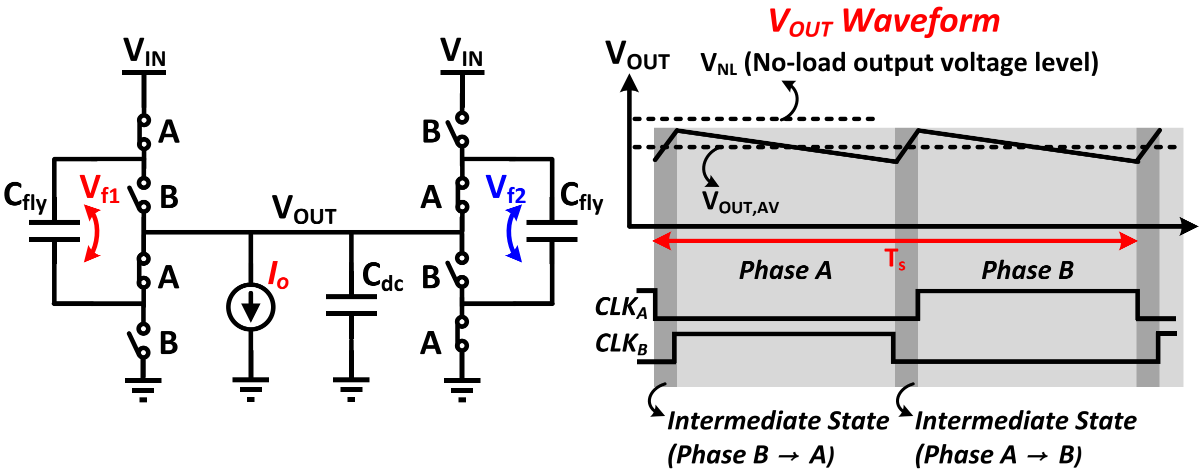


Figure 4.3. A 2:1 SC converter with two-phase interleaving (left), and V_{OUT} and clock waveforms (right).

varies periodically with input clocking and current flow during non-transition phases in addition to instantaneous charge flow during clock transition states must be considered for R_{SSL} as opposed to the analysis in [40, 44]. Thus, a different approach is required to derive R_{SSL} with non-constant output voltage. This section covers such an R_{SSL} derivation and compares R_{SSL} with non-constant output voltage level across various SC converter topologies.

R_{SSL} is inversely proportional to switching frequency and capacitance, so R_{SSL} can be generalized as $\frac{K_{SSL}}{F_s C_{tot}}$, where F_S is switching frequency, C_{tot} is total flying capacitance, and K_{SSL} is the R_{SSL} coefficient, determined by the topology of the given SC converter and stated as a function of flying capacitance (C_{fly}) and decoupling output capacitance (C_{dc}). K_{SSL} determines performance of the SC converter. For instance, SC topologies with small K_{SSL} values have small output voltage drop due to R_{SSL}, and they can supply larger output currents for a given switching frequency and flying capacitance than ones with large K_{SSL}. In general, output voltage of an SC converter can be written as equation 4.9.

$$V_{OUT,AV} = V_{NL} - R_{SSL}I_o = V_{NL} - \left(\frac{K_{SSL}}{F_s C_{tot}}\right) \times I_o \qquad (4.9)$$

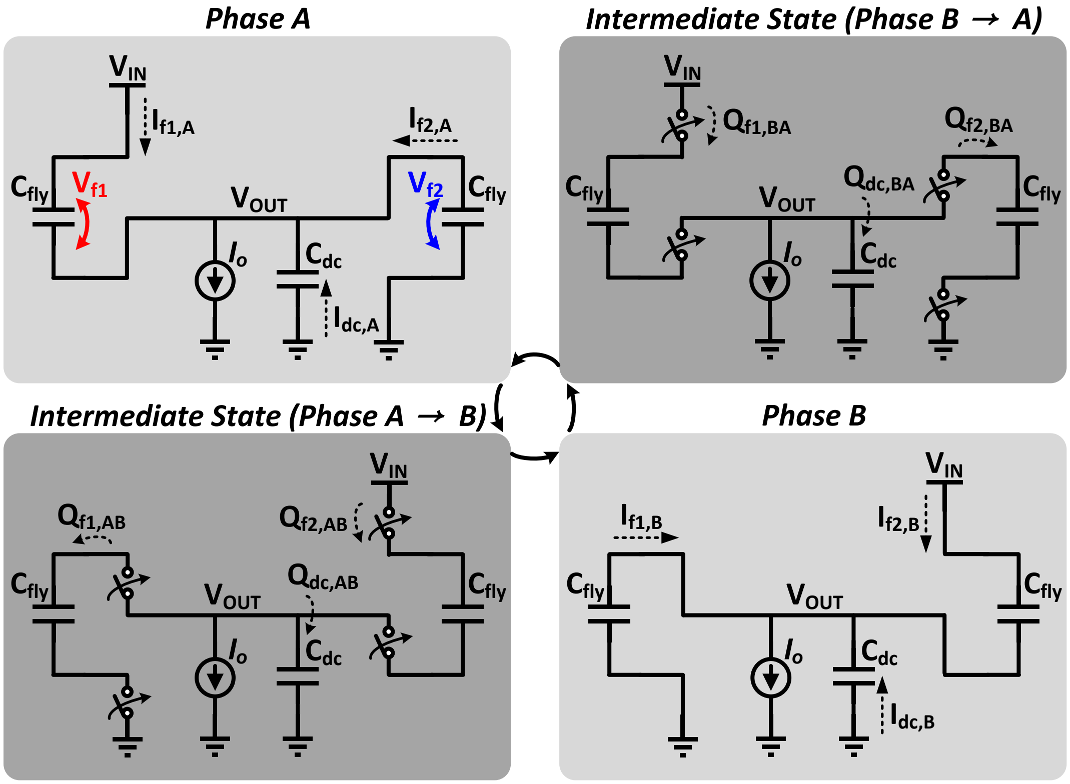


Figure 4.4. Current and instantaneous charge flow of a 2:1 SC converter with two-phase interleaving at each phase and intermediate state.

where V_{NL} is no-load output voltage, and I_o is output current.

Given the SC converter topology, the expression for K_{SSL} can be derived mathematically. A 2:1 SC converter with two-phase interleaving can be taken as an example, as illustrated in Figure 4.3. The V_{OUT} waveform along with two clocks (CLK_A and CLK_B) driving the converter switches are shown. The flying capacitor (C_{fly}) connection is periodically switched; it is connected between V_{IN} and V_{OUT} for one phase, and then connected between V_{OUT} and VSS for the next phase. To avoid short-circuit current that results in charge loss, an intermediate state between the two phases exists, where CLK_A and CLK_B are both zero and C_{fly} is floating. In practice, the SC converter stays in the intermediate state for a short period of time, but it is important to consider it for understanding charge flow during SC converter operation. Figure 4.4 illustrates the current (denoted as I) and instantaneous charge flow (denoted as Q) of the 2:1 SC converter with two-phase interleaving, and Figure 4.5 provides detailed waveforms of V_{OUT} and voltages across C_{fly}'s (V_{f1} and V_{f2}).

In deriving the K_{SSL} expression, the following naming conventions are used. The first subscript indicates what element the current or charge relates to, namely “f1” indicates left C_{fly}, “f2” indicates right C_{fly}, and “dc” indicates C_{dc}. The second subscript denotes the relevant phase or intermediate state. “A” is phase A, “B” is phase B”, “AB” is the intermediate state when transitioning from phase A to B, and “BA” represents the intermediate state when moving from phase B to A.

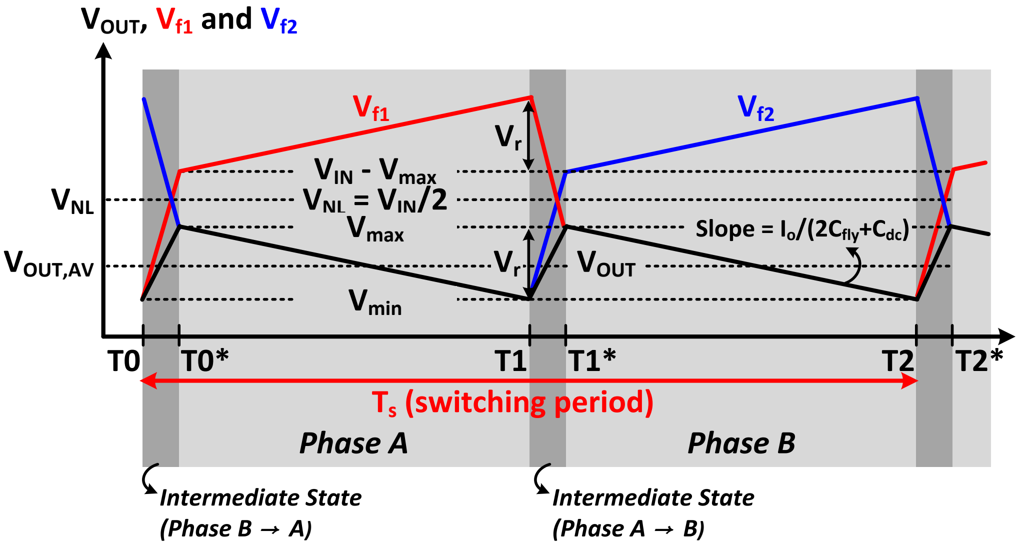


Figure 4.5. Waveforms of output voltage (V_{OUT}) and voltages across flying capacitors (V_f1 and V_{f2}) of 2:1 SC converter with two-phase interleaving.

As the number of interleaved phases is two, the following equations are valid.

$$I_{f1,A} = I_{f2,B} \tag{4.10}$$

$$I_{f1,B} = I_{f2,A} \tag{4.11}$$

$$I_{dc,A} = I_{dc,B} \tag{4.12}$$

$$Q_{f1,AB} = Q_{f2,BA} \tag{4.13}$$

$$Q_{f1,BA} = Q_{f2,AB} \tag{4.14}$$

$$Q_{dc,AB} = Q_{dc,BA} \tag{4.15}$$

In addition, the following current equations can be derived:

$$I_{f1,A} = \frac{C_{fly}}{2C_{fly}+C_{dc}} I_o \tag{4.16}$$

$$I_{f2,A} = \frac{C_{fly}}{2C_{fly}+C_{dc}} I_o \tag{4.17}$$

$$I_{dc,A} = \frac{C_{dc}}{2C_{fly}+C_{dc}} I_o \tag{4.18}$$

$$\text{Output voltage ripple: } V_r = \frac{T_s I_o}{2(2C_{fly}+C_{dc})} \tag{4.19}$$

Let us assume that the duration of the intermediate state is negligibly small (T0 ≅ T0*, T1 ≅ T1*, T2 ≅ T2*), and derive expressions for $Q_{f2,AB}$, $Q_{dc,AB}$, and $Q_{f1,AB}$. Since average input current $(I_{in})_{AV} = I_o/2$, charge influx from V_{IN} during half period of T_s is $\frac{I_o}{2} \times \frac{T_s}{2} = Q_{f2,AB} + I_{f1,A} \times \frac{T_s}{2}$, and thus $Q_{f2,AB}$ can be written as in Equation 4.20:

$$Q_{f2,AB} = \left(\frac{I_o}{2} - I_{f1,A}\right) \times \frac{T_s}{2} = \frac{C_{dc}}{2C_{fly}+C_{dc}} \times \frac{T_s I_o}{4} = \frac{V_r C_{dc}}{2} \tag{4.20}$$

In steady state, the charge lost from C_{dc} during phase A or B is replenished during intermediate states, hence $Q_{dc,AB}$ and $Q_{f1,AB}$ are written as in Equations 4.21 and 4.22 below:

$$Q_{dc,AB} = V_r C_{dc} \tag{4.21}$$

$$\text{Thus, } Q_{f1,AB} = Q_{f2,AB} - Q_{dc,AB} = -\frac{V_r C_{dc}}{2} \tag{4.22}$$

Now, let us define the minimum and maximum output voltage as V_{min} and V_{max}. Then,

$$V_{OUT}(T0) = V_{OUT}(T1) = V_{min} \tag{4.23}$$

$$V_{OUT}(T0^*) = V_{OUT}(T1^*) = V_{max} \tag{4.24}$$

$$V_{f1}(T0^*) = V_{IN} - V_{max} \tag{4.25}$$

$$V_{min} = V_{max} - V_r \tag{4.26}$$

To derive V_{max} as a function of V_{IN}, C_{fly}, C_{dc}, T_s, and I_o,

$$V_{f1}(T1) = V_{f1}(T0^*) + V_r = V_{IN} - V_{max} + V_r \tag{4.27}$$

$$V_{f1}(T1^*) = V_{f1}(T1) + \frac{Q_{f1,AB}}{C_{fly}} = V_{IN} - V_{max} + V_r - \frac{V_r C_{dc}}{2C_{fly}} \tag{4.28}$$

Since $V_{f1}(T1^*) = V_{max}$,

$$V_{max} = \frac{1}{2} \times \{V_{IN} - \frac{V_r(-2C_{fly}+C_{dc})}{2C_{fly}}\} = \frac{V_{IN}}{2} - \frac{T_s I_o(-2C_{fly}+C_{dc})}{8C_{fly}(2C_{fly}+C_{dc})} \tag{4.29}$$

$$V_{OUT,AV} = V_{max} - \frac{V_r}{2} = \frac{V_{IN}}{2} - \frac{T_s I_o(-2C_{fly}+C_{dc})}{8C_{fly}(2C_{fly}+C_{dc})} - \frac{T_s I_o}{4(2C_{fly}+C_{dc})} = \frac{V_{IN}}{2} - \frac{T_s I_o C_{dc}}{8C_{fly}(2C_{fly}+C_{dc})} \tag{4.30}$$

As $2C_{fly} = C_{tot}$ (total flying capacitance) and $T_s = 1/F_s$, a 2:1 SC converter with two-phase interleaving has K_{SSL} given by Equation 4.31. K_{SSL} is dependent on flying capacitance and decoupling output capacitance.

$$K_{SSL} = \frac{C_{dc}}{4(2C_{fly}+C_{dc})} \tag{4.31}$$

Topology	K_{SSL} ($C_{dc} = 0$)
SAR	$\propto \log_2 N$
Series-Parallel	$\frac{k^2(N-k)^2(N-2k)^2}{2N^2(k^2+(N-k)^2)}$
Ladder	$\frac{(N-1)\times(4N^3k^2-8N^2k^3 + 8N^2k^2 - N^2k - 3N^2 + 4Nk^4 - 16Nk^3 + Nk^2 + 4Nk - 3N + 8k^4 - 4k^2)}{12N(N+2)}$

Table 4.1. Generalized K_{SSL} expressions for various topologies when output capacitance (C_{dc}) is zero. k/N is conversion ratio, V_{OUT}/V_{IN}. It is assumed that flying capacitances in a SC converter are sized equal, and two-phase is interleaved.

Topology	K_{SSL} ($C_{dc} = \infty$)
SAR	$\propto \log_2 N$
Series-Parallel	$\frac{k^2(N-k)^2}{2N^2}$
Ladder	$\frac{(N-1)\times k(N-k)(-2k^2+2Nk+1)}{6N}$

Table 4.2. Generalized K_{SSL} expressions for various topologies when output capacitance (C_{dc}) is very large (∞). k/N is conversion ratio, V_{OUT}/V_{IN}. It is assumed that flying capacitances in a SC converter are sized equal, and two-phase is interleaved.

Generalized K_{SSL} expressions for a SAR SC converter, a Recursive SC (RSC) converter [44],series-parallel SC converter, and ladder SC converter with 2-phase interleaving and no output capacitance can be derived similarly; the derived expressions are summarized in Table 4.1 and Table 4.2, respectively for $C_{dc} = 0$ and $C_{dc} = \infty$. K_{SSL}'s of SAR SC and Recursive SC converters are not closed-form expressions, and they can be found using recursive MATLAB code. The K_{SSL} expressions were verified by comparisons against HSPICE simulation results. Difference between HSPICE simulation results with ideal switches and expected results of MATLAB modeling is less than 1%, regardless of load current, total flying capacitance and conversion ratio.

Figure 4.6 plots K_{SSL} for SAR SC converters, Recursive SC converters, series-parallel SC converters, and ladder SC converters with various conversion ratios for the given number of stages bits, when $C_{dc} = 0$. The number of bits indicated with series-parallel SC and ladder SC converters in Figure 4.6 represents the denominator of the conversion ratio. For instance, 4-bit series-parallel SC and ladder SC converters generate $k/2^4$ (k = odd integers greater than 0 and smaller than 2^4), It is shown that series-parallel SC and ladder SC converters have K_{SSL} increasing with N^2 and with N^4 when N is large and $k \simeq N/2$, where N is the number of available conversion ratios ($2^{\text{number of bits}}$) and k/N is conversion ratio. Although series-parallel SC converters perform better than ladder SC converters with the same number of bits in terms of

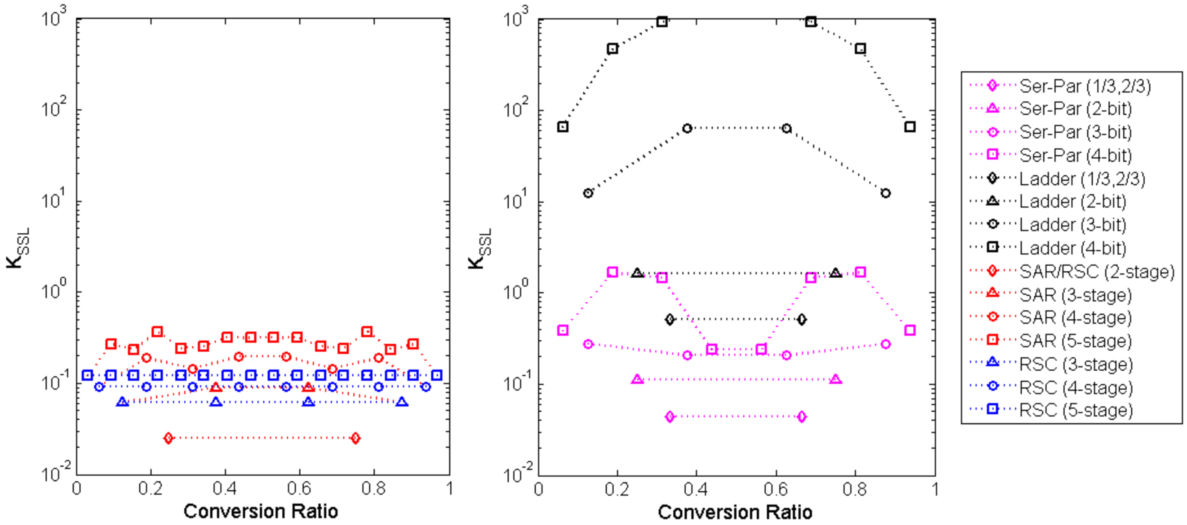


Figure 4.6. R_{SSL} coefficient (K_{SSL}) vs. conversion ratio for SAR SC and RSC converters (left); for series-parallel SC and ladder SC converters (right).

K_{SSL}, the irregular structure of series-parallel SC converters results in challenges when reconfiguring between multiple conversion ratios. In contrast, SAR SC converters have K_{SSL} that increases with $\log_2 N$, or the number of stages, and offer many conversion ratios with easy reconfiguration and less R_{SSL} overhead.

As R_o is an indicator of output power deliverable in SC converter, power density of an SC converter can be predicted with K_{SSL}. Equation 4.32 implies that power density is proportional to $C_{tot}/(\text{Area}\times K_{SSL})$, where C_{tot}/Area is defined as capacitance density determined by CMOS process technology, so $1/K_{SSL}$ can be used as a metric for comparison of power density in given CMOS process technology. In an SAR SC converter, power density degrades in proportion to $1/\log_2 N$, or 1/(number of stages).

$$\text{Power density} = \frac{P_L}{\text{Area}} = \frac{V_{OUT}I_L}{\text{Area}} \propto \frac{1}{R_o\times\text{Area}} \propto \frac{F_S C_{tot}}{K_{SSL}} \times \frac{1}{\text{Area}} \qquad (4.32)$$

4.1.3.2 Analysis of Switching Loss due to Bottom-Plate Parasitic Capacitors

Fully-integrated SC converters employ MIM or MOS capacitors for flying capacitors, and bottom-plate parasitic capacitors of these flying capacitors cannot be ignored in moderate load power condition where gate-driving loss and switch conductance loss are relatively small [54]. This is because switching frequency scales with load power, resulting in small gate-driving

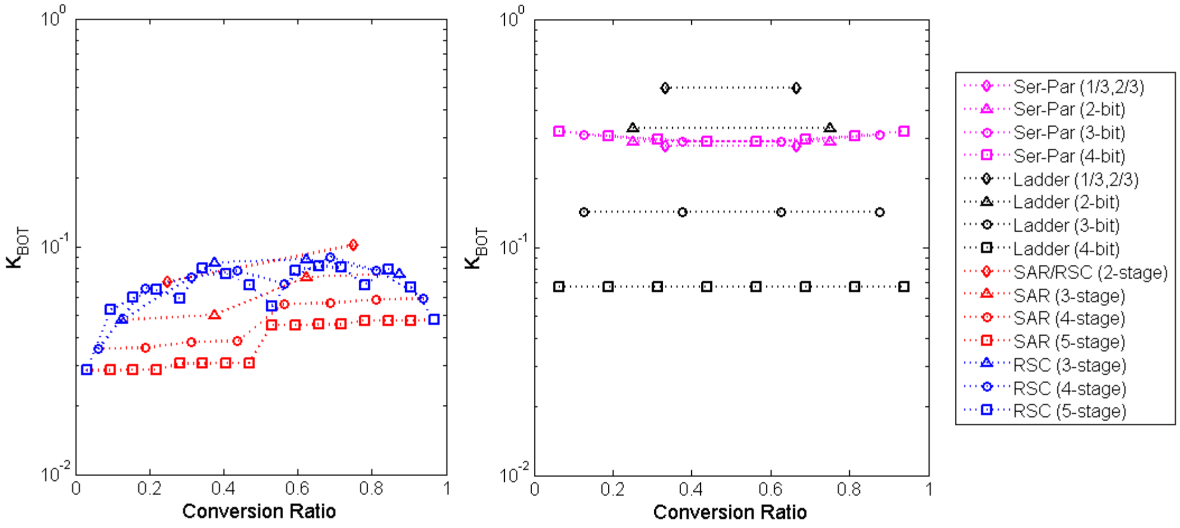


Figure 4.7. P_{BOT} loss coefficient (K_{BOT}) vs. conversion ratio for SAR SC and recursive RSC converters (left); for series-parallel SC and ladder SC converters (right).

loss in low load power, as discussed in [54]. Switching loss due to bottom-plate parasitic capacitors (P_{BOT}) can be generalized as Equation 4.33.

$$P_{BOT} = \sum_i \begin{matrix} C_{bot(i)}\{s_{ph1(i)}V_{bot(i),ph1}(V_{bot(i),ph1} - V_{bot(i),ph2}) \\ +s_{ph2(i)}V_{bot(i),ph2}(V_{bot(i),ph2} - V_{bot(i),ph1})\} \end{matrix} \times F_s \qquad (4.33)$$

where $C_{bot(i)}$ is a bottom-plate parasitic capcitance at node (i), $V_{bot(i),ph1}$ and $V_{bot(i),ph2}$ are potentials, $s_{ph1(i)}$ and $s_{ph2(i)}$ are scaling factors of $C_{bot(i)}$ at node (i) in phase 1 and 2, and F_s is switching frequency.

In Equation 4.32, scaling factors $s_{ph1(i)}$ and $s_{ph2(i)}$ are necessary because switching loss due to bottom-plate parasitic arises at internal node (i), where only a fraction of the loss comes from input voltage and the rest comes from ground. Equation 4.33 can be simplified to Equation 4.34 by using bottom-plate loss coefficient, K_{BOT}. K_{BOT} can be regarded as the normalized bottom-plate loss with ratio of bottom plate parasitic to a flying capacitor, total flying capacitance, squared input voltage, and switching frequency. Thus, K_{BOT} is determined by the topology of the given SC converter, and K_{BOT} limits power conversion efficiency of the SC converter.

$$P_{BOT} = K_{BOT}(a_{BOT}C_{tot})V_{IN}^2F_s \qquad (4.34)$$

where a_{bot} is the ratio of bottom plate parasitic to flying capacitors (C_{bot}/C_{fly}), C_{tot} is total flying capacitance, V_{IN} is input voltage, and F_s is switching frequency.

Based on Equation 4.34 and the R_{SSL} analysis, P_{BOT} can be obtained mathematically as plotted in Figure 4.7. Figure 4.7 shows K_{BOT} for SAR SC converters, Recursive SC converters, series-parallel SC converters, and ladder SC converters with various conversion ratios for the given number of stages bits. It is noted that SAR SC and ladder SC converters have K_{BOT} scaling with the number of stages, or the number of bits, but Recursive SC and series-parallel SC converters have K_{BOT} that is almost constant regardless of the number of stages, or the number of bits. This is because SAR SC converters decrease voltage swings, which is defined as $|V_{bot(i),ph1} - V_{bot(i),ph2}|$ of bottom-plate parasitic capacitors at the later stages, as the number of stages increases. Similarly, ladder SC converters reduce voltage swings of bottom-plate capacitors with increasing number of bits, and decrease K_{BOT}.

C-3. Loss Optimization and Comparison of Switched-Capacitor Converters

In moderate and low load power regime, total loss (P_{loss}) can be expressed as a sum of loss due to R_{SSL} (P_{SSL}), and loss due to bottom-plate switching (P_{BOT}): $P_{loss} = P_{SSL} + P_{BOT}$. P_{SSL} can be written with K_{SSL} as following equation:

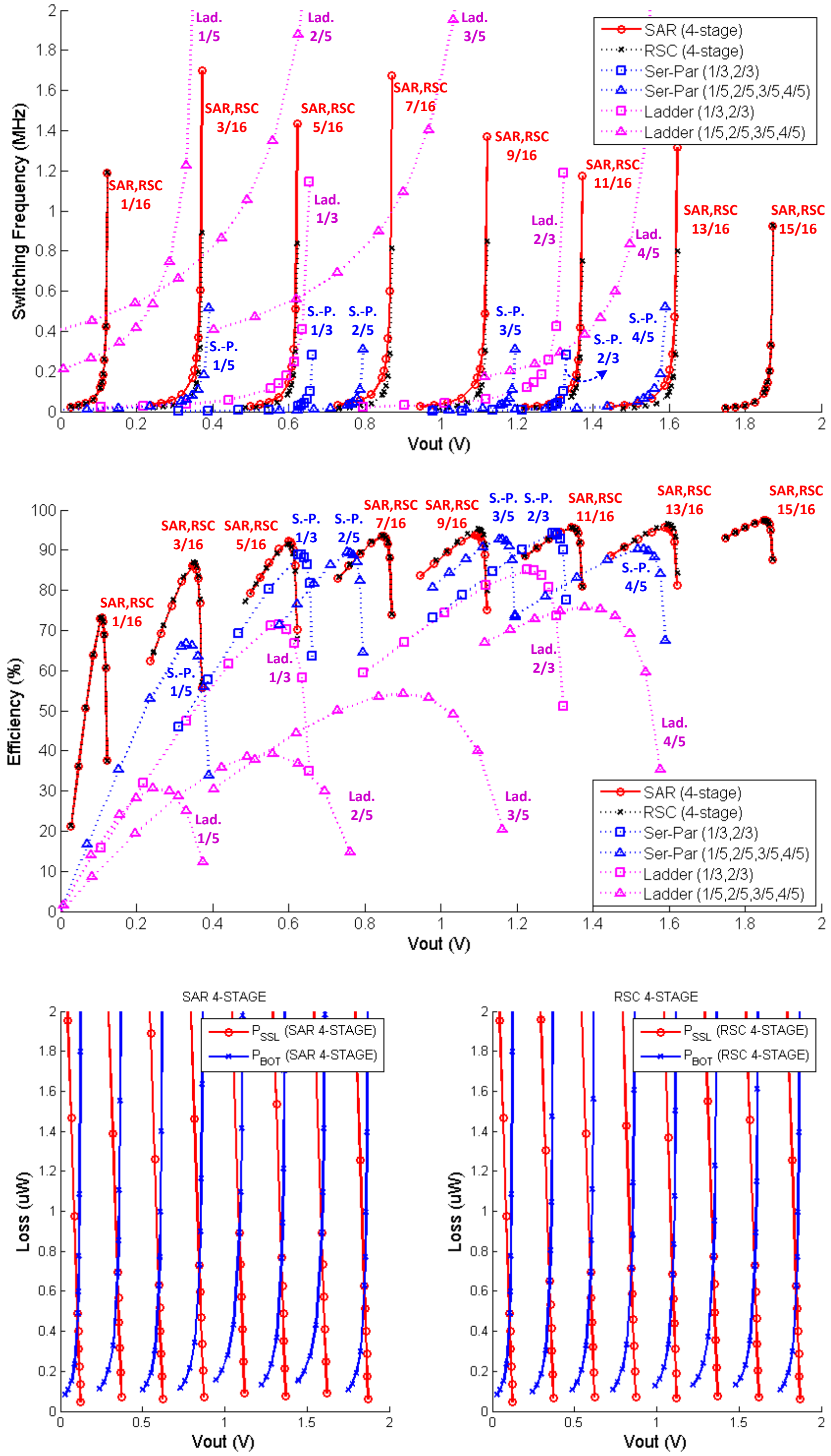


Figure 4.8. Efficiency vs. output voltage with frequency modulation (top), corresponding frequency vs. output voltage (middle), and corresponding changes of P_{SSL} loss and P_{BOT} loss in 4-stage SAR SC and RSC SC converters (bottom) when V_{IN} = 1V, total capacitance = 1nF.

$$P_{SSL} = R_{SSL} I_L^2 = \frac{K_{SSL}}{F_s C_{tot}} I_L^2 \tag{4.34}$$

where F_s is switching frequency, C_{tot} is total flying capacitance, and I_L is load current.

Based on Equations 4.33 – 4.34, optimal switching frequency ($F_{s,opt}$), optimal loss ($P_{loss,opt}$), and corresponding optimal efficiency (η_{opt}) can be obtained as following equations:

$$F_{s,opt} = \sqrt{\frac{K_{SSL}}{a_{BOT} K_{BOT}}} \times \frac{I_L}{V_{IN} C_{tot}} \tag{4.35}$$

$$P_{loss,opt} = 2\sqrt{P_{SSL} P_{BOT}} = 2\sqrt{a_{bot} K_{SSL} K_{BOT}} \times V_{IN} I_L \tag{4.36}$$

$$\eta_{opt} = \frac{P_{out}}{P_{out}+P_{loss,opt}} = \frac{1}{1+2\sqrt{a_{bot} K_{SSL} K_{BOT}} \times (\frac{V_{IN}}{V_o})} \tag{4.37}$$

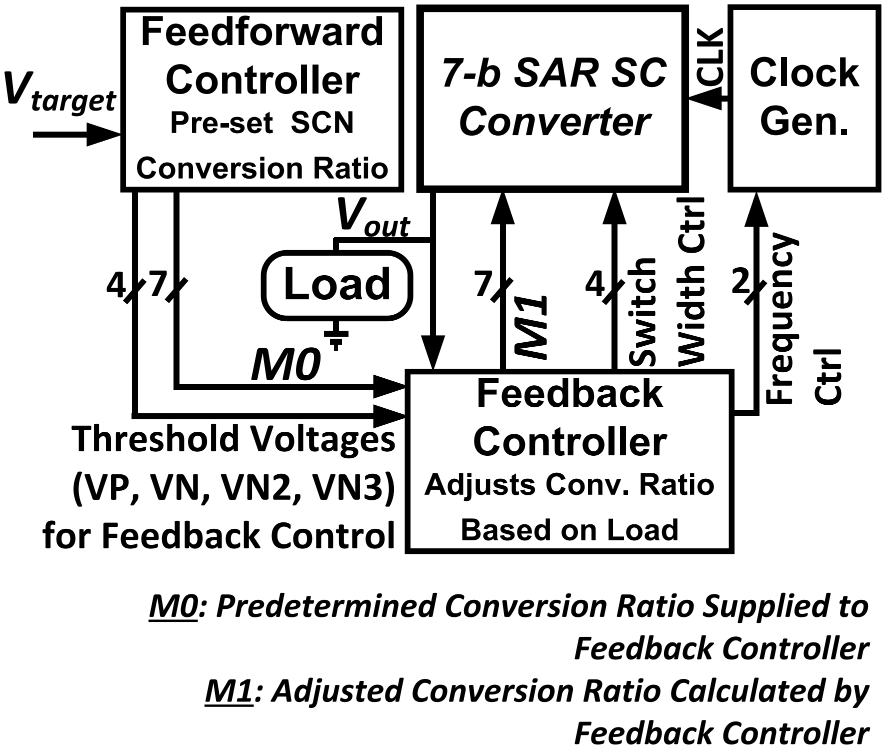
Equations 4.35 – 4.37 imply that frequency scaling with load current can track optimal efficiency, under the moderate and low load power regime, and regulation capabilities with frequency modulation for load regulation. Figure 4.8 shows simulation results based on mathematical models mentioned in Sections II-C-1 and II-C-2. Under simulation conditions of V_{IN} = 2V, C_{tot} = 1nF, a_{bot} = 3% and fixed I_L = 25uA, efficiency is plotted against output voltage as a result of switching frequency change for different topologies: 4-stage SAR SC, 4-stage Recursive SC, series-parallel SC and ladder SC converters with conversion ratios of 1/3, 2/3, 1/5, 2/5, 3/5 and 4/5, at the top of Figure 4.8. The corresponding frequencies are shown across output voltage in the middle of Figure 4.8. Also, the corresponding changes in P_{SSL} and P_{BOT} as a result of frequency control in SAR SC, and Recursive SC converters are plotted against output voltage in the bottom of Figure 4.8. It is noted that optimized SAR SC converters offer a large number of conversion ratios with comparable efficiency to series-parallel SC converters, and SAR SC converters achieve good efficiency over wide range of output voltages, as opposed to conventional series-parallel SC and ladder SC converters with comparable efficiency for a limited range of output voltage. It is also noted that smaller P_{BOT} of SAR SC converters compensates for P_{SSL} which is slightly larger than Recursive SC converters, resulting in similar optimal efficiencies.

4.2 Design of Prototype Successive-Approximation Switched-Capacitor Converter

4.2.1 Architecture of 7-b Successive-Approximation Switched-Capacitor Converter

A 7b SAR SC converter is fabricated in a test chip, with additional features such as closed-loop output regulation under load variation and line variation, as shown in Figure 4.9. The implemented 7b SAR SC converter is designed as a cascaded structure of one 4:1 converter and five 2:1 converters (Figure 4.10). Each converter is two-phase interleaved. To enable efficient low swing clocks, the first stage is constructed using a 4:1 converter, and the clock generation uses V_{BAT} and VDD3Q = 3/4 × V_{BAT} as its supply and ground. By using V_{BAT} and VDD3Q, clock swing and frequency automatically increase under heavy loading conditions as VDD3Q droops. This creates inherent negative feedback to automatically mitigate P_{SSL}. A conventional approach using VDD1Q and VSS would instead experience voltage droop on VDD1Q, yielding a clock frequency / swing reduction with positive feedback, limiting converter operating load range.

Capacitive level shifters are used in each converter to drive the switches. For example, the gate voltage G_1 of switch S_1 is referenced to its source (VH[3]) by the cross-coupled PFET



V_{target}
Feedforward Controller
Pre-set SCN Conversion Ratio
7-b SAR SC Converter
CLK
Clock Gen.
V_{out}
Load
4
7
M0
7
M1
4
Switch Width Ctrl
2
Frequency Ctrl
Feedback Controller
Adjusts Conv. Ratio Based on Load
Threshold Voltages (VP, VN, VN2, VN3) for Feedback Control
M0: Predetermined Conversion Ratio Supplied to Feedback Controller
M1: Adjusted Conversion Ratio Calculated by Feedback Controller

Figure 4.9. Top-level block diagram of proposed 7-b SAR SC converter.

structure R_1 and R_2. The gate voltage G_1 then swings low from this reference point through capacitive coupling driven by inverter I_1. Conversely, the gate voltage G_2 of S_2 is referenced to its source VM[3] and is coupled high by I_2. Since the cross-coupled PFET level shifter inherently generates two opposite polarities, a two-phase interleaving can be constructed with effectively no overhead. Four switch structures are connected in parallel, sized 1×, 1×, 2×, and 4× to implement binary-weighting using the thermometer control code SEN[3:0] for switch width modulation. The configuration switches toggle only when the configuration vector changes, and hence can be made large to limit resistive loss with minimal switching energy cost.

4.2.2 Feedforward and Feedback Control

Feedforward and feedback controllers provide fine grain control and react to load and line variations. As shown in Figure 4.11, the feedforward controller predetermines a conversion ratio M0 by comparing V_{target} with a ramp voltage (V_{RAMP}) that increases by $V_{BAT}/2^7$ for each cycle of CLKd, which is 32× slower than the converter switching clock (CLK). V_{RAMP} is

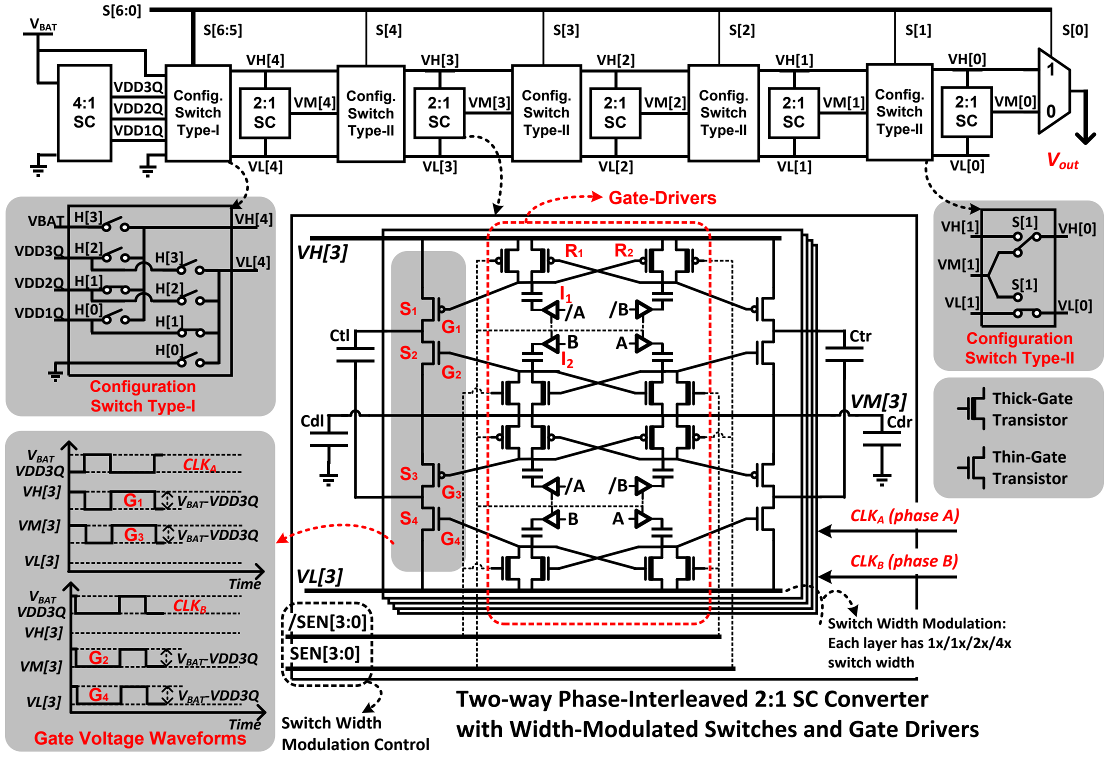


Figure 4.10. The proposed 7-b SAR SC DC-DC converter (top). Detailed schematics of configuration switch type-I, II, and 2:1 SC converter (middle). Gate voltage waveforms of a 2:1 SC converter (bottom).

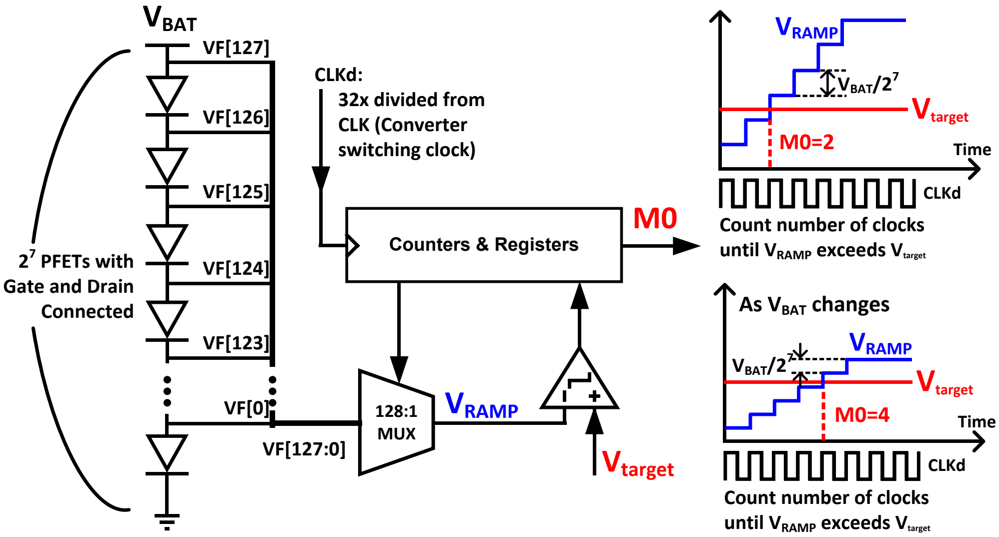
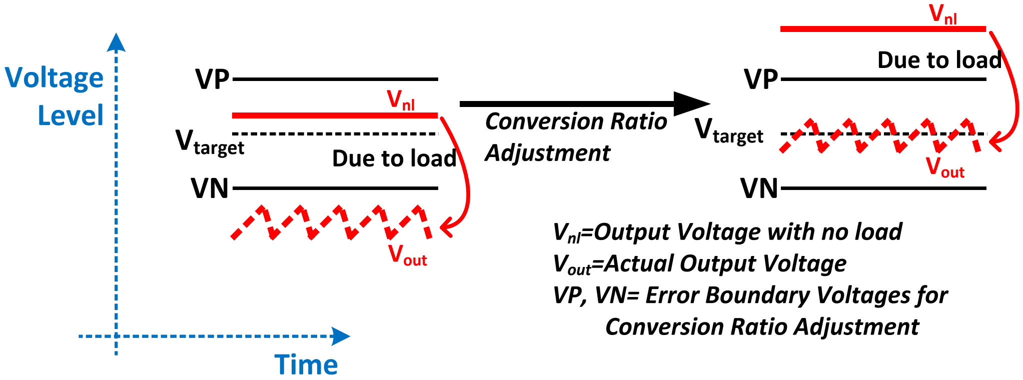
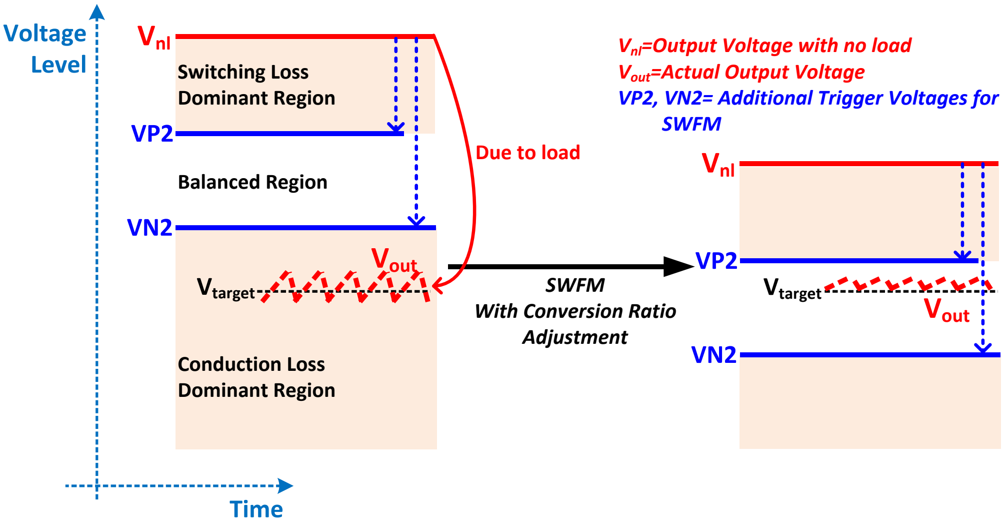


Figure 4.11. Feedforward controller details and operation.



Voltage Level
VP
V_{nl}
V_{target}
Due to load
VN
V_{out}
Time
Conversion Ratio Adjustment
V_{nl}
Due to load
VP
V_{target}
V_{out}
VN
V_{nl}=Output Voltage with no load
V_{out}=Actual Output Voltage
VP, VN= Error Boundary Voltages for Conversion Ratio Adjustment

Figure 4.12. Conversion ratio adjustment of feedback controller.



Voltage Level
V_{nl}
Switching Loss Dominant Region
VP2
Balanced Region
VN2
V_{out}
V_{target}
Conduction Loss Dominant Region
Due to load
Time
V_{nl}=Output Voltage with no load
V_{out}=Actual Output Voltage
VP2, VN2= Additional Trigger Voltages for SWFM
SWFM With Conversion Ratio Adjustment
V_{nl}
VP2
V_{target}
V_{out}
VN2

Figure 4.13. Switch width and frequency modulation.

generated from 2^7 diode-connected PFETs in series, and V_{RAMP} is not generated from the SC converter. M0 is obtained by counting the clock cycle at which V_{RAMP} exceeds V_{target}, and it is updated every 2^7 cycles of CLKd, so M0 is kept the same unless input voltage V_{BAT} changes. Note that V_{target} can be generated with ultra-low power reference voltages [42-43].

The M0 configuration code results in an SC output voltage (V_{OUT}) that matches V_{target} within one resolution of SAR SC converter, $V_s = V_{BAT}/2^7$ under no-load conditions (where V_s = 31.25mV with V_{BAT} = 4.0V). As shown in Figure 4.12, V_{OUT} droops in the presence of load, so the FB controller adjusts the conversion ratio to maintain a constant output voltage. For this, two trigger voltages VP and VN are generated from the diode stack with separate 2^7:1 muxes, where VP = VF[M0+Δ_{P1}] and VN = VF[M0−Δ_{N1}]. V_{OUT} is compared with VP and VN at each cycle and the conversion ratio is adjusted to maintain the condition VN < V_{OUT} < VP. By incrementing/decrementing a 7b counter and adding it to M0, the adjusted configuration code M1 (≥ M0) is obtained.

To prevent converter efficiency from being limited by conduction loss or series loss that consists of slow-switching limit impedance loss and fast-switching limit impedance loss [40, 54], the frequency and switch widths are dynamically modulated in a binary-weighted fashion by the

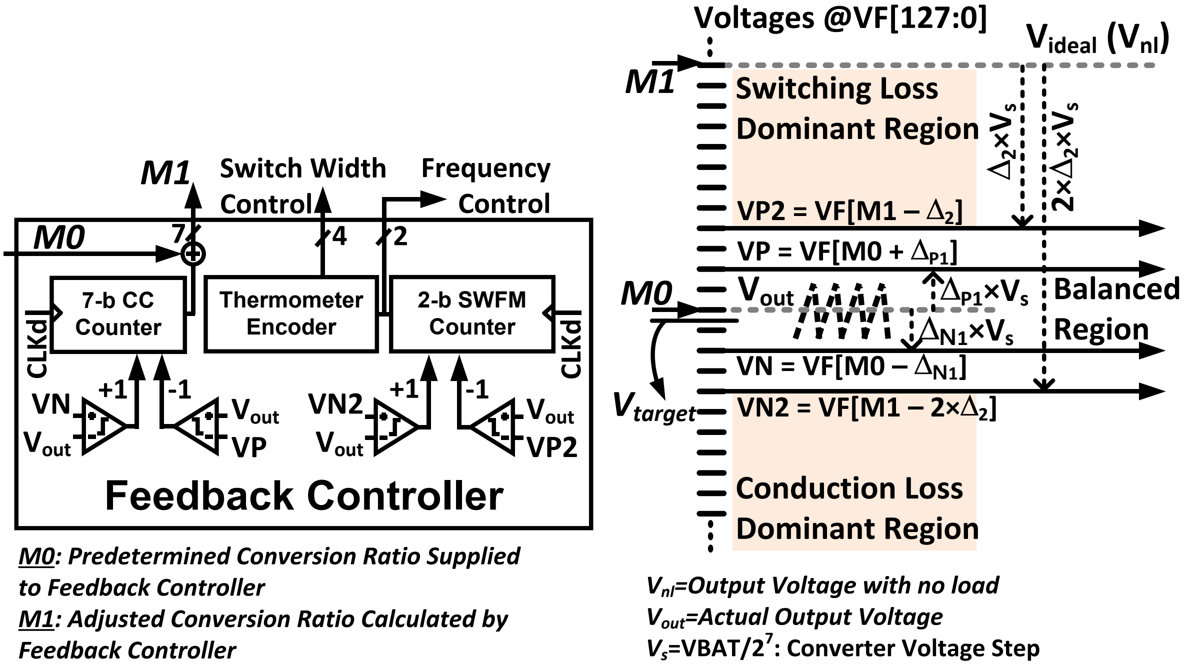


Figure 4.14. Feedback controller details and operation (left). Trigger voltage levels (right).

feedback controller. Two additional trigger voltages VP2 and VN2 are generated, where VP2 = VF[M1−Δ_2] and VN2 = VF[M1−2Δ_2] (using two additional 2^7:1 muxes). In this implementation, Δ_2 is set to 5 and hence $\Delta_2 \times V_s$ = 156.25mV. VP2 and VN2 are referenced to VF[M1], which is the ideal (no-load) voltage level of the SC converter at M1. As shown in Figure 4.13, when V_{OUT} lies within $\Delta_2 \times V_s$ of this no-load output voltage (V_{OUT} > VP2), switching loss dominates and switch size and frequency are reduced by decrementing switch width and frequency modulation (SWFM) counter shown in Figure 4.14 (left). Similarly, when V_{OUT} falls below the no-load output voltage V_{ideal} by more than $2 \times \Delta_2 \times V_s$ (V_{OUT} < VN2), conduction loss is dominant and switch size and frequency are increased. By correctly setting Δ_2, the switching and conduction losses remain balanced over a large load range, hence improving conversion efficiency. The feedback controller is implemented with digital counters, a thermometer encoder, and comparators, as seen in Figure 4.14 (left). Trigger voltage levels (VP, VP2, VN, and VN2) are visualized in Figure 4.14 (right).

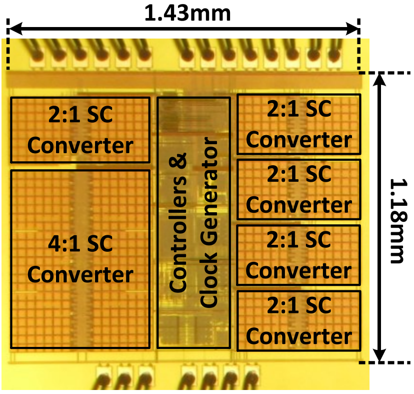


Figure 4.15. Die photograph.

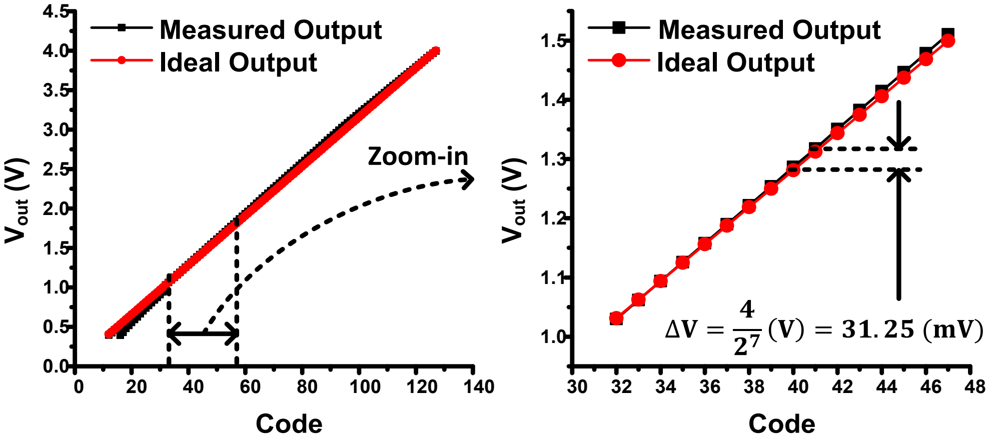


Figure 4.16. Measured output voltage levels and ideal output voltage levels of 7-b SAR SC converter at V_{BAT} = 4V and no load.

4.3 Measurement Results

The SAR SC converter is fabricated in 180nm CMOS process. Figure 4.15 shows the chip photograph. The fabricated test chip with 2.24nF on-chip capacitance occupies 1.69 mm^2. MIM capacitor and MOS capacitor are used for on-chip flying capacitors and de-coupling capacitors, respectively.

Figure 4.16 shows measured outputs and ideal outputs of the 7b SAR SC converter with a resolution of 31.25mV at V_{BAT}=4V[1]. Figure 4.17 shows V_{OUT} regulation results, code change, clock frequency change, and efficiency across load currents at V_{BAT} = 4V and V_{target} = 1.2V. V_{OUT} is continually compared against trigger voltage levels (VN, VP, VN2, and VP2). As the load current increases from 0 to 300μA, the conversion code increases to compensate for conduction loss. When $V_{OUT} - V_{ideal} > 2\times\Delta_2\times V_s$, the feedback controller increments the 2b SWFM counter, increasing switch width and clock frequency. Every step increase in the SWFM counter results in an approximately 1.7× increase in clock frequency. Note that when load

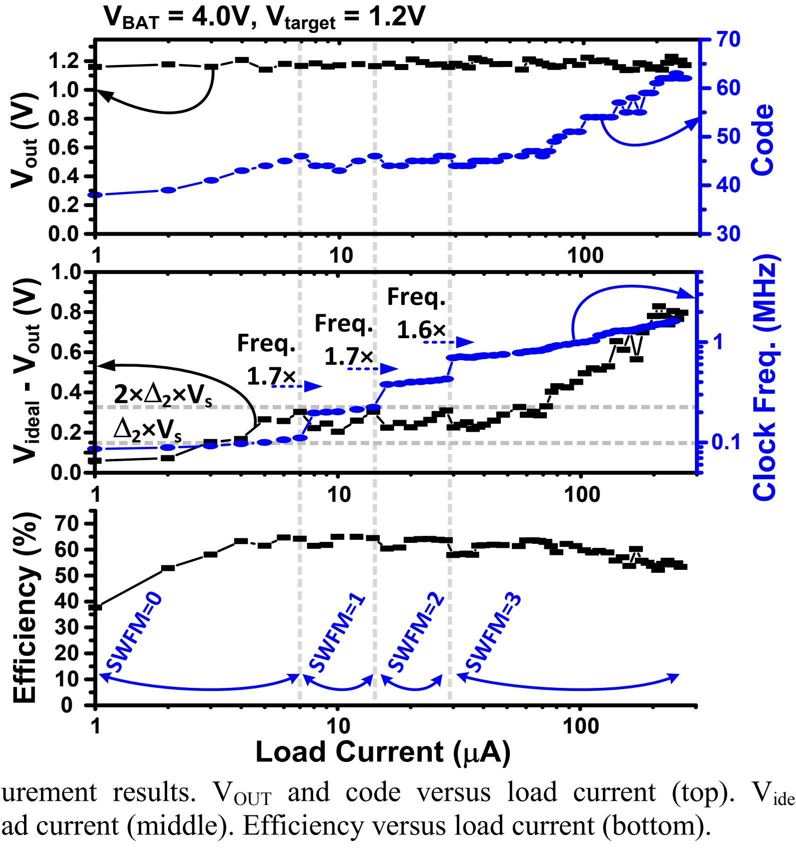


Figure 4.17. Measurement results. V_{OUT} and code versus load current (top). V_{ideal}–V_{OUT} and clock frequency versus load current (middle). Efficiency versus load current (bottom).

[1] In 180nm CMOS process, thick-gate transistors and thin-gate transistors are rated for maximum of 3.6V and 2.0V, respectively.

current decreases and $V_{OUT} - V_{ideal} < \Delta_2 \times V_s$, the feedback controller decrements the SWFM counter, decreasing switch width and clock frequency. Figure 4.18 shows a simplified schematic of the clock generator and 4:1 SC converter, and a plot of measured clock frequency and V_{BAT} – VDD3Q.

Figure 4.19 shows the effectiveness of dynamic SWFM; >50% efficiency is achieved across a load range of 2μA ~ 300μA using dynamic SWFM compared to 30~300μA when using a single setting (SWFM = 3). Efficiency fluctuates slightly at neighboring load currents. This arises due to varying SSL impedance with switch configurations, as was also seen in Figures 4.6. Figure 4.20 shows load regulation measurement results at V_{BAT} = 4V for V_{target} = 0.9V, 1.2V, and 1.5V. The converter achieves peak efficiency of 69%, 65%, and 72% with output voltage regulation within ±54mV, ±41mV, and ±81mV for V_{target} = 0.9V, 1.2V, and 1.5V, respectively. It

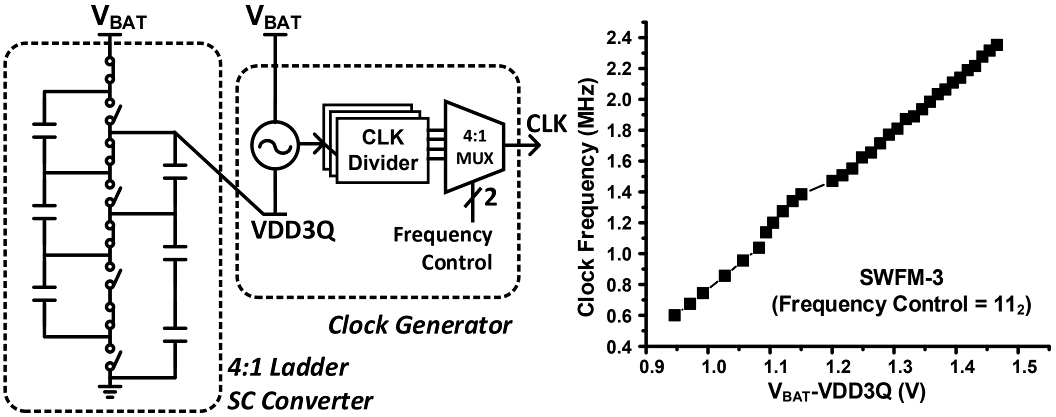


Figure 4.18. Simplified schematic of clock generator and 4:1 SC converter (left). Measured clock frequency versus V_{BAT}–VDD3Q (right).

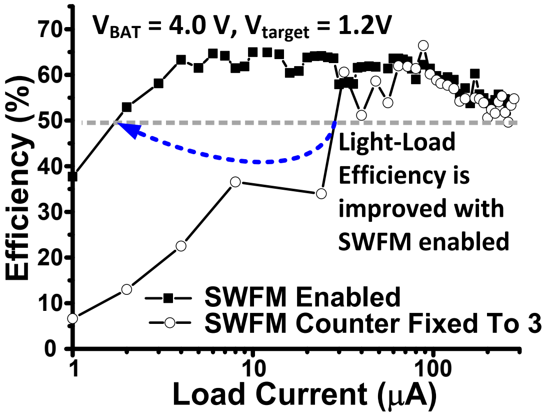


Figure 4.19. Measured efficiency versus load current with SWFM enabled, and SWFM fixed to 3.

is shown that SWFM effectively balances switching loss and conduction loss, enabling a flat efficiency curve across a wide range of load currents.

Figure 4.21 (left) shows that a wide range of arbitrary V_{OUT} can be generated due to the highly reconfigurable nature of SAR SC converters. Figure 4.21 (right) shows line regulation at target voltages of 0.9V, 1.2V, and 1.5V for V_{BAT} values ranging from 3.4 to 4.3V. V_{OUT} variation depends on VP − VN, which is set at $3\times V_s$ (93.75mV), as well as comparator offset in the feedback controller.

Figure 4.22 shows transient step response waveforms. Figure 4.22 (left) is the transient load step response when load current changes from 10μA to 0 and from 0 to 50μA, with V_{BAT} = 4V input and V_{target} = 1.2V. With the load current changes, SWFM is modulated from mode 1 to mode 3 and the code is adjusted. In this way, V_{OUT} is maintained close to V_{target}. Figure 4.22 (right) shows the transient output voltage response for a V_{target} change from 0.9V to 1.5V when

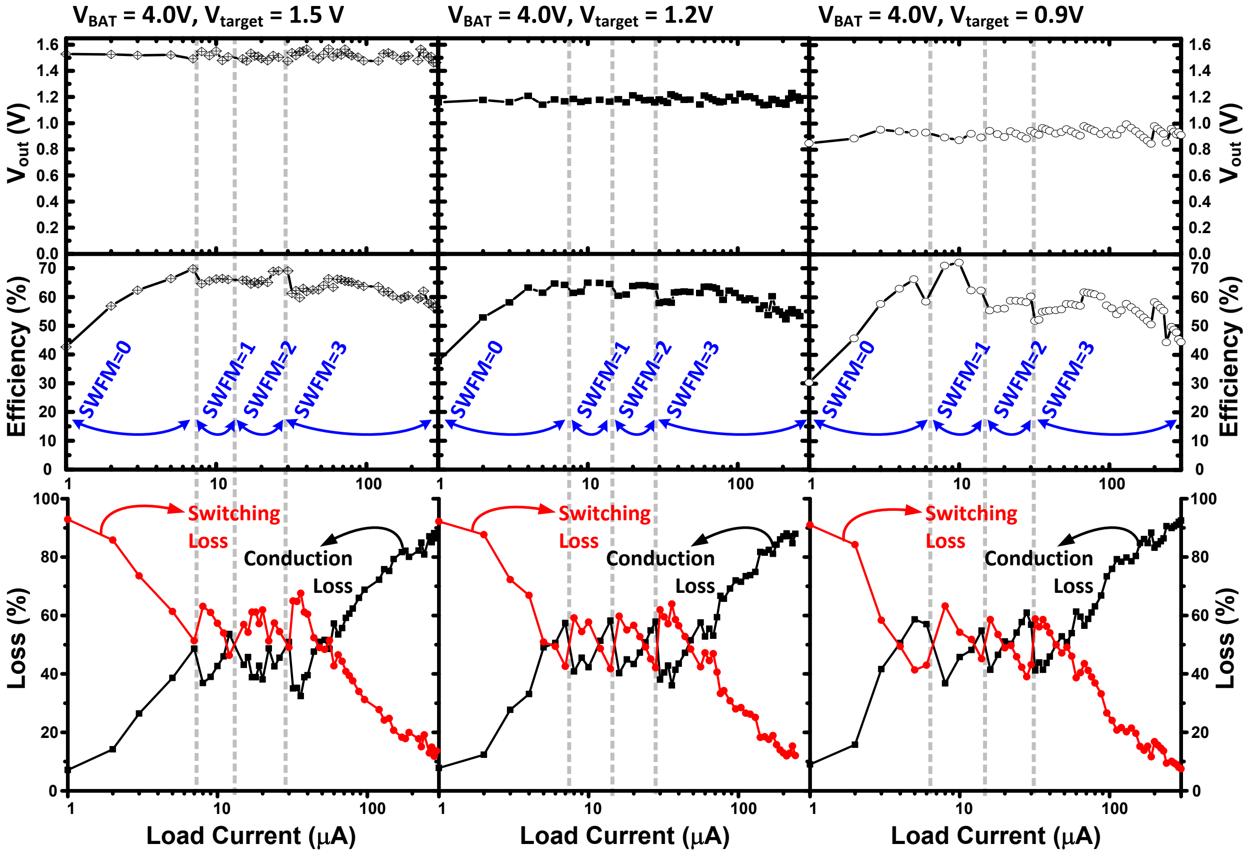


Figure 4.20. Measured V_{OUT}, efficiency, and switching/conduction loss versus load current for V_{target} = 1.5V (left), 1.2V (middle), and 0.9V (right).

load current is 10μA. It takes approximately 34ms to adjust V_{OUT} in the measurement. Response time depends on the feedforward clock frequency, which is divided by 32× from the converter switching clock. Table 4.3 provides a comparison of the implemented test chip to other works, including Recursive SC converters have recently expanded the proposed SAR-based approach by rearranging the switches to improve conversion efficiency [44-45]. The testchip achieved power density of 0.27mW/mm^2, which is relatively low because of slow switching frequency and switch sizes set for optimized efficiency at low power load. Faster switching frequency can improve power density.

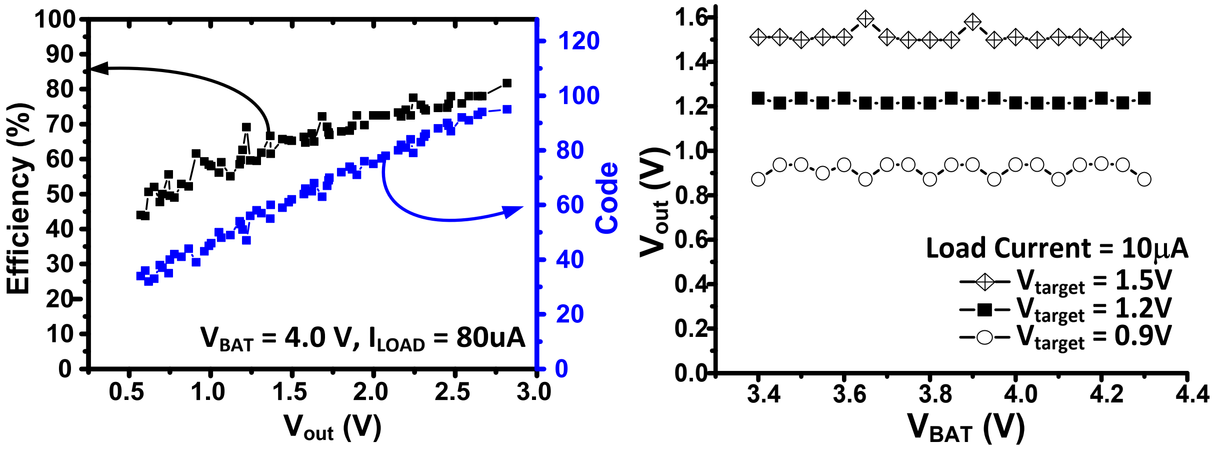


Figure 4.21. Efficiency and code versus V_{OUT} for the range of V_{OUT} = 0.55~2.85V (left), and line regulation for V_{target} = 1.5V, 1.2V and 0.9V (right).

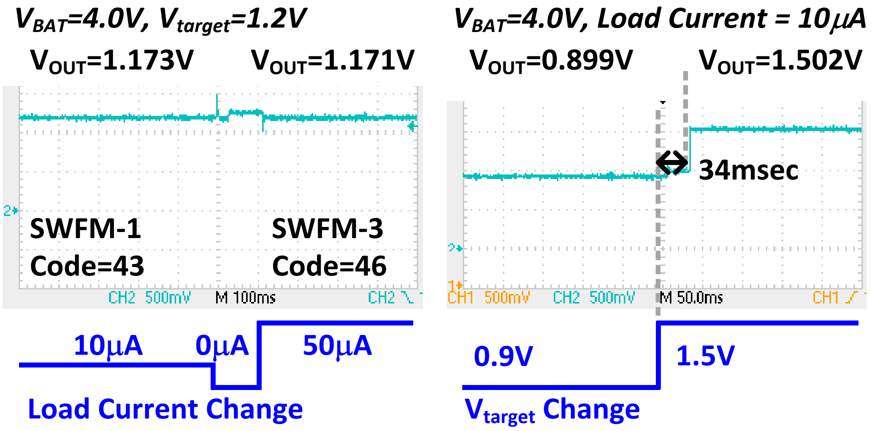


Figure 4.22. Transient load step response (left), and transient V_{OUT} response upon V_{target} change (right).

4.4 Summary

A SAR SC DC-DC converter is proposed for output regulation, providing conversion ratio resolution of $V_{IN}/2^{\text{Number of Stages}}$. The SAR SC converter has smaller R_{SSL} than conventional series-parallel SC and ladder SC converters with the same conversion ratio resolution respectively by $N^2/\log_2 N$ and $N^4/\log_2 N$ when N is large and $k \simeq N/2$, where N is the number of available conversion ratios and k/N is conversion ratio. Also, the SAR SC converter has P_{BOT} that scales with the number of stages. It was shown that in the moderate load power regime, optimal efficiency can be obtained by scaling frequency with load current change, so that P_{SSL} and P_{BOT} are balanced.

Metric	**This work**	[32]	[33]	[44]	[45]
Topology	**7-b SAR**	Series-Parallel	Series-Parallel	4-b Recursive Binary	3-stg Recursive Ternary
Cascaded SC?	**Yes**	No	No	Yes	Yes
Closed-loop?	**Yes**	Yes	Yes	No	No
Input Range	**3.4~4.3V**	1.2V	1.8V	2.5V	2.5V
Output	**>0.45V (In theory, >$V_{BAT}/2^7$)**	0.3~1.1V	0.8~1V	0.1~2.18V	0.1~2.24V
Number of Conversion Ratios	**117 (In theory, 127)**	6	1	15	45
Step Size	**31.25mV @V_{BAT} = 4V**	N/R	N/R	156mV	~55mV
I_{LOAD} Range	**300μA**	1mA	8mA	2mA	1.86mA
Clock Frequency	**80kHz~2.7MHz @V_{BAT} = 4V, V_{target} = 1.2V**	15MHz	30MHz	200kHz ~10MHz	N/R
Peak Efficiency	**69% @V_{BAT} = 4V, V_{target} = 1.5V 65% @V_{BAT} = 4V, V_{target} = 1.2V 72% @V_{BAT} = 4V, V_{target} = 0.9V**	>70%	69%	85%	86%
Process (nm)	**180**	180	45	250	250
Area (mm^2)	**1.69**	2.56	0.16	4.645	4.3
Capacitance (nF)	**2.24**	2.5	1.234	3	2.8

Table 4.3. Comparison table.

In a testchip fabricated in 180nm CMOS process, a 7-b SAR SC converter was designed with feedforward and feedback controllers that regulate output voltage with conversion ratio, frequency and switch width adjustment for operation across a wide range of input and output voltages. The proposed SAR SC converter can generate 2^N-1 conversion ratios, and the SC converter in each stage holds a fixed voltage, which can minimize configuration change and stabilization time upon conversion ratio adjustment. This property potentially makes the SAR SC converter a strong candidate for DVS voltage regulation in future work. This work was presented and published in [41, 62].

CHAPTER 5

Flying-Capacitance-Dithered Switched-Capacitor DC-DC Converter for Ripple Minimization

On-chip power delivery has become important for high performance circuits, in which power consumption is high and power loss due to parasitic resistance and package/board inductance is non-negligible. On-chip voltage regulation using on-chip DC-DC step-down converters can provide reduced package current which reduces power loss due to IR drop and Ldi/dt drop, fast load response, and per-block power management enabling for energy-efficient operation of load circuit [38, 55], as illustrated in Figure 5.1. Switched-capacitor voltage regulators (SCVRs) have gained popularity for on-chip regulation [33, 34, 47-49] as they are compatible with CMOS processes and do not require magnetic materials for inductors. SC converter is likely to suffer from ripple issue due to its inherent switching property enabling charge transfer for voltage conversion and power generation. In this chapter, flying-capacitance-dithered SC DC-DC converter is proposed for output voltage ripple minimization, and relation ripple and power conversion efficiency is investigated.

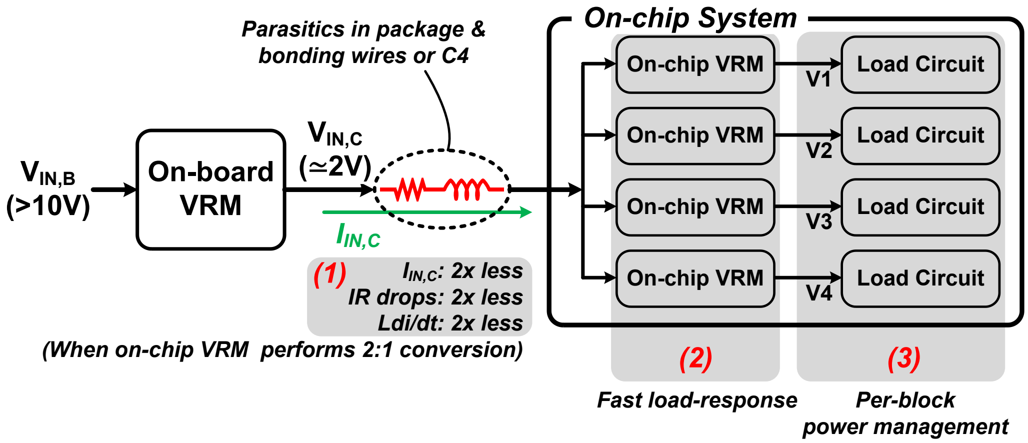


Figure 5.1 On-chip power delivery.

5.1 Tight Regulation and Closed-loop SC converter

Dynamic voltage and frequency scaling have been long regarded as effective technique for energy efficient operation of digital circuits [46]. Characteristics of given task at a time will demand different voltage and frequency to complete within fixed time window, which is designed to have timing margin to avoid timing violation due to uncertain variation factors such as temperature, process, signal integrity, voltage and delay model. Accumulation of these can cause large power and performance overhead. Tight regulation of voltage regulator with small output variation and transient ripple can enhance load circuit performance, by reducing voltage variation and therefore the timing margin, as conceptually described in Figure 5.2. DVFS with reduced voltage margin can run circuits at faster frequency without timing violation.

A conventional approach to reduce ripple in SC converter is multi-phase interleaving, an example of which is shown in Figure 5.2 for 2-phase interleaving The open-loop ripple magnitude with NPH-phase interleaving can be expressed as:

$$V_r = \frac{I_L \times T_{SC}}{N_{PH} \times C_{tot}} = \frac{I_L}{F_{SC} \times N_{PH} \times C_{tot}} \tag{5.1}$$

where I_L is the load current, T_{SC} is the switching period (F_{SC}=1/T_{SC}), N_{PH} is the number of interleaved phases, and C_{tot} is the flying capacitance plus any additional AC-equivalent decoupling capacitance seen at the output.

Total flying capacitance of C_{tot}, or ac-equivalent capacitance seen at output, determines output voltage slope between neighboring switchings, in steady-state. In other words, output

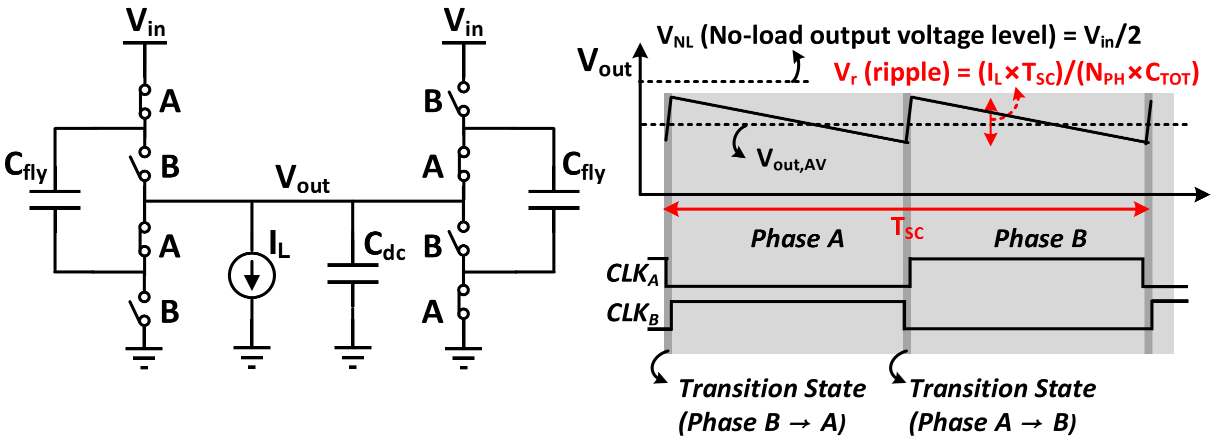
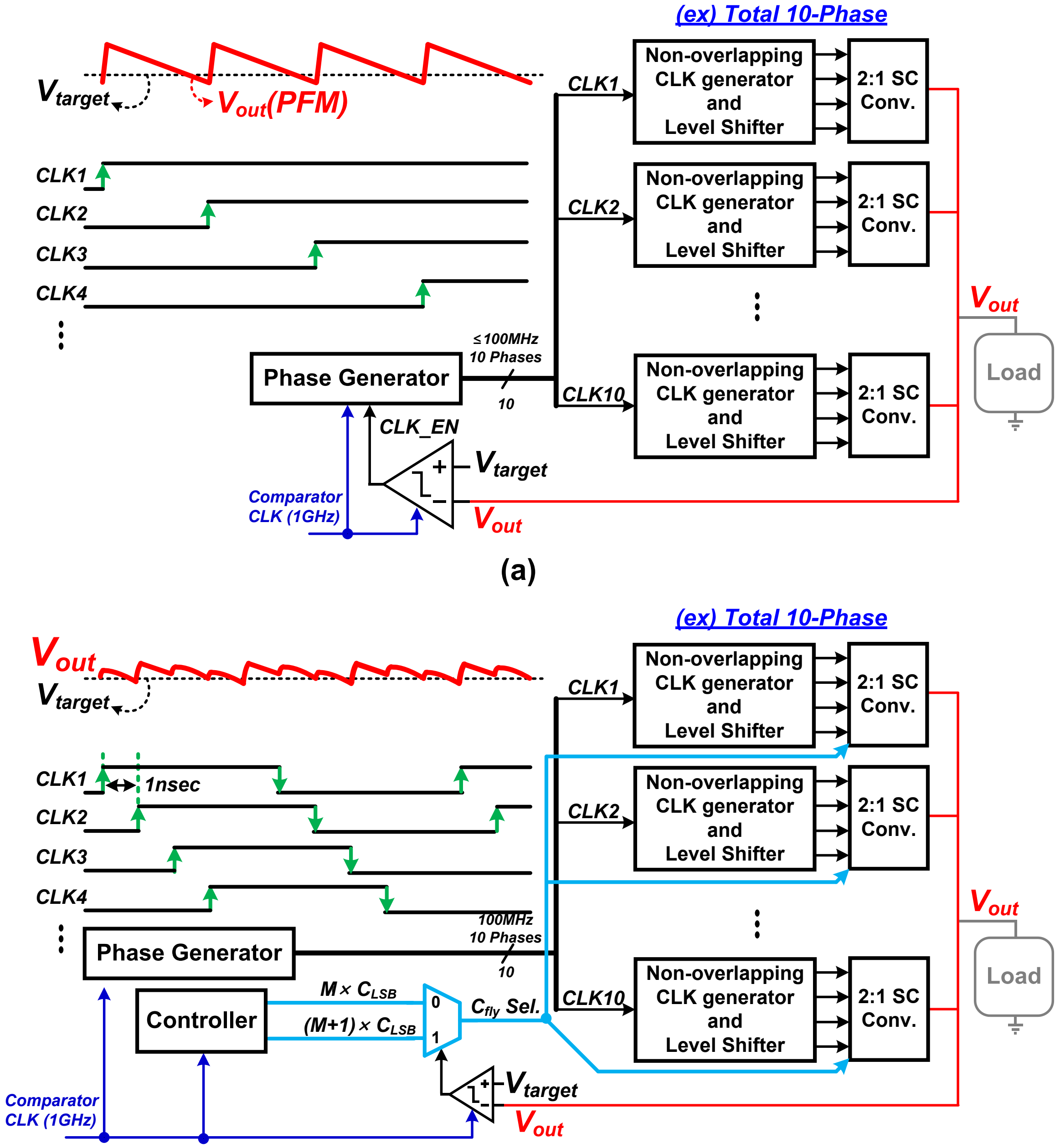

V_in
V_in
A
B
B
A
C_fly
C_fly
V_out
A
B
I_L
C_dc
B
A
B
A
V_out
V_NL (No-load output voltage level) = V_in/2
V_r (ripple) = (I_L×T_SC)/(N_PH×C_TOT)
V_out,AV
Phase A
T_SC
Phase B
CLK_A
CLK_B
Transition State (Phase B → A)
Transition State (Phase A → B)

Figure 5.2. Schematic of a 2:1 switched-capacitor converter with 2-phase interleaving (left), and waveform of output voltage (right).

voltage slope of $-I_L/C_{tot}$, it first decreases for T_{SC}/N_{PH} with N_{PH}-phase interleaving, and it recovers instantaneously the drop voltage during switching of capacitors, maintaining stable output level in steady state.

However, equation (5.1) is valid only when the SC converter operates in open loop, where the output voltage is not regulated to a target voltage in response to load current or input



(a)

(b)

Figure 5.3. (a) Baseline scheme: pulse-frequency modulation (PFM), and (b) proposed scheme: on-demand capacitance modulation.

voltage changes. In prior SC converter designs, many closed-loop output voltage regulation techniques with phase interleaving have been introduced, including single-boundary multi-phase control (SB-MC), multi-phase pulse frequency modulation (PFM), digital capacitance modulation, and conversion ratio adjustment [34-41]. However, most of these efforts do not address the output voltage ripple issue in SC converters, particularly as the output ripple magnitude in SB-MC and PFM schemes can unfortunately be even larger than that in open loop operation.

Figure 5.3 (a) shows an example of an SC converter with PFM and 10-phase interleaving. In a PFM design, a phase generator creates clocks of N_{PH} phases with frequency of up to F_{CMP}/N_{PH}, where F_{CMP} is comparator frequency and N_{PH} is number of interleaved phases (e.g. a 1GHz can generate ten-phase clocks with frequency of up to 100MHz as shown in Figure 5.3 (a)). PFM toggles the clocks in a round-robin fashion on demand when the output voltage (V_{out}) falls below the target voltage (V_{target}), thus frequency of the ten-phase clocks vary depending on load current. This results in ripple due to both excessive charge transfer and the inherent voltage drop below V_{target} as illustrated in Figure 5.4 (a).

There has been work on output ripple reduction in SC converters [18, 51], but with control techniques external to the primary regulation loop. In [49], the flying capacitance is modulated for ripple mitigation with the single bound hysteretic controller proposed in [34]. However, in addition to inheriting limitations from SB-MC, this work adjusts the number of phases for capacitance modulation, where the number of phases must be reduced to achieve a small flying capacitance, which can actually degrade output ripple. As an alternative approach, a hybrid converter was proposed in [58] where SC converter and linear regulator were connected in series. However, this approach results in area increase due to separate load capacitors required at output of linear regulator, and efficiency degradation due to linear regulator dropout voltage. Furthermore, in case of an SC converter with phase-interleaving and high load current, ripple with very high frequency > 1GHz can occur, and good PSRR of linear regulator at such a high frequency is required, resulting in design complication and power consumption overhead.

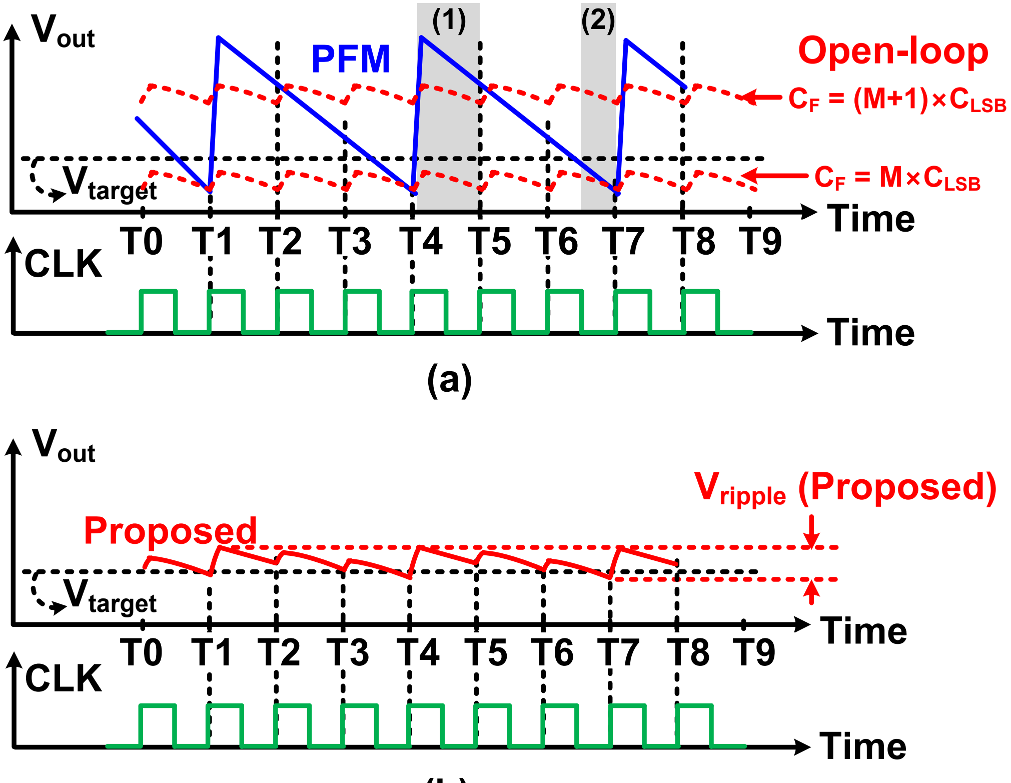
To minimize ripple in closed-loop SC converters, this work proposes dithered-capacitance-modulation, in which the flying capacitance is adjusted on-demand with a fixed frequency used for the SC converter. Clocks with a large number of phases are generated using a delay-locked-loop (DLL), so significant phase-interleaving can be used to maximize temporal distribution of the flying capacitance charge transfer [57].

5.2 Proposed Technique: Dithered-Capacitance Modulation

5.2.1 On-demand Capacitance Modulation

The operation and output voltage waveforms for the proposed on-demand capacitance modulation scheme are illustrated in Figures 5.3-5.5 as compared with traditional PFM schemes. On-demand capacitance modulation triggers SC converters at every clock edge, but changing the size of flying capacitor on a cycle-by-cycle basis allows modulation of the amount of the transferred charge. By splitting the SC into parallel structures with binary-sized flying capacitors, a sufficient range of capacitance modulation can be achieved.

As shown in Figure 5.3 (b), in the proposed on-demand capacitance modulation scheme,

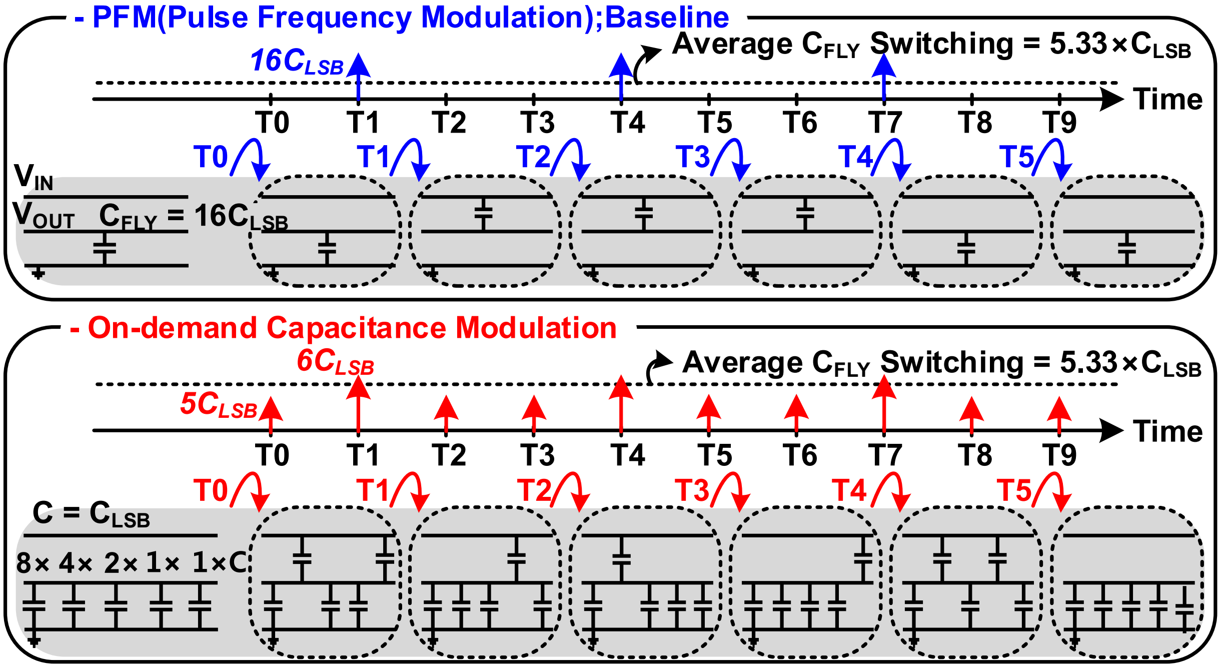


(b)

Figure 5.4. (a) Waveform of SC converter with pulse frequency modulation (PFM) overlaid with open-loop ripple, and (b) waveform of SC converter with on-demand capacitance modulation, output of which is lying between two open-loop voltage levels of $C_{FLY} = (M+1)\times C_{LSB}$ and $M\times C_{LSB}$.

a digital controller finds the required flying capacitance (C_{REQ}) to be switched at a given load current. With the flying capacitance quantized to a unit capacitance of C_{LSB}, there must exist an integer M such that $M \times C_{LSB} < C_{REQ} \leq (M+1) \times C_{LSB}$, where $0 \leq M \leq C_{MAX}/C_{LSB}$ and C_{MAX} is maximum flying capacitance assigned to a phase. Thus, the open-loop voltage level at flying capacitance $C_{FLY} = (M+1) \times C_{LSB}$ and $C_{FLY} = M \times C_{LSB}$ are above and below V_{target}, respectively, as illustrated in Figure 5.4. For both cases, open-loop ripple is in accordance with equation (5.1) and not affected by the control scheme. Figure 5.4 shows that PFM with a fixed flying capacitance degrades ripple characteristics as compared with the open-loop ripple when the load current is less than its maximum value, due to excessive charge transfer. However, with on-demand capacitance modulation, by switching $(M+1) \times C_{LSB}$ when $V_{out} < V_{target}$ and $M \times C_{LSB}$ when $V_{out} > V_{target}$, V_{out} regulation can be achieved with low ripple.

Figure 5.5 shows a comparison of the proposed scheme with PFM for 1/3 of the maximum load current. In PFM, clock pulses are generated on demand and the full C_{FLY} is switched each time a clock pulse is generated. To source 1/3 the maximum load current, PFM will generate a clock pulse every third cycle (in Figure 5.5 at T1 and T4 across 6 clock cycles). However, the full C_{FLY} switching followed by zero C_{FLY} switching results in high ripple. By



- PFM(Pulse Frequency Modulation);Baseline
16C_LSB
Average C_FLY Switching = 5.33×C_LSB
Time
T0 T1 T2 T3 T4 T5 T6 T7 T8 T9
T0 T1 T2 T3 T4 T5
V_IN
V_OUT C_FLY = 16C_LSB
- On-demand Capacitance Modulation
6C_LSB
5C_LSB
Average C_FLY Switching = 5.33×C_LSB
Time
T0 T1 T2 T3 T4 T5 T6 T7 T8 T9
T0 T1 T2 T3 T4 T5
C = C_LSB
8× 4× 2× 1× 1×C

Figure 5.5. Baseline scheme: pulse frequency modulation (top). Proposed scheme: on-demand capacitance modulation. (bottom).

contrast, on-demand capacitance modulation sets 5×C_{LSB} at time T0, which results in 5/16 of maximum charge transfer in that cycle under assumption that the maximum flying capacitor available in an SC converter is 16×C_{LSB}. Since 5/16 < 1/3, the output voltage drops slightly and falls below the target voltage, adjusting on-demand capacitance modulation to 6×C_{LSB}. At T1, the transferred charge increases to 6/16 of maximum, increasing the output voltage. Since the output voltage now exceeds the threshold, on-demand capacitance modulation then switches back to 5×C_{LSB} for T2 and T3, creating a repeating pattern every 3 cycles with an average charge transfer of 1/3 × 6/16 + 2/3 × 5/16 = 1/3 the maximum charge, which corresponds to flying capacitance of 5.33×C_{LSB}. Since the difference between actual and ideal flying capacitance is always less than 1×C_{LSB}, the charge transfer is kept largely proportional to the load current, resulting in low output ripple.

5.2.2 Dithered-Capacitance Modulation

In on-demand capacitance modulation, the ripple is induced by the C_{LSB} quantization of C_{FLY}. To further minimize the ripple beyond that, a dithering-like feature is proposed to obtain the effective C_{FLY} with finer granularity than C_{LSB}. Figure 5.6 shows an example where C_{FLY} of 5.2×C_{LSB} and 5.4×C_{LSB} are used for switching instead of 5×C_{LSB} and 6×C_{LSB}. Resolution smaller

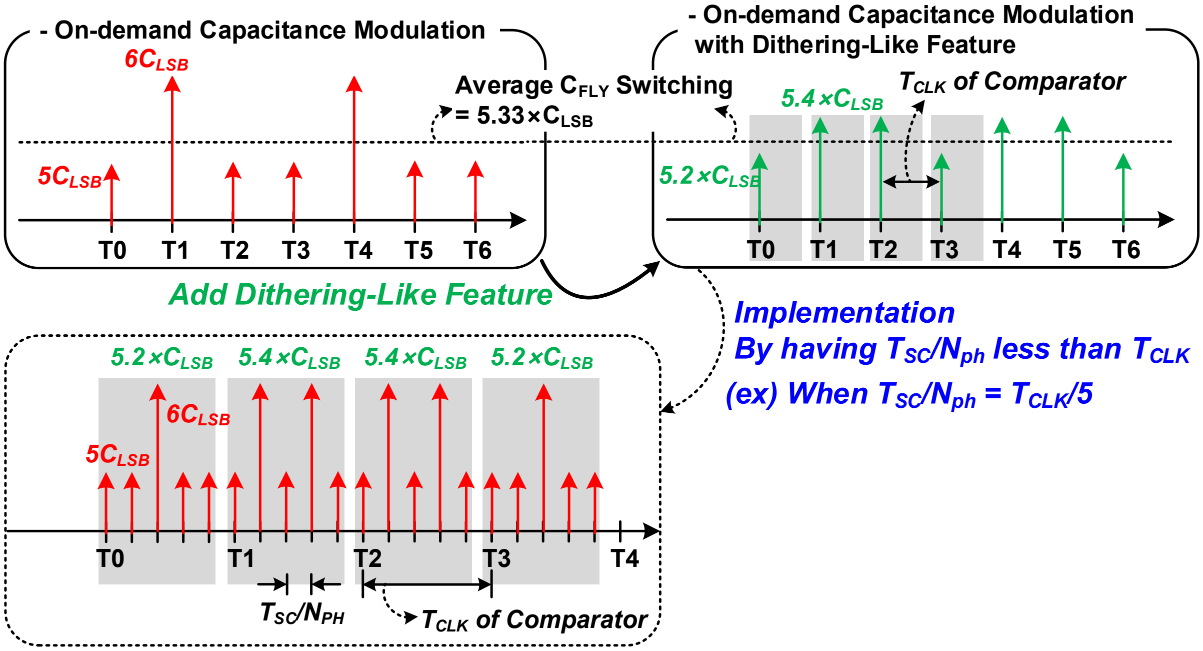
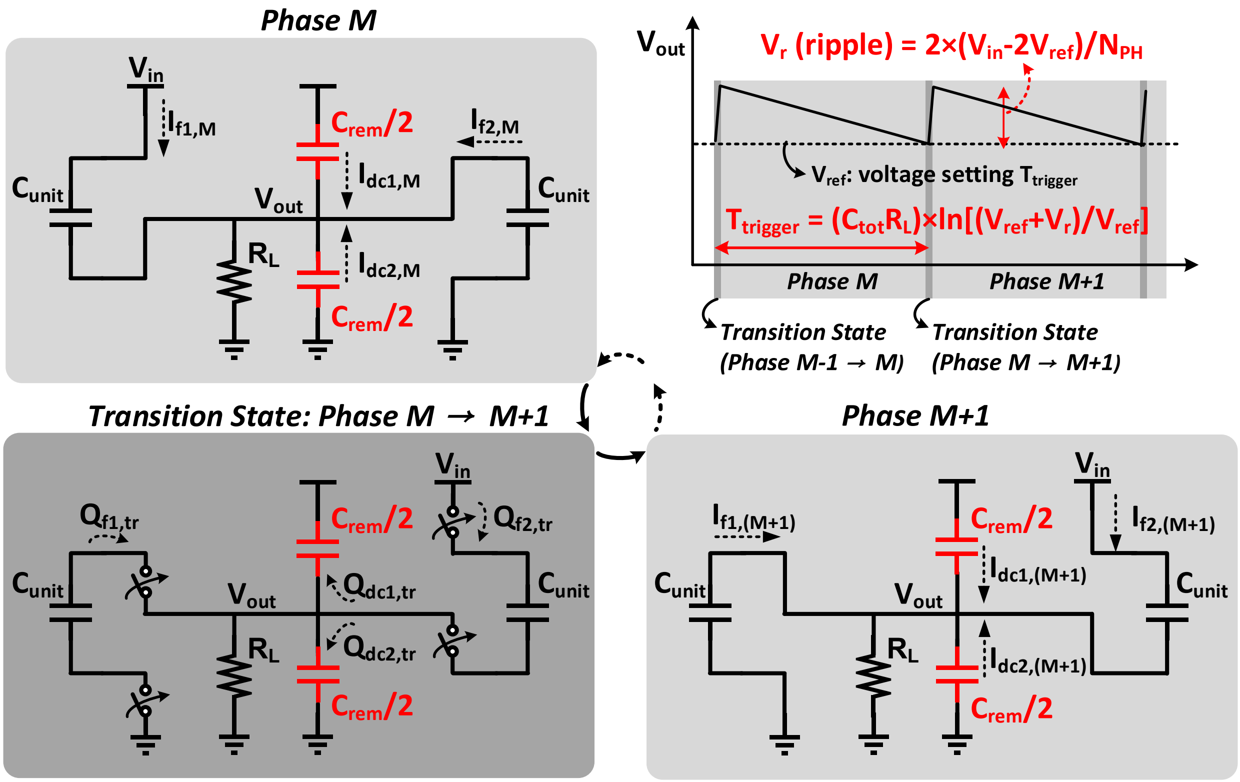

- On-demand Capacitance Modulation
6C_LSB
5C_LSB
T0 T1 T2 T3 T4 T5 T6
Average C_FLY Switching
= 5.33×C_LSB
- On-demand Capacitance Modulation with Dithering-Like Feature
5.4×C_LSB
T_CLK of Comparator
5.2×C_LSB
T0 T1 T2 T3 T4 T5 T6
Add Dithering-Like Feature
Implementation
By having T_SC/N_ph less than T_CLK
(ex) When T_SC/N_ph = T_CLK/5
5.2×C_LSB 5.4×C_LSB 5.4×C_LSB 5.2×C_LSB
6C_LSB
5C_LSB
T0 T1 T2 T3 T4
T_SC/N_PH
T_CLK of Comparator

Figure 5.6. Dithered-capacitance modulation (DCM) is combination of on-demand capacitance modulation and dithering-like feature.

than discrete C_{LSB} values can be conceptually obtained by averaging 5×C_{LSB} and 6×C_{LSB} in time. This can be realized by allowing phase resolution of the SC converter switching periods (T_{SC}/N_{PH}) smaller than clock period of the comparator (T_{CLK}) and considering the average C_{FLY} during this time window as an effective C_{FLY}. In Figure 5.6, where phase resolution (T_{SC}/N_{PH}) is set as $T_{CLK}/5$, 5.2×C_{LSB} is obtained by consecutively switching [5 5 6 5 5]×C_{LSB} in T_{CLK}, and 5.4×C_{LSB} is obtained by consecutively switching [5 6 5 6 5]×C_{LSB} in T_{CLK}. This is similar to the dithering concept in an oversampled analog-to-digital converter (ADC) where toggling between neighboring quantized values can be used to obtain fine resolution and good linearity [56]. We thus achieve dithered capacitance modulation (DCM) by combining this dithering feature with on-demand capacitance modulation.



C_{tot} = N_{PH} × C_{unit}, where C_{tot} = total capacitance & N_{PH} = number of interleaved phases
C_{rem} (remaining capacitors that do not switch at a transition state) = C_{tot} - 2×C_{unit}

Figure 5.7. Simplified operation of SC converter with N_{PH}-phase interleaving, resistor load (R_L) and clock period of N_{PH}×$T_{trigger}$, which sets minimum peak of V_{out} equal to V_{ref}, and corresponding output waveform.

5.2.3 Implication of Ripple Reduction: Power Conversion Efficiency (PCE), Load Power Utilization Factor (PUF), and Effective Power Conversion Efficiency (PCEeff)

With the load circuit modeled as a resistor, Figure 5.7 illustrates simplified operation of an SC converter with N_{PH}-phase interleaving, where the SC converter clock period (T_{SC}) is set so that the minimum output voltage, V_{min}, equals V_{ref}. Output voltage ripple (V_r) can be derived using a procedure as outlined next. Output voltage (V_{out}) changes from V_{ref} to V_{ref}+V_r during the phase transition state. If C_{rem} is the remaining capacitance that does not switch during the transition state, half of C_{rem} is connected between V_{out} and the input voltage (V_{in}) while the other half is connected between V_{out} and ground. Then, the instantaneous charge influx to the latter C_{rem}/2, $Q_{dc2,tr}$, is (C_{rem}/2)×V_r. As shown in Figure 5.7, when the state changes from phase M to phase M+1 (transition state), the voltage across the left C_{unit} changes from V_{in}-V_{ref} to V_{ref}+V_r, and the left C_{unit} loses charge of $Q_{f1,tr} = C_{unit}\times\{(V_{in} - V_{ref} - (V_{ref}+V_r)\} = C_{unit}\times(V_{in} - 2V_{ref} - V_r)$ in this process. After solving $Q_{dc2,tr} = Q_{f1,tr}$, we can represent the output voltage ripple as:

$$V_r = \frac{2\times(V_{in}-2V_{ref})}{N_{PH}} \tag{5.2}$$

where N_{PH} is the number of phases, V_{in} is the input voltage, and V_{ref} is the desired V_{min}.

Trigger period ($T_{trigger} = T_{SC}/N_{PH}$) can be written as below because the RC time constant is R_LC_{tot}, and $T_{trigger}$ is the time it takes to discharge the output node from V_{ref}+V_r to V_{ref}:

$$T_{trigger} = C_{tot}R_L \ln\left\{\frac{(V_{ref}+V_r)}{V_{ref}}\right\} \tag{5.3}$$

Based on equations (5.2) and (5.3), we can quantify the power delivered to load (P_L), input power (P_{in}), and power conversion efficiency (PCE = P_L/P_{in}) by the following equations:

$$P_L = \frac{V_{rms}^2}{R_L}, \text{where } {V_{rms}}^2 = \frac{(V_{ref}+V_r)^2+(V_{ref}+V_r)V_{ref}+V_{ref}^2}{3} \tag{5.4}$$

$$P_{in} = \frac{V_{in}Q_{in}}{T_{trigger}} = \frac{V_{in}}{T_{trigger}}\frac{C_{tot}V_r}{2} \tag{5.5}$$

$$\text{Power Conversion Efficiency (PCE)} = \frac{P_L}{P_{in}} = \frac{V_{rms}^2}{R_LP_{in}} \tag{5.6}$$

From the perspective of the digital circuit load, since the minimum output voltage (V_{min}) level determines the maximum operating frequency, we can define a load power utilization factor (PUF) as:

$$\text{Load Power Utilization Factor (PUF)} = \frac{P_{min}}{P_L} = \frac{V_{min}^2}{R_LP_L} = \frac{V_{ref}^2}{R_LP_L} \tag{5.7}$$

Then, we can obtain the effective PCE (PCE_{eff}) as:

$$\text{Effective PCE } (\text{PCE}_{\text{eff}}) = \text{PCE} \times \text{PUF} = \frac{V_{\text{min}}^2}{R_L P_{\text{in}}} = \frac{V_{\text{ref}}^2}{R_L P_{\text{in}}} \tag{5.8}$$

where P_{min} is defined as the load power consumption when the load circuit (modeled with R_L) operates at V_{min}. Similar concepts to PUF and P_{min} are also discussed in [53].

Intuitively, equations (5.3-5.6) imply that PCE degrades with reduction of V_r because a decrease of V_r results in more decrease in P_L than in P_{in}. However, as implied by equations (5.7-5.8), with reduction of V_r, PUF is improved and P_{in} is reduced. resulting in improvement of PCE_{eff}. Therefore, PCE_{eff} can be used as an indicator of both voltage regulation capability and power conversion efficiency, while PCE cannot represent voltage regulation capability. Also, it is noted in equation (5.5) that reduction of V_r attains power consumption saving.

Based on equations (5.2)–(5.8), the effect of ripple on PCE, PUF, and PCE_{eff} can be obtained. In Figure 5.8, by setting C_{tot} = 3.7nF, V_{in} = 2.3V, and V_{ref} = 1V and assuming that $T_{trigger}$ is set according to equation (5.3), PCE, PUF, and PCE_{eff} are plotted against ripple magnitude, which is governed by the number of phases (N_{PH}). Due to a decrease in V_{rms}, it is shown that PCE degrades with ripple reduction, but PUF is improved with smaller ripple: when

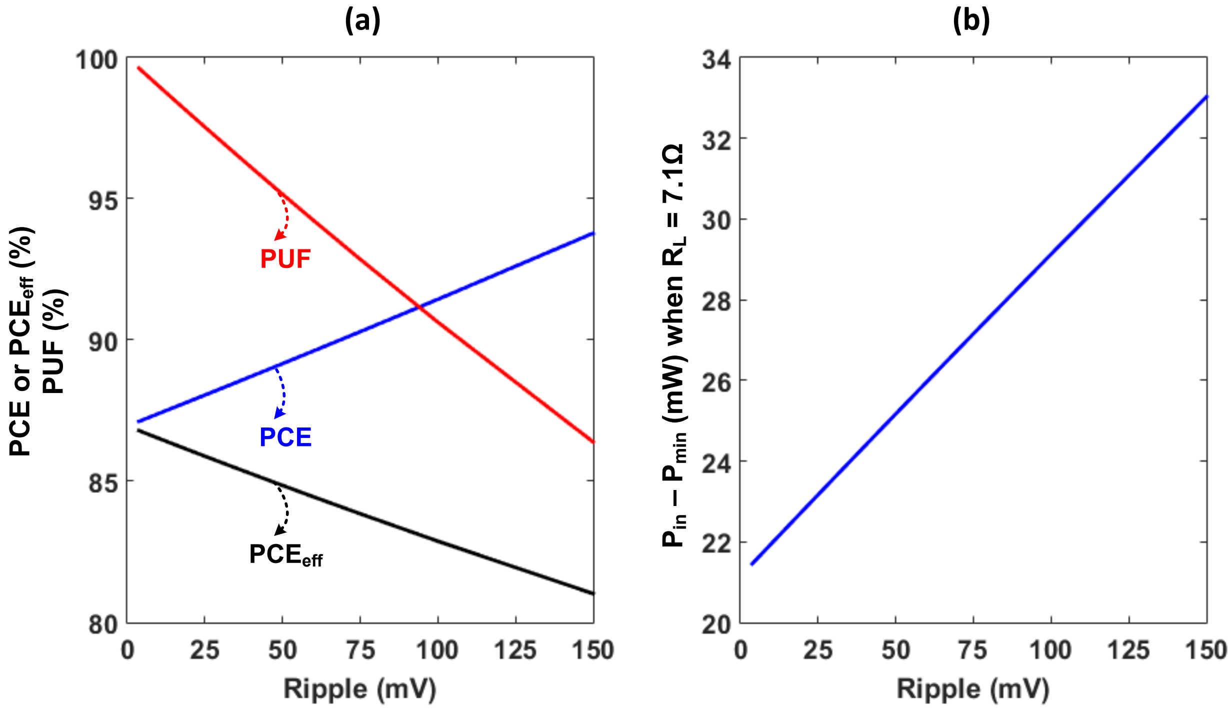


Figure 5.8. Impact of ripple reduction in SC converter (C_{tot} = 3.7nF, V_{in} = 2.3V, V_{ref} = 1V): (a) power conversion efficiency, effective power conversion efficiency, load power utilization factor vs. ripple, (b) power overhead (P_{in} - P_{min}) due to ripple and SC converter vs. ripple.

ripple is reduced from 150mV to 3.7mV, PCE decreases from 93.8% to 87.1%, and PUF increases from 86.4% to 99.6%. Therefore, it is noted that PCE_{eff} is improved with ripple reduction: when ripple is reduced from 150mV to 3.7mV, PCE_{eff} is improved from 81.0% to 86.8%. In other words, input power consumption decreases with ripple reduction. For instance, as shown in Figure 5.8 (b), when R_L = 7.1Ω (load current is approximately 141mA), improving ripple from 150mV to 3.7mV reduces input power consumption by 11.5mW. These results imply that the theoretically attainable PCE of SC converter can degrade with ripple reduction, but ripple reduction decreases load power consumption, thereby improving PUF and PCE_{eff}.

5.3 Implementation of Dithered-Capacitance Modulation

5.3.1 Selection of the Number of Interleaved Phases

The proposed dithered-capacitance modulation (DCM) approach is implemented in a 40-phase switched-capacitor voltage regulator (SCVR) with 4b DCM. The number of phases and the modulation resolution presents a trade-off between voltage ripple and area utilization of capacitance. As the number of phases and modulation resolution increases, the achievable ripple reduces; however the total capacitance is divided among a larger number of individual units,

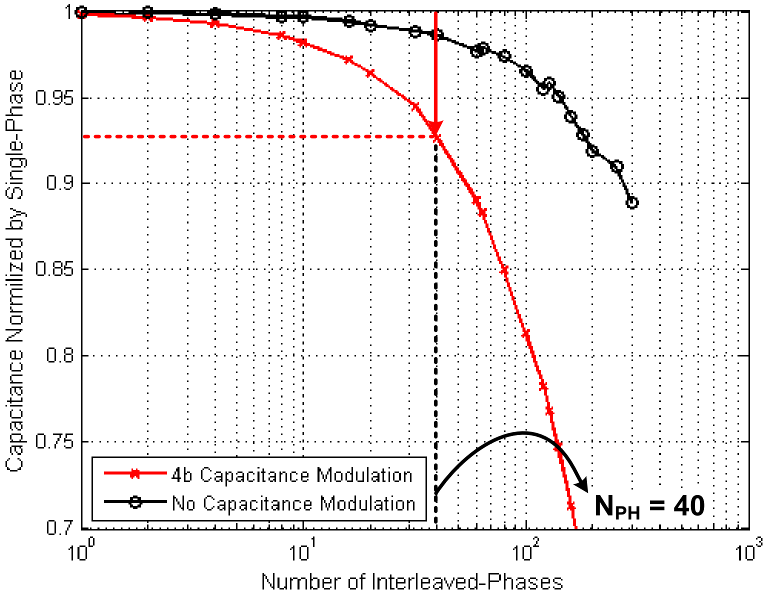


Figure 5.9. Normalized available flying capacitance as a function of number of interleaved phases with and without 4b capacitance modulation under area constraint of 880 × 830 μm^2.

resulting in area overhead due to capacitor spacing requirements in layout and peripheral circuits. Figure 5.9 shows that the available capacitance decreases as the number of interleaved phases increases under a fixed area constraint (in this case, 880×830μm^2). To keep the area overhead below 10%, 40 phases with 4b capacitance modulation was chosen, which has the same area utilization as 160-phase interleaving without capacitance modulation. However, 160-phase interleaving would increase the power consumption and implementation complexity due to the required clock generation by up to 4×. Moreover, clocks with excessive number of phases are susceptible to variation, where the ripple reduction benefits could be diminished.

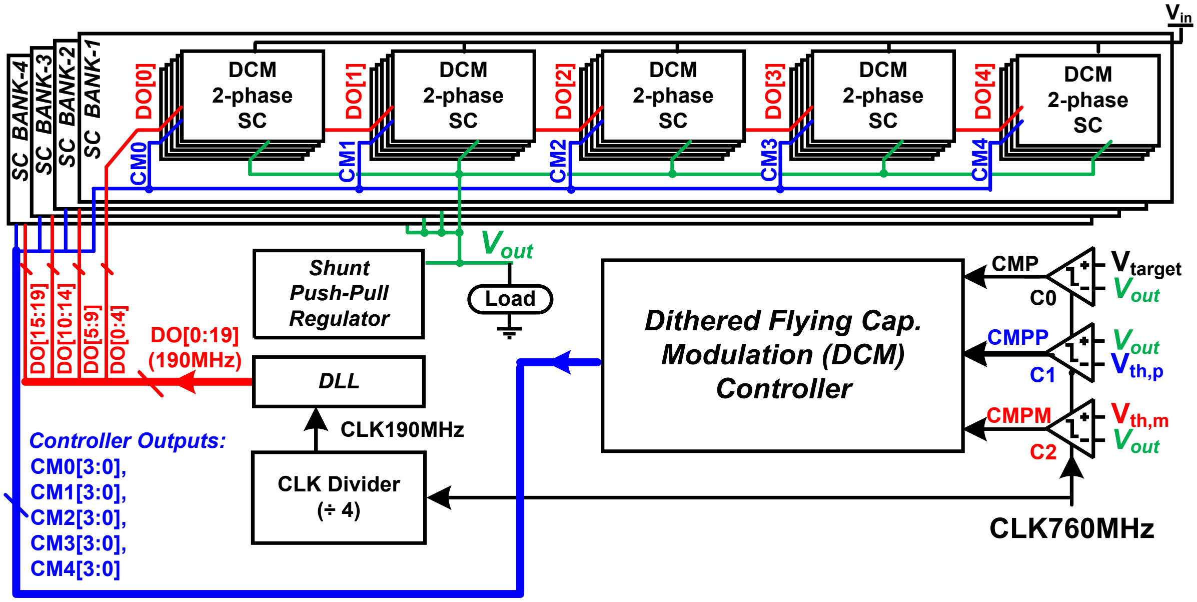


Figure 5.10. Top-level diagram of 40-phases switched-capacitor voltage regulator (SCVR) with 4b dithered-capacitance modulation (DCM).

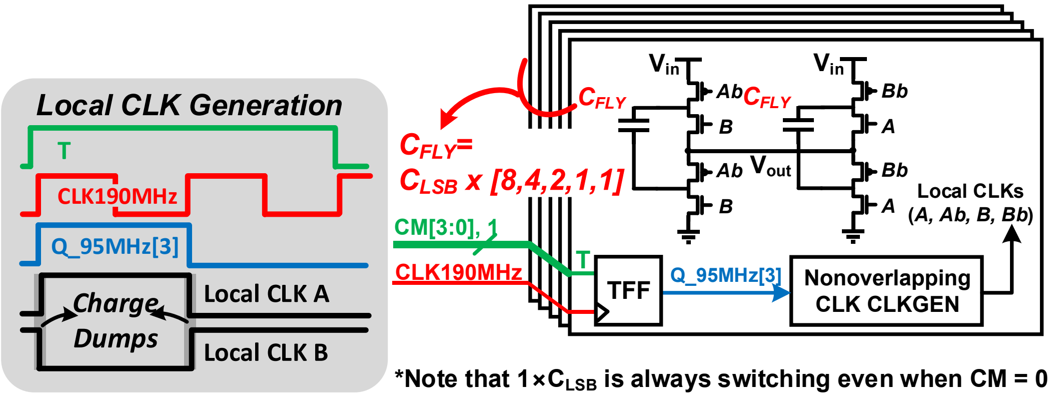


Figure 5.11. Two-phase SC converter block for 4b dithering-capacitance modulation (DCM).

5.3.2 40-Phase Switched-Capacitor Voltage Regulator with 4b Dithered-Capacitance Modulation

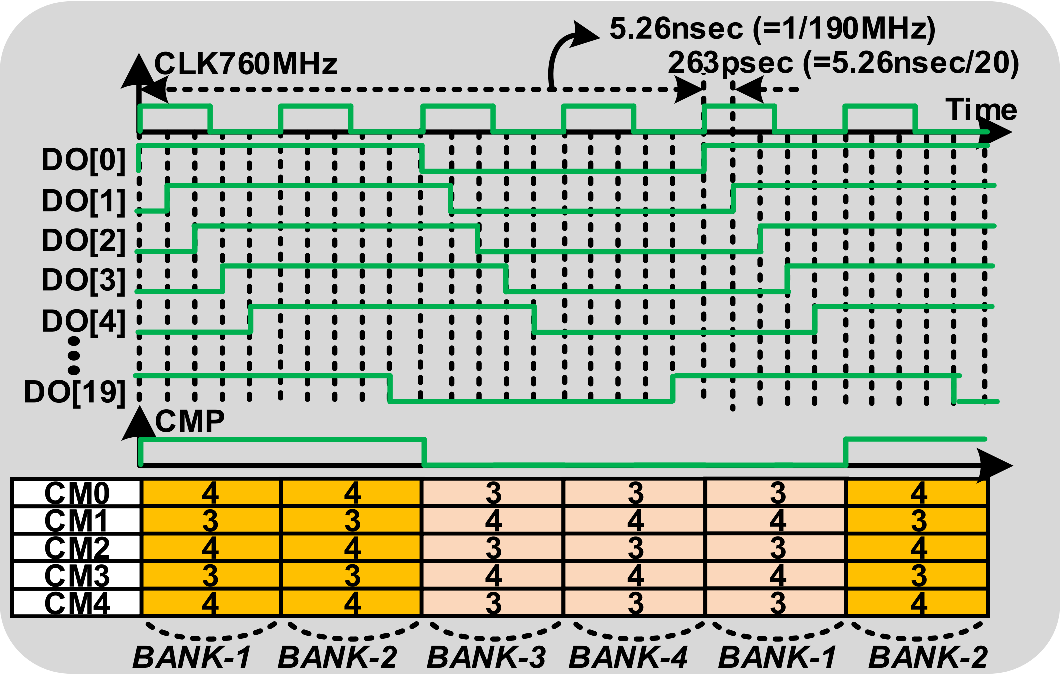
The 40-phase SCVR with 4b DCM is composed of four SC-banks as shown in Figure 5.10. Each SC-bank comprises five 2-phase SC converter blocks with 4b DCM, and each 2-phase SC converter block consists of SC converters with 4+1 binary-weighted flying capacitors where C_{LSB} = 5.8pF to provide a discrete flying capacitance value for each SC phase as shown in Figure 5.11. No explicit output capacitance is used. To ensure a fixed SCVR frequency and minimum ripple, one additional flying capacitor with C_{LSB} is always switching. The local clock generation waveform in the 2-phase SC converter is illustrated in Figure 5.11. Using a toggle flip-flop (TFF), a non-overlapping clock generator, and level converters, local 95MHz clocks are generated from the 190MHz input clock when the input signal to the TFF is asserted. Note that the SC converter recovers output voltage droop by dumping charge to the output node during the transition of the local clocks. A DCM controller adjusts inputs to TFFs, which is represented by CM[3:0], for capacitance modulation.

As shown in Figure 5.10, the 760MHz master clock is first divided down to 190MHz. A delay-locked-loop (DLL) then generates twenty 190MHz clocks with 263ps resolution between phases, and each of the 190MHz clocks drives a 2-phase SC converter block. Each 2-phase SC converter locally generates two non-overlapping, half-frequency clocks (95MHz) and allows 5.2ns between charge transfers. In total, twenty charge transfers occur in a 190MHz clock cycle, resulting in an effective operation at 3.8GHz.

Three comparators (C0-C2) operate off the 760MHz master clock, generating a comparison output every five clock phases. References $V_{th,p}$ and $V_{th,m}$ are used to adjust CM upon load current changes, and V_{target} is used to regulate V_{out} in steady state. The steady-state example in Figure 5.12 shows CM dithering between CM[3:0] = 3.6, when CMP = 1 (cycles 1-2) and CM[3:0] = 3.4, when CMP = 0 (cycles 3-5), resulting in an average CM value of 3.48.

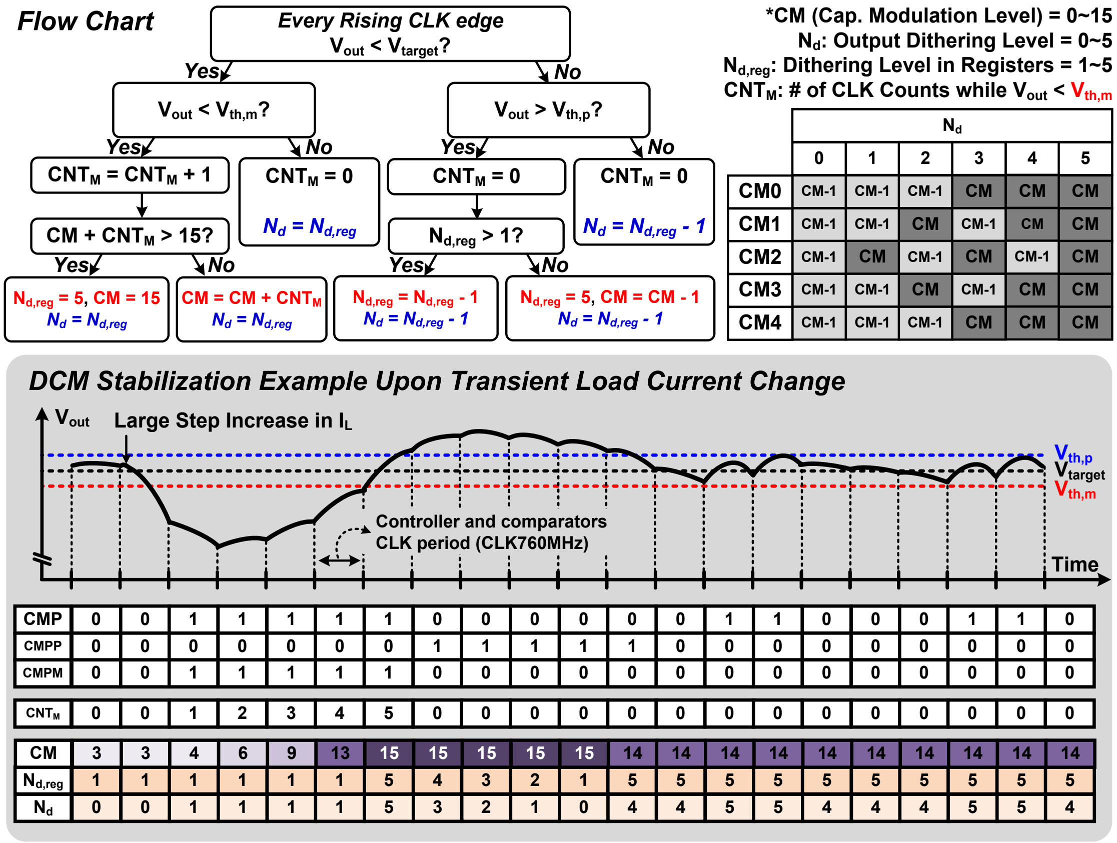
5.3.3 Dithered-Capacitance Modulation Controller

Figure 5.13 shows a flow chart of the dithered-capacitance modulation (DCM) controller, and a table of dithered capacitance modulation output generation and a stabilization example under transient change in load current. The DCM controller adjusts the capacitance modulation level (CM) for a given load current, and CM increases with a load current increase. Five



	BANK-1	BANK-2	BANK-3	BANK-4	BANK-1	BANK-2
CM0	4	4	3	3	3	4
CM1	3	3	4	4	4	3
CM2	4	4	3	3	3	4
CM3	3	3	4	4	4	3
CM4	4	4	3	3	3	4

Figure 5.12. Operation of 40-phase SCVR with 4b DCM in steady state.



*CM (Cap. Modulation Level) = 0~15
N_d: Output Dithering Level = 0~5
$N_{d,reg}$: Dithering Level in Registers = 1~5
CNT_M: # of CLK Counts while $V_{out} < V_{th,m}$

	N_d					
	0	1	2	3	4	5
CM0	CM-1	CM-1	CM-1	CM	CM	CM
CM1	CM-1	CM-1	CM	CM-1	CM	CM
CM2	CM-1	CM	CM-1	CM	CM-1	CM
CM3	CM-1	CM-1	CM	CM-1	CM	CM
CM4	CM-1	CM-1	CM-1	CM	CM	CM

CMP	0	0	1	1	1	1	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0
CMPP	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
CMPM	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

CNT_M	0	0	1	2	3	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0

CM	3	3	4	6	9	13	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14
$N_{d,reg}$	1	1	1	1	1	1	5	4	3	2	1	5	5	5	5	5	5	5	5	5	5
N_d	0	0	1	1	1	1	5	3	2	1	0	4	4	5	5	4	4	4	5	5	4

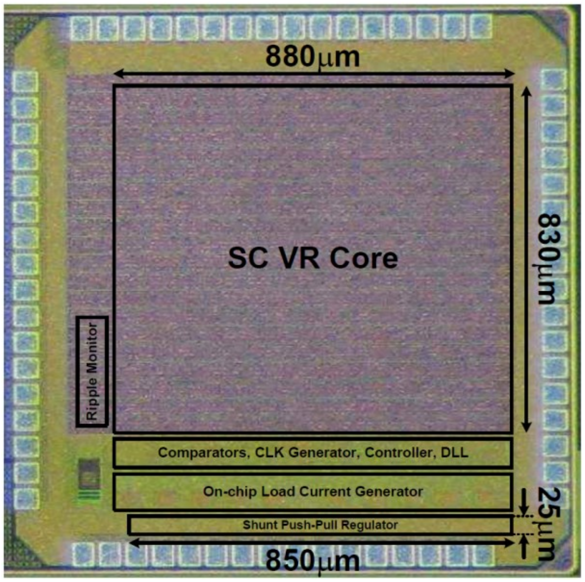
Figure 5.13. Flow chart of DCM controller (top). Stabilization example under transient load current change (bottom).

modulation signals CM0-CM4 are generated as a function of base discrete CM and output dithering value N_d, as shown in the top right part of Figure 5.13. N_d is obtained from $N_{d,reg}$, base dithering level stored in registers. N_d is set $N_{d,reg}$ - 1 for $V_{out} > V_{target}$, and $N_{d,reg}$ for $V_{out} < V_{target}$. In transient condition, the controller adjusts the base discrete CM and $N_{d,reg}$, and after entering stead-state, only N_d is adjusted between $N_{d,reg}$ - 1 and $N_{d,reg}$ depending on V_{out}. For instance, to generate CM[3:0] = 3.6 and 3.4 in Figure 5.12, the controller sets the base discrete CM = 3, and $N_{d,reg}$ = 3 in transient condition. In steady-state, for $V_{out} < V_{target}$, the controller sets N_d = 3, and it sets CM0 = 4, CM1 = 3, CM2 = 4, CM3 = 3, and CM4 = 4. For $V_{out} > V_{target}$, the controller sets N_d = 2, and it sets CM0 = 3, CM1 = 4, CM2 = 3, CM3 = 4, and CM4 = 3. The maximum value of N_d and $N_{d,reg}$ were set at 5. This is because (comparator clock period)/(phase resolution in switching clocks for SCVR) = $T_{CMP}/(T_{SC}/N_{PH})$ = 5 and one SC bank has five SC converter blocks. As CM[3:0] changes between two neighboring values to regulate V_{out} close to V_{target}, limit-cycle oscillation and output ripple ensue in steady-state. This locked-state of the closed loop of DCM behaves similar to bang bang PLL (BBPLL), in that it has inevitable limit cycles, and it cannot be analyzed in the traditional Laplace domain because of nonlinearity introduced by a 1-bit quantizer or a comparator in the loop. DCM, which is digitally controlled with a comparator, regulates V_{out} by asymptotically settling around V_{target} [59].

Operation sequence of the controller in transient condition is as follows: upon a large load current change, if $V_{out} < V_{th,m}$, then counter CNT_M is incremented and added to the base discrete CM (e.g. CM = CM+CNT_M). This increases the base discrete CM geometrically for each subsequent cycle with $V_{out} < V_{th,m}$, and it can lead to an overshoot in the output voltage. When $V_{out} > V_{th,p}$, the controller decrements the stored dithering level $N_{d,reg}$ by one in each cycle until $N_{d,reg}$ reaches zero, at which point the base discrete CM is decreased by one and $N_{d,reg}$ is reset to 5. The much stronger adjustment of C_{FLY} when $V_{out} < V_{th,m}$ than when $V_{out} > V_{th,p}$ minimizes the likelihood of an undershoot in favor of an overshoot. In steady state, where V_{out} lies between $V_{th,p}$ and $V_{th,m}$, only the dithering level (N_d) is adjusted. One example for transient stabilization of DCM controller is illustrated in the bottom part of Figure 5.13. Finally, a shunt push-pull regulator proposed in [53] is used to mitigate undershoot and overshoot against transient load current change.

5.4 Measurement Results

A test chip was fabricated in a 65nm CMOS process to compare the DCM and PFM schemes. A die photo and area summary are shown in Figure 5.14. MIM and MOS capacitors are



880μm
SC VR Core
830μm
Ripple Monitor
Comparators, CLK Generator, Controller, DLL
On-chip Load Current Generator
Shunt Push-Pull Regulator
25μm
850μm

	Area (μm^2)	Relative Ratio
SC VR Core	730,400	95.8%
Shunt Push-Pull Regulator	21,250	2.8%
Comparators, Controller, and DLL	10914	1.4%
Total	762,564	100%

Figure 5.14. Die photo and area summary.

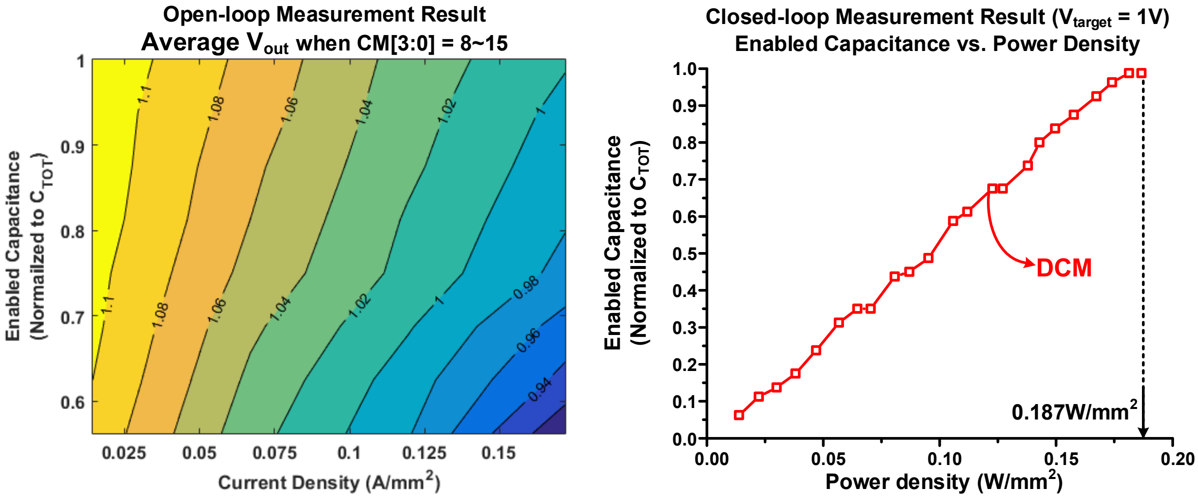


Figure 5.15. Contour plot of measured output voltage in open-loop as a function of current density and normalized enabled capacitance when CM[3:0] = 8~15 (left). Normalized enabled capacitance in closed-loop versus power density when V_{target} = 1V (right).

used for flying capacitors in this SCVR to achieve a total capacitance of 3.7nF (MIM capacitors = 0.93nF, and MOS capacitors = 2.77nF). Due to pn-junction diodes formed by p-substrate and n-well, MOS capacitors have larger bottom-plate parasitic capacitance compared to MIM capacitors, but they have good capacitance density. MOS and MIM capacitors respectively have k_{bot} = 5% and k_{bot} = 1% where k_{bot} = ratio of bottom-plate parasitic capacitance to non-parasitic capacitance, and MOS capacitors have > 2× capacitance density in comparison to MIM capacitors. With utilization of MOS capacitors in addition to MIM capacitors, die space can be more efficiently utilized as MOS capacitors can be placed beneath MIM capacitors, and higher power density can be achieved.

Parameters V_{in}, V_{target}, $V_{th,m}$, and $V_{th,p}$ are set to 2.3V, 1V, 0.985V, and 1.015V, respectively. Figure 5.15 shows the measurement results for open-loop operation and closed-loop operation of the DCM scheme. While the total capacitance is kept constant, a load current increase leads to an increase in the modulated flying capacitance for voltage regulation as expected. Open-loop measurements confirm that adjustment of V_{target}, $V_{th,m}$, and $V_{th,p}$ enables regulation of the output voltage to different voltage levels.

For accurate ripple measurement, we implemented an on-chip monitoring circuit [52] that consists of a comparator, which is asynchronously clocked, and 15b counters, which record the fraction of cycles with $V_{out} < V_{RMC}$, as shown in Figure 5.16. The measurement sequence is as follows:

1) Sweep V_{RMC} (voltage at plus-terminal of comparator) for the voltage range of interest.

2) At each V_{RMC}, start over after resetting two counters (CNT and CNT_{REF} = 0).

3) When DONE = 1, calculate probability = $CNT/(2^{15}-1)$.

4) Find V_{prob99} and V_{prob1} (with V_{probN} defined as voltage with probability equal to N%).

5) Calculate ripple from the difference between V_{prob99} and V_{prob1}.

Figure 5.16 shows an example plot for this on-chip ripple measurement. The probability calculated using outputs of the ripple monitor is plotted against V_{RMC} for DCM and PFM at a constant load current. DCM achieves 6mV ripple, compared to 140mV for PFM at I_L = 11mA. It should be noted that pulse frequency modulation (PFM) mode can be obtained by reconfiguring the existing DCM controller, as illustrated in Figure 5.17. In PFM mode, only comparator C0 is used to trigger the controller and CM0-CM4 are all set to 15 or 0, depending on the output

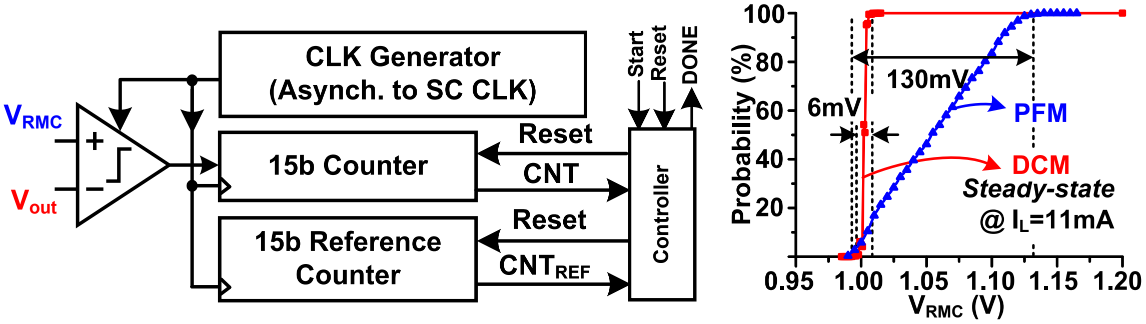
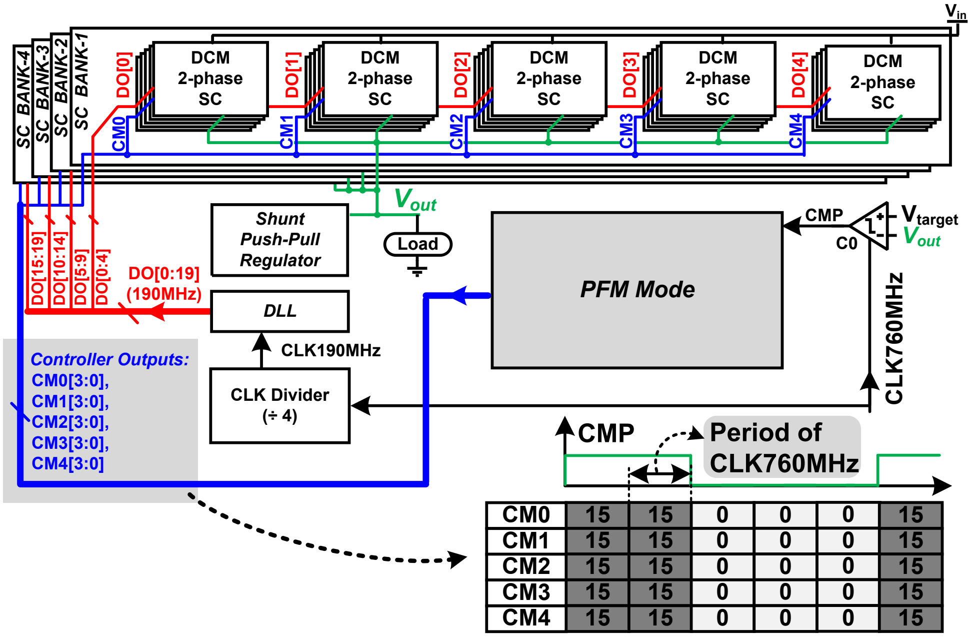


Figure 5.16. On-chip ripple measuring circuit (left) and on-chip ripple measurement at load current of 11mA (right).



CM0	15	15	0	0	0	15
CM1	15	15	0	0	0	15
CM2	15	15	0	0	0	15
CM3	15	15	0	0	0	15
CM4	15	15	0	0	0	15

Figure 5.17. Block diagram of test-chip and change of capacitance modulation as a function of comparator output when configured to PFM mode.

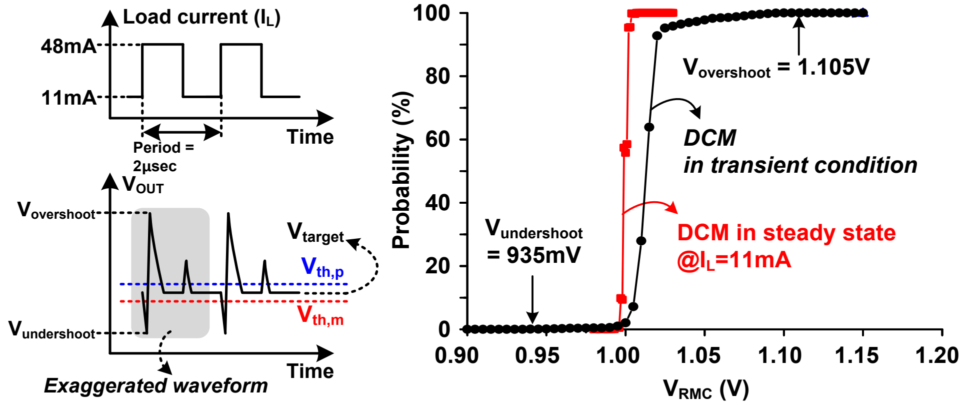


Figure 5.18. Periodic load current change between 11mA and 48mA with period of 2μsec (left), and corresponding overshoot and undershoot measurement results (right).

voltage level. DCM and PFM were both measured with the same on-chip ripple monitor for comparison.

Figure 5.18 shows the output voltage response in DCM to a periodic load current change between 11mA and 48mA (period = 2µs). Undershoot and overshoot voltages under these conditions were also measured with the on-chip ripple monitor. It was observed that the periodic load current change results in 65mV undershoot and 105mV overshoot, which are obtained by finding V_{RMC} with probability = 0% and 100%, respectively.

The fact that the overshoot is larger than the undershoot is a direct result of the more aggressive controller response to $V_{out} < V_{th,m}$. Figure 5.19 (a) shows the measured DCM and PFM ripple versus power density and Figure 5.19 (b) shows the average output voltage versus power density, both for the load current range from 11mA to 142mA. DCM ripple ranges from

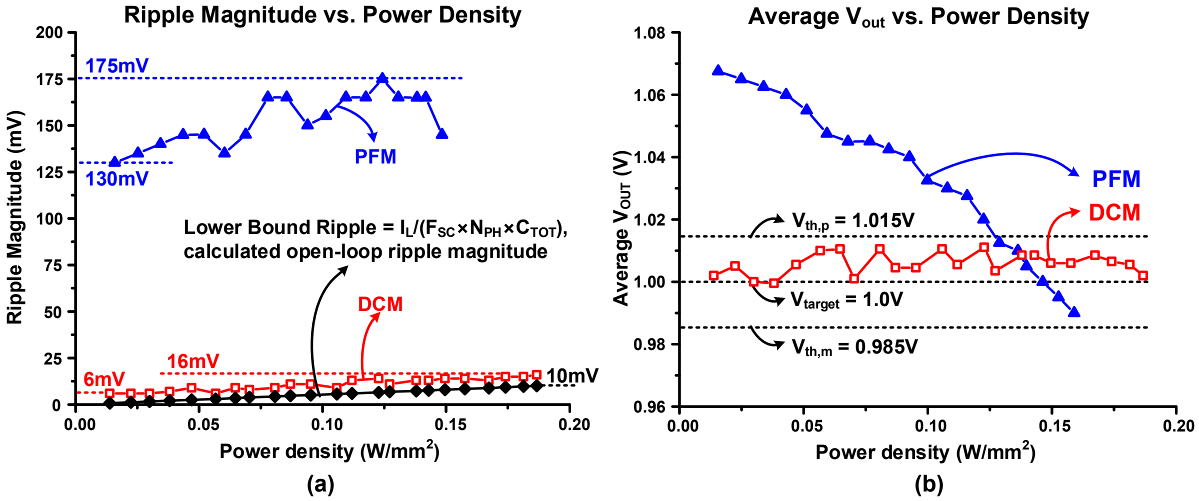


Figure 5.19. (a) Measured ripple magnitude versus power density for DCM and PFM, overlaid with calculated open-loop ripple, and (b) measured average output voltage (V_{out}) of DCM and PFM.

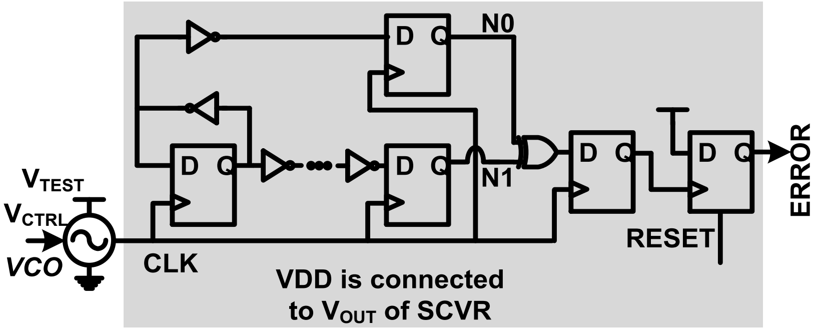


Figure 5.20. On-chip digital load circuit performance monitor.

6mV to16mV and scales with load current, as expected. DCM ripple closely tracks the open-loop ripple expression $I_L/(F_{SC}\times N_{PH}\times C_{TOT})$, and the output voltage of DCM is tightly regulated to V_{target}=1V.

The load performance monitor proposed in [53] was implemented in the test chip to investigate the impact of output ripple on digital load circuit performance. As shown in Figure 5.20, the monitor circuit measures propagation time of a path with a voltage controlled oscillator, the frequency of which can be observed externally. By sweeping control voltage (V_{CTRL}) and

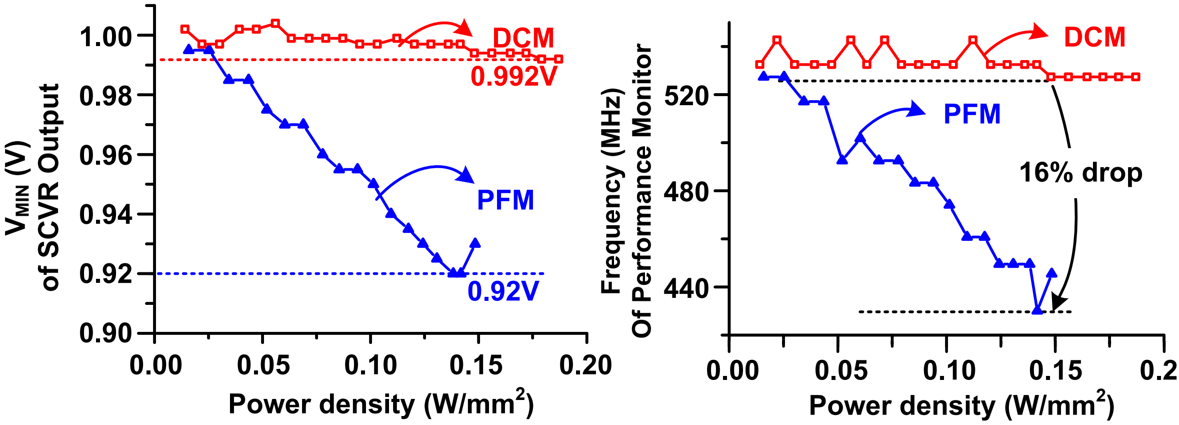


Figure 5.21. Measured minimum peak output voltage (V_{MIN}) versus power density (left). Impact of V_{MIN} on frequency of on-chip digital load circuit performance (right).

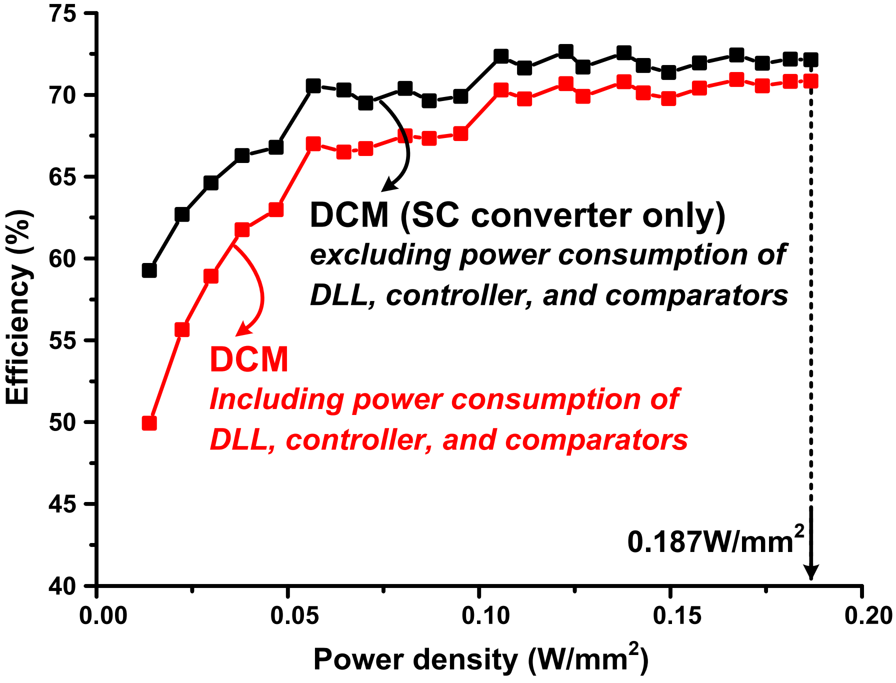


Figure 5.22. Measured efficiency versus power density.

changing clock frequency until the monitor flags an error signal, the maximum frequency, or the propagation time of the path, is found. The supply voltage of the load performance monitor is connected to output voltage of the SC converter in the test chip, such that the impact of the minimum output voltage (V_{MIN}) on the digital-load performance can be compared for DCM and PFM modes. In Figure 5.21, V_{MIN} is plotted against power density for both DCM and PFM. Using the on-chip digital load performance monitor, it is found that the PFM-driven circuit exhibits 16% slower performance than DCM at large load current. Figure 5.22 shows that DCM achieves a peak efficiency of 70.8% at a power density of 0.187W/mm^2, which corresponds to I_L = 142mA. DLL, controller and comparators consume 3.3mW. Excluding power consumption of the DLL, controller, and comparators, the SC converter itself achieves peak efficiency of 72.6%.

Table 5.1 summarizes power loss breakdown of the implemented DCM SC converter at I_L = 142mA. Main power loss is attributed to switching loss, because (1) the number of clock-driving circuits such as toggle flip-flops, level shifters and non-overlapping clock generators increase in proportion to the number of phase and capacitance modulation, and (2) switching loss due to bottom-plate parasitic capacitors of MOS capacitors is significant. To improve efficiency, the number of phases and capacitance modulation could be reduced with a relaxed constraint on ripple, and switching frequency could be reduced while trading off power density. Fabrication in advanced CMOS process can also help resolve switching loss of clock driving circuits and parasitic switching loss of flying capacitors. The bottom-plate parasitic switching loss (P_{bot}) can be written as equation (5.9). If MIM capacitors of 3.7nF had been used instead of the combination of MIM and MOS capacitors, P_{bot} can be significantly reduced with power density degradation, and peak efficiency could be improved by approximately 4%.

$$P_{bot} = k_{bot} C_{FLY} V_{sw}{}^2 F_{sw} \tag{5.9}$$

where C_{FLY} = flying capacitance, and V_{sw} = swing voltage of bottom-plate parasitic capacitor.

A comparison to prior work is summarized in Figure 5.23 and Table 5.2. Figure 5.23 plots peak efficiency and power density of prior works and the proposed work in a similar fashion to [51]. The dotted line in Figure 5.23 indicates that peak efficiency and power density have trade-off relation. Prior works fabricated in 65nm process were grouped. Compared to the other 65nm prior work, this work shows slightly lower peak efficiency because of split flying capacitors and reduced capacitance utilization per area. Also, if flying capacitors had not been split in this work, higher power density could have been obtained at the same peak efficiency.

On-chip capacitance density directly affects power density and parasitic capacitance affects achievable peak efficiency. As on-chip capacitance density and parasitic capacitance are determined by fabrication technology, the proposed work achieved higher power density than prior works in process with inferior technology than 65nm, but less power density or less peak efficiency than in process with advanced technology such as deep trench capacitor, ferroelectric capacitor, and SOI technology.

Component	Sub-component	Power Loss	Percentage
Conduction Loss		18.5mW	33.2%
Switching Loss		33.8mW	60.8%
	P_{bot} of MIM-capacitors (sim.)	0.88mW	1.6%
	P_{bot} of MOS-capacitors (sim.)	13.2mW	23.7%
Controller, Comparator, and DLL		3.3mW	5.9%
Total		55.6mW	100%

Table 5.1 Power loss breakdown of DCM SC converter (at IL = 142mA).

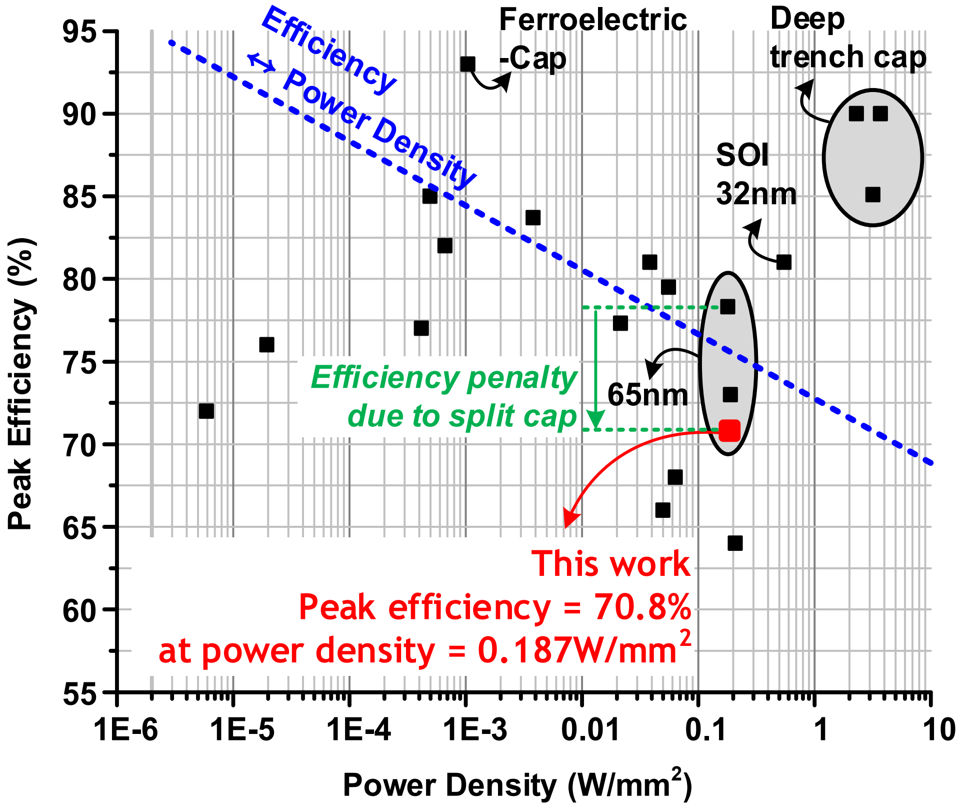


Figure 5.23. Comparison of peak efficiency versus power density: prior works are indicated with black square symbols, and this work is indicated with red square symbol.

5.5 Summary

In this work, dithered capacitance modulation (DCM) is proposed to minimize closed-loop ripple in SCVRs with multi-phase interleaving. It is shown that ripple reduction improves load power utilization and effective input power conversion efficiency (PCE_{eff}), thereby reducing power consumption at a constant minimum peak output voltage (V_{min}). An SC converter with 40-phase interleaving and 4b DCM level was implemented in 65nm CMOS technology to achieve a ripple magnitude of 6-16mV for load currents ranging from 11mA to 142mA as was measured with an on-chip ripple measurement circuit. This work was presented and published in [57, 63].

Metric	[37]	[36]	[33]	**This Work**
Technology	22nm Tri-gate	32nm SOI	45nm Bulk	**65nm Bulk**
Capacitor Type	MIM	Deep Trench	MOS	**MIM + MOS**
Conversion Ratio (M)	2:1, 3:2, 5:4, 1:1	2:1, 3:2	3:2, 2:1	**2:1**
Closed-loop Modulation	Frequency	Frequency	Frequency (PFM), Capacitance	**Capacitance Dithering**
Comparator Frequency	2GHz	4GHz	30MHz	**760MHz**
SC Converter Frequency	≤ 250MHz	≤ 125MHz	≤ 30MHz	**95MHz**
Number of Phases	8	16	2	**40**
C_{FLY} / C_{OUT}	[1]N/R / 0.1nF	16nF / 0	0.534nF / 0.7nF	**3.7nF / 0**
V_{in} / V_{out}	1.23V / 0.45 ~ 1V	1.8V / 0.7 ~ 1.1V	1.8V / 0.8 ~ 1V	**2.3V / 1V**
Ripple Measurement	Off-chip	Off-chip	Off-chip	**On-chip**
$V_{ripple,pp}$	60mV @≤88mA	30mV @365mA	50mV @≤10mA	**6~16mV @11~142mA**
V_{droop}	<25mV @15→30mA	94mV @30→365mA	155mV @2→10mA	**65mV @11→48mA**
Power Density (ρ)	0.064 W/mm[2]	2.71 W/mm[2]	0.050 W/mm[2]	**0.187 W/mm[2]**
Efficiency @ ρ	68%	86.4%	66%	**70.8%**

Table 5.2. Comparison table ([1]N/R = Not reported, [2]Power density = reported @ M=2:1).

CHAPTER 6

Low-Power Deep Learning Co-Processor

6.1 Deep Learning Co-Processor Design for Internet of Things (IoT) Devices

Deep learning is also known as neural network, which is named after neural because connections between layers resemble ones between neuron and synapse in nerve system. Figure 6.1 shows fully-connected layers of a deep neural network (DNN), a neural network consists of many hidden layers of neurons and many synapses that connect neurons in neighboring layers, in addition to input layer and output layer. The number of hidden layers, neurons, synapses and synapse information can be configured, depending on application and available memory resources. For instance, [71, 72] shows similar DNN for different application such as keyword speech recognition or hand-written digits recognition. In the proposed work, for operation in the IoT platform, we utilize small neural networks that were optimized for memory-limited condition in [71].

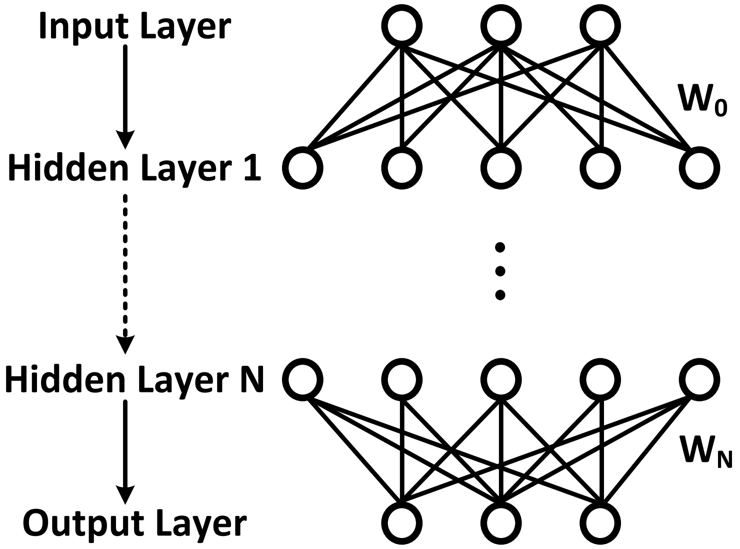


Figure 6.1.A deep neural network with N fully-connected hidden layers.

This chapter proposes a low-power configurable deep learning co-processor which performs fully-connected layer operation of DNN and FFT, aiming for keyword speech recognition, handwritten digits detection, and other generic digital signal processing tasks. The simplified diagram of proposed co-processor is shown in Figure 6.2. Low-power digital signal processor (DSP) with four processing elements (PEs) and hierarchical memory performs the fully-connected layer in addition to FFT. The inference operation of deep learning is only implemented as training operation of deep learning is expensive to be performed in the untethered IoT device, and it can be easily done in personal computers with better computing resources, earlier than inference operation.

6.1.1 Proposed Techniques and Data Flow for Deep Learning and Benefits

Connections between two layers in the deep neural network shown in Figure 6.1 can be regarded as a vector matrix multiplication, as all the neurons in the previous layer have connections to all the neurons in the next layer. Neurons of the previous layer and the next layer can be regarded as vectors, and a synapse between two neurons is regarded as an element of weight matrix, which is formed by all the synapses between the two neighboring layers. In Figure 6.1, W_0 represents weight matrix between input layer and hidden layer 1, and W_N represents one between hidden layer N and output layer. In addition, the neural network applies

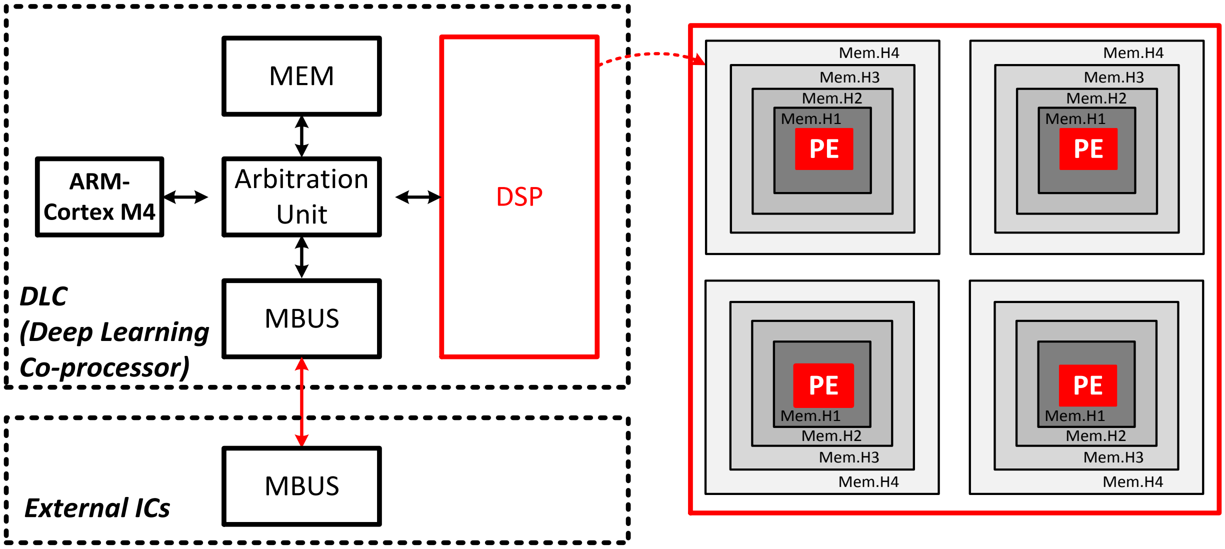


Figure 6.2. Simplified block diagram of proposed deep learning co-processor architecture with four PEs, and non-uniform memory structures.

offset addition and a transfer function such as tanh(x), max(0,x), etc., to the results of the vector-matrix multiplication. The arithmetic operation of a fully-connected layer can be described as the following equation:

$$out(\mathrm{p}) = \mathrm{NLI}\left(\sum_{q=0}^{N_{in}} w_{p,q} \times in[\mathrm{q}] + b[\mathrm{p}] \right) \qquad (6.1)$$

where NLI() is a transfer function, *out* is an output neuron layer, *w* is a synapse matrix or a weight matrix, *in* is an input neuron layer, and *b* is offset.

In order to perform energy efficient and low power operation, we propose deep learning processor utilizing the following techniques.

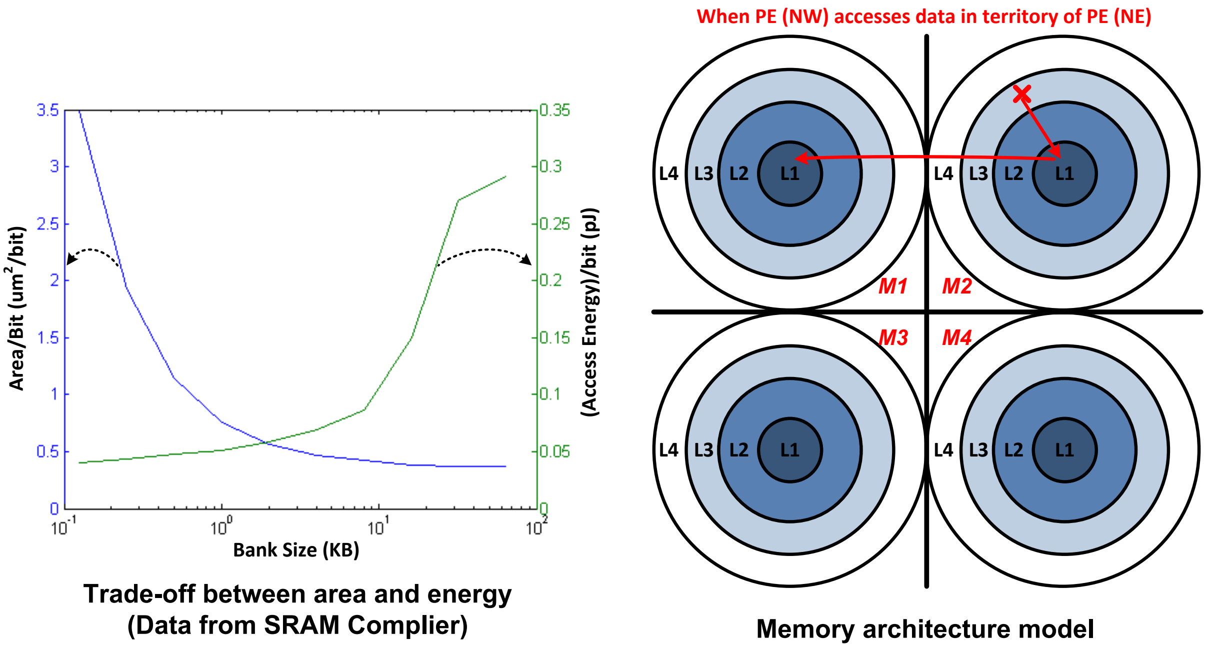
(1) Non-uniform memory architecture (NUMA).

(2) Temporal and spatial locality.

(3) Weight matrix tiling.

(4) Variable data precision, variable multiplier width, and long one line of memory.

As implied by equation (6.1), in a fully-connected layer, it is noted that input is repeatedly used to obtain output layer, and that weight matrix is access only once for the entire



Trade-off between area and energy (Data from SRAM Complier)

Memory architecture model

Figure 6.3. Trade-off between area and access energy is plotted against SRAM bank size. Data were taken from 40nm SRAM compiler datasheet (left), and non-uniform memory architecture model (right).

operation. Also, input size is much smaller than weight matrix. As shown in Figure 6.3, SRAM memory has trade-off between area efficiency and memory access energy according to bank size. We utilize NUMA for temporal spatial locality. Four hierarchical memories form one NUMA. By designing NUMA such that memory in the 1st hierarchy (H1) has the smallest access energy and the worst area efficiency in the 4th hierarchy (H4) has the best area efficiency and the largest access energy, for temporal locality, we can store input to H1 memory, and weight to H4. Also, for spatial locality, each PE is surrounded by its own NUMA, and thus assignment of data closer to corresponding PE is enabled.

In a similar context, weight matrix tiling is proposed. Figure 6.4 illustrates operation sequence of vector-matrix multiplication in baseline and in the proposed matrix tiling. In the baseline shown in the top of Figure 6.4, element-wise multiplications of one row of weight matrix and input vector are obtained and added to obtain an output element. This way, each input element is accessed by eight times. In contrast, in the proposed tiling approach shown in the bottom of Figure 6.4, by tiling the weight matrix into 2 × 2, each input element is only twice. In this example, we process tiles in the order from top left, to top right, bottom left, bottom right. Thus, when an element in the leftmost column of a tile is processed, an accumulated partial sum is written into temporary output memory, and then the temporary data is accessed later for computation. We call the accumulated partial sum temporary output, from now on.

We define parameters N and K as the number of rows and columns, and also define P and Q as the number of tiles in horizontal directions and in vertical direction, respectively. The parameter P is determined by size of nearby temporary input buffer size, and Q is determined size of temporary output memory.

Variable data precision is supported to utilize memory storage more efficiently in the memory constrained condition of the IoT application. The approach with short data precision has been already investigated in prior works [71]. Table 6.1 summarizes available precisions for weight, input, output and temporary output, and number of elements per word depending on data precision, where one word, or one line of SRAM memory is 96b. Considering that the fully-connected layer operation requires intensive memory accesses, as design decision, 96b is chosen to save memory access requests, and avoid excessive memory transactions. We have designed custom low-power SRAM with one line of 96b.

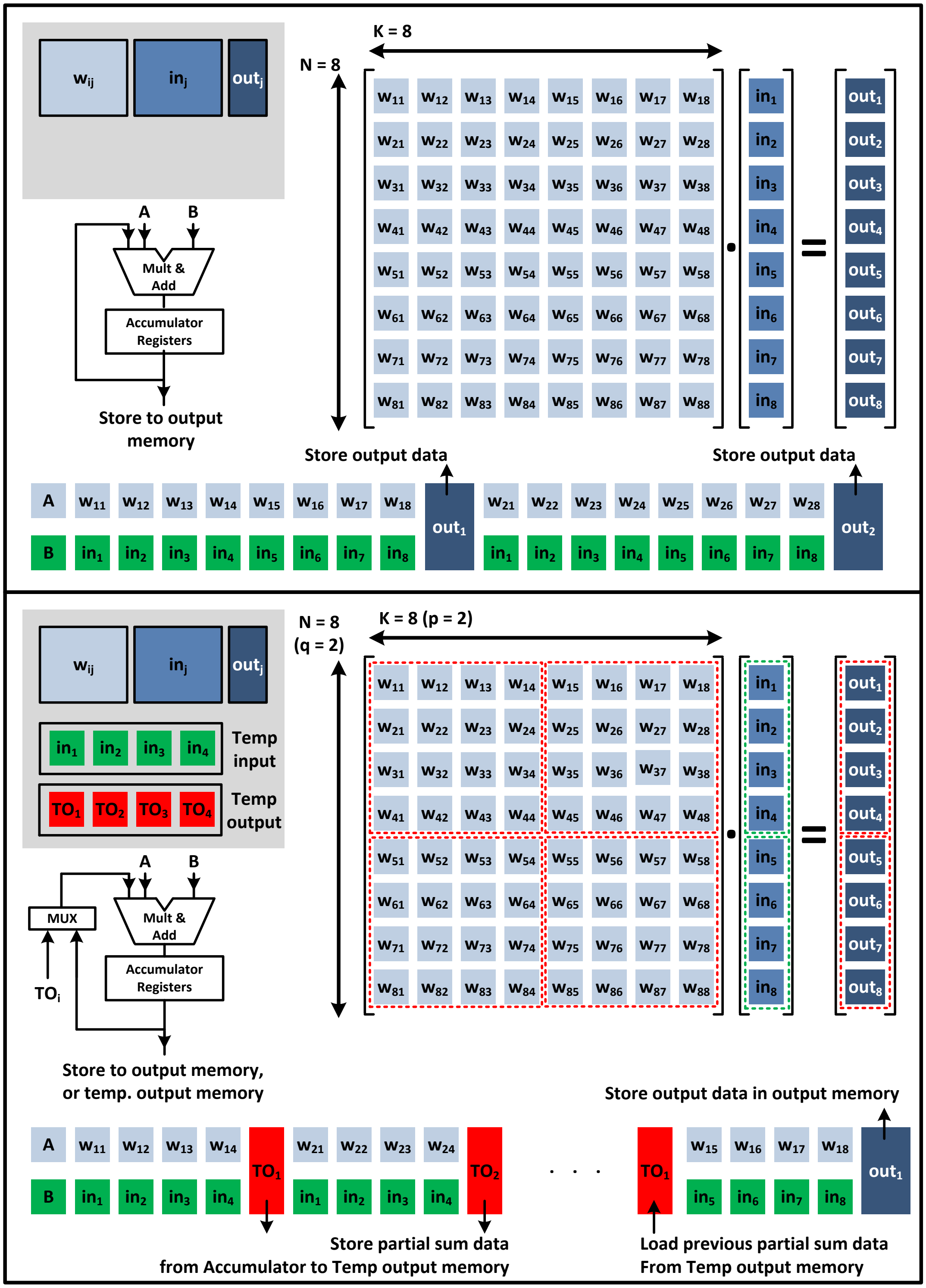

w_ij
in_j
out_j
A
B
Mult & Add
Accumulator Registers
Store to output memory
K = 8
N = 8
w11 w12 w13 w14 w15 w16 w17 w18
w21 w22 w23 w24 w25 w26 w27 w28
w31 w32 w33 w34 w35 w36 w37 w38
w41 w42 w43 w44 w45 w46 w47 w48
w51 w52 w53 w54 w55 w56 w57 w58
w61 w62 w63 w64 w65 w66 w67 w68
w71 w72 w73 w74 w75 w76 w77 w78
w81 w82 w83 w84 w85 w86 w87 w88
in1 in2 in3 in4 in5 in6 in7 in8
out1 out2 out3 out4 out5 out6 out7 out8
Store output data
Store output data
A w11 w12 w13 w14 w15 w16 w17 w18 out1 w21 w22 w23 w24 w25 w26 w27 w28 out2
B in1 in2 in3 in4 in5 in6 in7 in8 in1 in2 in3 in4 in5 in6 in7 in8
w_ij
in_j
out_j
in1 in2 in3 in4
Temp input
TO1 TO2 TO3 TO4
Temp output
A
B
MUX
TO_i
Mult & Add
Accumulator Registers
Store to output memory, or temp. output memory
N = 8 (q = 2)
K = 8 (p = 2)
Store output data in output memory
A w11 w12 w13 w14 TO1 w21 w22 w23 w24 TO2 . . . TO1 w15 w16 w17 w18 out1
B in1 in2 in3 in4 in1 in2 in3 in4 in5 in6 in7 in8
Store partial sum data from Accumulator to Temp output memory
Load previous partial sum data From Temp output memory

Figure 6.4. Operation sequence of a vector-matrix multiplication in baseline (top), and in the proposed weight matrix tiling approach (bottom). N is the number of rows of a weight matrix, K is the number of columns of a weight matrix, and P and Q are parameters of weight matrix tiling, respectively representing the number of tiles in horizontal direction and in vertical direction.

Data type	Available precisions
Weight	6b, 8b, 12b, 16b
Input	6b, 8b, 12b, 16b
Output	6b, 8b, 12b, 16b
Temporary Output (TO)	16b, 24b, 32b

Precision	Number of elements per word
6b	16
8b	12
12b	8
16b	6
24b	4
32b	3

Table 6.1. Available precisions of existing data types (left), and the number of data elements per word (right). In the propose system, one word is defined as 96bit.

Data type	Weight	Input	Output	Temporary Input	Temporary Output
Total number of entries	N×K/WpW	K	N/OpW	K/(P×IpW)	N/(Q×TOpW)
Number of access per entry	1	1	1	(Q-1)×P	2×Q×(P-1)

Table 6.2. Total entries and the number of access per entry with the proposed weight matrix tiling approach: WpW is the number of weight elements per word, IpW is the number of input elements per word, OpW is the number of output elements per word, and TOpW is the number of temporary output elements per word.

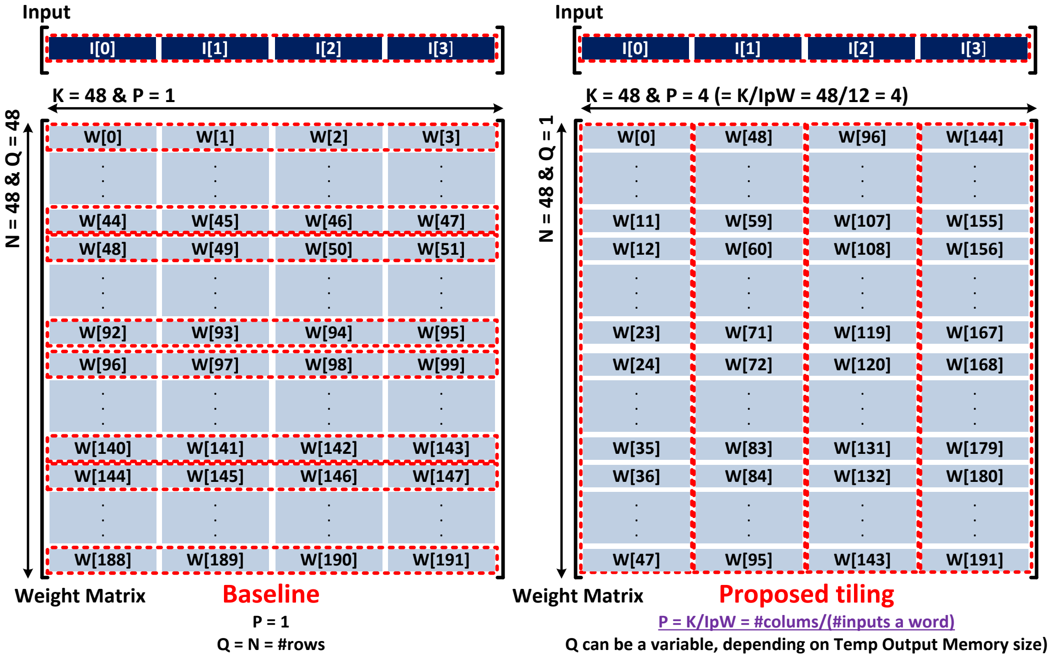


Figure 6.5. Operation sequence of a vector-matrix multiplication in baseline (left), and in the proposed weight matrix tiling (right).

	Proposed Weight Tiling	Baseline Operation
P	K/IpW (fixed)	1
Q	Q (variable)	N

Table 6.3. P and Q parameters for weight matrix tiling and baseline operation.

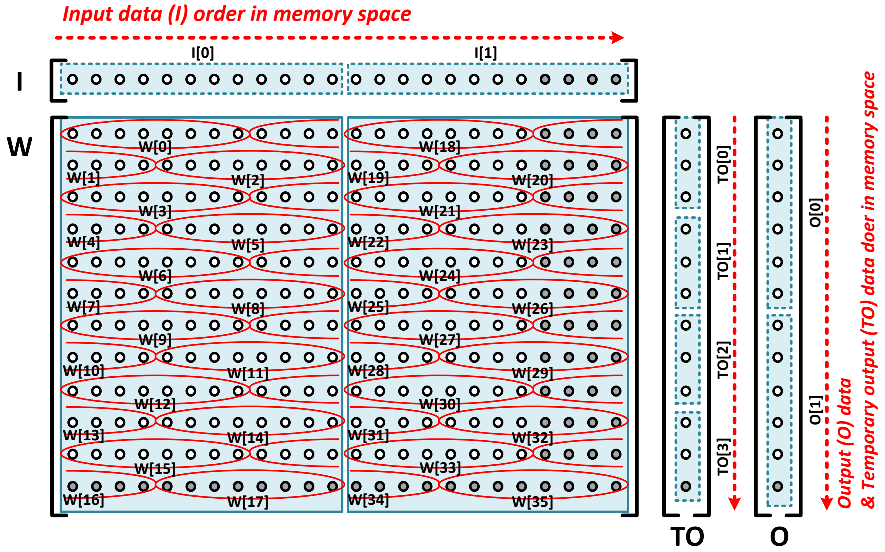
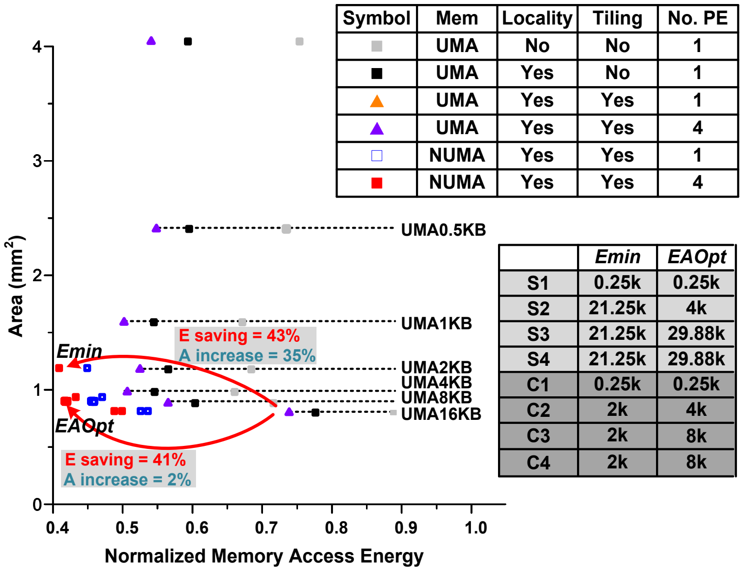


Figure 6.6. Data store/read order in memory space for MAC instruction: weight elements are first divided into tiles, depending on input data precision, and weights in each tile are consecutively stored/read in contiguous memory region.



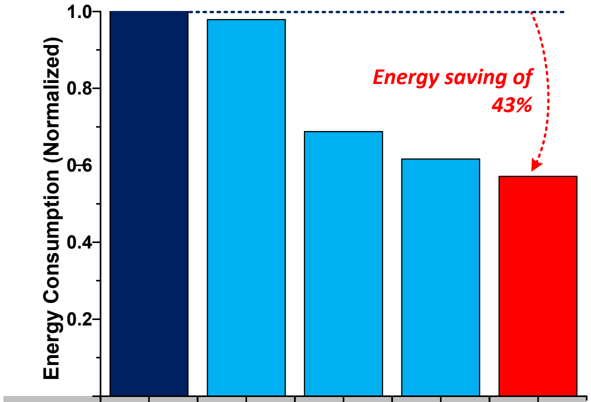
Symbol	Mem	Locality	Tiling	No. PE
■	UMA	No	No	1
■	UMA	Yes	No	1
▲	UMA	Yes	Yes	1
▲	UMA	Yes	Yes	4
□	NUMA	Yes	Yes	1
■	NUMA	Yes	Yes	4

	Emin	EAOpt
S1	0.25k	0.25k
S2	21.25k	4k
S3	21.25k	29.88k
S4	21.25k	29.88k
C1	0.25k	0.25k
C2	2k	4k
C3	2k	8k
C4	2k	8k

Figure 6.7. Area of 256kB with different memory structures, and corresponding energy consumption when the proposed techniques such as NUMA, locality, tiling and four PEs are utilized for computation of three layers of classifier with dimension of (N, K) = (400, 403), (400, 400), (13, 400), where N and K represents weight matrix dimensions.

Table 6.2 shows total number of entries and number of access per entry for five different data types, as function of number of data elements per word (WpW, IpW, OpW and TOpW) and weight matrix tiling parameters (P and Q). It is assumed is that input is first copied to temporary input space located nearby PE. It is observed that temporary input and temporary output are accessed frequently. Plugging P and Q parameters for the proposed tiling and baseline summarized in Table 6.3, the total number of entries and the number of access per entry can be to temporary output and temporary input. Compared to baseline, the proposed tiling effectively reduces the number of accesses.

Figure 6.5 illustrates the difference of the proposed tiling and baseline in an example of a vector-matrix multiplication with dimension of N × K = 48 × 48 for 8b data precision, or 12 elements in one word. P = K/IpW = 48/12 = 4 and Q = 1 is used for the proposed tiling, so the tiled matrix has four column tiles. The tiled weight matrix also determines the order of weight matrix data in memory, although input, output and temporary output are stored in order. The basic idea of the order of weight data is to sort that in the order they are accessed in operation. In the left of Figure 6.5, the weight matrix of baseline stores data in horizontal direction, row by row. In contrast, as shown in the right of Figure 6.5, the tiled weight matrix stores data in vertical



Technique	S1*	S2	S3	S4	S5**
NUMA	UMA 8kB bank	NUMA	NUMA	NUMA	NUMA
Locality	No	No	Yes	Yes	Yes
Tiling	No	No	No	Yes	Yes
PEs	1	1	1	1	4

S1* :baseline structure
S5**: proposed structure

Figure 6.8. Energy consumption of UMA with 8kB unit memory size, and NUMA structure (EA_{Opt}, summarized in Figure 6.x) with the proposed techniques for computation of a single layer of classifier with dimension of (N, K) = (500, 500).

direction. Variable data precision complicates the storing order of weight data. Figure 6.6 illustrates an example of weight matrix of 11 × 20, where 8b precision of input, 12b precision of weight, 32b precision of temporary output, and 16b precision of output are assumed. As the width of a tile is as long as the number of input elements per word, one row of a tile results in non-integer number of weight words. Also, the original dimension (11 × 20) of the matrix is not multiples of IpW, or OpW & TOpW, so zero-padded weight matrix with dimension of 12 × 24 should be considered. The zero-paddings are indicated in grey color, and the storing order of input, weight matrix, temporary output and output is labeled in Figure 6.6

6.1.2 Expected Benefits

Figure 6.7 and 6.8 plots expected energy savings and area overhead for fully-connected layers with dimension of (N, K) = (400, 403), (400, 400), (13, 400), or (N, K) = (500, 500), when the proposed techniques are used. The simulation data is based on CMOS 40nm SRAM compiler's datasheet. The bottom left table in Figure 6.7 shows two NUMAs (E_{min} and EA_{Opt})

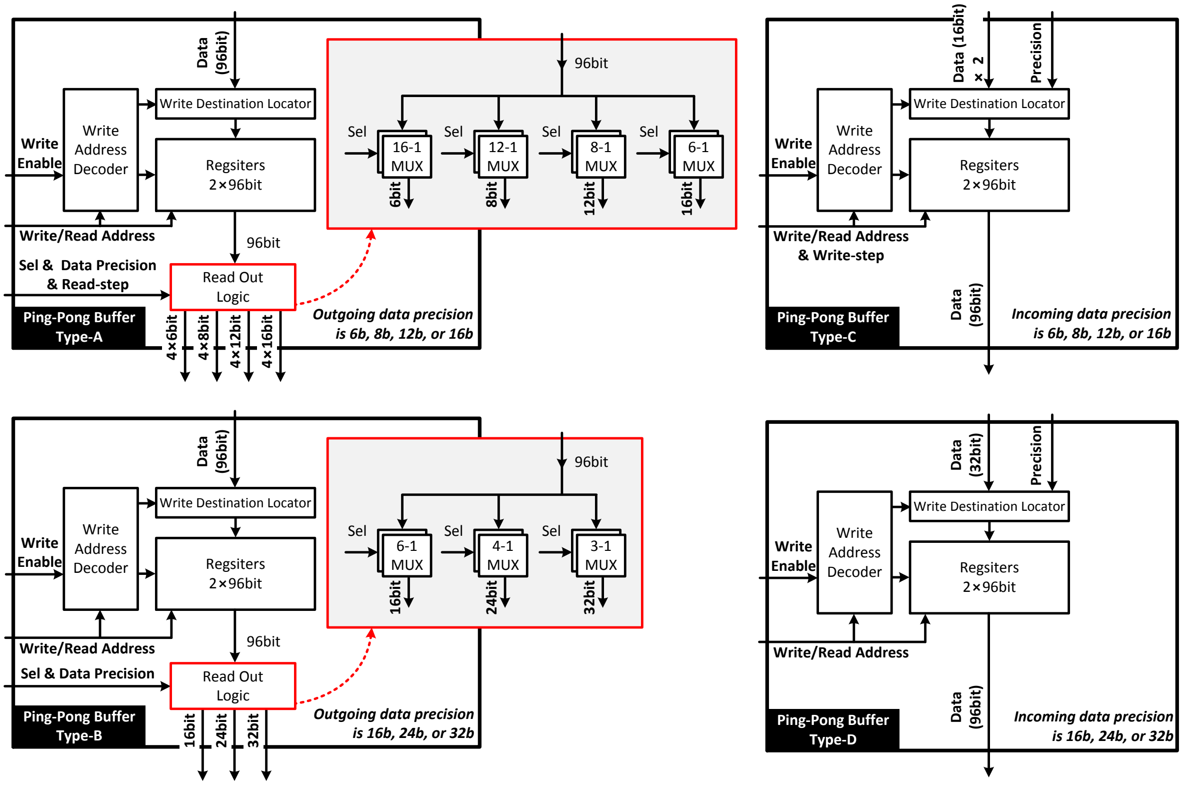


Figure 6.9. Ping-pong data buffers.

with the smallest energy consumption. S1~S4 represent size of each level memory hierarch, and C1~C4 represent unit memory size composing each level memory. It is shown that EA_{Opt} NUMA which optimize energy and area both results in energy saving of 41% and area increase of 2%, in comparison to UMA8kB (total 256kB, unit bank size of 8kB) with none of the proposed technique. E_{min} NUMA shows energy saving of 43%, but it has area overhead of 35%.

6.1.3 Implementation

Four different types (type-A, B, C and D) of ping-pong data buffers (PPBUF) are designed to unpack incoming 96b data, pack processed data into 96b data, and store data as buffers. Type-A and type-B have two-word registers for data buffering, and unpack 96b data loaded from memory into many pieces of elements with precision of 6b, 8b, 12b, 16b, 24b, and 32b. Type-C and type-D have two-word registers for data buffering, receive data with precision of 6b, 8b, 12b, 16b, 24b, or 32b from arithmetic units in a PE, and pack into data to be sent to memory. PPBUFs receive write/read address, write/read step, and precision information in addition to incoming data. Depending on write/read step and precision, the number of cycles required to complete one word (96b) and the number of cycles to read all elements from one word can vary, and only parts of outputs can be enabled in PPBUF-A and B while the remaining outputs are zeroed to avoid unnecessary switching.

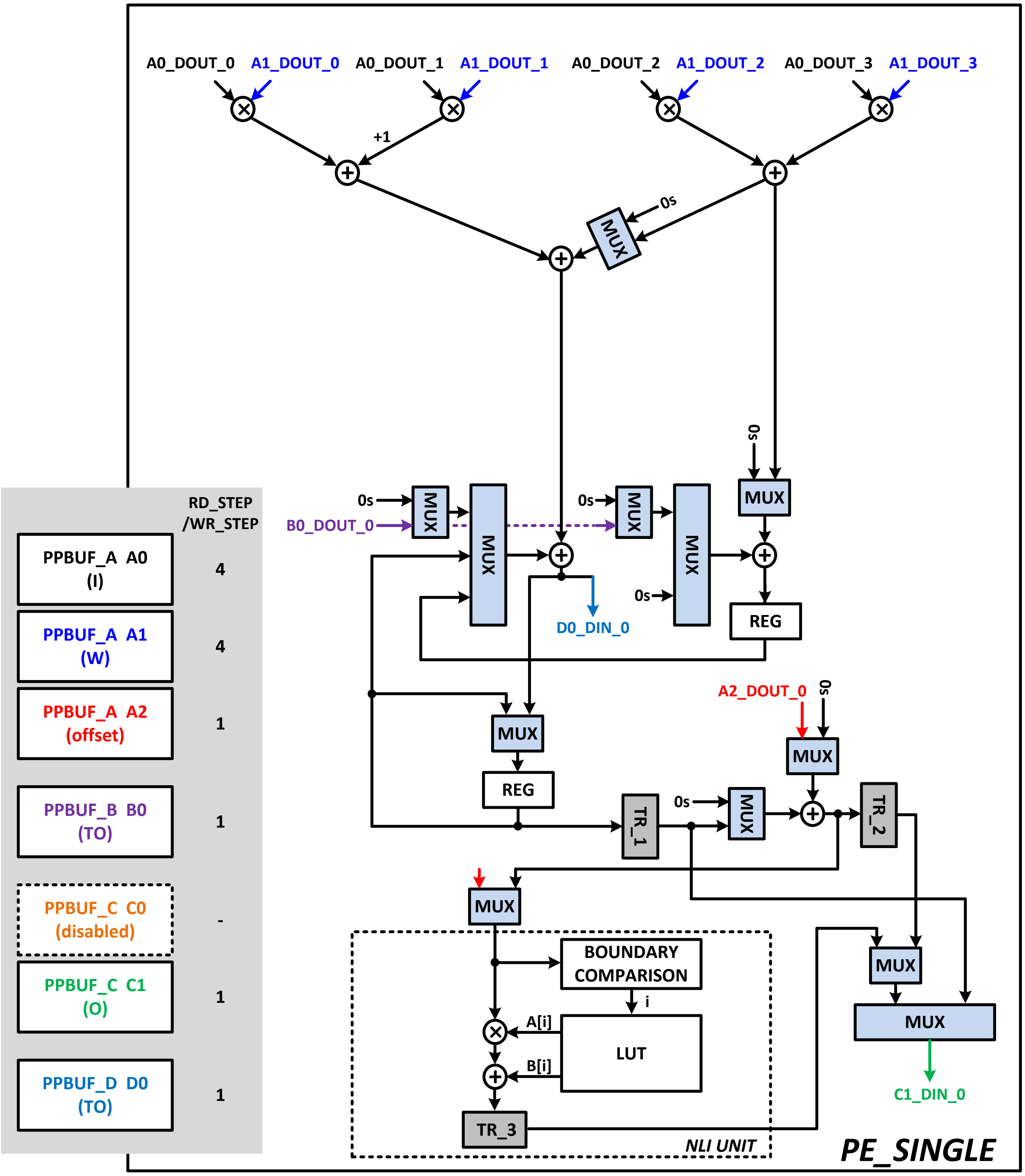
```
// Following counts are used to control MAC FSMs
//                              compared with IN.SIZE, W.SIZE, O.SIZE, TO.SIZE, which all are in WORD unit
CNT.IN = 0
CNT.W = 0
CNT.O = 0
CNT.TO.WR = 0
CNT.TO.RD = 0
CNT.OS = 0
LOAD IN[0] WORD to PPBUF_A0
CNT.IN++
LOAD W[0] WORD to PPBUF_A1
CNT.W++
LOAD OFFSET[0] to PPBUF_A2
CNT.OS++
// Following pointers are used to pack data from PE into one word (96b)
//                                  or to unpack one word (96b) into data used in PE
p = 0               // pointer for PPBUF_A0 that stores incoming IN (input)
q = 0               // pointer for PPBUF_A1 that stores incoming W (weight)
r = 0               // pointer for PPBUF_A2 that stores incoming OS (offset)
s = 0               // pointer for PPBUF_B0 that stores incoming TO (temporary output)
x = 0               // pointer for PPBUF_C1 that stores outgoing O (output)
```

```
y = 0           // pointer for PPBUF_D0 that stores outgoing TO (temporary output)
DO {
    IF (p == 0) {
        MAC = 0
    }
    MAC += A0[p]×A1[q] + A0[p+1]×A1[q+1] + ... + A0[p+3]×A1[q+3]
    // PE has four multipliers, so pointer p & q have strides of 4
    p += 4
    q += 4
    IF (!LEFT_COLUMN_TILES) {
        MAC += B0[x]
        s++
    }
    IF (CNT.TO.RD ==  TO.SIZE) {
        LOAD IN[CNT.IN] WORD to PPBUF_A0
        CNT.IN++
        CNT.TO.WR = 0
    }
    IF (RIGHT_COLUMN_TILES) {
        STORE TRUN_2 to PPBUF_C1
        x++
        IF (x == OpW/4){
            // OpW is the number of output elements per word
            O[CNT.O] = SHIFTED & TRUNCATED MAC RESULTS
            O[CNT.O] = O[CNT.O] + A2[r]
            r++
            WRITE O[CNT.O] from PPBUF_C1 to MEM
            CNT.O++
            x = 0
        }
    }
    IF (!RIGHT_COLUMN_TILES && p == IpW/4)  {
        STORE MAC to PPBUF_D0
        y++
    }
    IF (!RIGHT_COLUMN_TILES && p == IpW/4)  {
        LOAD TO[CNT.TO.RD] to PPBUF_B0
    }

    IF (p == IpW/4)  {
        // IpW is the number of input elements per word
        p = 0
    }
    IF (q == WpW/4) {
        // IpW is the number of weight elements per word
        LOAD W[CNT.W] WORD to PPBUF_A1
        CNT.W++
        q = 0
    }
    IF (r == OpW) {
        // OpW is same as OSpW, the number of offset elements per word
        LOAD OFFSET[CNT.OS] WORD to PPBUF_A2
        CNT.OS++
        r = 0
    }
    IF (s == TOpW) {
        // TOpW is the number of temporary output elements per word
        LOAD TO[CNT.TO.RD] WORD to PPBUF_B0
        CNT.TO.RD++
        s = 0
    }
    IF (y == TOpW)  {
        WRITE TO[CNT.TO.WR] from PPBUF_D0 to MEM
        CNT.TO.WR++
        LOAD TO[CNT.TO.RD] to PPBUF_B0
        y = 0
    }
    IF (CNT.TO.RD == TO.SIZE) {
        CNT.TO.RD = 0
    }
    IF (CNT.TO.WR == TO.SIZE) {
        CNT.TO.WR = 0
    }
} WHILE (CNT.O < O.SIZE)
```

Figure 6.10. Pseudo code for MAC operation for P = N/IpW, and Q = K.

Pseudo code of MAC is shown in Figure 6.10, and PE arithmetic units and ping-pong data buffers are illustrated in Figure 6.11. Due to variable data precision, weight matrix tiling approach and many data types such as input, weight, offset, incoming temporary output, outgoing temporary output and output, many pointers and counts are used to control MAC operation sequence, as shown in Figure 6.10. As shown in Figure 6.11, input, weight and offset data are respectively loaded in PPBUF-A A0, A1 and A2. While processing tiles in the leftmost or middle column of weight matrix, MAC results or temporary outputs are stored in PPBUF-D, and only when processing tiles in the rightmost column of weight matrix, MAC results are shifted, truncated and stored in PPBUF-C. With offset and nonlinear function enabled, before data are stored in PPBUF-C, offset are added, and nonlinear function is applied. Also, when processing tiles in the middle or rightmost column of weight matrix, the temporary outputs are loaded in PPBUF-B, and added. Data path of PE arithmetic units forms a single-stage pipeline with MAC registers and PPBUFs, unless the leftmost tiles are processed and final results are written back to PPBUF-C. Read step of PPBUF-A A0 and A2 is set four, and read/write steps of all the other PPBUFs are set one because four multipliers in PE take four input and weight data in one cycle, and only one result is generated at a time.



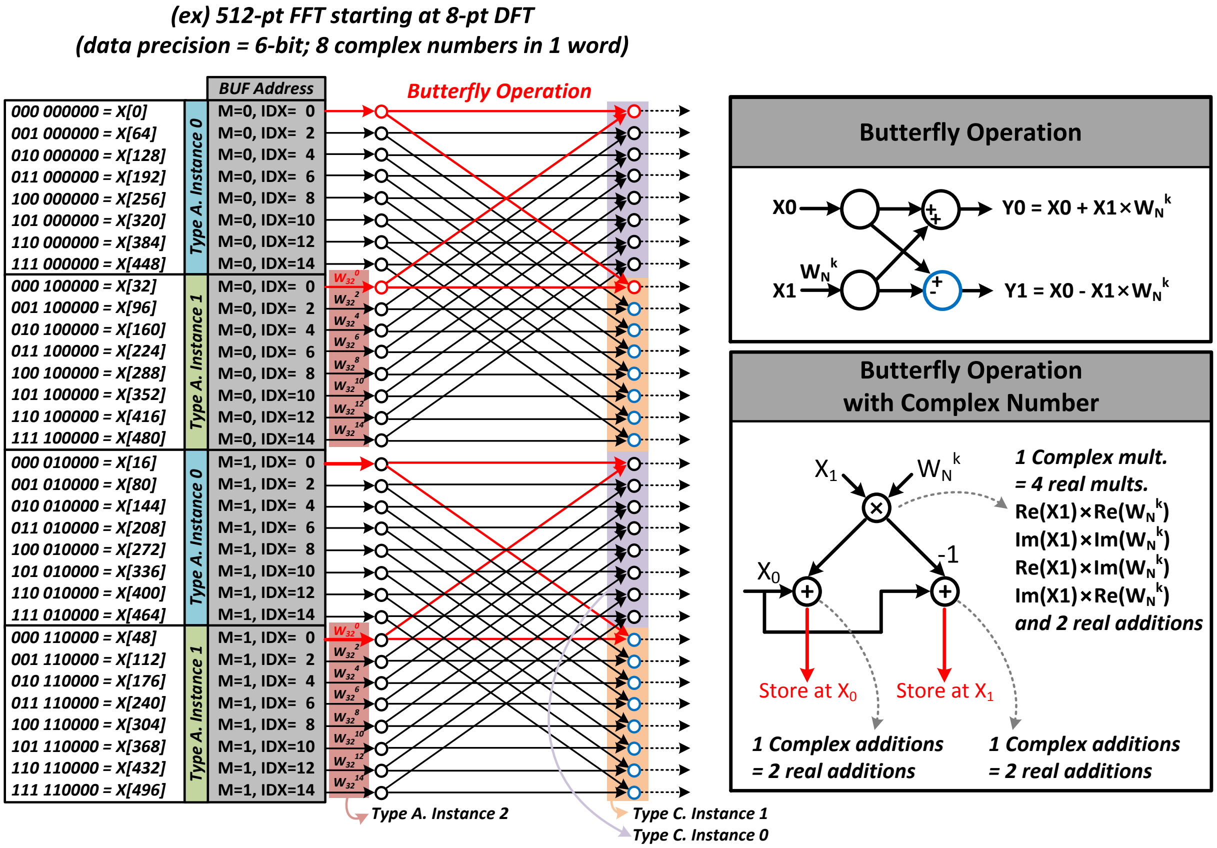
	RD_STEP /WR_STEP
PPBUF_A A0 (I)	4
PPBUF_A A1 (W)	4
PPBUF_A A2 (offset)	1
PPBUF_B B0 (TO)	1
PPBUF_C C0 (disabled)	-
PPBUF_C C1 (O)	1
PPBUF_D D0 (TO)	1

Figure 6.11. PE arithmetic units and ping-pong data buffers that are configured to MAC mode.

6.2 Proposed Data Flow for FFT and Implementation

FFT is an energy efficient and fast way to compute DFT. Basic idea is to (1) break N-pt DFT into two N/2-pt DFT, and (2) apply butterfly operation upon the two sets of N/2-pt DFT. For example, for N = 512 = 2^8, 512-pt DFT can be broken into two 256-pt DFTs, and the two 256-pt DFTs can be broken into four 128-pt DFTs. By repeating similar sequence, starting from 256 2-pt DFTs and applying butterfly operation at each level, 512-pt DFT computation can be obtained in the end. Figure 6.13 shows butterfly operation performed onto incoming several sets of 8-pt DFT data to generate 16-pt DFTs.

Figure 6.13 shows operation sequence performed for FFT for an example of 512-pt FFT starting at 16-pt FFT, the partial butterfly operation of which is illustrated in Figure 6.12, assuming in-place computation, in which outputs are written back into input memory. In Figure



		BUF Address
000 000000 = X[0]	Type A. Instance 0	M=0, IDX= 0
001 000000 = X[64]		M=0, IDX= 2
010 000000 = X[128]		M=0, IDX= 4
011 000000 = X[192]		M=0, IDX= 6
100 000000 = X[256]		M=0, IDX= 8
101 000000 = X[320]		M=0, IDX=10
110 000000 = X[384]		M=0, IDX=12
111 000000 = X[448]		M=0, IDX=14
000 100000 = X[32]	Type A. Instance 1	M=0, IDX= 0
001 100000 = X[96]		M=0, IDX= 2
010 100000 = X[160]		M=0, IDX= 4
011 100000 = X[224]		M=0, IDX= 6
100 100000 = X[288]		M=0, IDX= 8
101 100000 = X[352]		M=0, IDX=10
110 100000 = X[416]		M=0, IDX=12
111 100000 = X[480]		M=0, IDX=14
000 010000 = X[16]	Type A. Instance 0	M=1, IDX= 0
001 010000 = X[80]		M=1, IDX= 2
010 010000 = X[144]		M=1, IDX= 4
011 010000 = X[208]		M=1, IDX= 6
100 010000 = X[272]		M=1, IDX= 8
101 010000 = X[336]		M=1, IDX=10
110 010000 = X[400]		M=1, IDX=12
111 010000 = X[464]		M=1, IDX=14
000 110000 = X[48]	Type A. Instance 1	M=1, IDX= 0
001 110000 = X[112]		M=1, IDX= 2
010 110000 = X[176]		M=1, IDX= 4
011 110000 = X[240]		M=1, IDX= 6
100 110000 = X[304]		M=1, IDX= 8
101 110000 = X[368]		M=1, IDX=10
110 110000 = X[432]		M=1, IDX=12
111 110000 = X[496]		M=1, IDX=14

Figure 6.12. FFT butterfly operation: an example of 512-pt FFT computation starting at 16-pt FFT with 6b data precision (left), butterfly operation (top right), and required arithmetic operation for butterfly operation with complex number (bottom right).

6.13, it is observed that same weight data is repeatedly used many times especially when small-pt FFT is performed to obtain large-pt FFT. For instance, in Figure 6.13, 16-pt FFT is performed by 32 times, and its weights are used 32 times. Similarly, 32-pt FFT is performed by 16 times, and its weights are used 16 times.

To save energy consumption due to weight data memory access, we propose processing small-pt FFTs together to use each weight data multiple times, instead of processing small-pt FFTs group by group. In Figure 6.13, we can group and process C00~C31 for 16-pt FFTs, C32~46 for 32pt-FFTs, C48~C55 for 64pt-FFTs, C56~C59 for 128-pt FFTs, and C60~C61 for 256-pt FFTs. Pseudo codes for FFT operation are shown in Figure 6.14 and Figure 6.15, respectively corresponding to baseline data flow, and proposed data flow. Expected saving in the number of weight data access is plotted in Figure 6.16. The saving increases with FFT point and it can be written as the following equation:

$$\text{Saving in the number of weight access} = 1 - \frac{2^{M-1}-1}{2^{M-1} \times M} \qquad (6.2)$$

where $\mathrm{M} = \log_2(\text{number of FFT point})$.

```
// Perfom 16-pt FFT by 32 times
C00: FFT 016-PT X[00-00] X[01-01] W[00-00]
...
C31: FFT 016-PT X[62-62] X[63-63] W[00-00]
// Perfom 32-pt FFT by 16 times
C32: FFT 032-PT X[00-01] X[02-03] W[01-02]
...
C47: FFT 032-PT X[60-61] X[62-63] W[01-02]
// Perfom 64-pt FFT by 8 times
C48: FFT 064-PT X[00-03] X[04-07] W[03-06]
...
C55: FFT 064-PT X[56-59] X[60-63] W[03-06]
// Perfom 128-pt FFT by 4 times
C56: FFT 128-PT X[00-07] X[08-15] W[07-14]
...
C59: FFT 128-PT X[48-55] X[56-63] W[07-14]
```

```
// Perfom 256-pt FFT by twice
C60: FFT 256-PT X[00-15] X[16-31] W[15-30]
C61: FFT 256-PT X[32-47] X[48-63] W[15-30]
// Perfom 512-pt FFT, finally
C62: FFT 512-PT X[00-31] X[32-63] W[31-62]
```

Figure 6.13. Pseudo code for FFT operation: Proposed data flow.

```
For (q = 0 q < I_SIZE/I_UNIT_SIZE q++){
    // (ex) #FFTs = number of N-PT FFTs to be performed
    //      #FFTs is same as I_SIZE/I_UNIT_SIZE
    //      when 512-PT FFT is final goal, #FFTs @16-PT = 32
    For (r = 0 r < W.SIZE r++)
        LOAD I[q + r] WORD from MEM to PPBUF_A_2
        LOAD I[q + r + I_UNIT_SIZE/2] WORD from MEM to PPBUF_A_0
        // twiddle factor (W[r]) is loaded multiple times
        LOAD W[r] WORD from MEM to PPBUF_A_1
        For (p=0 p < #DATA/WORD p++){
            PERFORM BUTTERFLY OPERATION: calculate O[q + r] and O[q + r + O_UNIT_SIZE/2]
            STORE to PPBUF_C_0
            STORE to PPBUF_C_1
        }
        WRITE O[q + r] to MEM
        WRITE O[q + r + O_UNIT_SIZE/2] to MEM
    }
}
```

Figure 6.14. Operation sequence accompanied by operation of FFT instruction for an example of 512-pt FFT starting at 16-pt FFT with 6b precision data. X and W respectively represent input and weight data words, and each data word is assumed to have 8 elements.

```
For (r = 0 r < W.SIZE r++){
    // twiddle factor (W[r]) is loaded only once @N-PT
    LOAD W[r] WORD from MEM to PPBUF_A_1
    For (q = 0 q < I_SIZE/I_UNIT_SIZE q++){
        // (ex) #FFTs = number of N-PT FFTs to be performed
        //      #FFTs is same as I_SIZE/I_UNIT_SIZE
        //      when 512-PT FFT is final goal, #FFTs@16-PT = 32
        LOAD I[q + r] WORD from MEM to PPBUF_A_2
        LOAD I[q + r + I_UNIT_SIZE/2] WORD from MEM to PPBUF_A_0
        For (p=0 p < #DATA/WORD p++){
            PERFORM BUTTERFLY OPERATION: calculate O[q + r] and O[q + r + O_UNIT_SIZE/2]
            STORE to PPBUF_C_0
            STORE to PPBUF_C_1
        }
        WRITE O[q + r] to MEM
        WRITE O[q + r + O_UNIT_SIZE/2] to MEM
    }
}
```

Figure 6.15. Pseudo code for FFT operation: Baseline data flow.

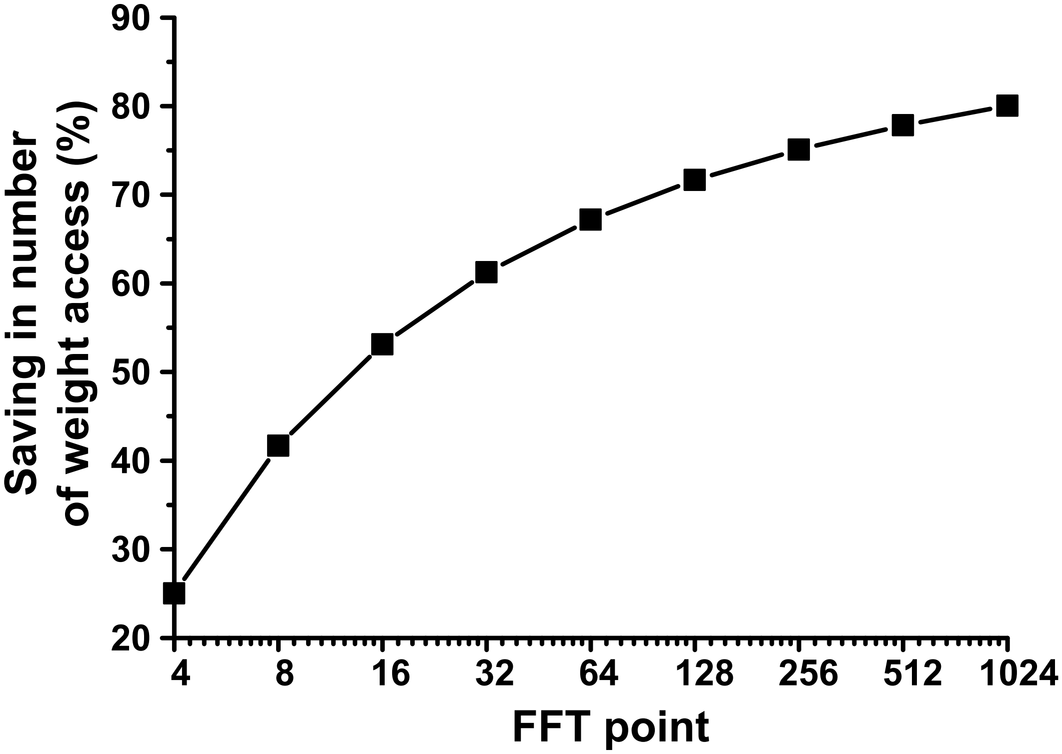


Figure 6.16. Saving in the number of weight data access with the proposed data flow.

Precision	6b	8b	12b	16b
Number of available real number per word	16	12	8	6
Number of available complex number per word	8	6	4	3
Number of complex number per word for aligned computation	**8**	**4**	**4**	**2**
Required zero-padding	-	32b	-	32b
Required DFT as the 1st stage inputs	8-pt	4-pt	4-pt	2-pt
Bit-reversal radix	3b	2b	2b	1b

Table 6.4. FFT data packing strategy for aligned computation.

6.2.1 Implementation

To implement variable data precisions and aligned computation in word unit, only 2^N complex numbers are allowed to be packed in one word. This enables neat FSM controls, avoiding excessive control complexity, in despite of potential data waste. The FFT data packing strategy is summarized in Table 6.4. Considering complex numbers, one word can have up to eight 6b data elements, six 8b data elements, four 12b data elements, or three 16b data elements. When precision is 6b or 12b, the number of data elements is power of two. When precision is 8b or 16b, to allow 2^N complex numbers, four data elements or two data elements are respectively used.

PE arithmetic units and ping-pong data buffers for FFT are illustrated in Figure 6.17. Four multipliers, six adders, three PPBUF-A, and two PPBUF-B are used for complex number butterfly operation, as also shown in bottom right of Figure 6.12. Two input data are loaded into PPBUF-A A2 and A0, and weight data is loaded into PPBUF_A A1. Output results are stored first in PPBUF-C C0 and C1 before they are sent to memory. As one butterfly operation takes real part and imaginary parts of data, read and write steps of PPBUF-A and PPBUF-B are set two.

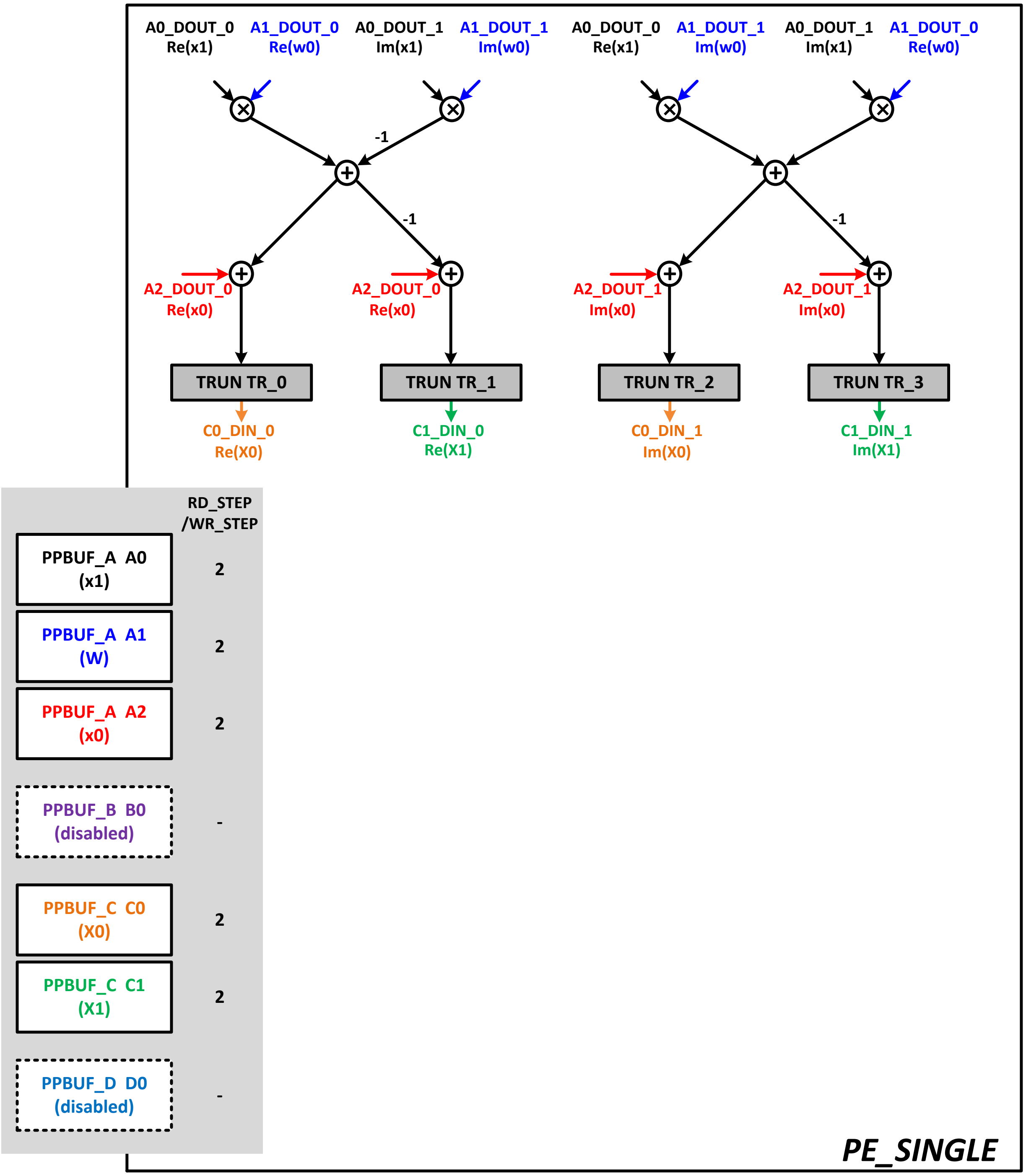


Figure 6.17. PE arithmetic units and ping-pong data buffers that are configured to FFT mode.

6.3 Co-Processor Design and Future Works

To design co-processor which supports energy-efficient operation of deep learning algorithm and FFT, configurable PE and VLIW instruction set architecture are designed, as described in Figure 6.18 and Table 6.5. A single PE is composed of five 16b multipliers, four 8b multipliers, eleven 32b adders, two 32b registers, three PPBUF-A, one PPBUF-B, two PPBUF-C, one PPBUF-D, and four truncation units. Also, it has registers for look-up table (LUT) of nonlinear-function. Depending on instruction, parts of PE arithmetic units are enabled, and unnecessary parts are disabled by zeroing inputs to corresponding units. In Figure 6.11 and 6.17, it is illustrated how PE is configured for MAC and FFT instruction. As also summarized in Table 6.5, individual instructions for nonlinear function (NLI) and data move operation (MOV) are supported. When NLI instruction operates, NLI unit indicated with dotted box in Figure 6.18 is only enabled to perform arithmetic operation, and inputs to the other units are zeroed. Similarly, MOV instruction operates, only temporary buffers are enabled to load and store data from memory before they are sent to destination memory, and most of PE arithmetic units is disabled. Depending on instruction, data read step or data write step of PPBUF-A and PPBUF-C are respectively configured between 2, 4 and 1, and between 2 and 1, and PPBUFs are selectively enabled, as summarized in Figure 6.18.

The proposed co-processor VLIW instruction set and most of the instruction fields are used to represent data memory address information such as data starting address and data size. TI, TO, and offset data address fields (TI data address start, TO data address start, TO data size, TO data precision, and offset data address start) are only used for MAC. For operation sequence control of PE FSM, it is required to have information such as the number of rows in MAC weight matrix (MAC row number. Also, the number of FFTs to be performed upon input data with given weight data (FFT number), FFT input unit size and FFT output unit size should be specified, as PE performs same FFT multiple times many number of input sets. Furthermore, there are MAC offset enable and MAC NLI enable fields such that MAC can be used for vector-matrix multiplication. Three shift operands for truncation units are used in truncation units that performs shift and truncation to outputs of butterfly operation of FFT instruction, multiply-accumulation registers of MAC instruction, and adder of NLI function. Some instruction fields such as TI data address start, TO data address start, TO data size, TO data precision, FFT number,

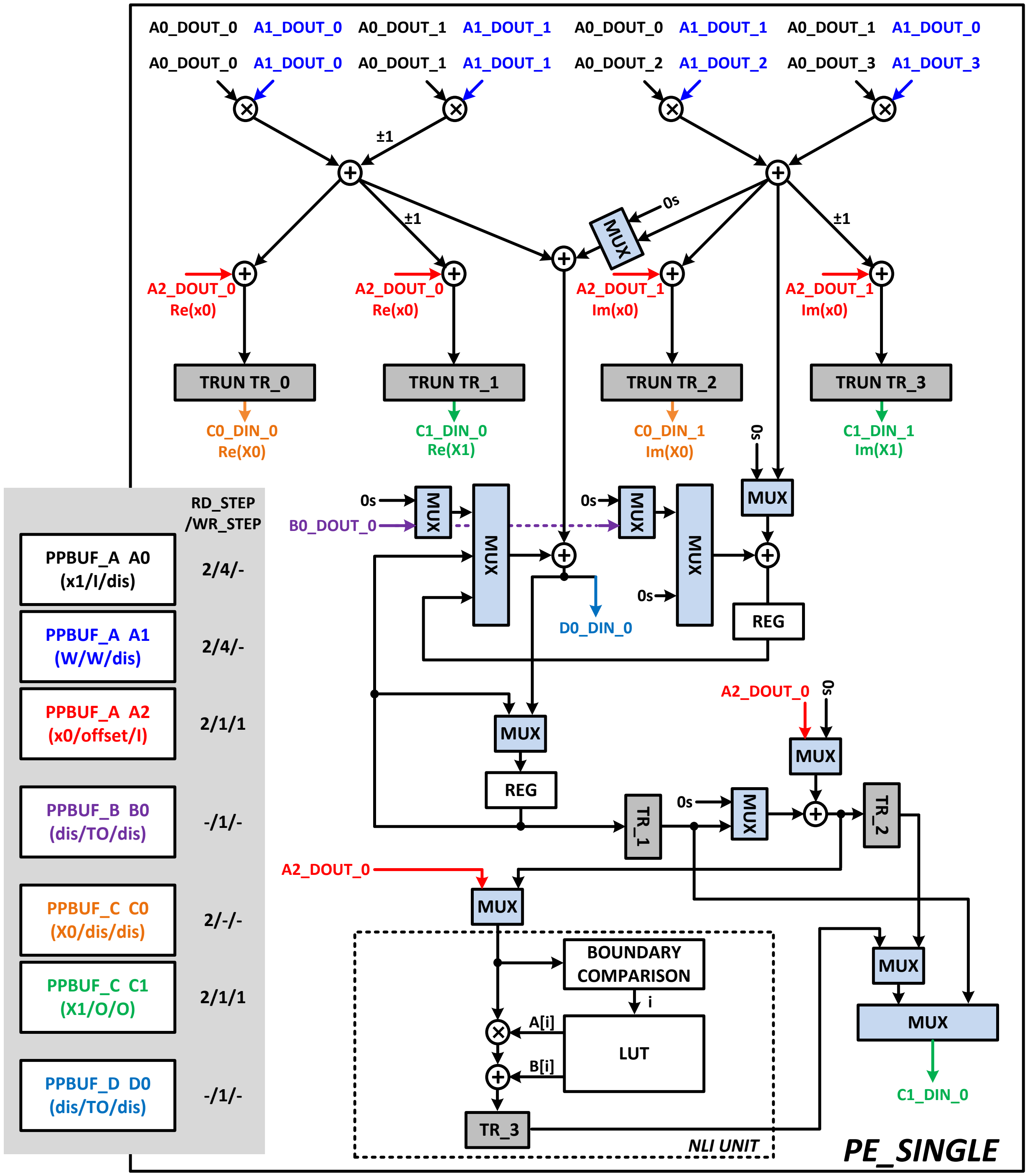


Figure 6.18. PE arithmetic units and ping-pong data buffers that are configurable with instruction: MAC, FFT, NLI, and MOV.

Region	Bits	Note	Use
OP code	2		MAC, FFT, NLI, MOV
PE ID	2		MAC, FFT, NLI, MOV
TI data address start	13	Address for 0~64kB	MAC
TO data address start	13	Address for 0~64kB	MAC
TO data size	13		MAC
TO data precision	2	TO precision = 16, 24, 32b	MAC
Input data address start	15	Address for 0~256kB	MAC, FFT, NLI, MOV
Input data size	15		MAC, FFT, NLI, MOV
Input data precision	2	Input precision = 6, 8, 12. 16b	MAC, FFT, NLI, MOV
Output data address start	15	Address for 0~256kB	MAC, FFT, NLI, MOV
Output data size	15		MAC, FFT, NLI, MOV
Output data precision	2	Output precision = 6, 8, 12. 16b	MAC, FFT, NLI, MOV
Offset data address start	15	Offset has same size and precision as output data	MAC
Weight data address start	15	Address for 0~256kB	MAC, FFT, NLI, MOV
Weight data size	15		MAC, FFT, NLI, MOV
Weight data precision	2	Weight precision = 6, 8, 12. 16b	MAC, FFT, NLI, MOV
MAC offset enable	1		MAC
MAC NLI enable	1		MAC
MAC row number	15	Used in MAC FSM	MAC
FFT number	13	One FFT instruction takes multiple input sets with identical size, generating multiple output sets.	FFT
FFT input unit size	13		FFT
FFT output unit size	13		FFT
Shift	5	Input to 32b shifters	MAC
Shift after offset addition	5		MAC
Shift after NLI	5		MAC, NLI
Total	188 (227)	TI data, TO data fields and FFT parameters are exclusively used only for MAC and FFT, so they share physical instruction regions.	

Table 6.5. Designed instruction set architecture (ISA) for MAC (multiply-accumulation), NLI (nonlinear-function), FFT, MOV (move).

FFT input unit size, and FFT output unit size are exclusively used only for MAC or FFT, so they can share physical instruction fields, and the total number of bits of one instruction is 188b, instead of 227b.

Details of the proposed co-processor design are summarized in Figure 6.19. It mainly consists of four PEs, and on-chip ARM-Cortex M4 microprocessor. Four PEs enables for parallel processing and higher throughput. To allow for memory access to a PE memory from other PEs, memory arbitration unit (MEM ARB) is located between PE and PE memory, and handles memory access requests from owner PE, external PEs, and the ARM processor. The ARM processor is used to program instructions of PEs, write input data to PE memory sectors, read data from PE memory sectors, and perform general DSP computations which are not supported in the proposed PE.

In addition, peripheral controllers such as layer controller and MBus controller are implemented on-chip, to communicate with external ICs which have a low-power bus protocol

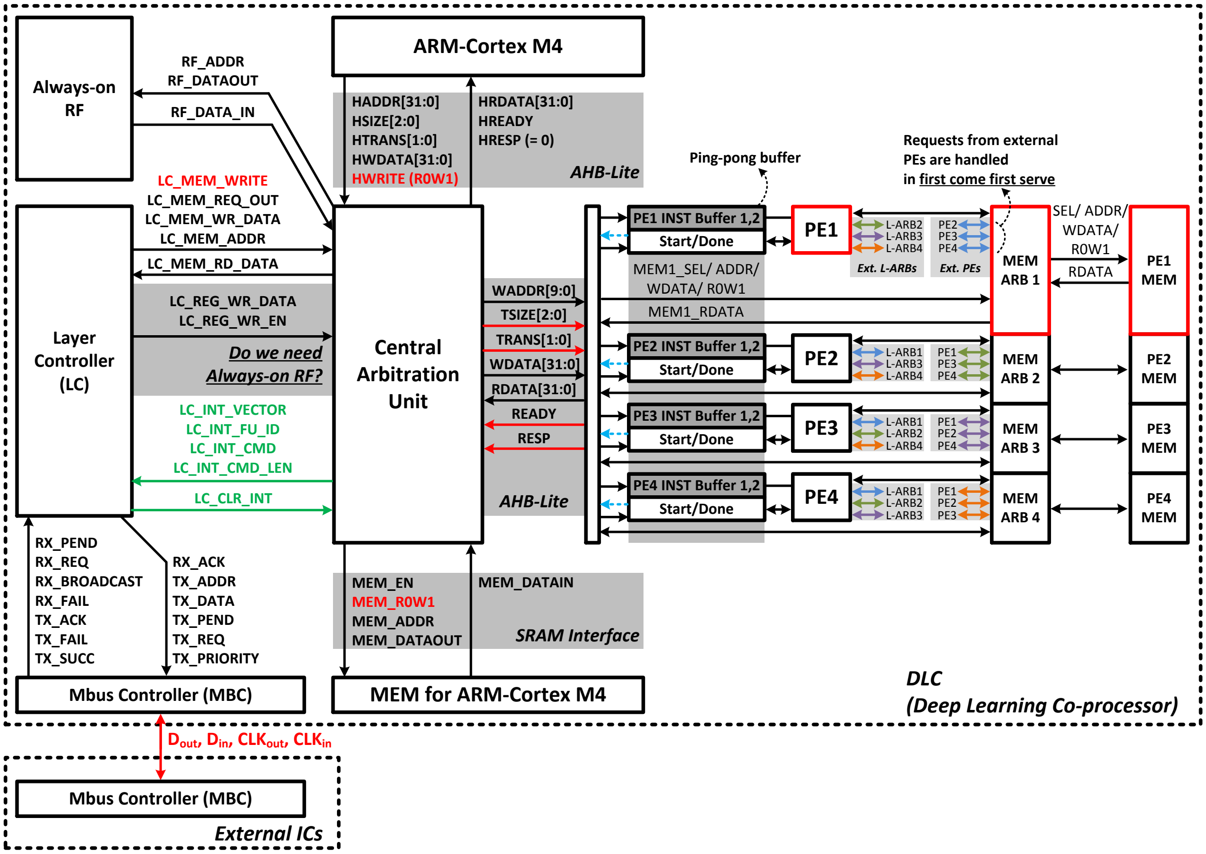


Figure 6.19. Detailed top block diagram of co-processor with four PEs, dedicated memory arbitration units, PE instructions buffers, ARM-Cortex M4, central arbitration unit, and miscellaneous blocks for communication with external ICs.

called MBus [73-75]. Bus protocols of the ARM processor (AHB-Lite), MBus, memory for ARM processor and four PEs are different, so central arbitration unit is designed to interpret and perform communication requests, and to arbitrate memory access from many sources.

So far, arithmetic units and FSM controller of a PE and a PE memory arbitration unit have been designed, and functionality was verified with instruction sets described in Table 6.5. In the near future, central arbitration units will be designed, and functionality of the proposed co-processor including four PEs, four memory arbitration units, ARM-Cortex M4 microprocessor, central arbitration units, and peripheral MBus controllers will be verified in simulation. The proposed deep learning co-processor will be fabricated in April, 2016.

CHAPTER 7
Conclusions

Efficient power conversion and power management continue to be important in the era of the Internet of Things (IoT). IoT devices with small form factors and untethered property will benefit from efficient power conversion and fully integrated techniques, accomplishing longer battery lifetime. In this thesis, the challenges such as volume constraints and limited battery capacity have been tackled with the following solutions: ambient energy extraction, dual mode operation (active and standby mode), and energy-efficiency operation through voltage down-conversion in addition to energy-efficient deep-learning co-processor design. The specific contributions made are listed below.

For ultra-low-power power management unit design, fully-integrated switched-capacitor-based power management unit (PMU) with self-adaptive conversion ratio for ultra-low power sensor platform was proposed/fabricated, and tested with varying input voltage ranges and with a variety of harvesting energy sources such as solar, microbial fuel cell, and thermal energy sources.

Also, in an attempt to reduce GIDL current that worsens leakage of sleep transistor with high input voltage of ~ 4V, reconfigurable sleep transistors was proposed and fabricated. It was demonstrated that the reconfigurable sleep transistor are effective in reducing leakage in the domain where GIDL is dominant.

Furthermore, we addressed issues of SC converters such as limited number of conversion ratios and output ripple characteristics, For a wide range of output generation in fine-grained conversion ratio resolution, successive-approximation switched-capacitor (SAR SC) DC-DC converter was proposed/ fabricated, and power conversion efficiency and SC converter losses were studied and compared for a variety of topologies in SC converters.

For ripple minimization of SC DC-DC converter, on-chip flying-capacitance-dithered SC converter was proposed/fabricated, and relation of ripple and power conversion efficiency was studied.

The last contribution made in this thesis is low-power configurable deep learning co-processor design. For the energy efficient and low power operation, a deep learning processor utilizing the following techniques was proposed and investigated: (1) non-uniform memory architecture (NUMA), (2) temporal and spatial locality, (3) weight matrix tiling, (4) variable data precision, variable multiplier width, and long one line of memory, and (5) dedicated data flow.

Continuing demand for IoT devices with small form factor, light-weight better battery lifetime is expected, and further researches on power conversion techniques and energy efficiency deep learning processor for IoT application are required. For better SC converter design, novel topologies that improve P_{SSL} and P_{bot} can be further investigated, and fundamental techniques to minimize bottom-plate parasitic capacitors MOS-cap in CMOS process can be studied for further efficiency improvement and power density improvement. Also, the emerging algorithm, deep learning is lack of data set designed for IoT application, where limited memory capacity is available, so the optimization of the inference performance and data set size can be investigated further.

APPENDIX A

Mathematical Derivation of Slow-Switching Limit Impedance for Step-Down Switched-Capacitor DC-DC Converter with 2-Phase Interleaving

This chapter demonstrates mathematical derivation steps of slow-switching impedance (R_{SSL}) and corresponding voltage drop, discussed earlier in Chapter. Series-parallel, conventional ladder, variant ladder, and SAR topologies are investigated. For simple analysis, two-phase interleaving is assumed, and notation is summarized for convenience in Table A.1.

C1	Unit flying capacitance
C2	Output decoupling capacitance (*C2=xC1*)
Ctot	Total flying capacitance
Ts	Switching period
Io	Load current
k/N	Conversion ratio, where k and N are integers such that $0 \leq k \leq N$

Table A.1. Notation for derivation of mathematical expression for slow switching impedance.

A.1 Series-parallel SC Converter

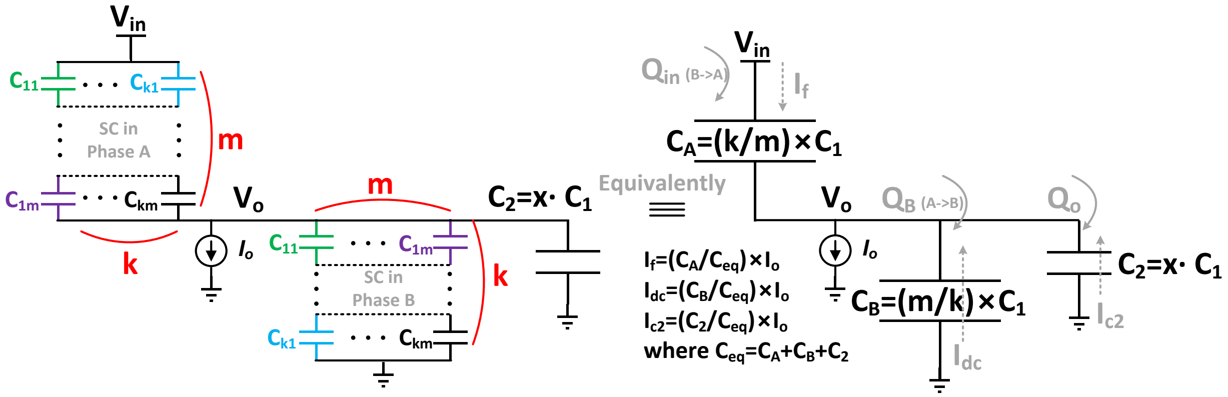


Figure A.1. Series-parallel SC for k/N × Vin generation (N=m+k).

Note that Q_o is the amount of instantaneous charge influx toward C2 at phase transition. Charge flow and current flow direction is shown in Figure A.1. Let us start with equivalent capacitance expression.

$$\text{Output capacitance } C_2 = xC_1 \tag{A.1}$$

$$C_A = \frac{k}{m}C_1 \text{ and } C_B = \frac{m}{k}C_1, \tag{A.2}$$

$$\text{thus } C_{eq} = C_A + C_B + C_1 = \left(\frac{m^2+k^2+mkx}{mk}\right)C_1 \tag{A.3}$$

$$\frac{C_A}{C_{eq}} = \frac{k^2}{k^2+m^2+mkx} \tag{A.4}$$

$$\frac{C_B}{C_{eq}} = \frac{m^2}{k^2+m^2+mkx} \tag{A.5}$$

To find Q_{in}, the amount of instantaneous charge influx from V_{in} at phase transition,

$$\text{Since } (I_{in})_{AV} = \frac{k}{m+k}I_o, \tag{A.6}$$

$$\frac{k}{m+k}I_o \times \frac{T_s}{2} = Q_{in} + I_f\frac{T_s}{2} = Q_{in} + \frac{C_A}{C_{eq}}I_o\frac{T_s}{2} \tag{A.7}$$

$$\text{Hence, } Q_{in} = \frac{I_oT_s}{2}\left(\frac{k}{m+k} - \frac{C_A}{C_{eq}}\right) \tag{A.8}$$

Now, to find $V_o(A^s)$, output voltage at the starting point of phase A,

$$V_o(A^s) = \frac{k}{m} \times V_{FLY}(A^e) + \frac{Q_B}{C_B} \tag{A.9}$$

$$V_o(A^s) = \frac{k}{m} \times V_{FLY}(A^e) + \frac{Q_B}{C_B} \tag{A.10}$$

$$V_o(A^s) = \frac{k}{m} \times (V_{in} - V_o(A^s) + \frac{I_o}{C_{eq}}\frac{T_s}{2}) + \frac{Q_B}{C_B} \tag{A.11}$$

$$\text{Thus, } V_o(A^s) = \frac{k}{m+k}V_{in} + \frac{I_oT_s}{2C_{eq}}\left(\frac{k}{m+k}\right) + \frac{m}{m+k}(\frac{\mathrm{Q_B}}{\mathrm{C_B}}) \tag{A.12}$$

As $V_o(A^s)$ is found, we can easily find $V_o(AV)$

$$V_o(AV) = V_o(A^s) - \frac{\mathrm{I_o}}{Ceq}\frac{T_s}{2}\frac{1}{2} = \frac{k}{N}Vin - \mathit{\Delta}_k \tag{A.13}$$

$$\text{where } \mathit{\Delta}_k = \frac{IoTs}{2C1}(\frac{k(N-k)(N^2-2k^2x-4Nk+4k^2+2Nkx)}{2N^2(N^2-k^2x-2Nk+2k^2+Nkx)}) \tag{A.14}$$

In summary,

$$\text{When } C2 = 0\ (x = 0),\ \mathit{\Delta}_k = \frac{IoTs}{2C1}(\frac{k(N-k)(N-2k)^2}{2N^2(k^2+(N-k)^2)}) \tag{A.15}$$

$$\text{(When k = 1); } \frac{IoTs}{2C1}(\frac{1(N-1)(N-2)^2}{2N^2(1^2+(N-1)^2)}) \approx \frac{IoTs}{2C1}(\frac{1}{2N}) \text{ for large N} \tag{A.16}$$

$$\text{When } C2 \text{ is very large } (x = \infty),\ \mathit{\Delta}_k = \frac{IoTs}{2C1}(\frac{k(N-k)}{N^2}) \tag{A.16}$$

$$\text{(When k = 1); } \frac{IoTs}{2C1}(\frac{1(N-1)}{N^2}) \approx \frac{IoTs}{2C1}(\frac{1}{N}) \text{ for large N} \tag{A.17}$$

Finally, to make the equation as a function of total flying capacitance, *C_{tot}*,

When $C2 = 0\ (x = 0)$

$$C1 = \frac{Ctot}{2km} = \frac{Ctot}{2k(N-k)} \tag{A.18}$$

$$\text{Thus, } \mathit{\Delta}_k = \frac{IoTs}{2\frac{Ctot}{2k(N-k)}}\left(\frac{k(N-k)(N-2k)^2}{2N^2(k^2+(N-k)^2)}\right) = \frac{IoTs}{2Ctot}(\frac{2\,k^2(N-k)^2(N-2k)^2}{2N^2(k^2+(N-k)^2)}) \tag{A.19}$$

$$\mathit{\Delta}_1 \approx \frac{IoTs}{2Ctot}(1) \text{ for large N} \tag{A.20}$$

When $C2$ is very large $(x = \infty)$,

$$\mathit{\Delta}_k = \frac{IoTs}{2\frac{Ctot}{2k(N-k)}}\left(\frac{k(N-k)}{N^2}\right) = \frac{IoTs}{2Ctot}(\frac{2k^2(N-k)^2}{N^2}) \tag{A.21}$$

$$\Delta_1 \approx \frac{IoTs}{2Ctot}(2) \text{ for large N} \tag{A.22}$$

A.2 Conventional Ladder SC Converter for Conversion Ratio of k/N (N>3)

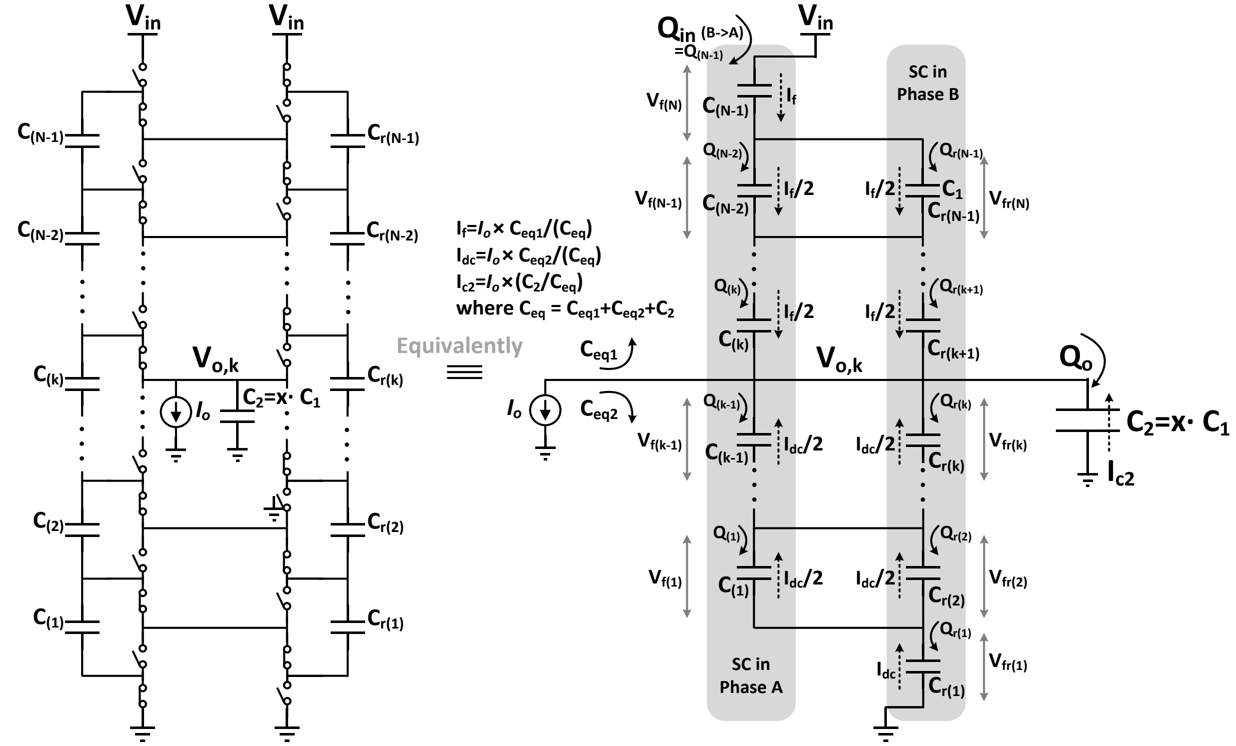


Figure A.2. Conventional ladder SC for k/N × Vin generation.

Let us start with equivalent capacitance expression

$$C_{eq1} = \frac{1}{\frac{(N-k-1)}{2C_1}+\frac{1}{C_1}} = \frac{2C_1}{N-k+1} \tag{A.23}$$

$$C_{eq2} = \frac{1}{\frac{(k-1)}{2C_1}+\frac{1}{C_1}} = \frac{2C_1}{k+1} \tag{A.24}$$

$$C_{eq} = C_{eq1} + C_{eq2} + C_2, \text{ where } C_2 = xC_1 \tag{A.25}$$

To find Q$_{in}$, the amount of instantaneous charge influx from V$_{in}$ at phase transition,

$$\text{Since } (I_{in})_{AV} = \frac{k}{N} I_o, \tag{A.26}$$

$$\frac{k}{N} I_o \times \frac{T_S}{2} = Q_{in} + I_f \frac{T_S}{2}, \text{ thus } Q_{in} = \frac{I_o T_S}{2}\left(\frac{k}{N}\right) - \frac{I_f T_S}{2} \tag{A.27}$$

Now, to find *Q*$_{r(i)}$ and *Q*$_{(i)}$, which indicate instantaneous charge influx to i-th flying cap on the right, and to i-th flying cap on the left, respectively. The charge flow direction is illustrated in Figure A.2. To derive expressions, net charge influx of any flying capacitor is zero in steady state.

For i = N-1; $\quad Q_{r(N-1)} = -Q_{in} - \frac{3}{2}\frac{I_f T_s}{2}$ (A.28)

For N-1 > i > k; $\quad Q_{r(i)} = -Q_{(i)} - \frac{I_f T_s}{2}$ (A.29)

For i = k; $\quad Q_{r(k)} = -Q_{(k)} - (\frac{I_f}{2} - \frac{I_{dc}}{2})\frac{T_s}{2}$ (A.30)

For k > i > 1; $\quad Q_{r(i)} = -Q_{(i)} + \frac{I_{dc} T_s}{2}$ (A.31)

For i = 1; $\quad Q_{r(1)} = -Q_{(k)} + \frac{3}{2}\frac{I_{dc} T_s}{2}$ (A.32)

For the i-th flying cap on the left, following equations are valid

For i = N-1; $\quad Q_{(N-1)} = Q_{in}$ (A.33)

For N-1 > i > k; $\quad Q_{(N-1)} = Q_{in} - Q_{r(i+1)}$ (A.34)

For k > i; $\quad Q_{(i)} = Q_{in} - Q_o - Q_{(i+1)}$ (A.35)

where $Q_o = \frac{I_o T_s}{2}(\frac{C_2}{C_{eq}})$, the amount of instantaneous charge influx toward C2 at phase transition.

In order to find expression for Q$_{(i)}$ with Q$_{r(i)}$, for *N-3* ≥ *i* ≥ *k*,

$$Q_{(N-2)} = Q_{in} - Q_{r(N-1)} = \frac{k}{N} I_o T_s - \frac{1}{4} I_f T_s \tag{A.36}$$

$$Q_{(N-3)} = Q_{in} - Q_{r(N-2)} = Q_{in} - \left(-Q_{(N-2)} - I_f \frac{T_s}{2}\right) = Q_{in} + Q_{(N-2)} + I_f \frac{T_s}{2} \tag{A.37}$$

Thus, for *N-3* ≥ *i* ≥ *k*, $Q_{(i)} = \left(Q_{in} + I_f \frac{T_s}{2}\right)(N - 2 - i) + Q_{(N-2)}$ (A.38)

Since

Thus, $\left(Q_{in} + I_f \frac{T_s}{2}\right) = \frac{I_o T_s k}{2N}$ (A.39)

$$Q_{(i)} = \left(\frac{I_o T_s k}{2N}\right)(N - 2 - i) + Q_{(N-2)} \tag{A.40}$$

To find expression for Q$_{(i)}$ with Q$_{r(i)}$, for *k* > *i* ,

$$Q_{(k-1)} = (Q_{in} - Q_o) - Q_{r(k)} = (Q_{in} - Q_o) — \{-Q_{(k)} - \tfrac{T_s}{4}\left(I_f - I_{dc}\right)\} \tag{A.41}$$

$$Q_{(k-1)} = Q_{in} - Q_o + Q_{(k)} + \frac{T_s}{4}\left(I_f - I_{dc}\right) \tag{A.42}$$

Since

$$Q_{(k)} = \left(Q_{in} + I_f\,\tfrac{T_s}{2}\right)(N - 2 - k) + Q_{(N-2)} \tag{A.43}$$

$$\text{Thus, } Q_{(k-1)} = Q_{in} - Q_o + \tfrac{T_s}{4}\left(I_f - I_{dc}\right) + \left(\tfrac{I_o T_s k}{2N}\right)(N - 2 - k) + Q_{(N-2)} \tag{A.44}$$

Similarly,

$$Q_{(k-2)} = (Q_{in} - Q_o) - Q_{r(k-1)} = (Q_{in} - Q_o) — (-Q_{(k-1)} + \frac{T_s}{2} I_{dc}) \tag{A.45}$$

$$Q_{(k-2)} = (Q_{in} - Q_o - \frac{I_{dc}T_s}{2}) + Q_{(k-1)} \tag{A.46}$$

$$\text{Thus, for } k\text{-}2 \geq i,\; Q_{(i)} = \left(Q_{in} - Q_o - \tfrac{I_{dc}T_s}{2}\right)(k - 1 - i) + Q_{(k-1)} \tag{A.47}$$

To find $V_{fr(1)}(A^s)$ and $V_{o,k}(AV)$, the following equation (1) will be used in the end

$$\{\sum_{i=1}^{N-1} V_{f(i)}(A^s)\} + V_{fr(1)}(A^s) = V_{in} - (1) \tag{A.48}$$

As it is always valid that $V_{f(i)} = V_{fr(i+1)}$ for $N\text{-}2 \geq i \geq 1$

$$V_{f(1)}(A^s) = V_{fr(1)}(A^e) + \frac{Q_{(1)}}{C_1} \tag{A.49}$$

$$V_{f(2)}(A^s) = V_{fr(2)}(A^e) + \tfrac{Q_{(2)}}{C_1} = V_{f(1)}(A^e) + \tfrac{Q_{(2)}}{C_1} \tag{A.50}$$

$$\text{And } V_{f(1)}(A^e) = V_{f(1)}(A^s) - \tfrac{I_{dc}T_s}{4C_1} = V_{fr(1)}(A^e) + \tfrac{Q_{(1)}}{C_1} - \tfrac{I_{dc}T_s}{4C_1} \tag{A.51}$$

$$\text{Thus, } V_{f(2)}(A^s) = (V_{fr(1)}(A^e) + \tfrac{Q_{(1)}}{C_1} - \tfrac{I_{dc}T_s}{4C_1}) + \tfrac{Q_{(2)}}{C_1} = V_{fr(1)}(A^e) - \tfrac{I_{dc}T_s}{4C_1} + \tfrac{(Q_{(1)} + Q_{(2)})}{C_1} \tag{A.52}$$

For $k \geq i \geq 2,$

$$V_{f(i)}(A^s) = V_{fr(1)}(A^e) - (i - 1)\left(\tfrac{I_{dc}T_s}{4C_1}\right) + \tfrac{1}{C_1}\sum_{j=1}^{i} Q_{(j)} \tag{A.53}$$

$$V_{f(k+1)}(A^s) = V_{fr(k+1)}(A^e) + \frac{Q_{(k+1)}}{C_1} = V_{f(k)}(A^e) + \frac{Q_{(k+1)}}{C_1} \tag{A.54}$$

$$\text{And } V_{f(k)}(A^e) = V_{f(k)}(A^s) + \frac{I_f T_s}{4C_1} \tag{A.55}$$

$$V_{f(k+1)}(A^s) = (V_{f(k)}(A^s) + \frac{\mathrm{I_f T_s}}{4\mathrm{C_1}}) + \frac{Q_{(k+1)}}{C_1} \tag{A.56}$$

For *N-1* ≥ *i* ≥ *k+1*,

$$V_{f(i)}(A^s) = V_{f(k)}(A^s) + (i-k)\left(\frac{I_f T_s}{4C_1}\right) + \frac{1}{C_1}\sum_{j=k+1}^{i} Q_{(j)} \tag{A.57}$$

$$\text{As } V_{f(k)}(A^s) = V_{fr(1)}(A^e) - (k-1)\left(\frac{I_{dc} T_s}{4C_1}\right) + \frac{1}{C_1}\sum_{j=1}^{k} Q_{(j)} \tag{A.58}$$

Thus, for *N-1* ≥ *i* ≥ *k+1*,

$$V_{f(i)}(A^s) = V_{fr(1)}(A^e) - (k-1)\left(\frac{I_{dc} T_s}{4C_1}\right) + (i-k)\left(\frac{I_f T_s}{4C_1}\right) + \frac{1}{C_1}\sum_{j=1}^{i} Q_{(j)} \tag{A.59}$$

Now, to find $\sum_{i=1}^{N-1} V_{f(i)}(A^s)$,

$$\begin{aligned}\sum_{i=1}^{N-1} V_{f(i)}(A^s) = (N-1)V_{fr(1)}(A^e) - \left(\frac{I_{dc} T_s}{4C_1}\right) \cdot \textstyle\sum_{i=1}^{k-1} i + \left(\frac{I_f T_s}{4C_1}\right) \cdot \sum_{i=1}^{N-1-k} i \, + \\ \frac{1}{C_1}\textstyle\sum_{i=1}^{N-1}(N-i)Q_{(i)} - (\mathrm{N}-1-\mathrm{k})(\mathrm{k}-1)\left(\frac{I_{dc} T_s}{4C_1}\right)\end{aligned} \tag{A.60}$$

$$\text{where } V_{fr(1)}(A^e) = V_{fr(1)}(A^s) - \frac{I_{dc} T_s}{2C_1} \tag{A.61}$$

Solving for $V_{fr(1)}(A^s)$,

$$V_{fr(1)}(A^s) = \frac{V_{in}}{N} - (\frac{I_O T_s}{2C_1})(\frac{n(N,k,x)}{6(\ xk^2 + Nxk + 2N + x + Nx + 4)}) \tag{A.62}$$

$$\begin{aligned}\text{where } n(N,k,x) = 12Nk - 12N - 12Nk^2 + 8N^2k + 2Nk^3 + 4N^3k + k^3x - k^5x \, - \\ 12N^2 + 4k^3 - 6N^2k^2 - 3N^2k^2x - 5N^2k^3x + 2N^3k^2x - 3Nk^2x + 2N^2kx \, + \\ Nk^3x + 2N^3kx + 4Nk^4x\end{aligned} \tag{A.63}$$

Then, solving for $Vo(AV)$

$$Vo(AV) = \{V_{fr(1)}(A^s) + \sum_{i=1}^{k-1} V_{f(i)}(A^s)\} - \frac{I_o T_s}{4(C_{eq1} + C_{eq2})} \tag{A.64}$$

$$V_{o,k}(AV) = \frac{k}{N} Vin - \Delta_k \tag{A.65}$$

$$\Delta_k = \frac{IoTs}{2C1}\left(\frac{n(N,k,x)}{d(N,k,x)}\right) \tag{A.66}$$

where $n(N,k,x) = 4Nk - 3N + Nk^2 - N^2k - 16Nk^3 + 4Nk^4 - k^2x + 3k^4x - 2k^6x - 3N^2 - 4k^2 + 8k^4 + 8N^2k^2 - 8N^2k^3 + 4N^3k^2 + 3N^2k^2x - 4N^2k^3x + 2N^3k^2x - 6N^2k^4x + 2N^3k^3x + Nkx - Nk^2x + N^2kx - 6Nk^3x + 2Nk^4x + 6Nk^5x$, and $d(N,k,x) = 6(-xk^2 + Nxk + 2N + x + Nx + 4)$ (A.67)

In summary,

When $C2 = 0$ $(x = 0)$,

$$\Delta_k = \frac{IoTs}{2C1}\left(\frac{4N^3k^2 - 8N^2k^3 + 8N^2k^2 - N^2k - 3N^2 + 4Nk^4 - 16Nk^3 + Nk^2 + 4Nk - 3N + 8k^4 - 4k^2}{12N(N+2)}\right) \tag{A.68}$$

$$\text{(When k = 1); } \frac{IoTs}{2C1}\left(\frac{(N-2)(2N^2+2N-1)}{6N(N+2)}\right) \approx \frac{IoTs}{2C1}\left(\frac{N}{3}\right) \text{ for large N} \tag{A.69}$$

When $C2$ is very large $(x = \infty)$,

$$\Delta_k = \frac{IoTs}{2C1}\left(\frac{k(N-k)(-2k^2+2Nk+1)}{6N}\right) \tag{A.70}$$

$$\text{(When k = 1); } \frac{IoTs}{2C1}\left(\frac{(N-1)(2N-1)}{6N}\right) \approx \frac{IoTs}{2C1}\left(\frac{N}{3}\right) \text{ for large N} \tag{A.71}$$

This is attributed to $V_{fr(1)}(A^s)$, which can be approximated to a term proportional to N

Finally, to make the equation as a function of total flying capacitance, *C_{tot}*,

When $C2 = 0$ $(x = 0)$,

$$C1 = \frac{Ctot}{2(N-1)} \tag{A.72}$$

$$\Delta_k = \frac{IoTs}{2Ctot}\frac{2(N-1)\times(4N^3k^2 - 8N^2k^3 + 8N^2k^2 - N^2k - 3N^2 + 4Nk^4 - 16Nk^3 + Nk^2 + 4Nk - 3N + 8k^4 - 4k^2)}{12N(N+2)} \tag{A.73}$$

$$\Delta_1 \approx \frac{IoTs}{2Ctot}\left(\frac{2N^2}{3}\right) \text{ for large N} \tag{A.74}$$

When $C2$ is very large $(x = \infty)$,

$$\Delta_k = \frac{IoTs}{2Ctot}\left(\frac{2(N-1)\times k(N-k)(-2k^2+2Nk+1)}{6N}\right) \tag{A.75}$$

$$\Delta_1 \approx \frac{IoTs}{2Ctot}\left(\frac{2N^2}{3}\right) \text{ for large N} \tag{A.76}$$

A.3 Variant Ladder Topology

A.3.1 Variant Ladder SC Converter for Conversion Ratio of k/N (k>1, N>3)

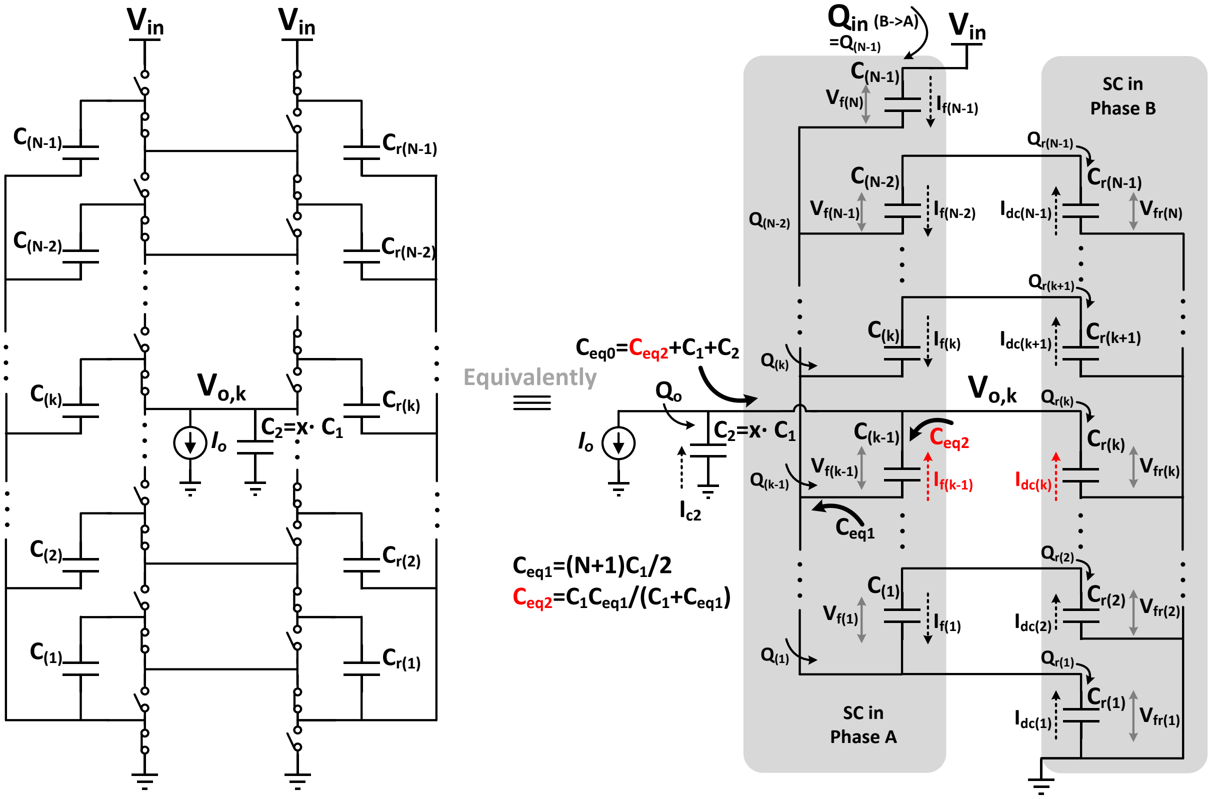


Figure A.3. Variant ladder SC for k/N × Vin generation (k ≠ 1).

As illustrated in Figure A.3, the following conditions are easily derived

$$Q_{r(i)} = Q_{(i-1)} \text{ for } N\text{-}1 \geq i \geq 2 \tag{A.77}$$

$$I_{dc(i)} = I_{f(i-1)} \text{ for } i \neq k \text{ and } N\text{-}1 \geq i \geq 2 \tag{A.78}$$

Now, let us start with equivalent capacitance expression

$$C_{eq1} = C_1 + \frac{C_1}{2}(N-3) = \frac{C1}{2}(N+1) \tag{A.79}$$

$$C_{eq2} = \frac{1}{\frac{1}{C_{eq1}}+\frac{1}{C_1}} = \frac{C_1(N+1)}{N+3} \quad \text{(A.80)}$$

$$C_{eq0} = C_{eq2} + C_1 + C_2, \text{ where } C_2 = xC_1 \quad \text{(A.81)}$$

Then, we can find $I_{f(i)}$, $I_{dc(k)}$, and $I_{dc(1)}$

$$I_{f(k-1)} = \frac{C_{eq2}}{C_{eq0}} I_o \quad \text{(A.82)}$$

$$I_{f(N-1)} = I_{f(k-1)} \times \left(\frac{C_1}{C_{eq1}}\right) = I_{f(k-1)} \times \left(\frac{2}{N+1}\right) \quad \text{(A.83)}$$

$$I_{f(i)} = I_{f(k-1)} \times \left(\frac{C_1/2}{C_{eq1}}\right) = I_{f(k-1)} \times \left(\frac{1}{N+1}\right) \text{ for } i \neq N\text{-}1 \text{ and } i \neq k\text{-}1 \quad \text{(A.84)}$$

$$I_{dc(k)} = \frac{C_1}{C_{eq0}} I_o \quad \text{(A.85)}$$

$$I_{dc(1)} = I_{f(k-1)} \times \left(\frac{C_1}{C_{eq1}}\right) = I_{f(k-1)} \times \left(\frac{2}{N+1}\right) \quad \text{(A.86)}$$

The current flow direction is illustrated in Figure A.3. Note that $I_{f(k\text{-}1)}$ is different from all the other $I_{f(i)}$ for $i \neq k\text{-}1$. To find Q_{in}, the amount of instantaneous charge influx from V_{in} at phase transition,

Since $(I_{in})_{AV} = \frac{k}{N} I_o$,

$$\frac{k}{N} I_o \times \frac{T_S}{2} = Q_{in} + I_{f(N-1)}\frac{T_S}{2}, \text{ thus } Q_{in} = \frac{I_O T_S}{2}\left(\frac{k}{N}\right) - \frac{I_{f(k-1)}T_S}{(N+1)} \quad \text{(A.87)}$$

Now, to find $Q_{r(i)}$ and $Q_{(i)}$, which indicate instantaneous charge influx to i-th flying cap on the right, and to i-th flying cap on the left, respectively. The charge flow direction is illustrated in Figure A.3. To derive expressions, net charge influx of any flying capacitor is zero in steady state.

$$\begin{aligned} Q_{(N-2)} &= Q_{r(N-1)} \\ &= -Q_{in} - \left(I_{f(N-1)} - I_{dc(N-1)}\right)\frac{T_S}{2} = -\left(\frac{I_O T_S}{2}\left(\frac{k}{N}\right) - \frac{I_{f(k-1)}T_S}{(N+1)}\right) - \frac{I_{f(k-1)}T_S}{2(N+1)} \end{aligned} \quad \text{(A.88)}$$

Thus,

$$Q_{(N-2)} = -\frac{k}{2N} I_o T_s + \frac{I_{(k-1)} T_S}{2(N+1)} \tag{A.89}$$

For *N-3* ≥ *i* ≥ *k*,

$$Q_{(i)} = Q_{(N-2)} \tag{A.90}$$

Then, to find $Q_{(k-1)}$,

$$Q_{(k-1)} = Q_{r(k)} + Q_o = \{Q_{(k)} - \left(I_{f(k)} - I_{dc(k)}\right)\frac{T_S}{2}\} + Q_o \tag{A.91}$$

Thus,

$$Q_{(k-1)} = Q_{(N-2)} + Q_o - (I_{f(k)} - I_{dc(k)})\frac{T_s}{2} \tag{A.92}$$

Then, to find $Q_{(k-2)}$,

$$Q_{(k-2)} = Q_{r(k-1)} = Q_{(k-1)} - (-I_{f(k-1)} - I_{dc(k-1)})\frac{T_s}{2} \tag{A.93}$$

$$\text{Thus,} \quad Q_{(k-2)} = Q_{(k-1)} + (\frac{N+2}{N+1})\, I_{f(k-1)} \frac{T_s}{2} \tag{A.94}$$

For *k-3* ≥ *i* ≥ *1*,

$$Q_{(i)} = Q_{(k-2)} \tag{A.95}$$

To find $V_{fr(1)}(A^s)$ and $V_{fr(k)}(AV)$, the following equation (2) will be used in the end

$$V_{f(N-1)}(A^s) + V_{fr(1)}(A^s) = V_{in} \;\; -(2) \tag{A.96}$$

Now, to find $V_{f(N-1)}(A^s) = V_{fr(N-1)}(A^e) + \frac{Q_{in}}{C_1}$

$$\text{Since } V_{fr(N-1)}(A^e) = V_{fr(N-1)}(A^s) - I_{dc(N-1)} \frac{T_s}{2C_1}, \tag{A.97}$$

$$\text{Thus, } V_{f(N-1)}(A^s) = V_{fr(N-1)}(A^s) - I_{dc(N-1)} \frac{T_S}{2C_1} + \frac{Q_{in}}{C_1} \tag{A.98}$$

To find $V_{fr(N-1)}(A^s)$,

$$V_{fr(i)}(A^s) = \left(V_{fr(i-1)}(A^s) - I_{dc(k-1)}\frac{T_s}{2C_1}\right) - \frac{Q_{f(k-1)}}{C_1} + V_{fr(1)}(A^s) \text{ for any } i \tag{A.99}$$

Thus,

$$V_{fr(N-1)}(A^s) = (N-1)V_{fr(1)}(A^s) - \left(\frac{T_s}{2C_1}\right) \cdot \sum_{i=1}^{N-2} I_{dc(i)} - \left(\frac{1}{C_1}\right) \cdot \sum_{i=1}^{N-2} Q_{f(i)} \tag{A.100}$$

Hence,

$$V_{f(N-1)}(A^s) = (N-1)V_{fr(1)}(A^s) - \left(\frac{T_s}{2C_1}\right) \cdot \sum_{i=1}^{N-1} I_{dc(i)} - \left(\frac{1}{C_1}\right) \cdot \sum_{i=1}^{N-2} Q_{f(i)} + \frac{Q_{in}}{C_1} \tag{A.101}$$

With equation (2), solving for $V_{fr(1)}(A^s)$,

$$V_{fr(1)}(A^s) = \frac{\mathrm{V_{in}}}{N} + \frac{I_o T_s}{2C_1}\left(\frac{4k-4N+2Nk-3Nx+3kx-N^2x+Nkx}{N^2(2N+3x+Nx+4)}\right) \tag{A.102}$$

Since

$$V_{o,k}(AV) = V_{fr(k)}(AV) = V_{fr(k)}(A^s) - \frac{I_o T_s}{4C_{eq0}} \tag{A.103}$$

$$\text{and } V_{fr(k)}(A^s) = kV_{fr(1)}(A^s) - \left(\frac{T_s}{2C_1}\right) \cdot \sum_{i=1}^{k-1} I_{dc(i)} - \left(\frac{1}{C_1}\right) \cdot \sum_{i=1}^{k-1} Q_{f(i)}, \tag{A.104}$$

$$\text{where } \sum_{i=1}^{k-1} I_{dc(i)} = \frac{k}{(N+1)} I_{(k-1)}, \text{ and } \sum_{i=1}^{k-1} Q_{f(i)} = I_o T_s \frac{n_Q(N,k,x)}{d_Q(N,k,x)} \tag{A.105}$$

$$\text{where } n_Q(N,k,x) = 4k - 7N + 7Nk - 3Nx + 3kx - 2Nk^2 + 2N^2k - N^2x - 3k^2x - 3N^2 - 4k^2 + 4Nkx - Nk^2x + N^2kx, \tag{A.106}$$

$$\text{and } d_Q(N,k,x) = 2N(2N + 3x + Nx + 4) \tag{A.107}$$

Hence,

$$V_{o,k}(AV) = \frac{k}{N} Vin - \Delta_k \tag{A.108}$$

$$\text{where } \Delta_k = \frac{IoTs}{2C1}\left(\frac{n(N,k,x)}{2N^2(2N+3x+Nx+4)}\right) \tag{A.109}$$

$$\text{where } n(N,k,x) = 16Nk - 12Nk^2 + 16N^2k + 4N^3k - 6N^2x - 2N^3x - 6k^2x - 11N^2 - 5N^3 - 8k^2 - 4N^2k^2 - 2N^2k^2x + 12Nkx - 8Nk^2x + 10N^2kx + 2N^3kx \tag{A.110}$$

In summary,

When $C2 = 0$ $(x = 0)$,

$$\Delta_k = \frac{IoTs}{2C1}\left(\frac{4N^3k-5N^3-4N^2k^2+16N^2k-11N^2-12Nk^2+16Nk-8k^2}{4N^2(N+2)}\right) \tag{A.110}$$

(When k = 2);

$$\Delta_2 = \frac{IoTs}{2C1}\left(\frac{3N^3+5N^2-16N-32}{4N^2(N+2)}\right) \approx \frac{IoTs}{2C1}\left(\frac{3}{4}\right) \text{ for } N \gg k, \text{ and } k = 2 \tag{A.111}$$

When $C2$ is very large $(x = \infty)$,

$$\Delta_k = \frac{IoTs}{2C1}\left(\frac{(N-k)(k-N+Nk)}{N^2}\right) \tag{A.112}$$

(When k = 2);

$$\Delta_2 = \frac{IoTs}{2C1}\left(\frac{(N-2)(N+2)}{N^2}\right) \approx \frac{IoTs}{2C1}(1) \text{ for } N \gg k, \text{ and } k = 2 \tag{A.113}$$

This is attributed to $\sum_{i=1}^{k-1} Q_{f(i)}$ which can be approximated to a term proportional to $\frac{IoTs}{2C1}(1)$

Finally, to make the equation as a function of total flying capacitance, C_{tot},

When $C2 = 0$ $(x = 0)$,

$$C1 = \frac{Ctot}{2(N-1)} \tag{A.114}$$

Thus,

$$\Delta_k = \frac{IoTs}{2\frac{Ctot}{2(N-1)}}\left(\frac{4N^3k-5N^3-4N^2k^2+16N^2k-11N^2-12Nk^2+16Nk-8k^2}{4N^2(N+2)}\right) \tag{A.115}$$

$$= \frac{IoTs}{2Ctot}\left(\frac{2(N-1)\times(4N^3k-5N^3-4N^2k^2+16N^2k-11N^2-12Nk^2+16Nk-8k^2)}{4N^2(N+2)}\right) \tag{A.116}$$

(When k = 2);

$$\Delta_2 = \frac{IoTs}{2Ctot}\left(\frac{2(N-1)\times(3N^3+5N^2-16N-32)}{4N^2(N+2)}\right) \approx \frac{IoTs}{2Ctot}\left(\frac{3}{2}N\right) \text{ for } N \gg k, \text{ and } k = 2 \tag{A.117}$$

When $C2$ is very large $(x = \infty)$,

$$\Delta_k = \frac{IoTs}{2\frac{Ctot}{2(N-1)}}\left(\frac{(N-k)(k-N+Nk)}{N^2}\right) = \frac{IoTs}{2Ctot}\left(\frac{2(N-1)\times(N-k)(k-N+Nk)}{N^2}\right) \tag{A.118}$$

(When k = 2); (A.119)

$$\Delta_2 = \frac{IoTs}{2Ctot}\left(\frac{2(N-1)\times(N-2)(N+2)}{N^2}\right) \approx \frac{IoTs}{2Ctot}(2N) \text{ for } N \gg k, \text{ and } k = 2$$

A.3.2 Variant Ladder SC Converter for Conversion Ratio of k/N (N>3)

As illustrated in Figure A.4, the following conditions are easily derived

$$Q_{r(i)} = Q_{(i-1)} \text{ for } N\text{-}1 \geq i \geq 2 \tag{A.120}$$

$$I_{dc(i)} = I_{f(i-1)} \text{ for } N\text{-}1 \geq i \geq 2 \tag{A.121}$$

Now, let us start with equivalent capacitance expression

$$C_{eq1} = C_1 + \frac{C_1}{2}(N-2) = \frac{C1}{2}N \tag{A.122}$$

$$C_{eq0} = C_{eq1} + C_1 + C_2, \text{ where } C_2 = xC_1 \tag{A.123}$$

Then, we can $I_{f(i)}$ and $I_{dc(1)}$

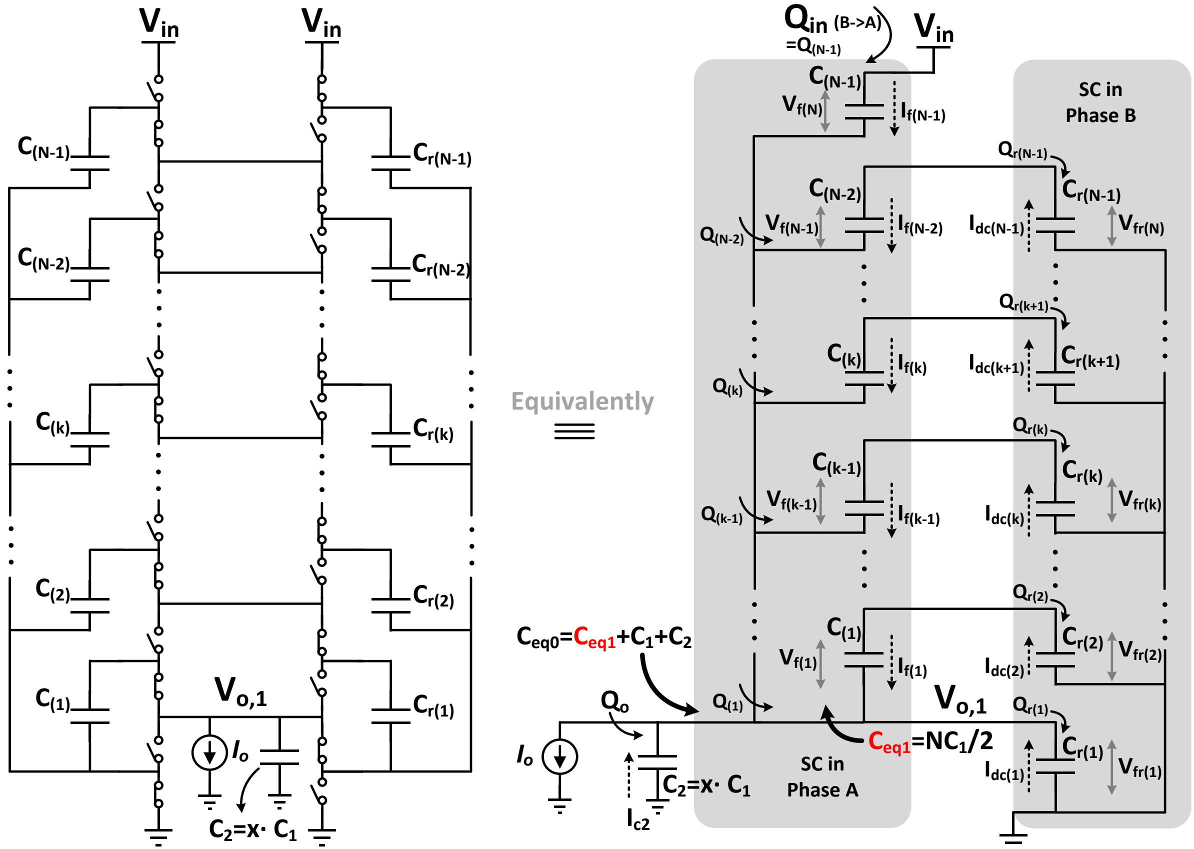


Figure A.4. Variant ladder SC for 1/N × Vin generation.

$$I_{f(N-1)} = I_o\left(\frac{C_1}{C_{eq0}}\right) \tag{A.124}$$

$$I_{f(i)} = I_o\left(\frac{C_1/2}{C_{eq0}}\right) = \frac{I_{f(N-1)}}{2} \text{ for } i \neq N\text{-}1 \tag{A.125}$$

$$I_{dc(1)} = I_o\left(\frac{C_1}{C_{eq0}}\right) = I_{f(N-1)} \tag{A.126}$$

The current flow direction is illustrated in Figure A.4. Note that $I_{f(i)}$ and $I_{dc(i)}$ has consistent direction as opposed to Figure A.3. To find Q_{in}, the amount of instantaneous charge influx from V_{in} at phase transition,

Since $(I_{in})_{AV} = \frac{1}{N} I_o$,

$$\frac{1}{N} I_o \times \frac{T_S}{2} = Q_{in} + I_{f(N-1)}\frac{T_S}{2}, \tag{A.127}$$

$$\text{Thus, } Q_{in} = \frac{I_o T_S}{2}\left(\frac{1}{N}\right) - \frac{T_S}{2} I_{f(N-1)} \tag{A.128}$$

Now, to find $Q_{r(i)}$ and $Q_{(i)}$, which indicate instantaneous charge influx to i-th flying cap on the right, and to i-th flying cap on the left, respectively. The charge flow direction is illustrated in Figure A.4. To derive expressions, net charge influx of any flying capacitor is zero in steady state.

$$Q_{(N-2)} = Q_{r(N-1)} = -Q_{in} - \left(I_{f(N-1)} - I_{dc(N-1)}\right)\frac{T_S}{2} = -Q_{in} - I_{f(N-1)}\frac{T_S}{4} \tag{A.129}$$

For $N\text{-}3 \geq i \geq 1$,

$$Q_{(i)} = Q_{(N-2)} \tag{A.130}$$

To find $V_{fr(1)}(A^S)$ and $V_{fr(k)}(AV)$, the following equation (3) will be used in the end

$$V_{f(N-1)}(A^S) + V_{fr(1)}(A^S) = V_{in} - (3) \tag{A.131}$$

To find $V_{fr(N-1)}(A^S)$,

$$V_{fr(i)}(A^S) = \left(V_{f(i-1)}(A^S)\right) + V_{fr(1)}(A^S) \text{ for any } i > 1 \tag{A.132}$$

And $V_{f(i-1)}(A^s) = V_{fr(i-1)}(A^e) - \frac{Q_{(k-1)}}{C_1} = (V_{fr(i-1)}(A^s) - I_{dc(k-1)}\frac{T_s}{2C_1}) - \frac{Q_{(k-1)}}{C_1}$ (A.133)

$$V_{f(i-1)}(A^s) = (V_{fr(i-1)}(A^s) - \frac{I_{f(N-1)}}{2}\frac{T_s}{2C_1}) - \frac{Q_{(k-1)}}{C_1} \quad \text{(A.134)}$$

Thus, $V_{fr(i)}(A^s) = \left((V_{fr(i-1)}(A^s) - \frac{I_{f(N-1)}}{2}\frac{T_s}{2C_1}) - \frac{Q_{(k-1)}}{C_1}\right) + V_{fr(1)}(A^s)$ for any i (A.135)

Now, we can find $V_{fr(N-1)}(A^s)$

$$V_{fr(N-1)}(A^s) = V_{fr(2)}(A^s) - (N-3)\frac{I_{f(N-1)}}{2}\frac{T_s}{2C_1} - \left(\frac{1}{C_1}\right)\cdot\sum_{i=2}^{N-2} Q_{f(i)} + (N-3)V_{fr(1)}(A^s) \quad \text{(A.130)}$$

Since $V_{fr(2)}(A^s) = V_{f(1)}(A^s) + V_{fr(1)}(A^s) = \{V_{fr(1)}(A^e) - \frac{Q_{(1)}}{C_1}\} + V_{fr(1)}(A^s)$, (A.131)

and $V_{fr(1)}(A^e) = V_{fr(1)}(A^s) - I_{dc(1)}\left(\frac{T_s}{2C_1}\right) - \frac{Q_{(1)}}{C_1} = V_{fr(1)}(A^s) - I_{f(N-1)}\left(\frac{T_s}{2C_1}\right) - \frac{Q_{(1)}}{C_1}$ (A.132)

$$\begin{aligned} V_{fr(2)}(A^s) &= \left[\left\{V_{fr(1)}(A^s) - I_{f(N-1)}\left(\frac{T_s}{2C_1}\right)\right\} - \frac{Q_{(1)}}{C_1}\right] + V_{fr(1)}(A^s) \\ &= 2V_{fr(1)}(A^s) - I_{f(N-1)}\left(\frac{T_s}{2C_1}\right) - \frac{Q_{(1)}}{C_1} \end{aligned} \quad \text{(A.133)}$$

Thus,

$$V_{fr(N-1)}(A^s) = (N-1)V_{fr(1)}(A^s) - (N-1)\frac{I_{f(N-1)}}{2}\left(\frac{T_s}{2C_1}\right) - \left(\frac{1}{C_1}\right)\cdot\sum_{i=1}^{N-2} Q_{f(i)} \quad \text{(A.134)}$$

Now, to find $V_{f(N-1)}(A^s)$

$$V_{f(N-1)}(A^s) = V_{fr(N-1)}(A^e) + \frac{Q_{in}}{C_1} \quad \text{(A.135)}$$

Since $V_{fr(N-1)}(A^e) = V_{fr(N-1)}(A^s) - I_{dc(N-1)}\frac{T_s}{2C_1}$, (A.136)

Thus, $V_{f(N-1)}(A^s) = V_{fr(N-1)}(A^s) - \frac{I_{f(N-1)}}{2}\left(\frac{T_s}{2C_1}\right) + \frac{Q_{in}}{C_1}$ (A.137)

Hence,

$$V_{f(N-1)}(A^s) = (N-1)V_{fr(1)}(A^s) - N\frac{I_{f(N-1)}}{2}\left(\frac{T_s}{2C_1}\right) - \left(\frac{1}{C_1}\right)\cdot\sum_{i=1}^{N-2} Q_{f(i)} + \frac{Q_{in}}{C_1} \quad \text{(A.138)}$$

With equation (3), solving for $V_{fr(1)}(A^s)$, and for $V_{fr(1)}(AV)$

Solving for $V_{fr(1)}(AV)$

$$V_{fr(1)}(AV) = V_{fr(1)}(A^s) - \frac{I_o T_s}{4C_{eq0}} = \frac{1}{N}V_{in} - \varDelta_1 \tag{A.139}$$

In generalized form,

$$V_{o,1}(AV) = V_{fr(1)}(AV) = \frac{1}{N}Vin - \varDelta_1, \text{ where } \varDelta_1 = \frac{IoTs}{2C1}(\frac{N-2x+2Nx-2}{N^2(N+2x+2)}) \tag{A.140}$$

When $C2 = 0$ $(x = 0)$,

$$\varDelta_1 = \frac{IoTs}{2C1}(\frac{N-2}{N^2(N+2)}) \approx \frac{IoTs}{2C1}(\frac{1}{N^2}) \text{ for large N} \tag{A.141}$$

When $C2$ is very large $(x = \infty)$,

$$\varDelta_1 = \frac{IoTs}{2C1}(\frac{N-1}{N^2}) \approx \frac{IoTs}{2C1}(\frac{1}{N}) \text{ for large N} \tag{A.142}$$

In summary,

Finally, to make the equation a function of total flying capacitance, *Ctot*,

When $C2 = 0$ $(x = 0)$,

$$C1 = \frac{Ctot}{2(N-1)} \tag{A.143}$$

Thus,

$$\varDelta_1 = \frac{IoTs}{2\frac{Ctot}{2(N-1)}}\left(\frac{N-2}{N^2(N+2)}\right) = \frac{IoTs}{2Ctot}(\frac{2(N-1)\times(N-2)}{N^2(N+2)}) \approx \frac{IoTs}{2Ctot}(\frac{2}{N}) \text{ for large N} \tag{A.144}$$

When $C2$ is very large $(x = \infty)$,

Thus,

$$\varDelta_1 = \frac{IoTs}{2\frac{Ctot}{2(N-1)}}\left(\frac{N-1}{N^2}\right) = \frac{IoTs}{2Ctot}(\frac{2(N-1)\times(N-1)}{N^2}) \approx \frac{IoTs}{2Ctot}(2) \text{ for large N} \tag{A.145}$$

A.4 MATLAB Simulation Methodology for SAR SC Converter

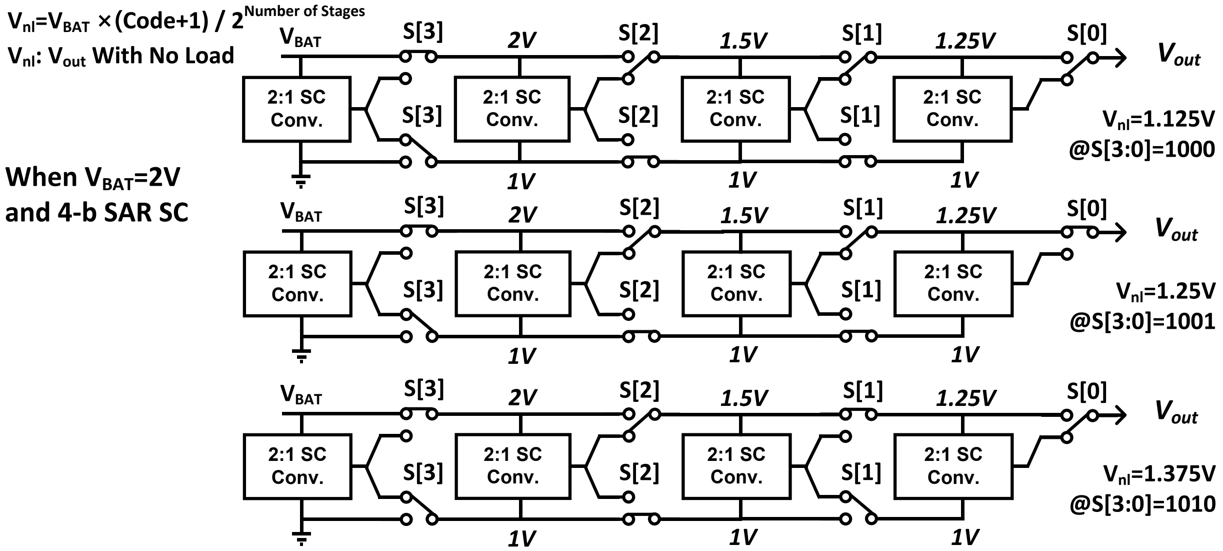


Figure A.5. Successive-Approximation (SAR) switched-capacitor: 4-b operation examples.

Equations for voltage drop due to conduction loss were formulated, but not simplified in a closed form. With specific values (k, N, Io, Ts, C1, C2) plugged in, MATLAB can calculate conduction loss, which is consistent with spice simulation result.

To find out expression for $V_{o,AV}$, we need to first find (1) current flow ($I_{dc(i)}$ and $I_{f(i)}$), and (2) instantaneous charge flow (Q_x, $Q_{dc(i)}$ and $Q_{f(i)}$) during phase transition in all flying capacitor and output capacitor. Then, we can find voltage across each capacitor, then output voltage. In order to find $I_{dc(i)}$ and $I_{f(i)}$, we need to find $I_{t(i)}$ and I_x as a function I_o as shown Figure A.6. Using loop analysis with KVL, we can build a following matrix multiplication equation, and solve it for $I_{t(i)}$ ($i = 2 \sim N$) and I_x. (N is the number of stages in SAR SC converter)

$$A \cdot I_M = 0, \tag{A.146}$$

$where\ A =$

$$\begin{bmatrix}
-\frac{1}{2C_1}-\frac{1}{C_1}L_1 & \frac{5}{2C_1} & -\frac{1}{C_1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
-\frac{1}{C_1}L_2 & -\frac{1}{C_1} & \frac{3}{C_1} & -\frac{1}{C_1} & 0 & \cdots & 0 & 0 & 0 & 0 \\
-\frac{1}{C_1}L_3 & 0 & -\frac{1}{C_1} & \frac{3}{C_1} & -\frac{1}{C_1} & \cdots & 0 & 0 & 0 & 0 \\
-\frac{1}{C_1}L_4 & 0 & 0 & -\frac{1}{C_1} & \frac{3}{C_1} & \cdots & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
-\frac{1}{C_1}L_{N-1} & 0 & 0 & 0 & 0 & \cdots & -\frac{1}{C_1} & \frac{3}{C_1} & -\frac{1}{C_1} & 0 \\
-\frac{1}{C_1}-\frac{1}{C_1}L_N & 0 & 0 & 0 & 0 & \cdots & 0 & -\frac{1}{C_1} & \frac{3}{C_1} & 0 \\
-\frac{3}{2C_1}-\frac{1}{C_1}\sum_{i=1}^{N}L_i+\frac{1}{C_2} & -\frac{1}{2C_1}-\frac{1}{C_1}L_1 & -\frac{1}{C_1}L_2 & -\frac{1}{C_1}L_3 & -\frac{1}{C_1}L_4 & \cdots & -\frac{1}{C_1}L_{N-2} & -\frac{1}{C_1}L_{N-1} & -\frac{1}{C_1}-\frac{1}{C_1}L_N & -\frac{1}{C_2}
\end{bmatrix}$$

$and\ I_M = \begin{pmatrix} I_x \\ I_{t2} \\ I_{t3} \\ \vdots \\ I_o \end{pmatrix}$. Here, for $i = 1 \sim N$, L_i is defined as S[N-i], for convenience.

Once $I_{t(i)}$ and I_x are found as a function I_o with the matrix equation above, we can find $I_{dc(i)}$ and $I_{f(i)}$ as a function of I_o, as following.

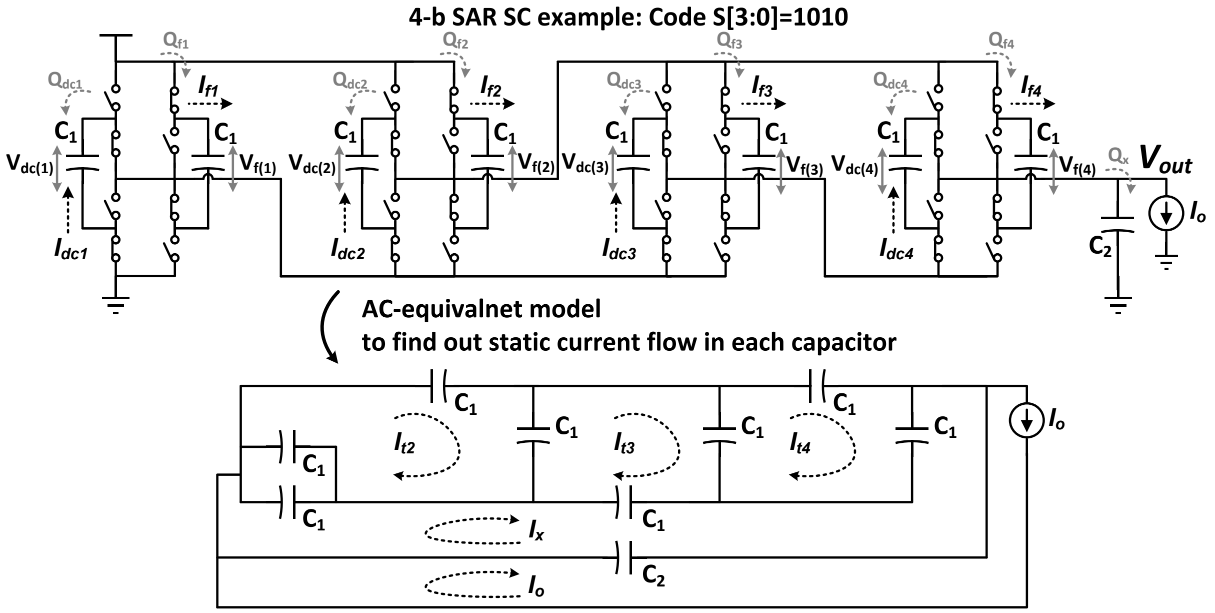


Figure A.6. 4-b SAR SC configuration and AC-equivalent model when S[3:0] = 1010.

If $L_N = 1$;

$$I_{f(N)} = -I_x + I_{tN} \tag{A.147}$$

$$I_{dc(N)} = I_x - I_{tN} \tag{A.148}$$

If $L_N = 0$;

$$I_{f(N)} = I_{tN} \tag{A.149}$$

$$I_{dc(N)} = I_x - I_{tN} \tag{A.150}$$

If $L_1 = 1$;

$$I_{f(1)} = \frac{1}{2}(I_x - I_{t2}) \tag{A.151}$$

$$I_{dc(1)} = \frac{1}{2}(I_x - I_{t2}) \tag{A.152}$$

If $L_1 = 0$;

$$I_{f(1)} = \frac{1}{2}I_{t2} \tag{A.153}$$

$$I_{dc(1)} = \frac{1}{2}I_{t2} \tag{A.154}$$

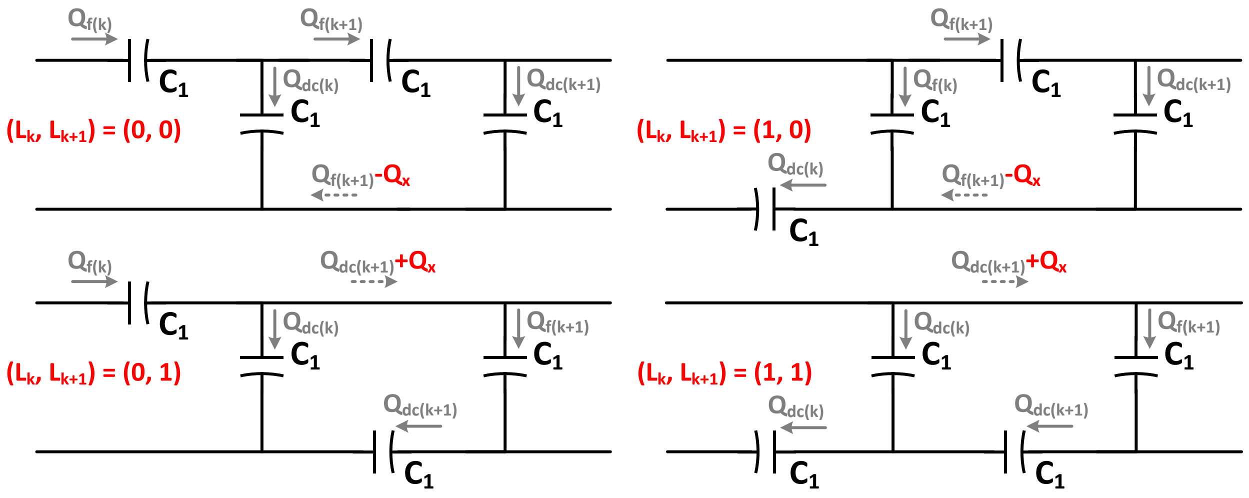


Figure A.7. Instantaneous charge flow, depending on configuration of two consecutive stages.

If $L_k = 1$ ($k \neq 1$ *or* N);

$$I_{f(N)} = I_{t(k)} \tag{A.155}$$

$$I_{dc(N)} = -I_{t(k)} + I_{t(k+1)} \tag{A.156}$$

If $L_k = 0$ ($k \neq 1$ *or* N);

$$I_{f(N)} = I_{t(k)} - I_{t(k+1)} \tag{A.157}$$

$$I_{dc(N)} = I_x - I_{t(k)} \tag{A.158}$$

Based on $I_{dc(i)}$ and $I_{f(i)}$, instantaneous charge flow (Q_x, $Q_{dc(i)}$ and $Q_{f(i)}$) during phase transition can be found as following. Refer to Figure A.7 for better understanding of how Q_x is involved.
Instantaneous charge flow into output capacitor (Q_x);

$$Q_x = \frac{(I_o - I_x)T_S}{2} \tag{A.159}$$

If $L_N = 1$;

$$Q_{f(N)} = \frac{T_S}{4}\left(-I_{f(N)} + I_{dc(N)}\right) + \frac{Q_x}{2} \tag{A.160}$$

$$Q_{dc(N)} = \frac{T_S}{4}\left(-I_{f(N)} + I_{dc(N)}\right) - \frac{Q_x}{2} \tag{A.161}$$

If $L_N = 0$;

$$Q_{f(N)} = \frac{T_S}{4}\left(-I_{f(N)} + I_{dc(N)}\right) \tag{A.162}$$

$$Q_{dc(N)} = \frac{T_S}{4}\left(-I_{f(N)} + I_{dc(N)}\right) \tag{A.163}$$

For $k \neq N$,

If $(L_k, L_{k+1}) = (0, 0)$;

$$Q_{f(k)} = \frac{1}{2}Q_{f(k+1)} + \frac{T_S}{4}\left(-I_{f(k)} + I_{dc(k)}\right) \tag{A.164}$$

$$Q_{dc(N)} = Q_{f(k)} - Q_{f(k+1)} \tag{A.165}$$

If $(L_k, L_{k+1}) = (0, 1)$;

$$Q_{f(k)} = \frac{1}{2}Q_{dc(k+1)} + \frac{T_s}{4}\left(-I_{f(k)} + I_{dc(k)}\right) + \frac{Q_x}{2} \tag{A.166}$$

$$Q_{dc(N)} = Q_{f(k)} - Q_{dc(k+1)} - Q_x \tag{A.167}$$

If (L_k, L_{k+1}) = (1, 0);

$$Q_{f(k)} = -\frac{1}{2}Q_{f(k+1)} + \frac{T_s}{4}\left(-I_{f(k)} + I_{dc(k)}\right) + \frac{Q_x}{2} \tag{A.168}$$

$$Q_{dc(N)} = Q_{f(k)} + Q_{f(k+1)} - Q_x \tag{A.169}$$

If (L_k, L_{k+1}) = (1, 1);

$$Q_{f(k)} = -\frac{1}{2}Q_{dc(k+1)} + \frac{T_s}{4}\left(-I_{f(k)} + I_{dc(k)}\right) \tag{A.170}$$

$$Q_{dc(N)} = Q_{f(k)} + Q_{dc(k+1)} \tag{A.171}$$

Now, we are ready to calculate $V_{dc(k)}$ and $V_{f(k)}$, based on current flow ($I_{dc(i)}$ and $I_{f(i)}$), and instantaneous charge flow ($Q_{dc(i)}$ and $Q_{f(i)}$) during phase transition in all flying capacitor.

$$V_{dc(1)}(A^s) = \frac{1}{2}V_{in} + \frac{1}{2}\left(\frac{I_{f(k)}}{C_1}\frac{T_s}{2} + \frac{Q_{dc(k)}}{C_1}\right) \tag{A.172}$$

$$V_{f(1)}(A^s) = V_{in} - V_{dc(1)}(A^s) \tag{A.173}$$

$$V_{dc(1)}(AV) = V_{dc(1)}(A^s) - \frac{1}{2}\left(\frac{I_{dc(k)}}{C_1}\frac{T_s}{2}\right) \tag{A.174}$$

$$V_{f(1)}(AV) = V_{dc(1)}(A^s) + \frac{1}{2}\left(\frac{I_{f(k)}}{C_1}\frac{T_s}{2}\right) \tag{A.175}$$

For k ≠ N,

If $L_{k-1} = 1$;

$$V_{dc(k)}(A^s) = \frac{1}{2}V_{f(k-1)}(A^s) + \frac{1}{2}\left(\frac{I_{f(k)}}{C_1}\frac{T_s}{2} + \frac{Q_{dc(k)}}{C_1}\right) \tag{A.176}$$

$$V_{f(k)}(A^s) = V_{f(k-1)}(A^s) - V_{dc(k)}(A^s) \tag{A.177}$$

$$V_{dc(k)}(AV) = V_{dc(k)}(A^s) - \frac{1}{2}\left(\frac{I_{dc(k)}}{C_1}\frac{T_s}{2}\right) \tag{A.178}$$

$$V_{f(k)}(AV) = V_{dc(k)}(A^s) + \frac{1}{2}\left(\frac{I_{f(k)}}{C_1}\frac{T_s}{2}\right) \tag{A.179}$$

If $L_{k-1} = 0$;

$$V_{dc(k)}(A^s) = \frac{1}{2}V_{dc(k-1)}(A^s) + \frac{1}{2}\left(\frac{I_{f(k)}}{C_1}\frac{T_s}{2} + \frac{Q_{dc(k)}}{C_1}\right) \tag{A.180}$$

$$V_{f(k)}(A^s) = V_{dc(k-1)}(A^s) - V_{dc(k)}(A^s) \tag{A.181}$$

$$V_{dc(k)}(AV) = V_{dc(k)}(A^s) - \frac{1}{2}\left(\frac{I_{dc(k)}}{C_1}\frac{T_S}{2}\right) \tag{A.182}$$

$$V_{f(k)}(AV) = V_{dc(k)}(A^s) + \frac{1}{2}\left(\frac{I_{f(k)}}{C_1}\frac{T_S}{2}\right) \tag{A.183}$$

Finally, we can find $V_{out}(AV)$ as following.

$$V_{out}(AV) = V_{dc(1)}(AV) + \sum_{i=1}^{N}[L_i\{V_{dc(i)}(AV) + V_{f(i)}(AV)\} + \overline{L_\iota}V_{dc(i)}(AV)] \tag{A.130}$$

where N is the number of stages in SAR SC, and L_i is defined as S[N-i], for $i = 1 \sim N$, and $\overline{L_\iota} = 1 - L_i$.

A.5 Summary Tables

Topology	$R_{SSL,k}$ (C2 = 0)
Series-Parallel	$\frac{Ts}{2Ctot}(\frac{2\,k^2(N-k)^2(N-2k)^2}{2N^2(k^2+(N-k)^2)})$
Conv. Lad.	$\frac{Ts}{2Ctot}(\frac{2(N-1)\times(4N^3k^2-8N^2k^3+8N^2k^2-N^2k-3N^2+4Nk^4-16Nk^3+Nk^2+4Nk-3N+8k^4-4k^2)}{12N(N+2)})$
Var. Lad. (k ≠ 1)	$\frac{Ts}{2Ctot}(\frac{2(N-1)\times(4N^3k-5N^3-4N^2k^2+16N^2k-11N^2-12Nk^2+16Nk-8k^2)}{4N^2(N+2)})$
Var. Lad. (***k* = *1***)	$\frac{Ts}{2Ctot}(\frac{2(N-1)\times(N-2)}{N^2(N+2)})$
SAR	$\propto log_2\,N$

Table A.2. Slow switching impedance (R_{SSL}) in various topologies when output capacitance (C2) is zero.

Topology	$R_{SSL,k}$ (C2 = ∞)
Series-Parallel	$\frac{Ts}{2Ctot}(\frac{2k^2(N-k)^2}{N^2})$
Conv. Lad.	$\frac{Ts}{2Ctot}(\frac{2(N-1)\times k(N-k)(-2k^2+2Nk+1)}{6N})$
Var. Lad. (k ≠ 1)	$\frac{Ts}{2Ctot}(\frac{2(N-1)\times(N-k)(k-N+Nk)}{N^2})$
Var. Lad. (***k* = *1***)	$\frac{Ts}{2Ctot}(\frac{2(N-1)\times(N-1)}{N^2})$
SAR	$\propto log_2\,N$

Table A.3. Slow switching impedance (R_{SSL}) in various topologies when output capacitance (C2) is very large (∞).

BIBLIOGRAPHY

[1] K. Ashton, “That ‘Internet of Things’ Thing,” *RFID Journal*, Jun. 2009.

[2] D. Blaauw, et al., "IoT design space challenges: Circuits and systems," in *IEEE Symp. VLSI Circuits* Dig., Jun. 2014, pp.1–2.

[3] http://store.apple.com/us/buy-watch/apple-watch

[4] https://www.fitbit.com/surge

[5] http://www.technologyreview.com/news/537596/internet-of-farm-things/

[6] http://www.technologyreview.com/news/521811/the-internet-of-things-unplugged-and-untethered/ and http://techcrunch.com/2015/02/25/iotera-1-million-seed/

[7] B. Zhai, et al., “Theoretical and Practical Limits of Dynamic Voltage Scaling,” July 2004, in *IEEE Design Automation Conference (DAC)* Dig. , pp. 868-873.

[8] B. Zhai, et al., “A 2.60pJ/Inst Subthreshold Sensor Processor for Optimal Energy Efficiency,” in *IEEE Symp. VLSI Circuits* Dig., Jun. 2006, pp.154–155.

[9] G. Chen, et al., “Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2011, pp. 288-289.

[10] G. Chen, et al., “A cubic-millimeter energy-autonomous wireless intraocular pressure monitor,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2010, pp. 310-312.

[11] Y. Lee, et al., “A modular 1mm^3 die-stacked sensing platform with optical communication and multi-modal energy harvesting,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2012, pp. 402-403.

[12] Y. Lee, et al., “A Modular 1 mm^3 die-stacked sensing platform with low power I^{2}C inter-die communication and multi-modal energy harvesting,” *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 229-242, Jan. 2013.

[13] M. D. Seeman., et al., “A comparative analysis of switched-capacitor and inductor-based DC-DC conversion technologies,” in *IEEE Workshop on Control and Modeling for Power Electronics (COMPEL)*, June 2010, pp. 1-7.

[14] S. R. Sanders, et al., “The road to fully integrated DC-DC conversion via the switched-capacitor approach,” *IEEE Transactions on Power Electronics (TPEL)*, vol. 28, no. 9, pp. 4146-4155, Sept. 2013.

[15] C. Huang and P. K. T. Mok, “A 100 MHz 82.4% efficiency package bondwire based four-phase fully-integrated buck converter with flying capacitor for area reduction,” *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 2977–2988, Dec. 2013.

[16] J. De Vos, and et al., “Switched-capacitor DC/DC converters for empowering Internet-of-things SoCs,” in *IEEE Faible Tension Faible Consommation (FTFC)*, May 2014, pp. 1-4.

[17] J. Xiao, and et al., “An ultra-low-power digitally controlled buck converter IC for cellular phone applications,” in *IEEE Applied Power Electronics Conference and Exposition*, 2004, pp. 383-391.

[18] X. Zhang, et al., "A 0.45-V input on-chip gate boosted (OGB) buck converter in 40-nm CMOS with more than 90% efficiency in load range from 2μW to 50μW," in *IEEE Symp. VLSI Circuits* Dig., Jun. 2012, pp.194-195.

[19] G. Villar-Pique, et al., "Survey and benchmark of fully integrated switching power converters: switched-capacitor versus inductive appraoch," *IEEE Transactions on Power Electronics*, vol. 28, no. 9, pp. 4156-4167, Sept. 2013.

[20] S. Bandyopadhyay, and A.P. Chandrakasan, "Platform architecture for solar, thermal and vibration energy combining with MPPT and single inductor," in *IEEE Symp. VLSI Circuits* Dig., June 2011, pp.238-239.

[21] I. Lee, et al., "A 635pW battery voltage supervisory circuit for miniature sensor nodes,"in *IEEE Symp. VLSI Circuits* Dig., June 2012, pp.202-203.

[22] J. Klassen, "A description of Cymbet battery technology and its comparison with other battery technologies," http://www.cymbet.com/

[23] Cymbet EnerChip CBD012 Datasheet: http://www.cymbet.com/pdfs/DS-72-02.pdf

[24] M. Alioto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: A Tutorial," *IEEE Trans. on Circuits and Systems – part I (invited)*, vol. 59, no. 1, pp. 3-29, Jan. 2012.

[25] Chandrakasan, et al., "Low-power CMOS digital design", *IEEE J. Solid-State Circuit*, vol. 27, pp. 473-484, Apr. 1992.

[26] Y. Ye, et al., "A New technique for standby leakage reduction in high-performance circuits," in *IEEE Symp. VLSI Circuits* Dig., June 1998, pp. 40-41.

[27] A. Keshavarzi, et al., "Effectiveness of reverse body bias for leakage control in scaled dual Vt CMOS ICs," *in Proc. Int. Symp. Low Power Electronic Design (ISLPED)*, Aug. 2001, pp. 207–212.

[28] J. W. Tschnz, et al., "Dynamic sleep transistor and body bias for active leakage power control of microprocessors," *IEEE J. Solid-State Circuits*, vol. 38, no. 11, Nov. 2003, pp. 1838-1845.

[29] S. Narenda, et al., "Full-chip subthreshold leakage power prediction and reduction techniques for Sub-0.18-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 39, no. 2, Feb. 2004, pp. 501-510.

[30] BSIM4 MOSFET model manual, http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf

[31] M. Wieckowski, et al., "A hybrid DC-DC converter for sub-microwatt Sub-1V implantable applications," in *IEEE Symp. VLSI Circuits* Dig., Jun. 2009, pp.166–167.

[32] Y. K. Ramadass and A. P. Chandrakasan, "Voltage scalable switched capacitor DC-DC converter for ultra-low-power on-chip applications," in *IEEE Power Electronics Specialists Conf.*, 2007, pp. 2353–2359.

[33] Y. K. Ramadass, et al., "A fully-integrated switched-capacitor step-down DC-DC converter with digital capacitance modulation in 45 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 45, no. 12, pp. 2557–2565, Dec. 2010.

[34] T. M. V. Breussegem and M. S. J. Steyaert, "Monolithic capacitive DC-DC converter with single boundary-multiphase control and voltage domain stacking in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1715–1727, Jul. 2011.

[35] D. El-Damak, et al., "A 93% efficiency reconfigurable switched-capacitor DC-DC converter using on-chip ferroelectric capacitors," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2013, pp. 374–375.

[36] T. M. Andersen, et al., “A sub-ns response on-chip switched-capacitor DC-DC voltage regulator delivering 3.7W/mm at 90% efficiency using deep-trench capacitors in 32 nm SOI CMOS,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 90–91.

[37] R. Jain, et al., “A 0.45–1 V fully-integrated distributed switched capacitor DC-DC converter with high density MIM capacitor in 22 nm trigate CMOS,” *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 917–927, Apr. 2014.

[38] H.-P. Le, et al., “A sub-ns response fully integrated battery-connected switched-capacitor voltage regulator delivering 0.19 W/mm at 73% efficiency,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2013, pp. 372–373.

[39] V. Ng, and S. Sanders, “A 92%-efficiency wide-input-voltage-range switched-capacitor DC-DC converter,” in *IEEE ISSCC Dig. Tech. Papers,* Feb. 2012, pp. 282-283.

[40] M. D. Seeman and S. R. Sanders, “Analysis and optimization of switched-capacitor DC-DC converters,” *IEEE Trans. on Power Electronics*, vol. 23, no. 2, pp. 841–851, Mar. 2008.

[41] S. Bang, et al, “A fully integrated successive-approximation switched-capacitor DC-DC converter with 31mV output voltage resolution,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2013, pp. 370-371.

[42] Y.-P. Chen, et al., "A 2.98nW bandgap voltage reference using a self-tuning low leakage sample and hold," in *IEEE Symp. VLSI Circuits Dig*., Jun. 2012, pp.200–201.

[43] M. Seok, et al., “A portable 2-transistor picowatt temperature-compensated voltage reference operating at 0.5V,” *IEEE J. Solid-State Circuits*, vol. 47, no. 10, pp. 2534-2545, Oct. 2012.

[44] L. G. Salem and P. P. Mercier, “An 85%-efficiency fully integrated 15-ratio recursive switched-capacitor DC-DC converter with 0.1-to-2.2V output voltage range,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 88–89.

[45] L. G. Salem and P. P. Mercier, “A 45-ratio recursively sliced series-parallel switched-capacitor DC-DC converter achieving 86% efficiency,” in *IEEE Custom Integrated Circuits Conference (CICC)*, Sep. 2014.

[46] T. Burd, et al., “A dynamic voltage scaled microprocessor system,” *IEEE J. Solid-State Circuit*, vol. 35, no. 11, pp. 1571-1580, Nov. 2000.

[47] R. Jain, et al., “A 0.45-1V fully integrated reconfigurable switched capacitor step-down DC-DC converter with high density MIM capacitor in 22nm tri-Gate CMOS,” in *IEEE Symp. VLSI Circuits* Dig, Jun. 2013, pp. 174-175.

[48] T. M. Andersen, et al., “A sub-ns response on-chip switched-capacitor DC-DC voltage regulator delivering 3.7W/mm^2 at 90% efficiency using deep-trench capacitors in 32nm SOI CMOS,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 90-91.

[49] S. S. Kudva, et al., “Fully integrated capacitive DC-DC converter with all-digital ripple mitigation technique,” *IEEE J. Solid-State Circuit*, vol. 48, no. 9, pp. 1920-1920, Aug. 2013.

[50] R. Jain, et al., “Conductance modulation techniques in switched-capacitor DC-DC converter for maximum-efficiency tracking and ripple mitigation in 22nm tri-gate CMOS,” in *IEEE Custom Integrated Circuits Conference (CICC)*, Sep. 2014.

[51] G. V. Pique, “A 41-phase switched-capacitor power converter with 3.8mV output ripple and 81% efficiency in baseline 90nm CMOS,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2012, pp. 98-100.

[52] K. Yang, et al., “A 23Mb/s 23pJ/b fully synthesized true-random-number generator in 28nm and 65nm CMOS,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 280-281.

[53] E. Alon, et al., “Integrated regulation for energy-efficient digital circuits,” *IEEE J. Solid-State Circuit*, vol. 43, no. 8, pp. 1795-1807, Aug. 2008.

[54] H.-P. Le, et al., “Design techniques for fully integrated switched-capacitor DC-DC converters,” *IEEE J. Solid-State Circuit*, vol. 46, no. 9, pp. 2120-2131, Sep. 2011.

[55] W. Kim, et al., “System level analysis of fast, per-core DVFS using on-chip switching regulators” in *IEEE Internaional Symp. HPCA*, Feb. 2008, pp. 123-134.

[56] P. Harpe, E. Cantatore, and A. van Roermund, “An oversampled 12/14b SAR ADC with noise reduction and linearity enhancements achieving up to 79.1dB SNDR,” in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 194-195.

[57] S. Bang, J. Seo, I. Lee, S. Jeong, N. Pinckney, D. Blaauw, D. Sylvester, and L. Chang, “A fully-integrated 40-phase flying-capacitance-dithered switched-capacitor voltage regulator with 6mV output ripple,” *IEEE Symp. VLSI Circuits Dig*, June 2015, pp. 336-337.

[58] M. Wieckowski, et al., “A hybrid DC-DC converter for sub-microwatt sub-1V implantable applications*,” IEEE Symp. VLSI Circuits Dig*, June 2009, pp. 166-167.

[59] N. D. Dalt, “A design-oriented study of the nonlinear dynamics of digital bang-bang PLLs,” *IEEE Trans. Circuits and Systems – I*, vol. 52, no. 1, pp. 21-31, Jan. 2005.

[60] S. Bang, et al, "A fully integrated switched-capacitor based PMU with adaptive energy harvesting technique for ultra-low power sensing applications," *IEEE International Symp. on Circuits and Systems (ISCAS),* May 2013, pp. 709-712.

[61] S. Bang, et al, "Reconfigurable sleep transistor for GIDL reduction in ultra-low standby power systems," *IEEE Custom Integrated Circuits Conference (CICC)*, September 2012, pp. 1-4.

[62] S. Bang, et al, “A successive-approximation switched-capacitor DC-DC converter with resolution of $V_{IN}/2^N$ for a wide range of input and output voltages,” *IEEE J. Solid-State Circuit*, vol. 51, no. 2, pp. 543-556, February 2016.

[63] S. Bang, et al, “A low ripple switched-capacitor voltage regulator using flying capacitance dithering,” *IEEE J. Solid-State Circuit* (to be published in 2016).

[64] C. G. Bell, et al., "The Effect of Technology on Near Term Computer Structures" Computer 2 (5), pp. 29-38, March/April 1972.

[65] Mark Horowitz, “Computing’s Energy Problem,” *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 10-14.

[66] H. Esmaeilzadeh, et al., “Dark Silicon and the End of Multicore Scaling,” *ACM International Symposium on Computer Architecture (ISCA)*, June 2011.

[67] N. D. Lane, et al., “Can deep learning revolutionize mobile sensing?” *ACM International Workshop on Mobil Computing Systems and Applications (HotMobile)*, 2015, pp. 117-122.

[68] T. Chen, et al., “DianNao: a small-footprint high-throughput accelerator for ubiquitous machine-learning, *ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, March 2014, pp. 269-284.

[69] Y. Chen, et al., “DaDianNao: a machine-learning supercomputer,” *IEEE/ACM International Symp. Microarchitecture*, Dec. 2014, pp. 609-622.

[70] Z. Du, et al., “ShiDianNao: shifting vision processing closer to the sensor,” *ACM International Symp. On Computre Architecture (ISCA)*, June 2015, pp. 92-104.

[71] M. Shah, et al., “A fixed-point neural network for keyword detection on resource constrained hardware,” *IEEE Workshop on Signal Processing Systems (SIPS)*, Oct. 2015, pp. 1-6.

[72] M. Tanaka et al, “A novel inference of a restricted boltzmann machine,” in *IEEE International Conference on Pattern Recognition (ICPR)*, Aug 2014, pp. 1526–1531.

[73] G. Kim, et al., “A millimeter-scale wireless imaging system with continuous motion detection and energy harvesting*,” IEEE Symp. VLSI Circuits Dig*, June 2014, pp. 1-2.

[74] P. Pannuto, et al., “MBus: An ultra-low power interconnect bus for next generation nanopower systems,” *ACM International Symp. On Computre Architecture (ISCA)*, June 2015, pp. 629-641.

[75] Y.-s. Kuo, et al, " MBus: A 17.5 pJ/bit/chip portable interconnect bus for millimeter-scale sensor systems with 8 nW standby power," *IEEE Custom Integrated Circuits Conference (CICC)*, September 2014, pp. 1-4.