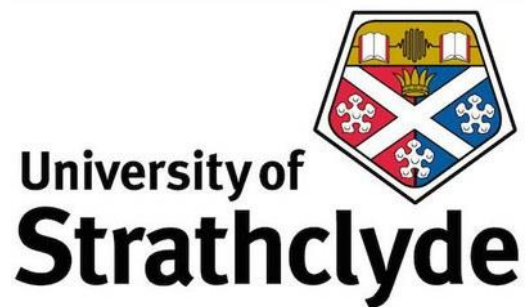**UNIVERSITY OF STRATHCLYDE**

Faculty of Science

Design and implementation of high linearity FPGA-TDCs and an integrated large scale TCSPC system for time-resolved applications

by

Haochang Chen

A thesis submitted for the degree of Doctor of Philosophy

August 2019

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for the examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as quailed by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: *HC Chen*   Date:2019-Aug-28

Abstract

The time-correlated single-photon counting (TCSPC) technology is a vital, advanced measurement and analytical tool for time-resolved biomedical, physics research and many industry areas because of its high temporal resolution and sensitivity. Analogue-based conventional TCSPC systems have been commercialised and applied in scientific experiments widely. However, the complicated system of conventional TCSPC equipment causes the bulky size, high cost, low conversion rate and limited channel number. With the recent rapid development of semiconductor technology, Field Programmable Gate Arrays (FPGA) become the promising platforms for high-performance digital TCSPC systems.

The time-to-digital converter (TDC) is the core component of a TCSPC system as it provides the temporal measurements with extremely high-resolution. For the scientific experiments, prototyping and high-end instruments, FPGA-based TDCs or TCSPC systems can provide excellent flexibility and compatibility with the much lower design and implementation costs. However, compared with ASIC and analogue implementations, the reported FPGA-TDCs have poor linearity performances with severe non-linearity problems such as missing-codes, ultra-wide bins and the bubbles problems. As a result, this study focuses on to improve the linearity performance by exploring the sources of non-linearity in the tapped delay line (TDL)-based FPGA-TDCs. This thesis proposes two novel FPGA-TDC designs to address the linearity drawbacks. The first TDC design proposes a combination architecture innovatively to restrain the differential non-linearity (DNL) to $<\pm 1$LSB (LSB = 10.5ps) with the complete removal of missing-codes. By developing and using a hardware-friendly bin-width calibration, the DNL has been reduced to $<\pm 0.1$LSB. Furthermore, getting benefits from the direct-histogram architecture, the proposed FPGA-TDC/TCSPC design has the capability of multi-event measurement and ultra-low dead-time (<100ps). The second TDC design proposes several new methods and architectures with minimised resource consumption for multiple-channel applications. The most significant contribution of this design is an ideal solution, the sub-TDL topology, for the 'bubble' problems has been proposed and verified. Next, the tap timing test method, which is derived from the sub-TDL topology, can provide more exact timing details

compared with the wide-used code density test. By applying the proposed compensation and calibration method, its linearity can be improved to $<\pm0.12LSB$ with 5ps temporal resolution. Two 96-channel TCSPC arrays have been implemented in different FPGA series which demonstrate the excellent potential for multi-channel designs.

Besides, an integrated, large-scale TCSPC array base on a world-leading SPAD sensor and TDC array is presented in this thesis for fast, wide-field, time-resolved applications. The system provides >24K independent TCSPC channels with both photon counting and time-correlated imaging mode and a tunable temporal resolution. For verification, this study applied the system in a typical fluorescence lifetime measurement. According to the CMM calculated results base on the measured data, the proposed system demonstrated the accurate and reliable measurement performances.

Acknowledgements

Contents

List of Figures

List of Tables

Abbreviations

ADC	**Analogue-to-Digital Converter**
ALICE	**A Large Ion Collider Experiment**
API	**Application Programmer's Interfaces**
ASIC	**Application-Specific Integrated Circuit**
BUFG	**Global clock buffer**
BRAM	**Block Random Access Memory**
CFD	**Constant-Fraction Discriminators**
CLB	**Configurable Logic Blocks**
CR	**Clock Region**
CM	**Centre-of-Mass**
CMM	**Centre-of-Mass Method**
CMOS	**Complementary Metal-Oxide-semiconductor**
CMT	**Clock Management Tile**
DAC	**Digital to Analog Converter**
DC-DC	**Direct Current to Direct Current**
DCR	**Dark Count Rate**
DDR	**Double Data Rate**
DPLL	**Digital Phase-Locked Loop**
DLL	**Delay-Locked Loop**
DNL	**Differential Non-Linearity**
DR	**Dynamic Range**
DSP	**Digital Signal Processor**
EPC	**Embedded Power Controller**
FLIM	**Fluorescence-Lifetime Imaging Microscopy**
FF	**Flip-Flop**
FIFO	**First In, First Out**
FIR	**Finite Impulse Response**

FPGA	Field Programmable Gate Arrays
FRET	Forster Resonant Energy Transfer
FMC	FPGA Mezzanine Card
FSR	Full-Scale Range
FSM	Finite State Machine
FWHM	Full-Width Half-Maximum
GUI	Graphical User Interface
HBUFG	Horizontal Clock Buffer
IC	Integrated Circuits
ICCD	Intensified Charge-Coupled Devices
ILA	Integrated Logic Analyzer
IOB	Input/Output Block
IP	Intellectual Property
INL	Integral Non-Linearity
IRF	Instrument Response Function
LDO	low Dropout Regulator
LIDAR	Light Detection And Ranging
LSB	Least Significant Bit
LUT	Look-Up Tables
MCB	Memory Controller Block
MR	Measurement Range
MMCM	Mixed-Mode Clock Manager
MUXCY/MUX	Multiplexer
NRE	Non-Recurring Engineering
PDE	Photon Detection Efficiency
PGA	pin grid array
PISO	parallel in serial out
PMT	Photomultiplier
PLL	Phase-Locked Loop
PVT	Process, Voltage and Temperature
RAM	Random Access Memory

RTL	Register-Transfer Level
PMIC	Power Management Integrated Circuit
RO	Ring Oscillators
SoC	System on a Chip
SPAD	Single-Photon Avalanche Diode
SPI	Serial Peripheral Interface
SNR	Signal to Noise Ratio
TI	Time Interval
TAC	Time-to-Amplitude Converter
TDC	Time-to-Digital Converter
TDL	Tapped Delay Line
TCSPC	Time-Correlated Single Photon Counting
ToF	Time-of-Flight
TM2OH	Thermometer code to One-hot-code
VDL	Vernier Delay Line
VHDL	VHSIC Hardware Description Language
VIO	virtual Input/Output
WU	Wave-Union
UART	Universal Asynchronous Receiver/Transmitter
USB	Universal Serial Bus
UTC	Coordinated Universal Time

Chapter 1: Introduction

1.1 Background

Time is one of the most critical physical quantified parameters, playing critical roles both in the industrial areas and many scientific research fields. Different from the coordinated universal time (UTC), which is characterised by astrometry [1], the time interval (TI) is a widely-used and pivotal parameter employed to describe the relative time information between two events [2]. The time interval meter (TIM) is a type of stopwatch with extremely high resolution and accuracy. It is designed to measure TIs between two electrical pulses or signal edges and to convert the measurement results to digital formats. The time-to-digital converter (TDC) is a typical TIM.

1.1.1 Time-correlated single-photon counting (TCSPC)

The time-correlated single-photon counting (TCSPC) is one of the most widely-used application approaches of the TIMs, especially for applications with low signal intensity, high repetition rates and high temporal resolution requirement [3]. TCSPC was first reported in 1984 by Desmond O'Connor and David Phillips and became the leading technology for fluorescence-lifetime imaging microscopy (FLIM) due to its high temporal resolution and sensitivity to light [4]. However, the research progress of TCSPC tended to be slow in the 80s and 90s, due to sluggish developments in electronic and laser techniques. With rapid developments of these techniques after the 90s, the advanced TCPSC technology emerged and has already become a multidimensional optical imaging technique with features such as fast measurement, high precision and excellent sensitivity. By using single-photon avalanche diode (SPAD) as detectors, the single-photon sensitivity can be achieved. First, compared with the photomultiplier (PMT), semiconductor detectors can achieve better quantum efficiency by using the internal photon effect [3]. Second, the avalanche effect is used in avalanche photodiodes (APDs) to obtain a stable gain from 10^2 to 10^3. Then, the SPADs work in the Geiger mode by setting the diode

reverse voltage exceed their breakdown voltage to achieve a higher gain and single-photon detection capability [3, 5]. Finally, a passive or active quenching mechanism is used to avoid damage and reset SPADs [6].

The basic principle of TCSPC is building a histogram based on measured timestamps or time-of-flight (ToF) of detected photons, shown in **Figure 1.1**. For many applications of TCSPC, dark environment, low light intensity and the limitations of photon detectors make the photon detection rate far less than one and contain various types of noise. Therefore, it is necessary to measure and recode a large number of photon events to build a histogram which contains sufficient information for the reconstruction of time-domain optical waveforms [7, 8].
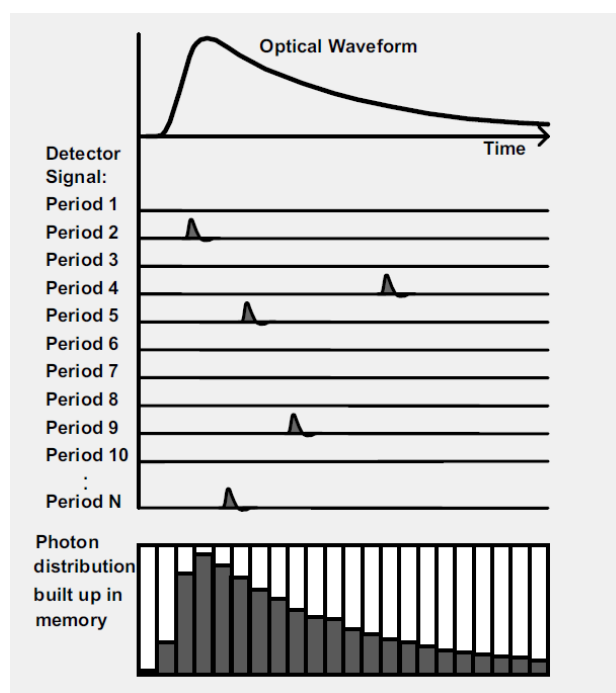
FIGURE 1.1 GENERAL PRINCIPLE OF TCSPC [3]

A traditional TCSPC system setup is shown in **Figure 1.2.** A photon detector generates a pulse signal once a photon is detected. The selection of detectors depends on the requirement of applications; PMTs and SPADs are two mainstream photon detectors for TCSPC systems. For detectors with low voltage outputs, signal amplifiers and constant-fraction discriminators (CFD) [9] are necessary to generate compatible, low-jitter pulse signals. A time-to-amplitude converter

(TAC) is used as the TIM device, and amplifiers and analogue-to-digital converters (ADC) are applied to convert measured results to a digital form as timestamps. Finally, the measured timestamps are used as the addresses of the memory to build up a histogram for further data processing.
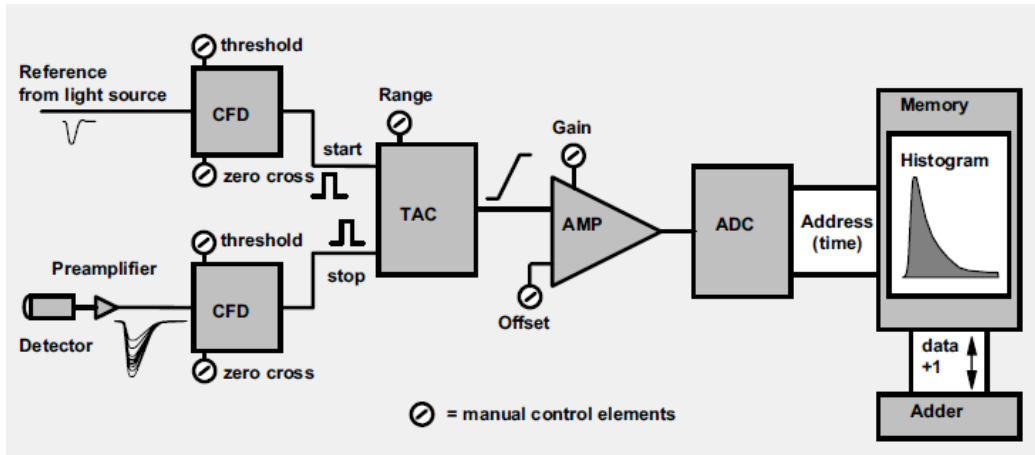
FIGURE 1.2 THE CLASSICAL TCSPC SYSTEM [3]

Various methods have been developed for multidimensional TCSPC. Considering low detection rates in many applications, multiple channel signals can be measured simultaneously, such as multidetector TCSPCs [10], multiplexed-TCSPCs, sequential recording techniques [11], and various scanning techniques have been presented. However, traditional TCSPC systems tended to have a sophisticated electronic system and the limitations of bulky size and high price.

By using full digital TIMs such as TDCs, a highly integrated TCSPC system can be implemented in an application specific integrated circuit (ASIC) or field-programmable gate array (FPGA) devices. This histogram memory can be easily implemented and connected with TDCs in ASIC or FPGA devices. Many publications show that the highly integrated TCSPC system has become one of the frontier researches subjects [12-14].
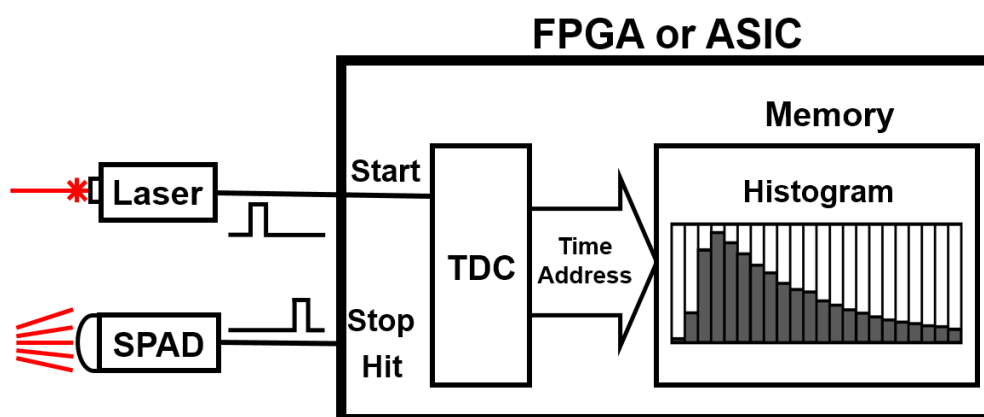
FIGURE 1.3 THE DIGITAL TDC BASED DIGITAL TCSPC SYSTEM

1.1.2 TDC

In the scientific research fields, one of the main TIM devices is time-to-digital converters (TDCs), which can be considered as a high-resolution stopwatch [3]. The TDC is a pivotal module for digital TCSPC systems and other time-domain applications. A block diagram of a basic TDC is shown in **Figure 1.4**. It is capable of measuring fast events with extremely high temporal resolution and generating digital outputs. The TIs to be measured are represented by pulse signals on "START" and "STOP" ports. If measurements are periodical, these two ports can be renamed as the "Hit" signal and the sampling clock.
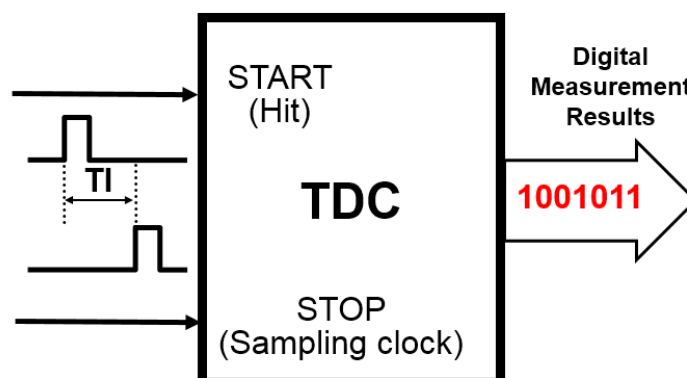
FIGURE 1.4 THE BASIC DIAGRAM OF A TYPICAL TDC

High-precision TDCs with a temporal resolution of picoseconds are vital for modern industrial and scientific progress such as laser ranging with millimetre precision or fluorescence lifetime measurements in biomedical applications. Conventional and commercial TCPSC devices mainly use analogue-based TIMs due to their excellent resolution and linearity performance compared with digital TDCs. In the 1990s, analogue TIMs achieved better than 10ps temporal resolution [15, 16]. However, the resolution of most full digital TDCs was limited to several tens of picosecond [17-19] in the 2000s. In recent years, the interest in full digital TDCs has increased rapidly, since the resolution and linearity have been greatly improved due to the development of semiconductor and manufacturing processes [20-22].

1.2   Application fields

With the benefit of extremely high temporal resolution, TDCs are widely used in different scientific research and industrial areas, such as biomedical, physics, quantum communications and metrology. **Figure 1.5** demonstrates the key applications of TDCs and TCSPCs.
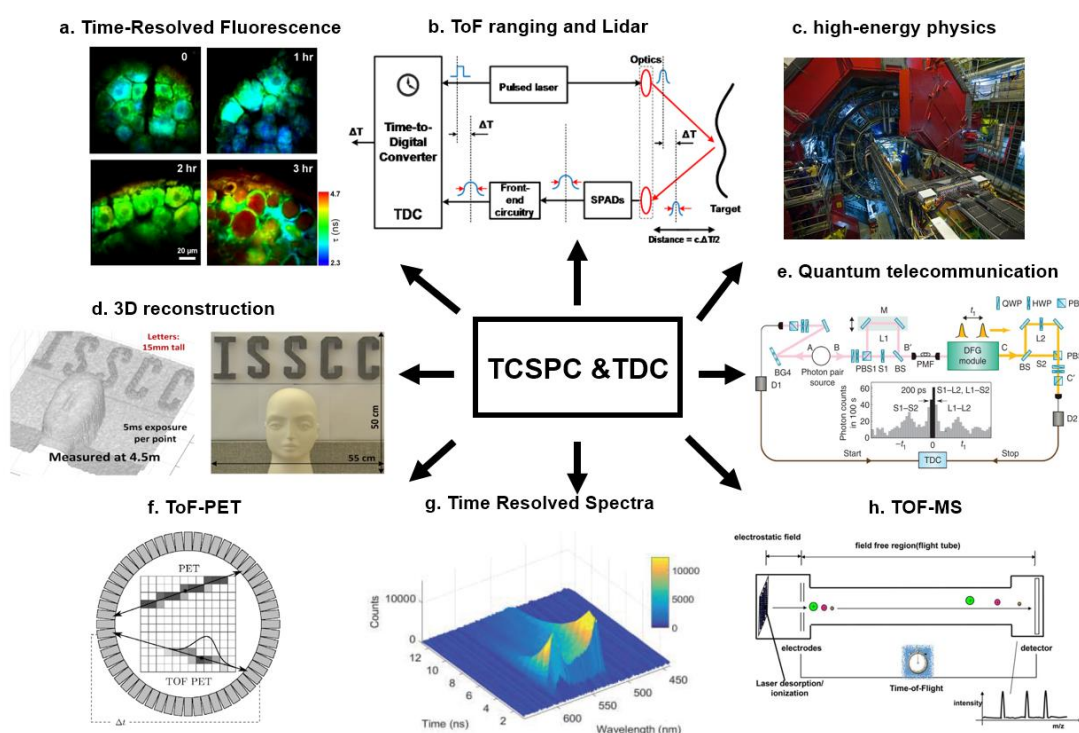
FIGURE 1.5 (A) FLUORESCENCE LIFETIME IMAGING MICROSCOPY (FLIM) [23], (B) SIMPLIFIED BLOCK DIAGRAM OF A D-TOF LIDAR SYSTEM, (C) A LARGE ION COLLIDER EXPERIMENT

(ALICE) EXPERIMENT [24], (D) 3D RECONSTRUCTION OF MULTIPLE TARGETS USING THE DUAL-AXIS SCANNER [25], (E) EXPERIMENTS FOR QUANTUM TELECOMMUNICATION [26], (F). DIAGRAM OF TOF-PET AND CONVENTIONAL PET PRINCIPLE [27], (G). TIME-RESOLVED 3-D RAMAN SPECTRUM [14], (H) MALDI ANALYSES WITH A TIME OF FLIGHT MASS ANALYSER [28].

a) Biomedical, disease diagnosis, biochemistry, and materials analysis

The time-resolved fluorescence technique is one of the leading imaging techniques in biochemistry, biophysics, clinical chemistry and genetic analysis. Fluorescence will decay exponentially after excitation. The lifetime of fluorescence can be defined as the time interval reduced to $1/e$ of the maximum fluorescence intensity [27][29]. Different from fluorescence intensity, the lifetime of fluorescence is a relatively absolute and stable parameter which cannot be affected by the intensity of excitation light, fluorophore concentration and photobleaching. Fluorescence lifetime is valuable when it comes to measuring various biochemistry and physics parameters including oxygen, ionic concentration, hydrophobicity and pH value. Fluorescence lifetime values can be further developed for fluorescence lifetime imaging microscopy (FLIM) [12, 30-33], fluorescence resonant energy transfer (FRET) [30, 33-35] and flow cytometry [36].

TCSPC is a typical time-domain method for FLIM applications and has a better measurement performance with picosecond temporal resolution [31, 33, 37]. Due to its high temporal resolution, the TCSPC method is more suitable for biomedical applications since the fluorescence lifetime of most widely used fluorescent proteins or fluorophores is around hundreds of picoseconds to a few nanoseconds. However, the conventional TCSPC instruments are, in general, complicated, expensive, bulky, and slow. The performance of TDCs in TCSPC systems determines the measurement accuracy directly. The TDCs, which are suitable for FLIM application, should have <100ps temporal resolution[38, 39].

Positron emission tomography (PET) imaging is a powerful nuclear medicine diagnosis technique; is widely used for cancer and tumour diagnosis [40-44]. By combining the ToF technique with the PET, significant improvements of imaging quality, lesion localisation and tomographic reconstruction have been achieved by increasing the signal to noise ratio (SNR) and sensitivity [43]. The current commercial ToF-PETs from Philips and GE achieve hundreds

of picoseconds temporal resolution [45-47]. To further enhance the temporal resolution of ToF-PET lower than 100ps or even 10ps for isolating annihilation events within a 3-mm voxel, there is a need to have both high-performance cooled detectors and TDCs with a temporal resolution of a few picoseconds.

Another potential application field of TCSPC is Raman spectrometers, which are non-contact, nondestructive material composition detection devices widely-used in biochemistry, material, food and water safety areas. Raman signals tend to be very weak and can be submerged easily by fluorescence signals and background noises in conventional products [48, 49]. However, Raman and fluorescence signals can be separated in the time domain, since Raman signals appear earlier with shorter duration (<1ns), while fluorescence signals appear later with much longer duration (a few nanoseconds to more than a dozen nanoseconds). Kerr gating [48] and timing-gated intensified charge-coupled devices (ICCD) [50] was presented to increase SNR by filtering out fluorescence signals. However, these two methods tend to be laboratory solutions in respect of price, size and system complexity. By combining fast response detectors, such as single-photon avalanche diodes (SPADs) and photomultiplier tubes, time-correlated Raman spectrometers, [14, 51] which offer time information and better integration, constitute a producible alternative of the Kerr gating and gated ICCD.

b) Nuclear and high-energy physics

For many nuclear applications [52] and high-energy physics experiments, TDCs are vital device to obtain high-resolution time information such as the collision time of a particle in a large ion collider experiment (ALICE) [24, 53]. To provide accurate identification information of particles, ToF detectors with TDCs with lower than 100ps temporal resolution are usually required for ALICE experiments.

c) Quantum communications and measurement

Time synchronisation and high precision timing measurement are essential to improve fidelity and reduce timing jitter in quantum entanglement, communication and measurement areas. The

temporal resolution of a timing system can directly influence the quantum bit error rate [26, 54].

d) Industrial applications

Light detection and ranging (LIDAR) is a widely used ToF technique in industrial areas [55, 56], space science [57-59] and commercial products such as industrial robots, laser altimeters, collision avoidance, speed measurement and automatic driving. Typical lidar hardware consists of a pulsed laser transmitter, optics systems, photon detectors and TDCs. TDCs convert TIs between stimulation and detection of laser pulses for distance calculation. TDCs with features of high resolution and low power consumption are also used in digital phase-locked loop (DPLL) designs which are widely employed in various electronic systems for synchronisation of signals such as in communication, transceivers, digital television, and broadcast [60].

1.3   Research aim

This research focuses on two aspects. One purpose of this study is to develop high-resolution and low-nonlinearity digital field-programmable gate arrays (FPGA)-based TDCs architecture for TCSPCs. Recently, high-integration, large-scale, multiple-channel advanced TCSPC systems have been gathering increased interest, especially when it comes to achieving fast time-corrected imaging [61]. Features such as flexibility and compatibility make FPGAs as one of ideal platforms to implement digital TDCs [62], and applying digital-TDCs reduce complexity, size and cost of advanced TCSPC systems [13, 63]. However, poor linearity of conventional FPGA-TDCs is the main limitation in terms of encumbering measurement accuracy and will become increasingly severe with higher resolution [64]. This thesis, therefore, examines nonlinearity sources in FPGAs and explores solutions with which to solve the missing-codes, ultra-wide bin and bubble problems to improve the linearity. This research aims to develop FPGA-TDCs with high temporal resolution ($\leq$10ps) and much lower nonlinearity compared with previous published FPGA-TDCs.

Conventional TCSPC systems typically consist of several independent hardware devices

including single-channel photon detectors, TDC or TCSPC cards, data processing modules and a workstation [3]. These devices cause a limited channel number and high system complexity, which is not ideal for wide-field TCSPC imaging applications [61]. Therefore, the second aim is to develop a compact, integrated, large-scale TCSPC array for wide-field time-resolved imaging. Hardware, firmware and software design are required for driving and configuring photon detectors and TDCs. A high-speed data link such as USB 3.0 is needed between firmware and software for data transmission. Histogramming modules of the TCSPC can be achieved either by the software or the firmware. Validation and characterisation of the detector and TDCs based on the presented TCSPC system are necessary. An essential fluorescence lifetime measurement is required to verify the reliability and accuracy of the proposed TCSPC system for biomedical applications.

1.4　　Main contributions

The main contributions of this thesis focus on the development of advanced TCSPC systems and contain three parts, shown in **Figure 1.6**. The first and second contributions are two high-resolution, low-nonlinearity FPGA-TDCs for various applications. The third contribution is a compact, multi-channel TCSPC system with a world-leading, large-scale sensor and TDC array which is presented for high-speed wide-field light intensity and time-correlated imaging.
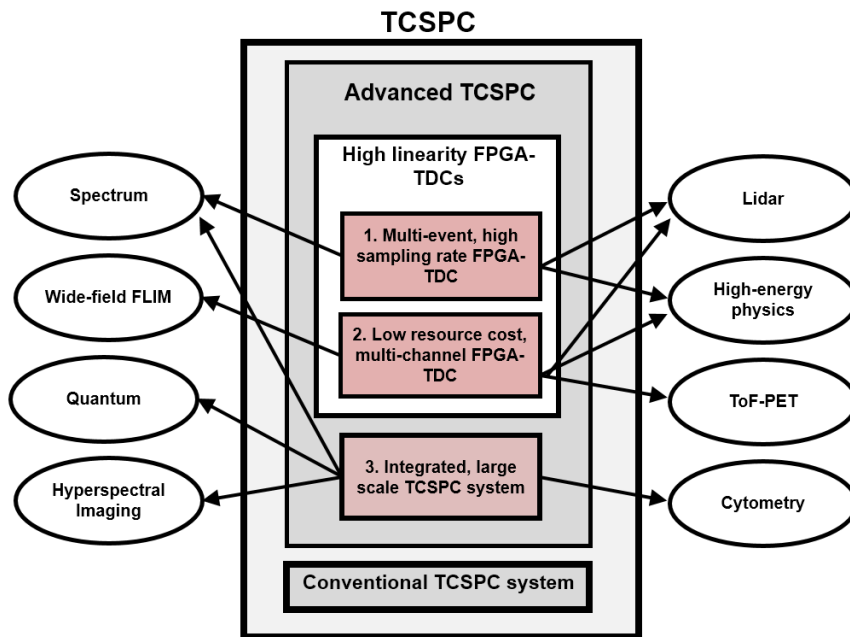
FIGURE 1.6 MAIN CONTRIBUTIONS OF THIS THESIS

Features, published oncomes, and involved projects are summarised below:

1. Multi-event, high sampling rate FPGA-TDC:

- ~10ps temporal resolution and significant improvement of the linearity performance compared with previously-published FPGA-TDCs, and even some ASIC-TDCs and analogue TDCs.

- The capability to measure multiple events using single-channel TDC each measurement and a significantly higher sampling rate is achieved.

- Published in a core journal from measurement and instrument field [65].

2. Low resource cost, multi-channel FPGA-TDC

- ~5ps temporal resolution and significantly improved linearity performance are achieved

- Minimised resource and area cost for multiple-channel architecture and 96-channel TDCs with the independent histogramming module were integrated into a single FPGA chip.

- Several novel architectures and methods were created. An innovative timing test method was also presented, which can be used to quantitate the temporal

characteristic.

- The proposed TDC was published in a high-impact IEEE journal [66].

3. Integrated, large scale TCSPC system

- A wide-field time-resolved imaging system with a world-leading 192x128 SPAD&TDC array is achieved and presented.

- Presented at a top solid-state circuit conference [67] and in a high-impact IEEE journal [68].

- This system has already been applied to different projects including single photon correlations for hyperspectral imaging [69], fast diagnosis for retinopathy based on autofluorescence lifetime imaging, and smart flow cytometry for cancer diagnosis using silicon single-photon sensors and nanoprobes.

1.5   Outline of the thesis

The summary of the following chapters of this thesis is given below:

Chapter 2: Literature reviews of TDC designs

Chapter 2 expounds the performance parameters of TIM devices. A critical analysis was conducted of various analogue and digital TDCs structures, mainstream methods of ASIC and FPGA-based TDCs from previous studies. Based on these previous studies, the research gaps were presented.

Chapter 3: Nonlinearity of FPGA-based TDC designs

Nonlinearity is one of the significant challenges in the FPGA-based TDC design. Chapter 3 focuses on this problem and discusses explicitly static nonlinearity and the solutions put forth in the previous studies. The issues, including ultra-wide bin, missing-codes and bubble problems due to clock skews and nonuniformity of the FPGA internal structure, were further analysed in detail.

Chapter 4: Low-nonlinearity, multiple events, direct histogram FPGA-TDC

In accordance with the critical analysis of the FPGA-TDC in the previous studies, Chapter 4 proposes the first design of a low-nonlinearity, multiple-events, direct histogram FPGA-TDC in terms of its architecture, firmware configuration, performance and evaluation. An new combinational architecture is presented, and its improvement on linearity is verified and evaluated in this chapter.

Chapter 5: Multi-channel, low-nonlinearity TDC based on 20nm and 28nm FPGAs

Chapter 5 presents the second FPGA-TDC design. This design proposed and applies several new methods to achieve linearity improvement effectively with minimum hardware resources and area cost. Notably, the proposed sub-TDL topology provides an ideal solution for the bubble problem. Two 96-channel TDC/TCSPC arrays were implemented in the two different FPGA series to verify the compatibility and feasibility of the proposed multiple-channel architecture design.

Chapter 6: An integrated 40nm CMOS 192 x 128 SPAD-TDC array sensor for time-correlated wide-field imaging

Chapter 6 further presents and discusses the innovative imaging time-correlated wide-field imaging system. The characteristics of measurement are also presented in this chapter. A basic fluorescence lifetime measurement is performed to verify the reliability and accuracy of biomedical applications.

Chapter 2: Literature review of TDCs

TDCs can be classified into two main categories: analogue methods and full-digital methods. The full-digital methods are mostly based on complementary metal-oxide-semiconductor (CMOS) process technologies. Early TDCs were usually built on analogue circuits. Since the 1990s, digital TDCs have improved rapidly with the development and enhancements of digital circuits and integrated circuits (IC).

This chapter discusses the performance parameters of TDCs and reviews different methods, architectures and implementation platforms of TDCs. The chapter then critically introduces and discusses various analogue and digital TDCs' structures and features. As the leading implementation platforms of digital TDCs, the characteristics of ASIC and FPGA devices are briefly introduced. Finally, previous studies of FPGA-TDC designs are critically analysed.

2.1   Performance parameters

The performance of TDCs needs to be characterised and evaluated by various technical parameters. These parameters mainly include temporal resolution, and nonlinearity parameters such as differential nonlinearity (DNL) and integral nonlinearity (INL), measurement range (MR), quantisation error and precision, and dead-time.

a)   Temporal resolution (LSB or bin-width)

The temporal resolution is one of the fundamental parameters that theoretically describes the minimum measurable difference of input TIs [2, 62, 70]. It is also called the least significant bit (LSB), averaged width of bin, or bin-width. The measurement or conversion of TDCs is a mapping and quantisation progress from input TIs to digital outputs. The quantisation progress can be represented as a transfer plot, as shown in **Figure 2.1.** The resolution is the step width or the code bin in the transfer plot, which should be uniform under the ideal conditions. The

resolution of an N-bit TDC can be calculated as the averaged bin-width:

$$LSB = MR / 2^N \qquad\qquad (2.1)$$
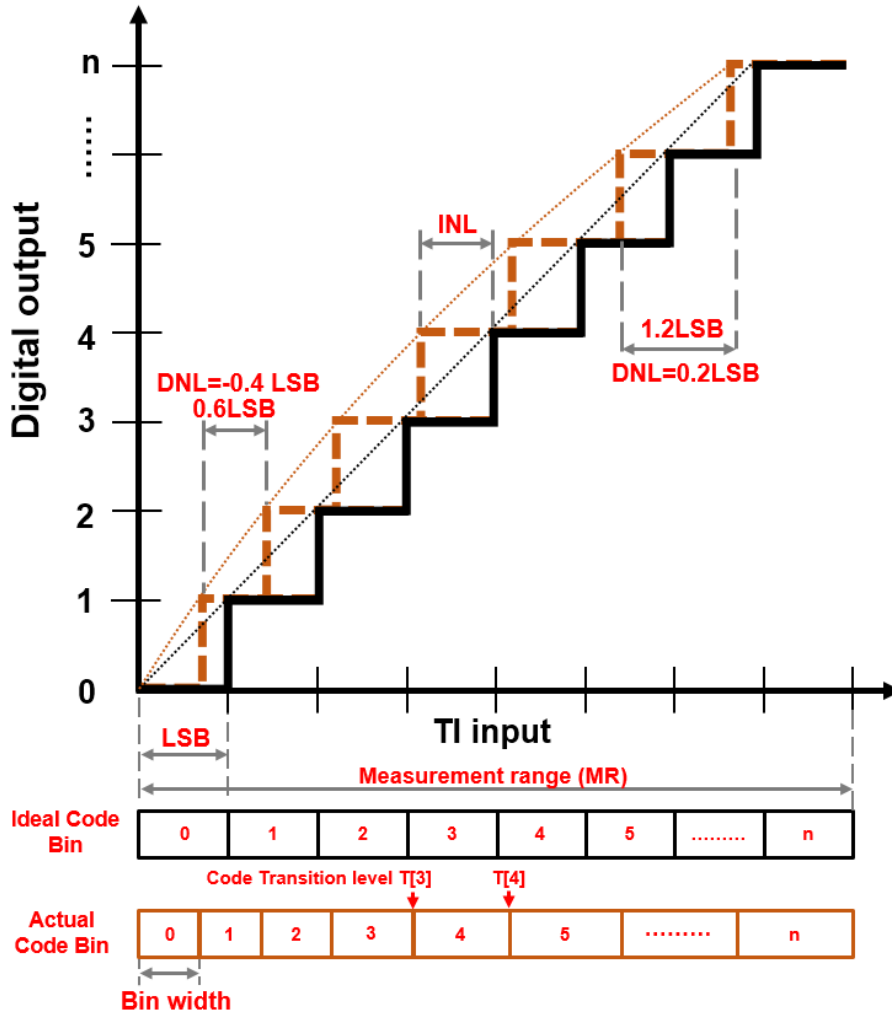
b) Nonlinearity

The nonlinearity of a time-to-digital conversion causes quantisation errors and deviations between the real and the measured results. The statistical code density test [71] is a widely used method for the quantification and evaluation of the TDC's linearity performance [2, 62, 72]. The test was based on the following two assumptions: all code bins have identical widths in an

ideal TDC, and there is an equal number of events captured by ideal bins with random inputs. During the code density test, a large number of random TIs with equal probability (for all TIs) over a full MR were fed into a TDC, and a histogram was generated. Based on the results, the differential nonlinearity (DNL) and the integral nonlinearity (INL) were calculated for the quantisation of linearity performance. DNL is the proportional relation of the difference between the actual size of a bin and the ideal bin-width. The formula of DNL is:

$$DNL[k] = \frac{W[k] - LSB}{LSB}$$
(2,2)

where $W[k]$ is the actual bin-width of the k-th bin. INL is the maximum accumulated deviation between the ideal code transition levels and the actual output values after gain and offsets correction [73]. It can be calculated by integrating DNL values as:

$$INL[k] = \sum_{n=0}^{k} DNL[n]$$
(2.3)

The maximum and minimum values or the peak-to-peak values (in LSBs) of the DNL and the INL are used to present the linearity performance of TDCs. The DNL and INL plots of a TDC with perfect linearity should be demonstrated as a horizontal line at $y=0$.

c) Measurement range (MR)

MR is also called full-scale range (FSR). MR is the time interval between maximum and minimum measurable time of TDCs. The MR is opposite to its temporal resolution, area or resource cost. With the same area or the cost of the resource, a finer resolution leads to a shorter MR. For a TDC with an N-bit output, MR can be expressed as:

$$MR = 2^N \times LSB$$
(2.4)

d) Quantisation error and precision:

Quantisation error, σ_q, describes the measurement distortion from the original values of TIs. This parameter is based on the temporal resolution, and can be calculated as below [74]:

$$\sigma_q = \sqrt{\frac{LSB_{START}^2}{12} + \frac{LSB_{STOP}^2}{12}}$$

(2.5)

where LSB$_{START}$ and LSB$_{STOP}$ are the temporal resolutions for the 'START' and the 'STOP' inputs of TDCs, respectively.

The measurement precision of TDCs needs to be evaluated with other parts in a TCSPC or ToF system [62, 74, 75]. During the evaluation, the noise sources need to be considered, such as the internal nonlinearity of TDCs, jitter from detectors and clocks in the system. The precision describes the degree of stability or reproducibility during continuous measurements of a fixed TI. It is generally determined by the standard deviation, σ, or the full-width half-maximum (FWHM) of the instrument response function (IRF). The precision of a TDC system can be expressed as:

$$\sigma_{rms} = \sqrt{\sigma_q^2 + \sigma_{clock}^2 + \sigma_{INL-START}^2 + \sigma_{INL-STOP}^2 + \sigma_{others}^2}$$

(2.6)

where σ_q is quantisation error, σ_{CLK} is clock jitter, σ_{INL_start} and σ_{INL_stop} are the standard deviations of the INL of two interpolators for the 'START' and the 'STOP' signals, and $\sigma_{Additional}$ stands for other related internal jitters of TDC and input signals [62, 70, 75].

e) Dead-time and sampling rate

The dead-time and sampling rate evaluate the measurement capability of TDCs for high repeat rate events. The dead-time of a TDC is defined as the minimum time interval from the current measurement to the start of the next measurement [2], or the shortest time interval for one conversion [62]. The sampling rate, or conversion rate, is reciprocal of the dead-time. The dead-time of TDCs, detectors and data readout codetermines acquisition speed and data throughout. Therefore, the dead-time and the sampling rate is crucial in high-speed applications such as FLIM, PET and ToF ranging [13, 76, 77].

2.2 Analogue-based TDC architectures

Many published studies have described various analogue-based TI measurement methods, including the time-to-amplitude converter (TAC) with ADCs [3, 15, 16], time-stretching [78] and streak camera [79]. These methods convert the TI information into other physical quantities such as voltage amplitude and spatial information to achieve high-performance TIMs [2]. Maturing of analogue TDCs earlier than digital methods and many commercial TCSPC systems based on analogue TDCs provide the excellent temporal resolution performance and linearity (such as PicoQuant TCSPC systems with 1ps resolution, and the femtosecond streak cameras). However, the limitations of analogue circuits and design principles cannot be avoided, such as being sensitive to external factors, challenging to miniaturise or integrate, and a long dead-time. Many analogue TDC products are bulky and expensive, requiring higher operation conditions which limit the number of channels.

2.3 Digital TDCs

Advanced by rapid developments in the integrated circuit (IC) technologies from the 90s, digital TDCs have gradually become mainstream. Digital circuits have advantages in terms of flexibility, integration and tolerance to interferences from temperature variations and electromagnetic noise. Furthermore, with the rapid growth of CMOS technologies, the temporal resolution of TDCs has been further improved from hundreds of picoseconds [80] to sub-picoseconds [22, 81]. Previous studies have discussed several methods to implement full-digital TDCs, including the direct counting method, Vernier method, tapped delay line, differential delay line, pulse-shrinking, ring oscillator (RO) TDC and interpolation architectures.

2.3.1 Direct counting method

The direct counting method uses a binary counter driven by a clock signal to count the number of clock periods within TIs. A counter is usually composed of flip-flops (FFs) and synchronised by a clock signal. The temporal resolution of direct counting equals to the period of a clock,

$T_{clk} = 1/f_{clk}$. However, the temporal resolution of the direct counting method is limited to hundreds of picoseconds, because the maximum clock frequencies of ASIC and FPGA are limited to several GHz and MHz, respectively. At the same time, the maximum error of single-shot measurements for the direct counting method can approach approximately $2T_{clk}$ when both edges of TIs ("START" and "STOP" signal) are asynchronous with the clock signal. The average standard deviation of an asynchronous TI is: $\sigma_{ave} = \pi T_{clk}/8 \cong 0.39T_{clk}$ [2]. The measurement error of this method can be reduced by averaging the results of multiple measurements. Assuming that the number of measurements is N, the average standard deviations of multiple measurements is: $\sigma_{N,ave} \cong 0.39T_{clk}/\sqrt{N}$. However, the direct counting method also has benefits. The well-developed counters with the simple structure mean that the MR can easily be extended by simply increasing the number of the counter bits as $MR = T_{clk} \times 2^{N}$ in an N-bit counter. The direct counting method is appropriate for applications with a high requirement of MR but a low requirement of temporal resolution. The method can also be used as a 'coarse' counter in an interpolation architecture for a longer MR.

2.3.2 Vernier method

The vernier method [82, 83] is a digital time stretching approach. The basic concept of this approach is similar to that of a Vernier calliper. The architecture of a typical Vernier method is shown in **Figure 2.2**:
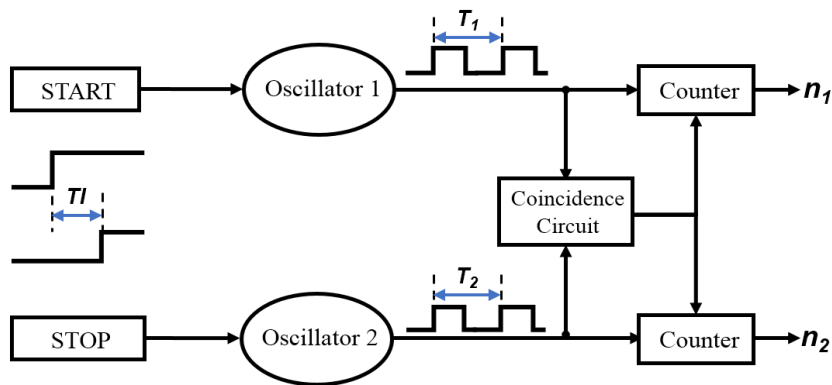
FIGURE 2.2 BLOCK DIAGRAM OF THE VERNIER METHOD

In this method, two startable oscillators, Oscillator 1 and 2, are used to generate two clock signals with slightly different periods, namely T_1 and T_2. Oscillator 1 and 2 are independently triggered by the 'START and 'STOP' signals. Therefore, the phase difference between the two oscillator outputs is equal to the TI between the 'START' and 'STOP' signals at the beginning. Following this, the phase difference will shrink gradually, and a coincidence of the two clock signals will occur finally, as shown in **Figure 2.3**.
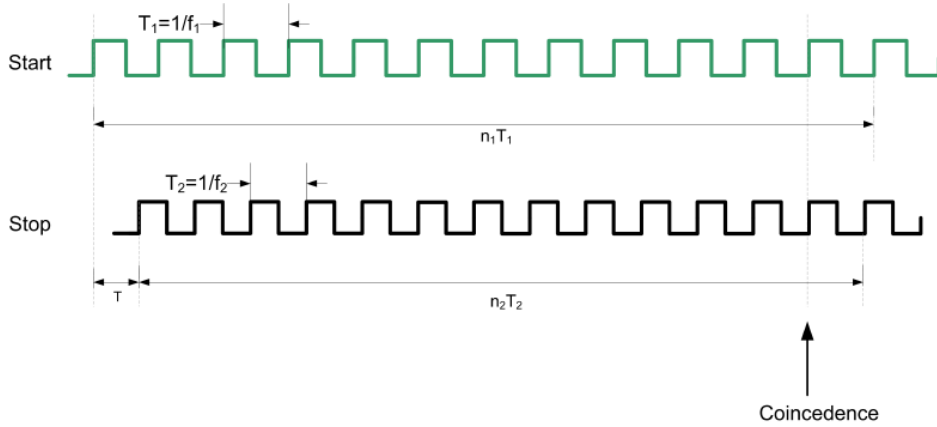
FIGURE 2.3 WAVEFORM OF THE CONCEPT OF THE VERNIER METHOD

The two counters are driven by Oscillator 1 and 2, respectively, and will be frozen when a coincidence circuit detects a coincidence. The TI can be calculated as:

$$T = (n_1 - 1) \times T_1 - (n_2 - 1) \times T_2 \qquad (2.7)$$

where n_1 and n_2 are frozen data in two counters respectively. The temporal resolution of the Vernier method is equal to the period difference of the two clocks, and the linearity depends on the stability and accuracy of the two oscillators. By using this method, it has been possible to achieve a resolution level lower than 10ps [84, 85]. However, to achieve a better resolution, a longer conversion time is usually required, limiting the applications of this method.

2.3.3   Tapped delay line

The tapped delay line (TDL) method is one of the most common full-digital TDC architectures, which can be implemented in both ASIC and FPGA devices. Integrated delay lines were used for TDL-based TDCs from 1982 [86], and were combined with phase-locked loops (PLLs) and delay-locked loops (DLLs) to calibrate and maintain stability [87-89]. This method feeds electric signals (the 'HIT' and 'START' signals) into a unidirectional delay line which is composed of cascaded delay cells (delay elements). The propagation distance of the electric signal is represented by sampled signal status in all delay cells when the 'STOP' or the sampling clock signal is activated and converted to fine binary codes by following encoders. A certain number of taps exist in a delay line as sampling points and are connected to a group of samplers, such as FFs or latches. Finally, the TIs between the two input signals are calculated according to the fine codes and the propagation speed of the delay line.
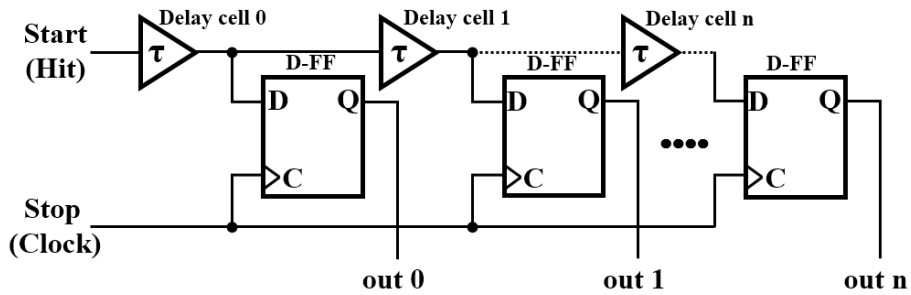
FIGURE 2.4 BLOCK DIAGRAM OF THE BASIC TDL-TDC

Figure 2.4 demonstrates a typical structure of a basic TDL using cascaded buffers as the delay cells and a group of D-type FFs (D-FFs). The signal 'START' or 'HIT' is fed into a TDL and passes through each delay cell in order and is delayed by τ each time. When the rising edge of the 'STOP' or the sample clock signal arrives, the D-FF group samples and stores the signal status at corresponding taps and outputs the stored status from the Q ports of D-FFs. The following formula can be used to calculate time intervals:

$$TI = b \times \tau \qquad (2.8)$$

where b is fine binary codes which are converted from outputs of the D-FF group. Due to the

limitations of IC technologies in the early stage, the temporal resolutions of TDL-TDCs were only achieved at a nanosecond or sub-nanosecond level [52, 80, 90, 91], which does not meet the high-resolution requirements of many applications, such as FLIM and Lidar. However, with the rapid development of ASIC and FPGA devices, the resolution of the typical TDL-TDC was significantly improved to a satisfactory level, which will be further discussed later in this chapter.

2.3.4   Vernier delay line

Vernier delay line (VDL), or differential delay line, was used in previous studies [85, 89, 92] to improve the resolution of TDCs. The VDL applies delay cells in both the 'START' ('HIT') and 'STOP' (the sampling clock) signals routes, as shown in **Figure 2.5**. Similar to the analogue Vernier method, the propagation delay of the 'STOP' path, τ', is slightly smaller than the delay of the 'START' path, τ. Thus, the TIs between the two signals are gradually shrinking down during the propagation. Instead of being captured simultaneously, the signal status is captured in turn by a D-FF group. **Figure 2.6** illustrates a measurement example of this process. The resolution is equal to the difference between τ and τ'. Through further appropriate configuration, the differential delay time achieves a resolution lower than 50ps [89]. However, the dead-time was longer than that of the basic TDL-TDCs, since the output was only valid once the 'STOP' signals reached the end of the TDLs.
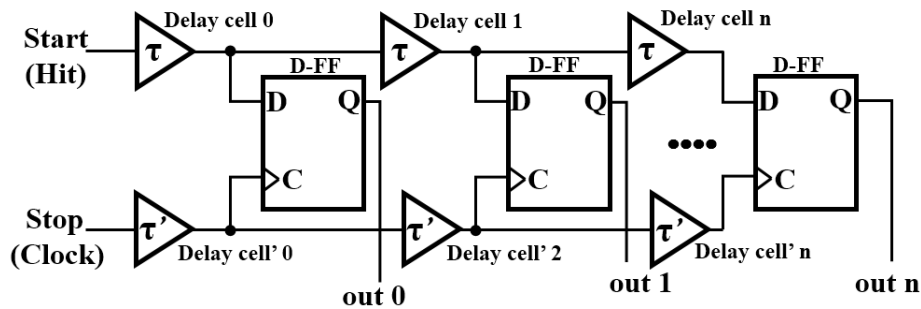
FIGURE 2.5 BLOCK DIAGRAM OF DIFFERENTIAL TDL-TDC
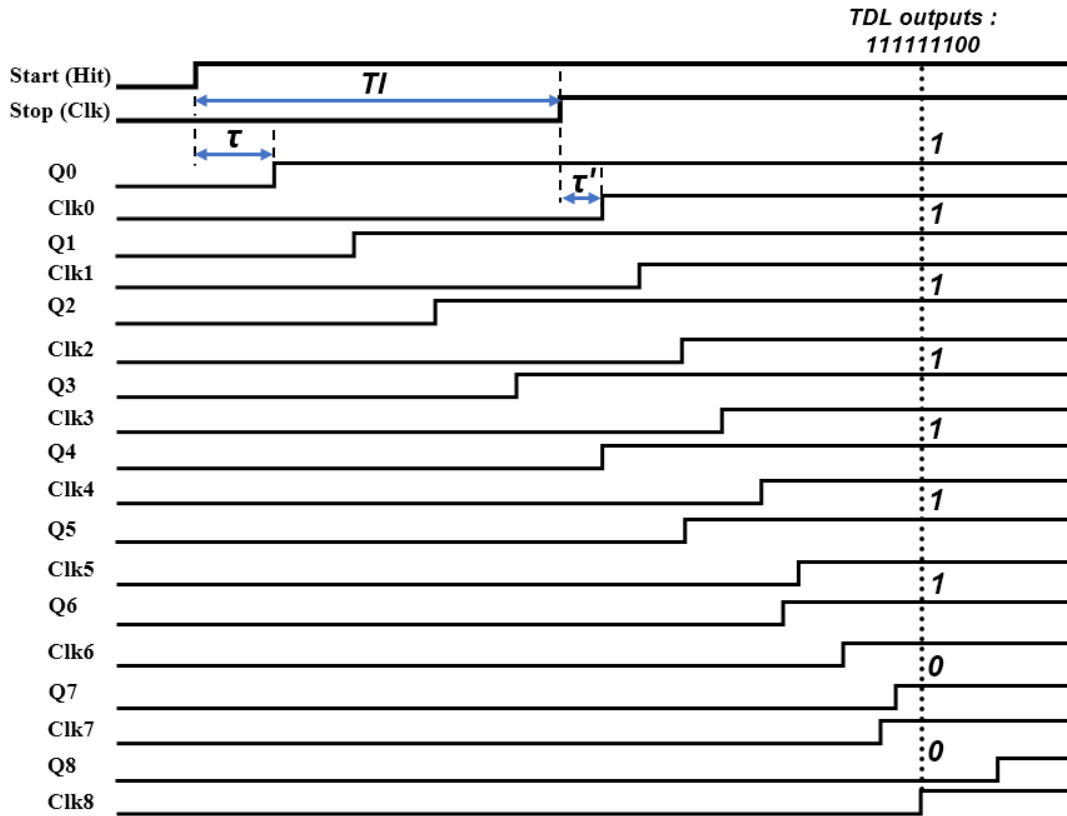
FIGURE 2.6   TIMING WAVEFORM OF A DIFFERENTIAL DELAY LINE

2.3.5   Pulse-shrinking delay line

The pulse-shrinking delay line method is modified from basic TDL architecture to achieve a better temporal resolution [93]. In this method, different from the basic TDL architecture, TIs are defined by the pulse width of a single-end signal in the pulse-shrinking delay line. In **Figure 2.7,** pulse-shrinking delay cells make the propagation speed of rising edges slower than falling edges by applying special-designed inverters [17, 18] or current-starving transistors [80]. Therefore, the width of input pulses will be reduced by T_r-T_f (T_r and T_f are the rising and falling time of the pulse, respectively) at each propagation until the pulse disappears. The positions in which when the pulse disappears can be detected and used for the TI calculation. The resolution of this architecture depends on T_r-T_f, which can be smaller than the gate delay of the delay cells [94]. Studies such as [17, 18, 95] used 0.35μm, 0.8μm and 0.13μm CMOS technology

respectively, and presented cyclic pulse-shrinking TDCs with 68ps, 20ps and 6ps time resolution. However, the dead-time of this method is increased for the same reason as applied to the VDL architecture.
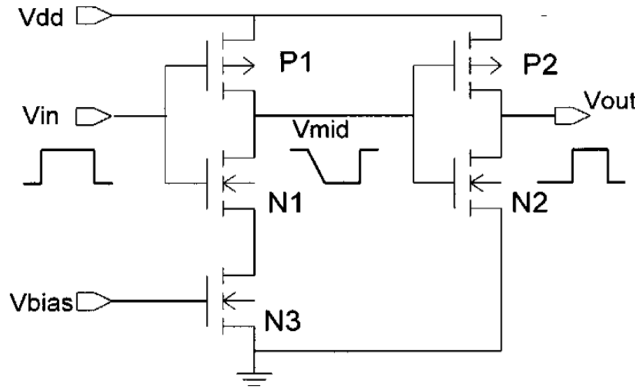
FIGURE 2.7 CIRCUIT OF PULSE-SHRINKING DELAY CELL [33]

2.3.6 Ring oscillator TDC

A ring oscillator (RO)-based TDC consists of a loop of inverters, a sampling logic module, and a loop counter [19, 96, 97], as shown in **Figure 2.8.**
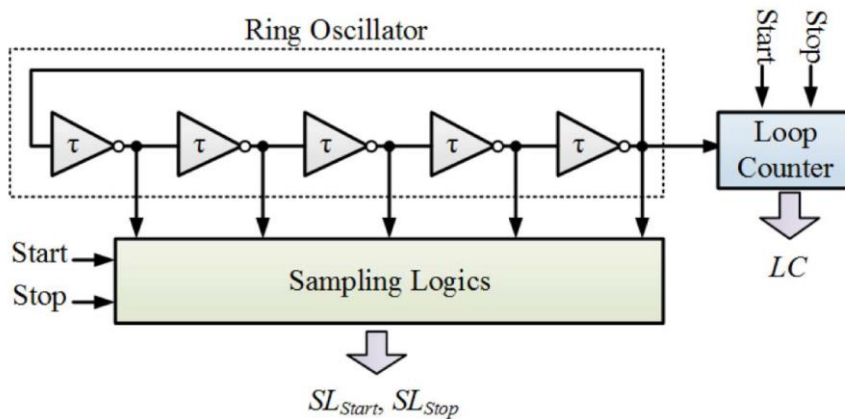
FIGURE 2.8 BLOCK DIAGRAM OF THE BASIC ARCHITECTURE OF A RING OSCILLATOR [62]

During running, multiple pulses or signal transmissions propagate along the inverter loop. The sampling logic module and the loop counter controlled by the 'START' and 'STOP' signals are used to record the signal status of the inverter loop and the total iteration number during each

measurement, respectively. The measured TI can be calculated as [62]:

$$TI = (SL_{STOP} - SL_{START}) \times \tau + \frac{CNT}{f_{RO}}$$ (2.9)

where SL_{Start} and SL_{Stop} are the outputs of the sampling logic module, τ is the propagation delay of the inverters, CNT and f_{RO} are the count result and the oscillation frequency of the loop counter, respectively. Besides the basic-RO, various RO architectures were developed to improve the temporal resolution and power consumption of Multipath RO [98, 99], Vernier RO [100], Gated-RO [98-100], and 2-D Vernier RO [101].

Compared with the delay line TDCs, one of the critical advantages of RO-TDCs is the number of the delay cell (inverters). At the same time, the space occupation can be significantly reduced because of its loop structure. The advantage also means that RO-TDCs has become one of the preferred architectures for a large-scale TDC array and can be integrated into various circuits and devices. Another RO-TDC advantage is that the MR can be easily extended by increasing the bit width of the loop structure. A disadvantage of RO architecture is the high power consumption during the free-running mode [62].

2.3.7   Interpolation Method

Base on the definition of MR or LSB, the level of temporal resolution and the length of MR of TDCs are contradictory if only limited resources are available. Instead of simply increasing the length of TDLs or the size of ROs, an interpolation method was proposed to achieve a high temporal resolution and a long MR at the same time, with limited additional resource utilisation. Nutt's interpolation architecture [102] is a typical interpolation method which has been implemented in ADC and TDC designs for many years. As shown in **Figure 2.9**, Nutt's method splits a TI into three segments, namely two fractional segments and one integral segment, in accordance with a coarse clock signal.
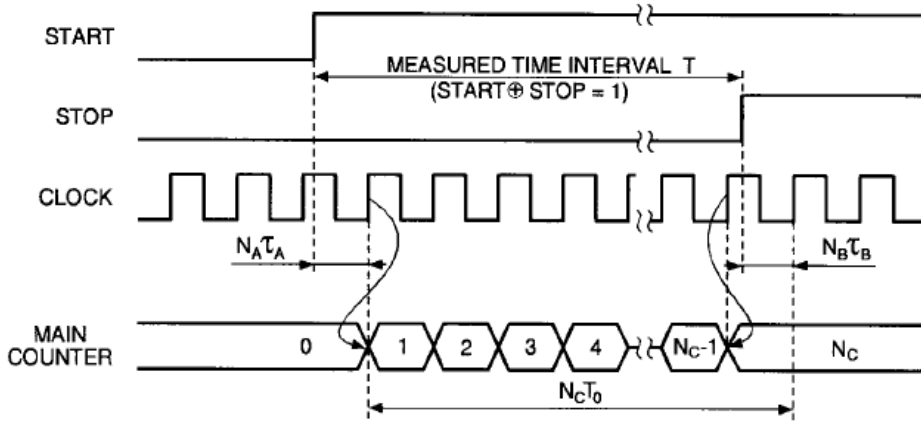
FIGURE 2.9 THE WAVEFORM OF SIMPLIFIED NUTT METHOD [2]

The first fractional segment starts from the rising edge of the 'START' signal and moves to the subsequent rising edge of the coarse clock. The second fractional segment is from the rising edge of the 'STOP' signal to the succeeding rising edge of the clock. The integral segment begins from the end of the first fractional segment and runs to the end of the second fractional segment, which can be measured directly by a digital counter, since it is synchronised to the coarse clock. Two high-resolution TDCs measure the two fractional segments as two interpolators. One of the examples of relevant architecture is shown in **Figure 2.10**. Interpolation circuits for the 'START' and 'STOP' channels are in the same structure. Assuming the period of the clock is T_0, and the TI, T, can be expressed by:

$$T = T_0 \times n_c + \tau_A \times n_A - \tau_B \times n_B \qquad (2.10)$$

where n_A and n_B are the fine codes of the two TDCs, τ_A and τ_B are the resolutions of the two TDCs respectively.
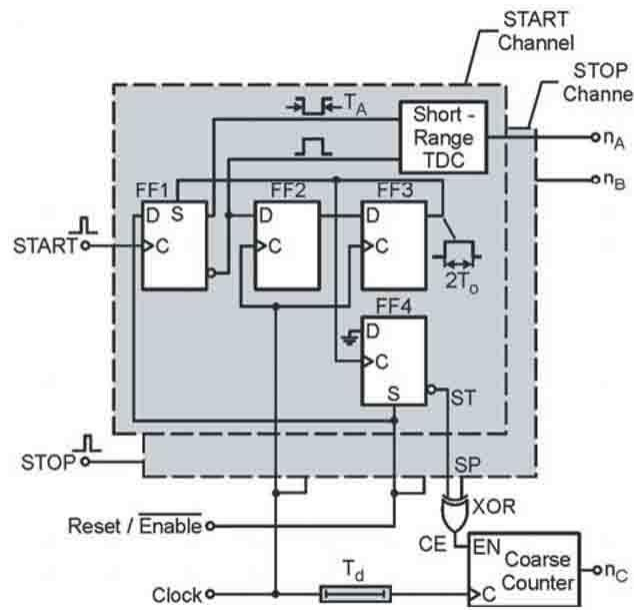
FIGURE 2.10   DIAGRAM OF INTERPOLATION [2]

Szplet simplified Nutt's architecture [103], which is shown in **Figure 2.11**. In this version, each channel only has two flip-flops in use with a buffer designed to provide the time offset.
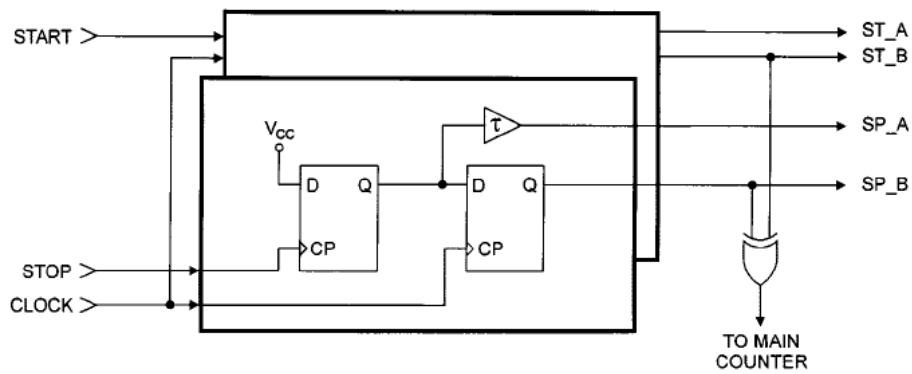
FIGURE 2.11 BLOCK DIAGRAM OF SIMPLIFIED NUTT METHOD [103]

For TCSPC applications, the interpolation architecture can be further simplified as a fine&coarse architecture by synchronising the reference clock to either the 'START' or 'STOP' signal. The architecture is shown in **Figure 2.12**, and only one fractional segment needs to be

measured by a high-resolution TDC each time as the fine time code. Therefore, the utilisation of high-resolution TDCs and related digital circuits is halved, and measurement precision is improved based on equations (5) and (6). This architecture is widely applied in the TDL- and RO-based TDC designs [104, 105].
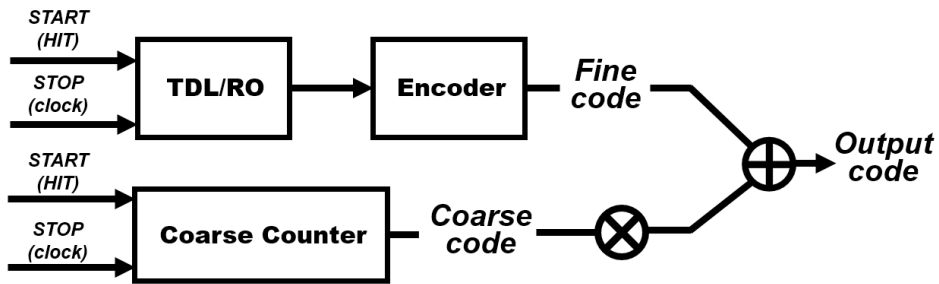
FIGURE 2.12 BLOCK DIAGRAM OF THE FINE & COARSE ARCHITECTURE

2.4   ASIC-Based TDC Design

ASIC and FPGA are the two foremost platforms when it comes to implementing full-digital TDCs. ASIC-based TDCs provide features of the full-custom design, which can achieve better precision and linearity [70] compared with FPGA-TDCs. The ASIC manufacturing process is hugely complicated, along with high design and production costs and high barriers to entry. Since the internal circuits, function and specification of ASIC devices are fixed, their initial development cost tends to be very high. Nevertheless, the unit cost will be reduced by an increasing production scale. Consequently, ASIC devices are more suitable for general-purpose and large-scale commercial products. ASIC-TDC target markets are the applications which require high design density, such as LIDAR systems for the automatic drive and machine version. However, this kind of product carries high risks due to its non-recurring engineering (NRE) cost, including one-time design and a test cost which is higher than that of FPGA applications.

Delay line and ring oscillator are the two mainstream architectures for ASIC-TDCs, and can also be combined with the Nutt or fine & coarse interpolation to extend the MR. **TABLE 2.1**

summarises the ASIC-TDCs designs in the previous studies over the last two decades. The delay line-based architectures (TDL, VDL and pulse-shrinking) have advantages in the temporal resolution aspect. Recently, it has been found that the resolution of CMOS TDC can achieve a sub-picosecond level [22]. The RO method is well implemented in ASIC to reduce the area occupied in large-scale TDC arrays. In 2014, the soft-injection locking ring oscillator was used for Vernier-TDC, which cost $95\mu m^2$, with a $0.18\mu m$ CMOS process [106]. In 2009, a 32x32 SPAD sensor with in-pixel RO-TDC array was implemented in a 130nm CMOS [12]. In this chip, each RO-TDC provided 50ps temporal resolution occupying an area of approximately 2200 μm^2, which was directly connected to an individual SPAD sensor for the parallel measurement. In 2018, the scale of the TDC array was extended to 128x192, with a smaller TDC size of 9.2x9.2um [67]. The time resolution of the RO-TDC was adjusted from 33ps to 120ps by controlling the dedicated voltage of RO-TDC.

Based on the full custom feature of ASIC, many methods and architectures were applied to improve the linearity and stability of ASIC-TCDs. Firstly, the uniformity of the gate delay in the delay line and RO can be minimised by manually fine-tuning the circuit layout. DLLs are applied to maintain the stability of the gate delay in the delay cells and against variations of the process, supply voltage and temperature (PVT) to correct INL and the long-term offset. The linearity of the delay line architecture can be stabilised by using a voltage-controlled buffer in the feedback loop [89, 107, 108].

Table 2.1 The summary of published ASIC-TDCs

Year	Process (μm)	Architecture	LSB(ps)	MR(ns)	Area	DNL(LSB)	INL(LSB)	Accuracy
			Delay line based ASIC-TDC					
1994[90]	1.0	TDL, interpolation	1560	204,800,000	25mm^2	-	-	-
1995[94]	1.2	pulse-shrinking, interpolation,	780	10,000	7.25 mm^2	-	-	±120-200ps
2000[17]	0.35	pulse-shrinking, interpolation,	68	NA	0.35x0.09mm	-	-	3ns
2000[89]	0.7	Vernier TDL, DLL	30	0.384-32	3.2x3.1mm	-	±30ps	20ps
2003[18]	0.8	pulse-shrinking	20	18	2x1mm	<0.5 LSB	-	-
2006[75]	0.35	TDL	12.2	204,000	2.5x3.0mm	-	±30ps	8.1ps
2013[109]	0.065	gated delay-lines	3.75	0.45	0.02 mm^2	0.9LSB	2.3LSB	
2013[74]	0.35	interpolation, Vernier TDL	10	160	-	0.9%LSB	-	17.2ps
2018[22]	0.065	interpolation, Vernier TDL	0.45ps	200ps	0.502x0.11mm	0.65LSB	1.2LSB	1.7LSB
			Ring Oscillator based ASIC-TDC					
1996[88]	0.5	RO	250(RMS)	2,560	6.4x6.4	<0.08LSB	<0.1LSB	-
2003[96]	0.35	Vernier RO	156	300ns	1.81x1.81mm	-	90ps	-
2009[19]	0.18	RO	61	80ns	0.53 ×0.8mm	~28ps	~22ps	50ps
2009[12]	0.13	RO	52-178	53ns-182ns	4.6x3.8mm (32x32 array)	±0.5LSB	2.4LSB	-
2012[97]	0.09	Basic RO	13.6	13-bit	0.021 mm^2	-	-	-
2012[20]	0.13	cyclic, time amplifier	1.25	0.32	0.07 mm^2	0.7LSB	3LSB	-
2013[101]	0.09	Gated-Vernier RO	5.8	40ns	0.03 mm^2	-	-	-
2013[99]	0.065	multi-path gated RO	4.22	1ns	0.02 mm^2	-	-	-
2014[110]	0.09	switched- RO	0.32	2ns	0.02 mm^2	-	-	-
2015[21]	0.065	interpolation, time amplifiers	1.2ps	0.614ns	-	0.67LSB(Simu)	0.62LSB(Simu)	-
2016[111]	0.13	Vernier RO	7.3ps	9ns	0.03 mm^2	3.2LSB	1.2 LSB (rms)	17.4ps
2017[81]	0.065	cyclic-ring Vernier, time amplifier	0.98ps	5.76ns	0.02 mm^2	±0.8LSB	±2.2LSB	-

2.5 FPGA-based TDC

FPGAs are a type of CMOS device with the capability of re-programmability and reconfigurability. As shown in **Figure 2.13**, an FPGA chip comprises the array of configurable logic blocks (CLBs), interconnected matrices and Input/Output Block (IOB).
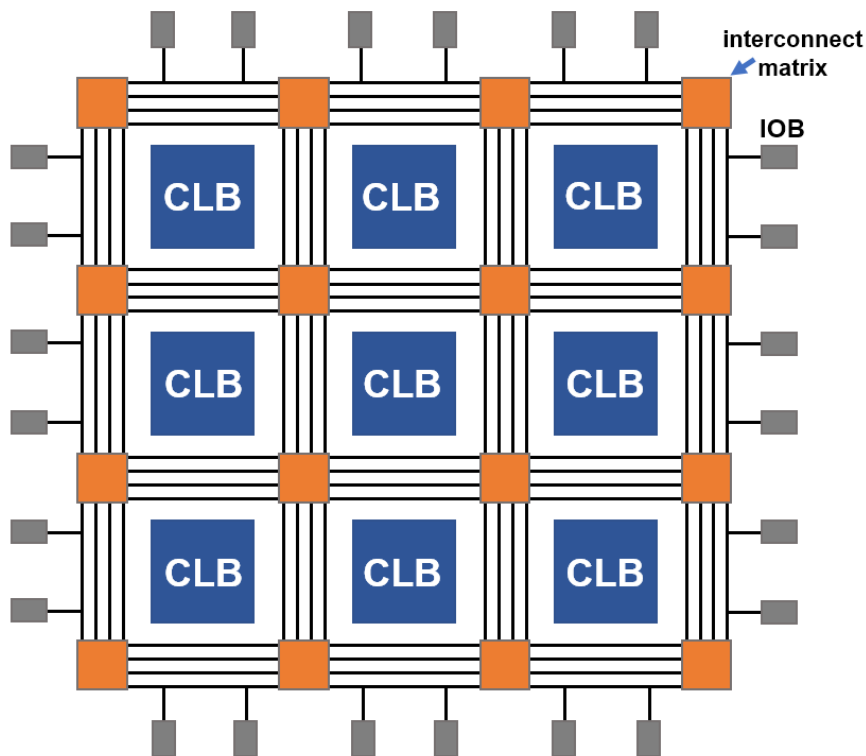
FIGURE 2.13 A SIMPLIFIED INTERNAL STRUCTURE OF FPGAS

The CLB contains multiple compact random access memory (RAM)-based Look-Up Tables (LUTs), carry-chains and flip-flops (FFs) [112-114]. The configuration, specification and function of FPGA-based designs can be modified or rebuilt repeatedly by downloading firmware into the chip after delivery. Several special software (such as ISE, Vivado and Quartus II) and hardware description languages (such VHDL and Verilog) are used to design, compile and generate downloadable configuration files of the firmware. The first FPGA device was invented in 1985 by Xilinx. During the early stage, FPGAs were mainly used in

low-speed, low-complexity and low-capacity applications. With the development of the manufacturing process, the latest generation of the commercial FPGA has much better performances, such as a higher system operating speed (>500MHz) and a competitive logical density [115, 116]. With the feature of re-programmability, FPGAs provide unique advantages, such as excellent flexibility, compatibility, and parallel processing capabilities. The FPGA firmware can be modified, upgraded and seamlessly and rapidly combined with other modules. With the support of the soft and hard intellectual property (IP) cores, mature predesigned functional modules which comply with the universal standards can be customised and inserted into an existing design to reduce the duration time and the difficulty of the development.

Reaping benefits from the features of FPGAs, TDCs can be embedded and seamlessly integrated into other designs. The temporal resolution, measurement range and the number of channels of TDCs are adjustable to match the requirements of various applications. Numerous architectures, calibration methods, data processing and compression algorithms can be inserted, tested and verified to improve the performance of FPGA-TDCs further. Therefore, these advantages make it possible for the FPGA-TDC to be applied to prototyping, scientific experiments and high-end instruments.

2.5.1   Main architectures and temporal resolution

a)   TDL-based FPGA-TDC

Since the internal circuits of FPGAs are predefined, only several feasible methods are available. As can be seen in **Table 3.1** from Chapter 3, most previous studies tended to use TDL-TDCs as their method during the research. In 1997, the first example of FPGA-TDCs was implemented in a 65nm CMOS FPGA [92]. A typical TDL architecture was applied to obtain the 200ps temporal resolution. The amorphous anti-fuse structures and the pASIC architectures were used as the delay cells. The local reset loop directly reset a typical thermometer output code to a one-hot code bit, following which the code bit was converted to a binary code. The block diagram of the logic design is presented in **Figure 2.14**.
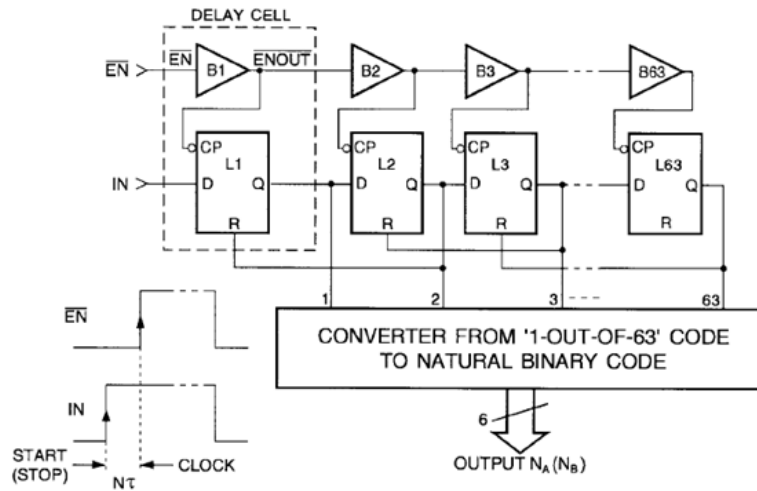
FIGURE 2.14 LOGIC DIAGRAM OF DELAY LINE WITH DIRECT CODING [92]
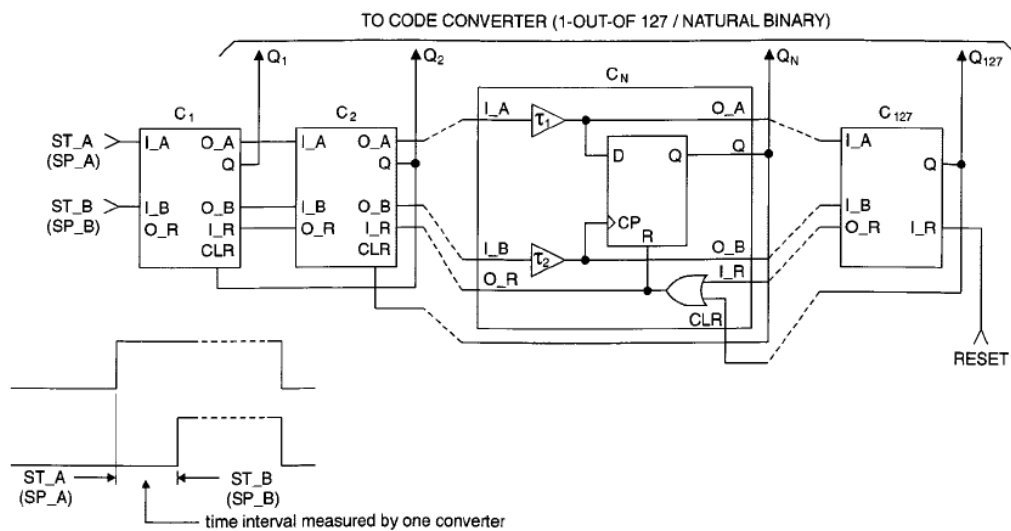
FIGURE 2.15 LOGIC DIAGRAM OF TIME CODING DELAY LINE [103]

A differential TDL method was also presented and tested in a QuickLogic FPGA, as shown in **Figure 2.15** [103]. Two differential TDLs were used as the two-stage fine interpolators for excellent temporal resolution. The structure was modified based on a previous work in 1997 to improve the temporal resolution from 200ps to 100ps. After software correction, the random error was reduced to 0.65 LSB (129ps). These two designs showed the vast potential of FPGA to implement the sub-hundred picosecond-level resolution TDC.

Following this, the mainstream TDL architecture of FPGA-TDCs depends on carry-chain modules which are widely provided by FPGA devices from Xilinx and Altera (Intel). The two companies have provided almost 90% of FPGA devices during the past two decades. The propagation delay of carry-chains decides the temporal resolution of TDL-TDC, and the resolution was improved rapidly with the upgrading of FPGA technologies. In 2013, the temporal resolution of TDL FPGA-TDC achieved to 10ps in Xilinx Virtex-6 FPGA, which was fabricated by a 40nm copper process [117]. When the TDL-TDC was implemented in the 20nm process Xilinx UltraScale FPGA, the time resolution of a raw TDL achieved around 5ps, and it was further improved to 2.32ps by using the dual-sampling method [118]. **Figure 2.16** summarises the temporal resolutions of FPGA-TDC in various devices from the previous studies.

Raw CARRY chain TDC:

QuickLogic 1997	QuickLogic 2000	ACEX 1K 2003	Virtex-II 2006	ACEX 1K 2006	Virtex 5 2009	Virtex 6 2013	Kintex 7 2016	Kintex 7 2017
200 ps	**100 ps**	**400 ps**	**46 ps**	**65 ps**	**17 ps**	**9.8 ps**	**10.6 ps**	**5 ps**

2000   **2005**   **2010**   **2015**
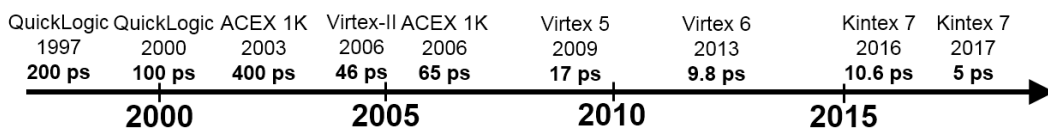
FIGURE 2.16 TIMELINE OF TEMPORAL RESOLUTION OF **FPGA-TDC** IN VARIOUS DEVICES

b) Multiple-chain averaging

The temporal resolution of TDL-TDCs is related to the signal propagation speed in delay lines, and it mainly depends on the manufacturing process of FPGA devices [62]. Previous studies implemented the multi-chain averaging approach, which uses multiple TDLs to measure the same signals to overcome this process-related limitation. The basic idea of the approach is shown in **Figure 2.17**. In ideal circumstances, assuming the bin-width of the delay line 0 and 1 are 30ps. If there is a half bin-width offset between the two delay lines, the averaged bin-width (temporal resolution) can be reduced from 30ps to 15ps by averaging two delay lines.
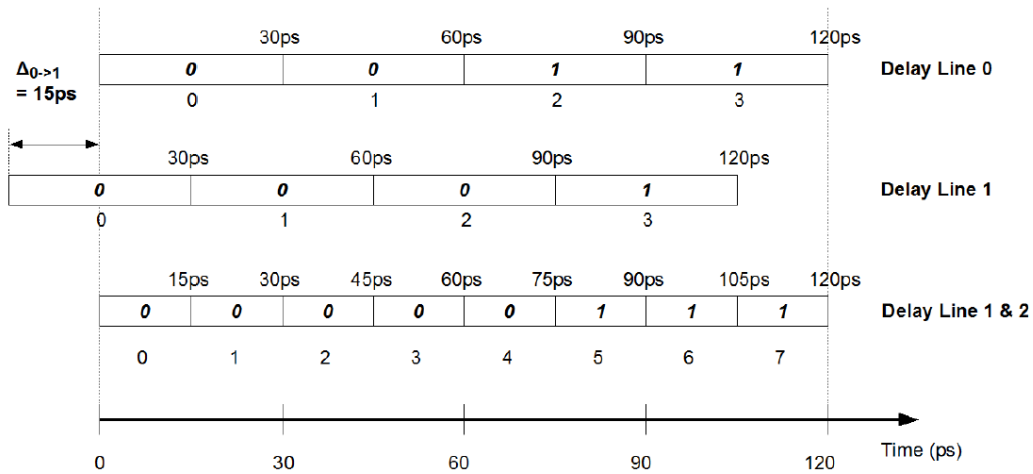
FIGURE 2.17 EXAMPLE ILLUSTRATING THE USE OF MULTIPLE TAPPED DELAY LINES. (A) IDEAL SITUATION WITH UNIFORM DELAYS AND OPTIMISED OFFSET VALUE [119].

In 2010, 10 TDL chains were combined in parallel as a single TDC channel in a 130nm Xilinx Virtex-II Pro FPGA. After averaging, the temporal resolution was improved to 10ps when the bin-width of a single TDL was around 27ps [119]. In 2013, the temporal resolution was improved from 18.38ps to 1.14ps by averaging 16 TDLs together in Xilinx Spartan-6 FPGA [120]. The multiple-channel method improved the bin-width significantly. However, the method also had two main drawbacks, namely the high logic resource cost and the complicated data processing procedure.

c) Wave-Union method

The Wave-Union (WU) method is another inventive method designed to break the process limitation. It was presented firstly in 2008 [121] and had been applied and modified in several works [105, 122, 123]. The WU method is based on the theories of multiple time measurements and averaging. Different from the multiple-chain averaging method, only one TDL is required to perform the multiple measurements each time. A WU is a pre-defined waveform pattern or a pulse train which contains multiple "1 to 0" or "0 to 1" signal transitions. A short segment of a carry-chain, the WU launcher, is used to store WUs and is connected with the main body of the TDL. Once the 'START' or 'HIT' signal triggers, a WU

will be released from the launcher and fed into a followed TDL. By sampling and processing the status of multiple signal transitions in the TDL, an equivalent multiple-channel averaging is performed. The simple version of the WU method, WU-A, has a fixed length and pattern of the pulse train with a limited number of transitions. After adding the feedback with a NAND gate and a delay buffer into the WU launcher, the WU-B, an improved version of the WU method, has an infinite-length repetitive pulse train with unlimited transitions. In this method, the WU launcher is turned into a controllable ring oscillator. The simplified circuits of the WU-A and the WU-B are shown in **Figure 2.18**.
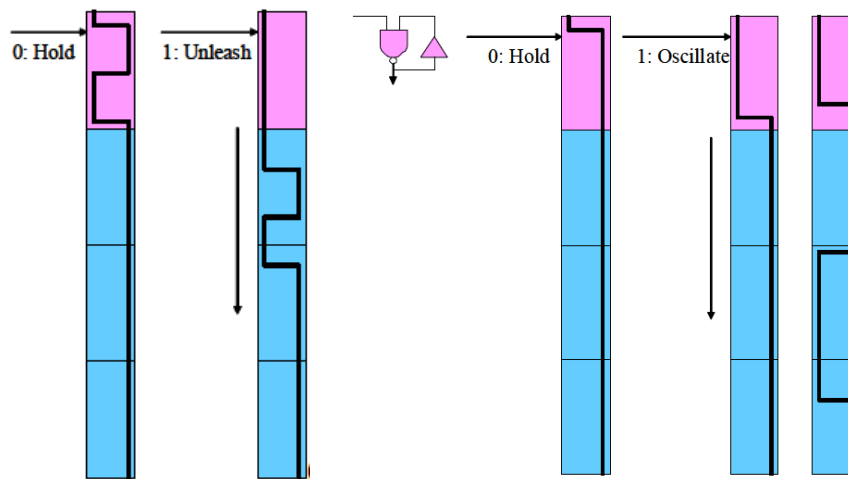
FIGURE 2.18 THE WAVE UNION LAUNCHER A AND THE WAVE UNION LAUNCHER B [121]

By applying the finite WU in the Altera Cyclone II FPGA device, the average bin-width was improved from 60ps to 30ps, and the RMS measurement error was improved from 40ps to 25ps. By using the WU-B, the RMS measurement error was further improved to 10ps [121]. In 2011, a modified WU was implemented in Xilinx Virtex-4 FPGA, and its temporal resolution was improved from 50ps to 12ps. Compared with the multi-channel averaging method, the WU method consumes far fewer logic resources. The disadvantage of the WU method is that the dead-time is increased. For example, in previous studies on WU, the dead-time of the WU-A increased from 2.5ns to 5ns, and that of the WU-B increased to 45ns [121].

d) Other methods

Besides the main methods discussed above, there are also several methods implemented in FPGA chips from the previous studies. In 2010, a two-stage pulse-shrinking TDC was implemented in Xilinx Spartan-3 FPGA with a 42ps time resolution and an 11.5ns MR [124]. Zhang (2017) used the interconnection routing instead of the carry-chains to implement the delay cells of TDLs [125]. In 2010, the IDELAY module, a programmable TDL with a delay step of approximately 75ps, was applied to achieve TDC in Xilinx Virtex-4 FPGA [126].

2.5.2 Dead-time and sampling rate

Full-digital TDCs, especially FPGA-TDCs, have a tremendous advantage in dead-time compared with the analogue solutions. For carry-chain based TDCs, the 'START' or 'HIT' signals are allowed to enter TDLs unimpeded and continuously. However, for a typical encoding procedure of FPGA-TDCs, each TDL is only capable of measuring one event in each sampling period. Thereby, TDLs are ready for the next signal immediately once the current signal status has been sampled. The D-FFs for sampling will be ready for the next sampling within a short period, which is related to the setup and the hold-on time of D-FFs, which is around a few hundred picoseconds [115, 116, 127]. This latency is far less than the period of the 'STOP' signal or the sampling clock, and it cannot block the next hit signal access and propagate along the TDL. Therefore, if the sampled data can be read out timely or buffered before the next sampling, the extra dead-time of a single TDL is negligible. The dead-time of FPGA-TDCs is related to the encoding and the readout operation speed. By using the pipeline structure and FIFO buffering operation [128], this effect can also be ignored for most of the applications, because the next measurement results can be buffered temporarily when the current results are processed. In employing this approach, the whole system will be immediately available for the next measurements if the total latency of the results encoding, processing and readout are shorter than the sampling period of the TDCs. Therefore, the maximum sampling rate of a single TDL is normally equal to the frequency of the sampling clock, which is limited to hundreds of MHz [13, 129, 130].

2.5.3　Multiple-channel FPGA-TDCs

The last few years have seen a growing interest in multiple-channel TDCs for high-speed parallel measurements and fast imaging applications such as LIDAR, wide-field FLIM and ToF-PET. In order to integrate a number of TDC channels into a single FPGA device, factors including circuit size, complexity, design density, resource consumption and Place&Route congestion must be considered. Most analogue TDCs or TCSPC systems have only achieved the single-digit numbers of channels, such as the PicoHarp 300 TCPSC system. This limitation is mainly since the core components of TCSPC, such as capacitors and ADCs, require dedicated circuits which restrict reducing of size and power consumption.

The full-digital TDCs can achieve much higher integration level in ASIC and FPGA devices compared with the analogue approaches. By applying the ring oscillator TDCs, a single ASIC can integrate up to tens of thousands of TDC channels [12, 67]. Nevertheless, temporal resolution, measurement range and linearity are limited. The readout speed is also a considerable challenge for a large-scale TDC array. Since the number of IOB is always limited to around tens to more than a hundred, the readout operation of the measurement results of TDCs is inevitably arranged in a serial mode, which encumbers the operation speed of the whole system.

FPGA TDCs can balance the number of channel and performance parameters to fit the applications. The carry-chain is an abundant logic resource in every CLB. By using carry-chains, the FPGA-based multiple-channel TDCs can integrate tens to hundreds of TDC channels depending on the scale of FPGA chips and the specification of TDCs [13, 130]. One of the key advantages of the FPGA-based multiple-channel TDCs is that the temporal resolution does not need to be compromised, and the high-throughput data can be pre-processed and compressed within the FPGA firmware to reduce the readout pressure effectively.

2.6　　　Summary

This chapter performed literature reviews of different implementation types of TDCs or TCSPCs. The advantages and disadvantages of analogue methods and digital methods base on ASIC and FPGA devices are summarised in **Table 2.2.**

Table 2.2 The summary of different implementation methods

	Advantages	**Disadvantages**
Analogue	Excellent resolution (<1ps) and linearity performances. Verified	Bulky, complicated and expensive. Long dead time (<95ns). Limited channel numbers (<10 channels).
ASIC	Good resolution (~1ps) or linearity Highest integration. Ideal for mass production and general-purpose devices. Capable for large-scale TCSPC array.	Long development period and high upfront cost. Non-updatable.
FPGA	Highest flexibility and compatibility. Short development period and lowe development cost. Good integration and design portability. Low dead time (<10ns).	Limited resolution (<10ps). Poor linearity.

The analogue methods mainly implement current commercial TDC and TCSPC devices used for scientific experiments. That is because the analogue methods provide best resolution and linearity performances, and their reliability and stability have been widely verified. However, the size and price of analogue TCSPC devices make they are mainly used in laboratories. ASIC-TDC or TCSPC has the highest integration, good resolution and linearity benefit from fully customised design approaches. Considering the considerable upfront cost and long developing period, ASICs are ideal for mass-produced general-proposed products.

Chapter 2: Literature review of TDCs

The FPGA based designs can provide great flexibility and compatibility, and much lower development cost and period. Moreover, FPGA-TDC/TCSPC can be integrated with other designs seamlessly. These features make FPGAs are perfect for scientific experiments, fast prototyping, and high-end instruments. The manufacturing processing of FPGAs limits the temporal resolution. With the rapid development of FPGAs during recent years, the resolution performance gap with ASIC-based TDC/TCSPCs has been narrowed significantly. The linearity performance of FPGA-based TDC/TCSPC designs is the main disadvantage compared with other types. Especially, their poor static linearity will cause various problems, including the missing codes, ultra-wide bins and 'bubble' problems. The nonlinearity impact the overall measurement precision and these accompanying problems will lead to encoding failures.

Chapter 3: The nonlinearity of FPGA-based TDC designs

The nonlinearity of FPGA-TDCs directly impacts on measurement accuracy, as illustrated by **Equation 2.2**, and cannot be removed fundamentally. Different from analogue and ASIC-based TDCs, which are entirely customisable, meaning their linearity can be improved significantly, FPGA-TDCs, due to their pre-defined architecture, have suffered more from the problem of poor linearity performance. The nonlinearity of FPGA-TDCs can be classified into two categories, namely dynamic and static nonlinearity. This chapter will mainly focus on the sources and solutions of dynamic and static nonlinearity, especially the static nonlinearity and its related issues, including missing-codes, ultra-wide bins, and the 'bubble' problem.

3.1 Dynamic nonlinearity

The dynamic nonlinearity of FPGA-TDCs is caused by the deviations of the manufacturing process and the drift of supply voltage and temperature (PVT) [19, 131]. In mature commercial FPGA devices, influences from manufacturing process variations are negligible and tend to be a static influence after production. FPGA devices are mass-produced by mature and stable manufacturing processes. Each FPGA will be thoroughly tested, characterised, and then classified into three speed grades. Tapped delay line (TDL)-based FPGA-TDCs with different speed grades have a slight difference in temporal resolution, which can be quantised using the time interval test or the double registration method.

3.1.1 Power supply

Regarding the voltage jitter of a power supply, FPGAs have extremely high standards.

Different modules and parts are supplied dividedly, with extremely high requirements in the voltage stability to maintain the expected performance of FPGAs. Take the latest Xilinx Ultrascale devices as an example; power supply inputs are classified into 3 main classes and 20 subclasses, while the tolerance ranges of supply voltages are from 22mV to 30mV [116]. In actual cases, the stability of the supply voltages should be better than the toleration range, since peripherals and external circuits will introduce extra interferences and jitters. By using integrated power management ICs (PMICs) and the power system on a chip (SoC) which contains low dropout regulators (LDOs), embedded power controllers (EPCs) and a DC-DC converter, power noise has been effectively restrained [132]. For example, an Intel EM2130 Power SoC can provide 0.5% output accuracy and lower than 10mV peak-peak output ripple. With 1V voltage, the steady-state accuracy of supply voltage is equal to $0.5\% \times 1V + 0.5 \times 10mV = 10mV$. Song (2006) also demonstrated that the effects on the linearity of FPGA-TDCs are negligible within a ±5% range of a normal supply voltage [133].

3.1.2   Temperature

Temperature drift will affect the propagation speed of internal FPGA circuits and the temporal resolution of TDCs. Various methods and procedures have been published to correct and calibrate the resolution drift caused by temperature variation [92, 103, 121, 134]. A TDL-TDC implemented in an FPGA was tested under a controlled ambient temperature range spanning from 0℃ to +50℃, and the resolution was changed from 196.1ps at 0 ℃ to 204.1ps at +50℃ [92]. Szplet increased the temperature drift range, from -20℃ to +60℃, and the resolution was increased linearly by around 0.5ps/℃, shown in **Figure 3.1**.
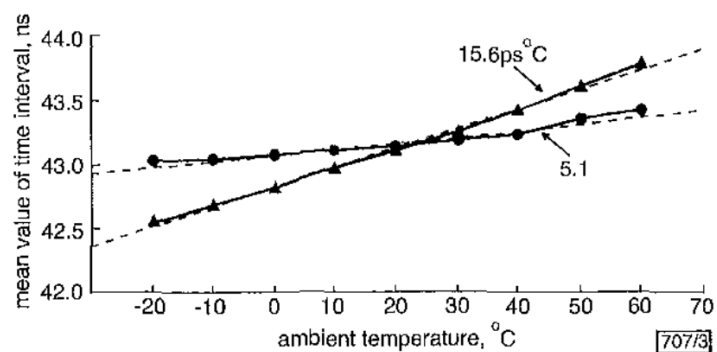
Fig. 3 *Mean value of measured time interval against ambient temperature*
Sample size: 5000
▲ without DLL
● with DLL

FIGURE 3.1 MEAN VALUE OF MEASURED TIME INTERVAL AGAINST AMBIENT TEMPERATURE [103]

The influence of the temperature drift can be reduced by using DLLs in FPGAs to detect the changes in the propagation speed of TDLs under different temperatures and adjust the FPGA power supply voltage as compensation. Since the propagation speed of delay cells in FPGAs or other CMOS devices is in inverse proportion to the supply voltage, the resolution drift can be corrected or compensated to some degree [103]. The test results of the method are demonstrated in **Figure 3.2**.
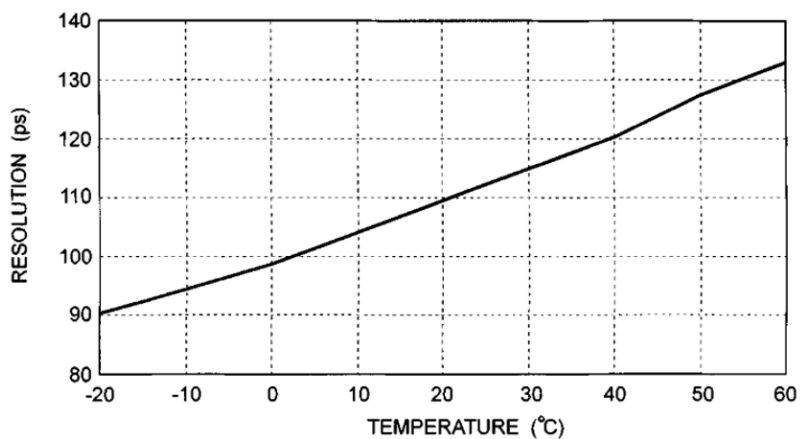
FIGURE 3.2 RESOLUTION OF THE TIME-CODING DELAY LINE AS A FUNCTION OF THE AMBIENT TEMPERATURE [103].

Digital calibration and compensation methods are feasible for FPGA-TDC designs. One simple calibration method, the double registration method, was proposed in 2003 [121, 134]. This method can detect the total propagation delays and calculate the resolution of TDLs in operation. The total propagation delay of TDLs needs to be configured so that it is slightly longer than the period of the coarse clock or the sampling clock, Pclk. Thereby, some 'START' or 'HIT' signals will be sampled and registered twice in the same TDL and generate two binary codes, N1 and N2. As such, the temporal resolution or averaged bin-width of the TDL can be calculated as LSB= Pclk /(N2-N1) and used to convert to actual time values. The advantages of this method include the capability of the on-the-fly resolution compensation and the fact that additional processing time and logical resource cost are minimal. However, this method is unable to detect resolution drifts smaller than 1 LSB.

3.2   Static nonlinearity

The static nonlinearity of TDL-based FPGA-TDCs is caused by the nonuniformity of TDLs, and clock distribution skews. It mainly impacts the short-term stability and the measurement accuracy of TDCs. Compared with ASIC-TDCs, FPGA-TDCs usually demonstrate much worse static linearity performances, since the placement and routing of FPGA internal circuits cannot be manually adjusted after delivery. Furthermore, the design philosophy of FPGAs is to achieve optimised performance for general-purpose applications. Therefore, many structures, such as clock distribution trees and lookahead-carry logic architectures, are used to reduce the overall signal delay and skews. However, these advanced structures introduce more serious nonlinearity problems to TDL-based FPGA-TDC designs.

3.2.1   Influence of clock distribution skews on TDC linearity

Clock signals are the most critical signals in FPGAs designs since the majority of FPGA applications are built by sequential logic circuits which are synchronised by the clock signals and distributed to FFs and hard-core modules in the whole area of FPGA chips. The distribution skews of clock signals in different locations should be minimised as far as possible to increase the operational frequency. To reduce the overall clock skews and propagation delays, FPGA chips are divided into many clock regions (CRs) and corresponding dedicated clock routes (clock distribution tree). Furthermore, various dedicated buffers are utilised to increase the number of fanouts and enhance drive capability.

Figure 3.3 shows the block diagram of the internal clock distribution tree and TDL structure in a Virtex-7 (XC7VX690tFFG1761) FPGA [135]. This device is segmented into 20 CRs (2x10 of CRs from X0Y0 to X1Y9), and each CR contains 50 rows of CLB. A vertical clock backbone is located in a joint between two CR columns. The FPGA chip has 32 global clock buffers (BUFGs and BUFGCTRLs) near the centre of the clock backbone. BUFGs and BUFGCTRLs are ideal, most frequently and primarily used global clock buffers for designs with the requirement of extensive area access. After accessing BUFGs, clock signals propagate along the backbone to the top and bottom parts of the chips. Following this, the clock signals can access each CR and propagate horizontally through the horizontal clock buffers (HBUFGs) at the clock backbone and the middle position of each CR. The clock signals are then branched vertically into the up and bottom side of each CR and finally delivered to each CLB.
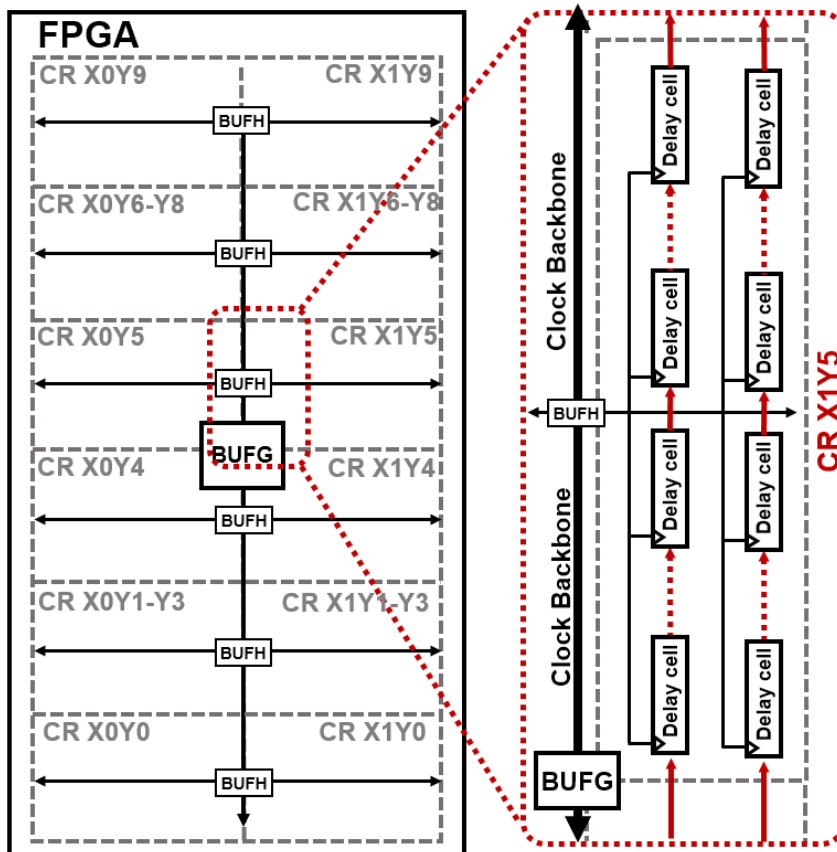
FIGURE 3.3 BRIEF DIAGRAM OF CLOCK REGIONS AND CLOCK DISTRIBUTION ROUTE IN AN FPGA

By applying dedicated and well-designed clock distribution trees with various clock buffers, the overall clock skews around the CR boundaries section are reduced to lower than a few hundred picoseconds, as demonstrated in **Figure 3.4**. For most FPGA-based digital designs with no more than 500MHz system clock frequency, the clock skews can meet the timing requirements. However, the temporal resolution of TDL-TDCs implemented in the recent FPGA devices has reached around 10ps or better. Therefore, the clock skews cause grievous nonlinearity problems, especially at boundaries of two adjacent vertical CRs, since the clock skews jump to more than 100ps at the two sides of the CR boundary [104, 117], as shown in **Figure 3.5**.
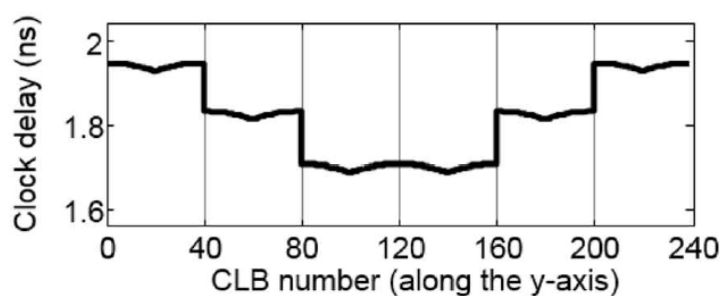
FIGURE 3.4 CLOCK DELAY FOR CLBS LOCATED ALONG THE Y-AXIS. [12]
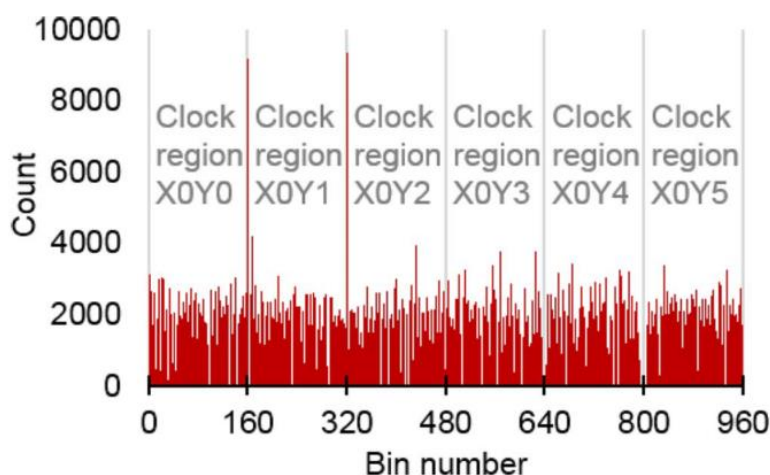
FIGURE 3.5 CODE DENSITY TEST RESULTS FOR A 960-BIN TDL ALONG THE Y-AXIS [136].

Besides the apparent nonlinearity around CR boundaries, the clock skews also exist in the rest of the delay cells in a TDL, and the clock skews between adjacent delay cells range from 0ps to 4ps according to the post place&route simulation of a Virtex-7 FPGA. Considering that the temporal resolution of Virtex-7 FPGAs is around 10ps, these clock skews also impact the uniformity of bin-width. From **Figure 3.4**, it can be seen that this type of clock skew has a particular pattern which is explicable. Delay cells of a TDL within a single CR can be separated into two groups base on the clock distribution: 1) from the bottom boundary up to the CR halfway line and 2) from the halfway line up to the top boundary of the CR. Following the clock distribution route, clock signals are first sent to delay cells located around the halfway line of a CR and arrive at the bottom and top delay cells of a CR in the end. In a TDL, the 'HIT' or 'START' signals are propagated uniaxially from the bottom to the top. Therefore, relative to the 'HIT' or 'START' signals, the clock propagation in the bottom part of the CRs is the

negative direction and is the same direction in the top part of the CRs. Therefore, the actual bin-width of a TDL in the bottom part will be slightly increased and will be slightly reduced in the top part of the CR. The bin-width drifts will contribute to overall static nonlinearity, especially of the INL values.

A basic method to mitigate the nonlinearity of clock skews is to reduce the length of TDLs. However, the frequency of the clock restricts the shortening of TDLs in FPGA-TDC designs. The fine&coarse interpolation [117, 118, 137] and Nutt interpolation [120, 125, 138] are widely utilised in FPGA-TDCs. As a fine interpolator, the total propagation time of TDLs should be not shorter than the sampling clock period to avoid incomplete measurements [136]. The sampling clock is also used to drive the coarse counter, encoding, and buffering modules, as well as the readout and data processing modules. Therefore, the frequency of the sampling clock cannot be increased limitlessly, and the timing requirements of routing and involved modules must be satisfied. Considering that TDLs are placed vertically, and the height of each CR is limited, it is frequent to place TDLs in different CRs and across the CR boundaries in conventional TDL-TDC designs.

With more advanced manufacturing processes (20nm and 16nm) or complicated structures (such as CARRY8), the averaged bin-width of FPGA-TDCs is reduced to 5ps, and even up to 2.5ps [64]. The maximum frequencies of various modules and components in FPGAS are still limited to hundreds of MHz (e.g. the $F_{pll\_out\_max}$ is 630MHz, and the F_{MAX_RF} is 400MHz) [116], while the higher clock frequency will lead to a higher risk of timing failures. Therefore, it is a challenge to shorten the length of a TDL and to ensure that the total propagation time of the TDL is longer than the period of the sampling clock at the same time.

Another method applies the "Dual-Phase" structure [136, 139] to shorten TDLs and avoid a CR boundary with large clock skews. The basic principle of this method is to use two parallel located TDLs to serve one TDC channel. These two TDLs are sampled by two clock signals which have the same frequency and 180-degree phase shift. With this method, the total propagation time of each TDL is only required to cover half the period of the sampling clock. The principle is shown in **Figure 3.6**. This design was implemented on a 40nm process Virtex-6 FPGA, following which a10ps resolution and DNL_{max}=1.91LSB were achieved [136].
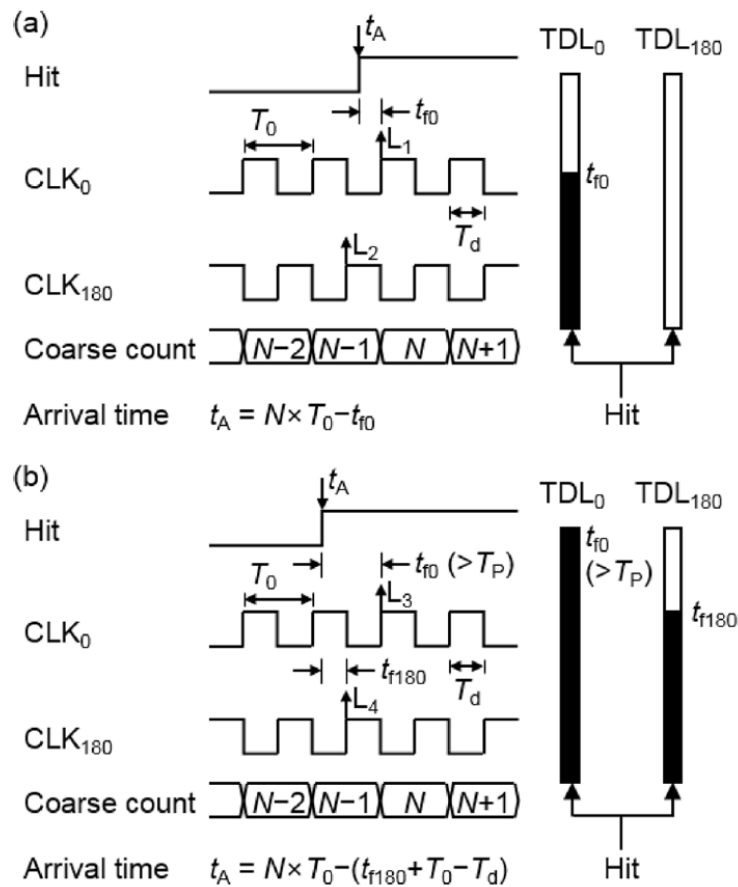
FIGURE 3.6 THE PRINCIPLE OF DUAL-PHASE METHOD. THE TIMING DIAGRAM AND THE SAMPLED STATES OF TWO TDLs ARE SHOWN. (A) THE CASE WHERE A HIT IS ASSERTED WHEN CLK0 IS AT A LOGICAL LOW. (B) THE CASE WHERE A HIT IS ASSERTED WHEN CLK0 IS AT A LOGICAL HIGH [136].

3.2.2 Nonuniformity of carry-chains of TDLs

Compared with the clock skews, serious nonuniformity in carry-chains is a leading resource of nonlinearity in the entirety of TDLs and frequently causes missing-codes, ultra-wide bins and bubble problems.

a) Basic structures of carry-chains in FPGA

Carry-chains consist of multiple cascaded carry-modules which are included in all CLBs [113,

114]. The carry-module uses the fast lookahead-carry logic circuit which, is designed specifically for arithmetic addition and subtraction. In each carry-module, several carry elements are cascaded to form an integrated carry-module. For example, as shown in **Figure 3.7**, each carry element has a multiplexer (MUXCY/MUX) and an XOR gate. A carry signal propagates upwardly through the MUXCY. Each carry element also has two independent output ports, namely 'CO' and 'O'. Previous FPGA (7 series and previous devices) devices have two separated carry-chains within the single CLB, and each carry-module (CARRY4) contains four carry elements and eight independent outputs per carry-module. The latest FPGAs (UltraScale and UltraScale+ series) have only one carry-chain within each CLB. Every chain module (CARRY8) contains 8 carry elements and 16 independent outputs per carry-module [114].
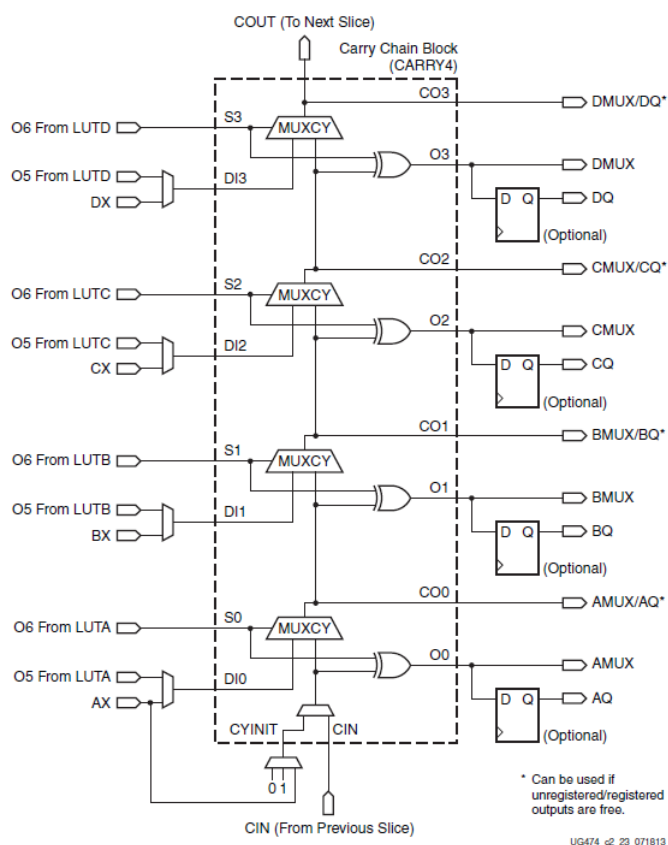
Figure 2-24: **Fast Carry Logic Path and Associated Elements**

FIGURE 3.7 FAST CARRY LOGIC PATH AND ASSOCIATED ELEMENTS [114].

In each CLB, a group of D-FFs is used as synchronisation elements and registers. The data

input port (D) of each D-FF can be connected with one of the two independent outputs of corresponding carry elements. The clock ports of D-FFs are connected with the 'STOP' signals or the sampling clocks to trigger the sampling of 'START' or 'HIT' signal status in carry-chains. By combining cascaded carry-modules (CARRY4/8) and D-FFs, a typical TDL can be implemented. However, the dedicate fast lookahead-carry architecture is widely applied in the carry-modules of FPGAs, which introduces a high nonlinearity into the TDCs.

b) Fast lookahead-carry architecture

The fast lookahead-carry architecture can significantly improve the operation speed compared with that in traditional ripple-carry adders circuits. As shown in **Figure 3.8,** classic ripple-carry architecture is a train of cascaded 1-bit full adders, and each adder has a carry-in and a carry-out port (C_0 to C_4) connecting with adjacent adders in order. An advantage of the ripple-carry adder circuits is that they consume less area and power. However, the operation speed of the ripple-carry adder circuits is limited since final carry output will not be valid until the carry-bit arrives at the final adder.
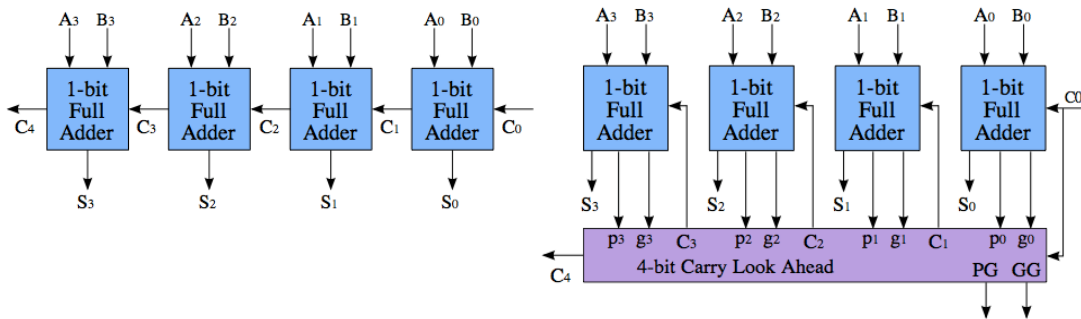
FIGURE 3.8   THE SIMPLIFIED DIAGRAM OF CLASSIC RIPPLE-CARRY ADDERS (LEFT) AND FAST LOOKAHEAD-CARRY ADDERS(RIGHT)

The lookahead-carry architecture creates an independent path for carry-bits (the 4-bit carry lookahead module), and all full adders are independent, as shown in **Figure 3.8**. Carry-bits can be calculated parallelly, and the propagation route and delay of carry-bits can be minimised in the lookahead-carry architecture. As a result, the propagation latency of the lookahead-carry architecture is relatively shorter than the classic ripple-carry structure, and this advantage is

more distinct in larger-scale cascaded adders. However, for FPGA-TDCs, the architecture generates more timing and linearity problems due to two factors. First, the carry-modules are separately placed in adjacent CLBs. Since the interconnection routes are located around every CLB, the propagation delays between two CLBs or two carry-modules (external delay) are much larger than the internal delay within carry-modules. The dedicated lookahead-carry architecture is applied to compensate for this sizeable difference between the external and internal delay of carry-chains in order to reduce overall operating latency [113]. However, more problems of linearity are introduced by this architecture.

The timing characters in carry-chains can be investigated via software simulation, the code density test, and the proposed tap timing test. Simulation is the simplest way to evaluate the timing of FPGA designs, which is based on the ideal and universal timing information provided by manufacturers. From the results of the simulation, researchers can infer a rough idea of the timing relationship in carry-chains. However, the results from the simulation cannot be used as an accurate quantitative reference for the actual timing situation, especially for high-resolution TDC designs. That is because the universal timing information excludes processing and device deviations, as well as timing deviations in individual delay modules.

The post place&route simulation considers the delays of circuit routes and can provide more realistic timing information compared with the behavioural simulation. **Figure 3.9** is a post-place&route simulation result of a TDL in a Xilinx Vertex-7 device, and the figure demonstrates part of the rising edge timing of carry-out signals. The result shows that the propagation order of these carry-out signals is nonmonotonic, which does not match with the circuit diagram in the datasheet (shown in **Figure 3.7**). The last carry-out signals of CARRY4s (CO[7], CO[11]…) are much faster than the first carry output port of CARRY4s (CO[4], CO[8]...). However, as shown in **Figure 3.7**, the cascaded structures of CARRY4s in the datasheets cannot explain this disorder phenomenon. According to the simulation, the average internal delay of CARRY4 is 26ps, and the external delay of CARRY4 is 127ps. One of the reasonable interpretations of the simulation result is that the actual routes in the CARRY4 lookahead-carry circuit are in parallel, and the last carry element has a specifically-designed route allowing carry-bit signals to be transmitted to the next CARRY4 module in advance. In this way, carry signals are fed into the external carry route earlier to compensate for the external delay. When multiple carry-modules

are cascaded as a carry-chain, this feature helps to reduce the total propagation time significantly.
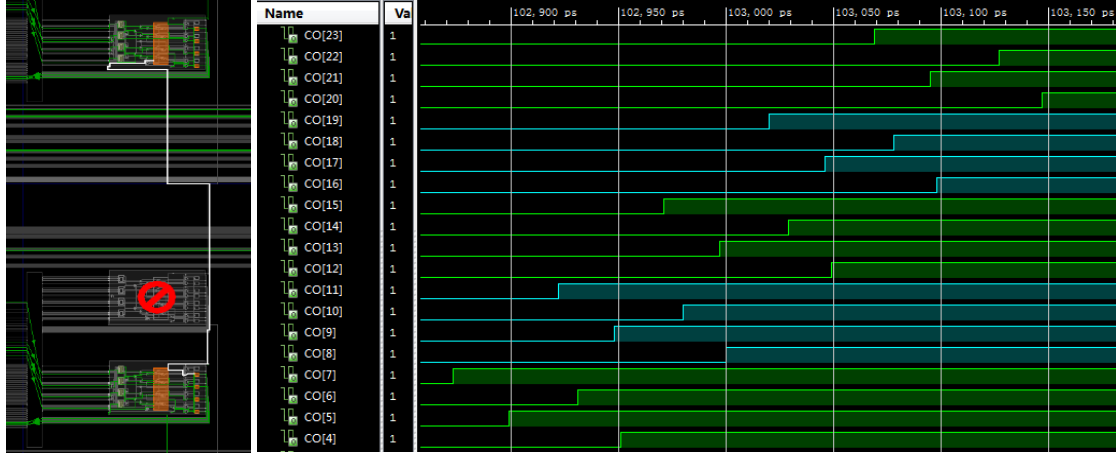
FIGURE 3.9 A ROUTE OF EXTERNAL CARRY (LEFT)A TYPICAL POST-PLACE&ROUTE SIMULATION RESULT OF A CARRY-CHAIN FRAGMENT (RIGHT)

c) Nonlinearity problems caused by the fast lookahead-carry architecture

For carry-chain-based TDL-TDCs, the fast lookahead-carry architecture improves the resolution by increasing the overall propagation speed. However, the nonuniformity of this dedicated architecture generates more serious nonlinearity problems, including missing-codes, zero-width bins, ultra-wide bins, and the bubble problem. Missing-codes appear when the width of a code bin is shorter than one-tenth of the mean width of all code bins, which can be defined as [73]:

$$DNL[k] \leq -0.9 LSB$$

Some of the missing-code bins will be classified as zero-width bins (DNL[k] = -1 LSB) when their width is not sufficient to capture any event, or some functional failures appear. Zero-width bins are noneffective bins, which will reduce the number of effective bins and can thus be abandoned. Regarding ultra-wide bins, there is no unified and precise definition. In this case, the bins which are wider than 3LSB (DNL>2LSB) can be classified as ultra-wide bins. The ultra-wide bins will lead to problems such as resolution and low measurement precision loss.

d) The bubble problems and basic bubble-removal and bubble-proof methods

Besides the nonlinearity described above, the disordered timing of TDLs will generate error codes in sampling results, which is termed the 'bubble' problem. The problem is caused by the non-monotonicity of TDLs with the dedicated lookahead-carry architecture [118, 123, 140]. A TDL with ideal monotonicity generates clean thermometer codes after samplings, such as '11110000' or '00001111'. The '1-to-0' or '0-to-1' bits in the thermometer codes identify the propagation distance or the position of captured signal transitions. Following this, the subsequent encoders are responsible for converting thermometer codes to one-hot codes and binary codes. However, the non-monotonicity of the dedicated lookahead-carry architecture will generate multiple signal transitions such as '1101000' or '00001011'. Therefore, these kinds of codes will confuse the actual propagation distance. Moreover, they can lead to the failures of encoding and conversion, since the typical data processes for thermometer codes require clean thermometer codes to avoid functional failures. Consequently, the bubble codes must be removed entirely or recognised.

In previous studies, several bubble-removal circuits or bubble-proof converters were published and applied in FPGA-TDCs to deal with the bubble problem [104, 118, 128, 141]. The simplest bubble-removal circuits consist of an array of OR gates or AND gates, as shown in **Figure 3.10** (a. and b.). These gate arrays are located between the TDLs and the thermometer code for one-hot-code (TM2OH) converters. They can remove 1-bit bubble codes (such as '1110100') directly.
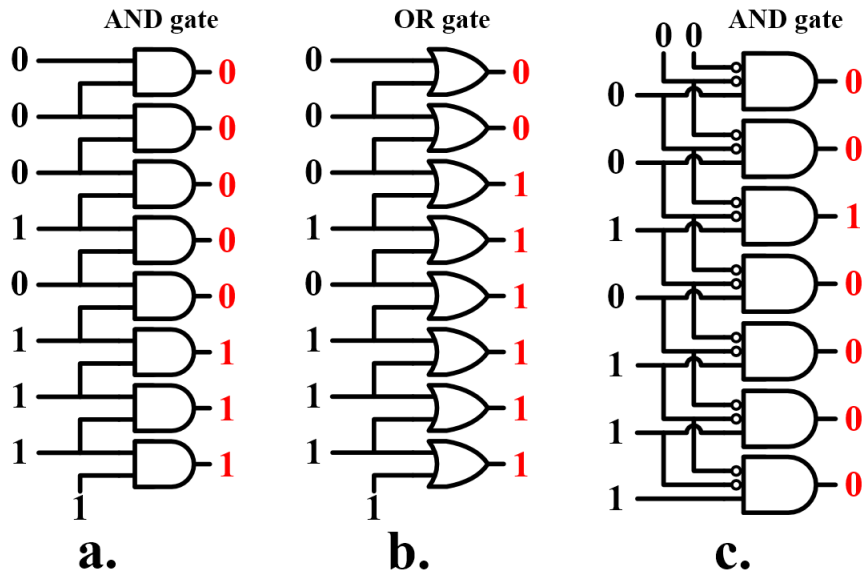
FIGURE 3.10 ARCHITECTURE OF BUBBLE-REMOVAL CIRCUITS (A AND B) OR BUBBLE-PROOF CONVERTERS (C).

Different from the bubble removal circuits, a bubble-proof TM2OH converter was presented in order to detect transition edges by ignoring bubble bits [140]. A basic TM2OH converter consists of an array of 2-input logic gates, and recognises single transition pairs such as '01' and '10'. This kind of TM2OH converter will be interfered with by bubble bits. The bubble-proof TM2OH converter applies stricter standards to identify the position of a single transition each time. **Figure 3.10(c)** shows a basic bubble-proof TM2OH converter which consists of an array of 3-input AND gates with inverters. By using this converter, 1-bit bubbles are ignored, since the 3-bit transition patterns such as '110' or '100' are identified as a real signal transition. The input bits of logic gates in the bubble-proof converter need to be extended when multi-bit bubble codes exist in sampling results.

The advantages of the above methods include simple structures, a low-logic source cost and no additional operation latency involved. Logic operations such as AND, OR and XOR, can be implemented directly by LUTs in CLBs. The delay of LUTs is neglectable, since the signal delay of a LUT is around 50-60ps [115], and the frequency of the synchronisation clock is hundreds of MHz. However, the methods also have a severe disadvantage, specifically that they bring about negative impacts on the linearity performance of TDLs.

These two methods will eliminate a part of measurement details and exacerbate the problems of missing-codes and ultra-wide bins. **Figure 3.11** shows the example of TDL timing to demonstrate how the bubble-proof conversion affects the sampling results and introducing extra missing-codes and ultra-wide bins. In **Figure 3.11**, time bins are obtained by the vertical mapping transition levels of the TDL. Three pseudo-events (labelled as **a, b,** and **c**) appear in order and fall into three adjacent time bins. These events are sampled by the TDL and represented by different thermometer codes, shown in **Figure 3.11**. From the table, it can be seen that the measurement results of events **a** and **c** contain bubble codes. Assuming that a bubble-proof TM2OH converter with an edge recognition pattern of '110' is used to generate the one-hot codes for the following binary conversion, the table in **Figure 3.11** also shows the one-hot code and binary code of the three pseudo-events after applying the bubble-proof converter. Event **b** is indicated wrongly by the binary code of '11(3)', and is stamped by 'Bin 3'. This mistake turns the 'Bin 2' into a zero-width bin since all events falling into the 'Bin 2' category will be redirected to 'Bin 3'. As a result, the equivalent width of the 'Bin 3' category is extended to the total width of 'Bin 2' and 'Bin 3', and an ultra-wide bin is created by this process. Consequently, the linearity of TDCs is damaged as the numbers of missing-codes and ultra-wide bins increase.

Events	a	b	c
thermometer codes	110100	111100	111101
3-input AND gate	010000	000100	000100
binary format	01 (1)	11 (3)	11 (3)
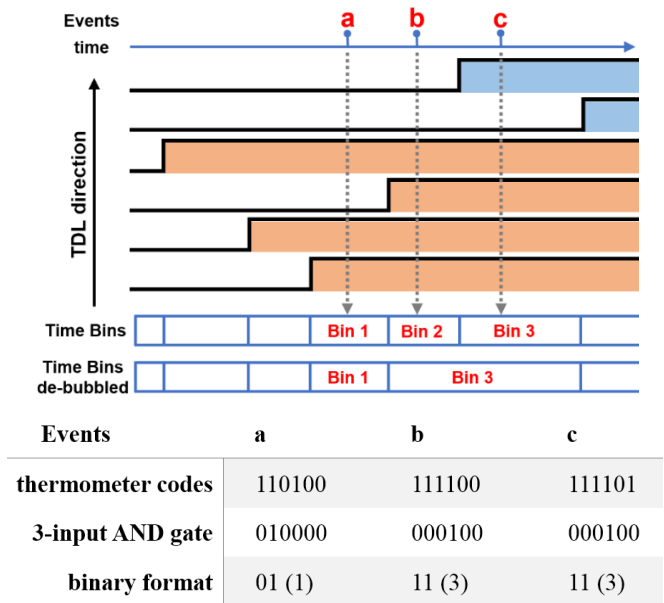
FIGURE 3.11 THE EXAMPLE OF MEASUREMENTS BY A PLAIN TDL WITH THE BUBBLE-PROOF

To reduce the negative effects of the bubble problem, several methods have been published, including downsampling, bin realignment, tuned-delay line and ones-counter encoding. The tuned-delay line is fully described in Chapter 4.

● **Downsampling method**

The downsampling method improves the linearity at the cost of increasing the width of TDL bins and reducing the time resolution. One of the approaches reduces the number of the output bit of TDC directly by ignoring its partial taps in TDLs and keeps the total preparation time of TDLs unchanged. In [104], the INL value was improved from [-3, 2.58] LSB to [-0.49, 1.18] LSB by taking one out of four taps of each CARRY4 and abandoning the remaining taps. Moreover, the missing-codes were almost removed according to the results of the code density test shown in **Figure 3.12**. As indicated by the equation of the resolution (**Equation 2.1**), the average bin-width is extended by four times after the downsampling. Therefore, this method is more suitable for applications whose temporal resolution can be compromised.
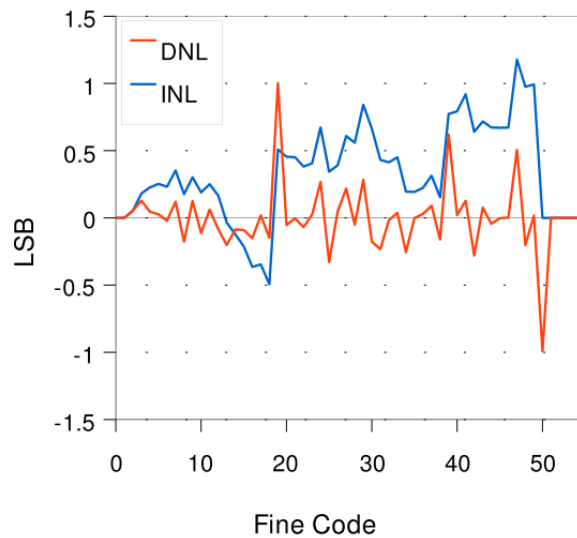
FIGURE 3.12 DNL AND INL AFTER DOWNSAMPLING BY 4 (1 TAP PER SLICE) [104].

In 2015, the bin decimation method as an improved version of the downsampling was published [142]. Instead of simply reducing the number of TDC bins, the bin decimation method reassigns and merges original bins into wider bin groups according to the original bins' time property.

The first step is to evaluate the width of the original bins by performing the code density test and calculating the transmit level of each bin as $T[k] = \sum_{k=0}^{i} W_k$. The second step is to calculate the extended bin-width according to the expected number of the bin after the bin decimation. Finally, the original bins can be readdressed and merged into the correct new bin group directly.

● **Bin realignment**

The bin realignment method is designed to solve the disorder problem in TDLs and attenuate the bubble codes without sacrificing temporal resolution [128]. When bubble codes appear in output thermometer codes, disorder positions in TDLs can be detected by sending raw thermometer codes out for software processing. Following this, the detected bubble bits can be exchanged with their adjacent bits until the bubble codes disappear. For example, a thermometer code such as '1110100' can be corrected to '1111000' by exchanging the positions of the 4th and 5th bits. A MATLAB program can be used to analyse the position of bubble bits according to the raw thermometer codes and to generate a look-up-table to readdress the bubble bits in TDL outputs. According to the published results in [128], as shown in **Figure 3.14**, this method can remove zero-width bins and improve the linearity to a certain degree. Both the maximum bin-width and the number of ultra-wide bins are reduced. However, it is difficult to further improve linearity by using this method, and the analysis process of the realignment is kind of an enumeration process which take an enormous amount of time.
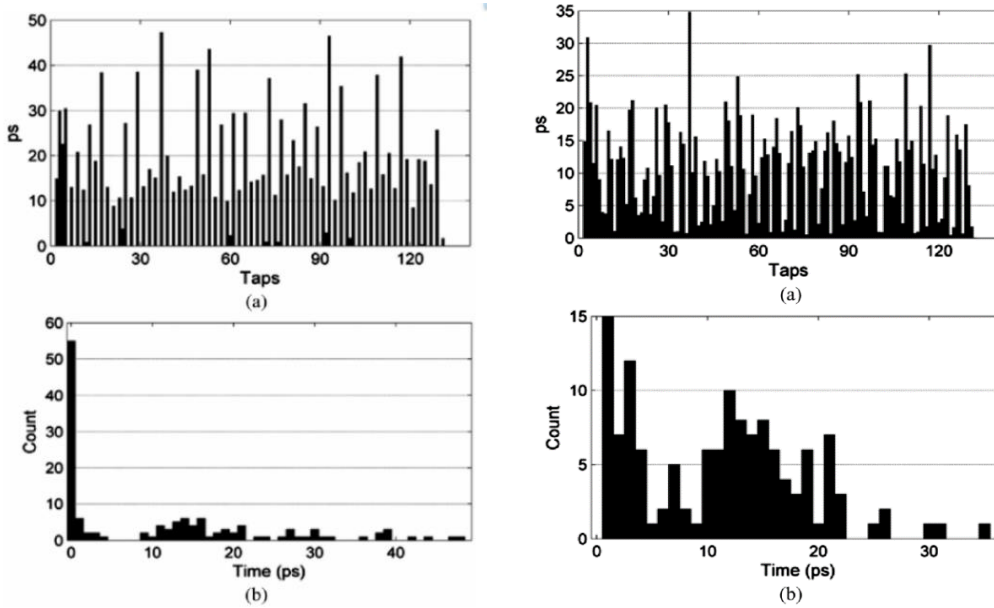
FIGURE 3.13 (LEFT)(A) BIN-WIDTH AND (B) HISTOGRAM OF THE BIN-WIDTH OF A TRADITIONAL FPGA-TDC. (RIGHT)(A) BIN-WIDTH AND (B) HISTOGRAM OF BIN-WIDTHS OF AN FPGA-TDC WITH BIN REALIGNMENT [128].

- **Ones-counter encoding**

The ones-counter encoding tool is a more complicated bubble-proof method which is inspired by a similar method applied in flash-ADCs [143, 144] and implemented by Xilinx Kintex-7 FPGA in 2017 [64]. The basic principle of the encoding is to counter the number of high-level bits in the output thermometer codes of TDLs. LUT primitives in CLBs are used to implement the ones-counter encoding in FPGA devices. Via this approach, the bubble problem can be solved correctly with the robustness of various bubble sizes, and all the bubble bits will be counted. Take the case shown in **Figure 3.11** for comparison; events **b** and **c** can be distinguished from each other by using the ones-counter approach instead of being mixed up, as would happen with those conventional bubble-proof encoders. Compared with other methods, the ones-counter approach can provide better linearity of TDCs since no extra nonlinearity problems or temporal details loss are introduced during de-bubbling. However, ones-counter encoding brings with it a notable additional logic resource cost and longer operation time. For example, in [64], the ones-counter encoder for a 10-bit TDC requires a 9-stage logic circuit, and alongside this, the operation of encoding costs nine clock cycles, as shown in **Figure 3.12**.
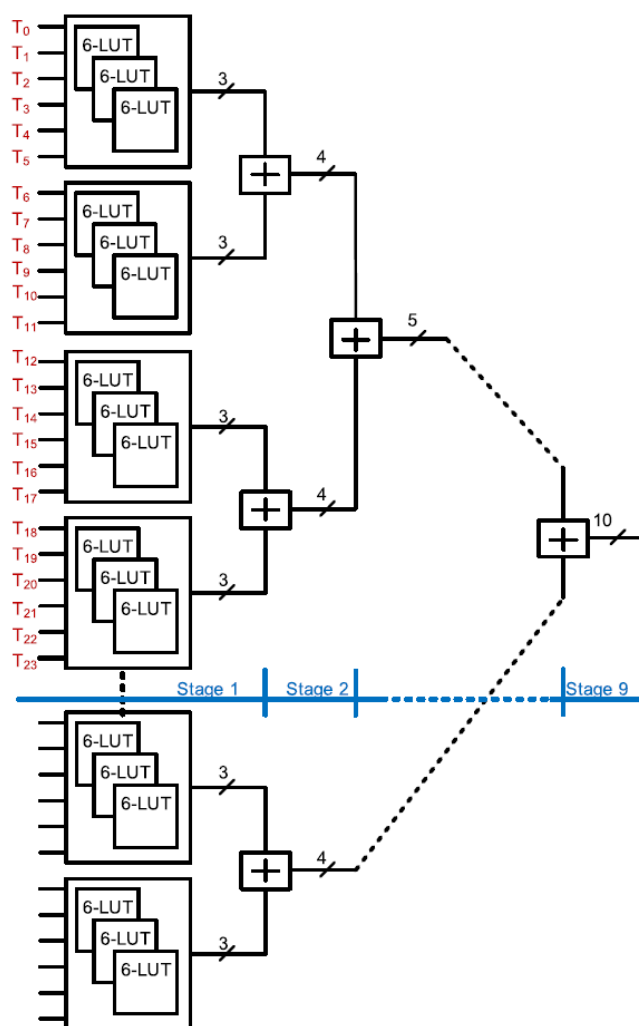
FIGURE 3.14   **FUNCTION BLOCKS OF THE ONES-COUNTER ENCODER WITH A PIPELINE STRUCTURE [64]**

3.3   Summary

This chapter has thoroughly analysed the dynamic and static nonlinearity of FPGA-based TDL-TDCs. The dynamic nonlinearity caused by the PVT variations is not a dominant factor for the poor linearity of FPGA-TDCs. The deviation of the manufacturing process is negligible since the quality, performance characteristics and specifications of FPGAs are strictly controlled and verified. By applying commercial power solutions, the voltage jitter of the power supply is

minimised to a relatively low level. The influences from the temperature drift are linear, and can be effectively monitored and compensated by the dual-register or LUT-based correction methods.

Compared with dynamic nonlinearity, static nonlinearity is a dominant reason for the unsatisfactory linearity performance of FPGA-TDCs. **Table 3.1** summarises part of published FPGA-TDC designs within the last two decades. The TDL is a mainstream structure designed to implement FPGA-TDCs. Of note here is severe nonlinearity caused by the nonuniformity of TDLs and their concomitant problems, including missing-codes, ultra-wide bins and the bubble problem. These existing methods for nonlinearity problems, such as bin realignment and tuned-TDL, are not able to provide significant improvements in the aspect of linearity.

Nevertheless, FPGA-TDCs are very competitive based on other performance parameters compared with ASIC or analogue TDCs. FPGA-TDCs with temporal resolutions lower than 10ps were presented recently. Measurement ranges can be effectively extended by using the Coarse&Fine counter architecture. FPGA-TDCs also show superiority with regard to sampling rate and dead-time. To expand the application fields of FPGA-TDCs, there is an urgent need to address the nonlinearity problems.

Table 3.1 the summary of published FPGA-TDC designs

Year	Architecture	Devices	LSB	Accuracy	DNL	INL	MR/DR	SR/Dead-time
1997[92]	VDL	QuickLogic	200.0	N/S	[-0.50, 0.50], 1.00	[-0.20, 1.40], 1.60	10ns	100MHz
2000[103]	Time-coding delay line	QuickLogic	100.0	70.00	-	-	43s	200MHz
2003[134]	TDL	ACEX 1K	400.0	130.00	-	-	19.2ns	70 MHz
2006[133]	TDL, fine&coarse	ACEX 1K	112.5 ps	129.40	[-0.42, 0.78], 1.20	[-0.60, 0.69], 1.30	-	5ns
		Virtex-II	69.5ps	93.10	[-0.95, 1.05], 2.10	[-2.00, 1.86], 3.90	-	10ns
2008[121]	TDL, WU-A	Cyclone-II	30.0	25.00	Max: 1.17LSB	-	1.92ns	400MHz
	TDL, WU-B	Cyclone-II	N/S	10.00	-	-	-	400MHz
2009 [145]	VDL, Nutt	Spartan-3	75.0	300.00	[-1.00, 2.50], 3.50	[-2.50, 3.00], 5.30	56.32us	125MHz
2009 [104]	TDL, Turbo Mode	Virtex-5	17.0	24.20	[-1.00, 3.55], 4.55	[-3.00, 2.58], 5.58	50ns	300MHz
2009 [146]	Vernier-RO	Stratix II	41ps	-	<±0.5 LSB	<±1 LSB	2.6ns - >22ns	22.4 ns
2010 [137]	TDL	Virtex-4	51.5	25.00(RMS)	[-0.40, 1.40], 1.80	[-1.30, 1.70], 3.00	-	10ns
2010 [124]	Pulse-shrinking	Spartan-3	42.0	56.00	[-0.98,~0.5]	[4.17,~3.5]	11.5 ns	710ns
2011 [147]	TDL, Dynamic reconfiguration	Virtex-II Pro	50.0	43.00	[-0.80, 1.90], 2.70	[-2.20, 1.60], 3.80	10ns	100Mhz
2012[148]	Vernier-TDL, Nutt	Kintex-7	22.7	-	2.6 LSB	3.4 LSB	5.24 us	30ns
2013 [117]	TDL	Virtex-6	9.8	19.60	[-1, 1.50], 2.50	[-2.25, 1.61], 3.86	>100us	300MHz
2014 [149]	TDL, direct histogram	Virtex-5	16.3	N/S	[-0.90, 3.00], 3.90	[1.50; 5.00], 6.50	2.86ns	6.17 Gs/s
2015 [150]	TDL	Virtex-6	24.0	42.3	[-1, 2.8]	[-3.7, 3.3]	-	(12.5, 18.75)ns
	16-TDL averaging	Virtex-6	1.5	4.2	[-0.70, 0.80], 1.50	[-1.00, 0.70], 1.70		
2015 [142]	TDL, Bin realignment	Kintex-7	17.6	12.70	[-1.00, 0.84], 1.84	[-0.81, 0.87], 1.68	1.4ns	710 MHz
2016 [136]	TDL, Dual-phase,	Virtex-6	10.0	11.03	[-1.00, 1.91], 2.91	[-2.20, 3.93], 6.13	>20ns	400MHz
		Kintex-7	10.6	8.13	[-1.00, 1.45], 2.45	[-1.23, 4.30], 5.53	-	5ns
2016 [151]	TDL, Tuned-TDL	Virtex-6	10.1	9.82	[-1.00, 1.18], 2.18	[-3.03, 2.46], 5.49		
		Spartan-6	16.7	12.75	[-1.00, 1.22], 2.22	[-0.70, 2.54], 3.24		
2016 [64]	TDL, Bin realignment, Dual-sampling	UltraScale	2.25	3.9	-	-	440ns	4ns
2017 [125]	TDL, Matrix of Counters	Virtex-5	7.4	6.8 ps	[-0.74, 0.74]	[-1.52, 1.57]	13.5ns	135.5MHz
2017 [138]	2D Vernier TDL	Stratix IV	2.5	6.72 ps	[-0.56, 0.46]	[-2.98, 3.23]	2~16.5ns	50MHz
2017 [152]	Multi-chain TDL integrated	Virtex-7	1.15	3.5	[-0.98, 3.5]	[-5.9, 3.1]	up to 1s	238.5Gs/s

Chapter 4: Low-nonlinearity, multiple events, direct histogram FPGA-TDC

4.1 Motivation

The temporal resolution of FPGA-TDCs has been greatly enhanced in recent years [64, 118, 138, 150, 152]. Based on the summarised performance of published FPGA and ASIC TDC designs in **Table 2.1** and **Table 3.1**, there is not a huge performance gap regarding the temporal resolution between FPGA and ASIC solutions. However, as described in Chapter 3, the nonlinearity performance of FPGA-TDCs is much worse than that of ASIC-TDCs. Conventional carry-chain based FPGA-TDCs suffer from missing-codes, ultra-wide bins and the 'bubble' problem. Previously published methods, such as the ones-counter, downsampling and bin realignment, have been used to solve the 'bubble' problem and improve linearity. However, the effects of these methods are unsatisfactory and require compromises on other parameters such as resolution, resources consumption and operation latency. As a result, there is a demand for new methods to improve linearity significantly with minimised sacrificing on other parameters.

This chapter presents a new combinational architecture and a hardware-friendly calibration to overcome the performance drawback of linearity in FPGA-TDCs. The development, testing and evaluation of the new architecture are also fully described.

4.2 Methods and architectures

This design, for the first time, combines the tuned-TDC [151], a modified direct-histogram architecture and a multi-sampling method to break through the nonlinearity problems in FPGA-TDCs. Novel hardware-friendly calibration methods are invented based on the proposed

62

architecture to improve linearity further. To extend the MR of TDCs, the Coarse&Fine interpolation structure is applied in this design.

4.2.1   Tuned-TDL structure

The tuned-TDL is designed to suppress the nonlinearity in carry-chains by trimming the timing of carry-modules [151]. The tuned-TDL selects the signal pathways of carry-elements in TDLs to improve the uniformity of carry-chains. This method neither requires any complex and additional encoding circuits nor a time-consuming analysis process.

Figure 4.1 presents the CARRY4 modules along with their circuit diagram as an example. Each CARRY4 contains four carry elements (MUXCYs) and can be cascaded with lower and upper CARRY4s as a carry-chain via a CIN and a COUT port. The MUXCY is a two-input multiplexer with a selection signal 'S'. Four MUXCYs can be connected in series by setting S signals (S0-S3 = '1'). Each MUXCY has two carry-out ports with different propagation delays. The port 'O' is outputted from a 2-input XOR gate. One input of the XOR gate is connected with the output of the lower MUXCY, and another input is connected to the signal 'S' of the corresponding MUXCY. Considering that signals 'S0' to 'S3' must be '1' to ensure that all of the MUXCYs are cascaded, the O-type outputs need to be inverted before used. The port 'CO' is connected to the output port of the MUX. These two pathways are multiplexed by a subsequent multiplexer, and only one of them can be registered by a D-FF located in the same CLB.
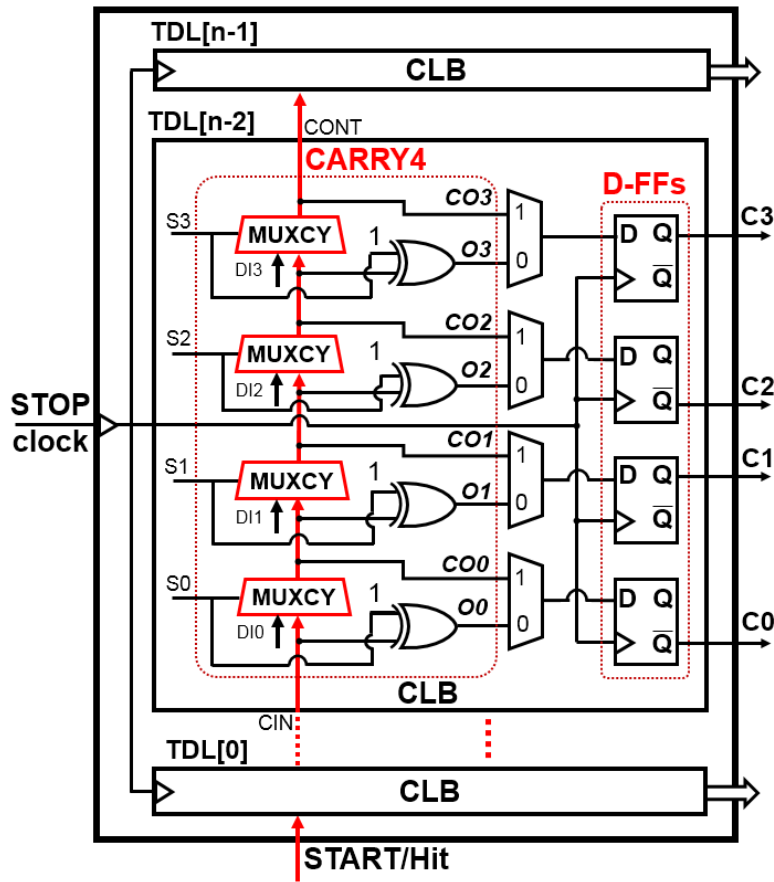
FIGURE 4.1 THE BLOCK DIAGRAM OF CARRY4 MODULES

A conventional carry-chain-based TDL selects the same type of carry-out ports (homogeneous) as its taps. The tuned-TDL method uses different carry-out types (heterogeneous) to achieve a better temporal uniformity. For CARRY4s, there are 16 possible combinations of output selection patterns. As discussed in Chapter 3, the simulation results cannot be used to precisely calculate the actual timing property of individual carry-out ports. Won et al. tested and evaluated the linearity of the different composite patterns in different FPGA devices in 2016 [151]. According to the DNL and INL results and their standard deviation, shown in **Figure 4.2 (left)**, the linearity was improved by a certain extent after applying the tuned-TDL method. From the bin-width distributions in **Figure 4.2 (right)**, it can be seen that the number of zero-width bins, missing-codes and ultra-wide bins has been reduced, and the shape of the bin-width distribution is closer to the Poisson distribution. However, the zero-width bins, missing-codes and ultra-wide bins still cannot be removed completely.
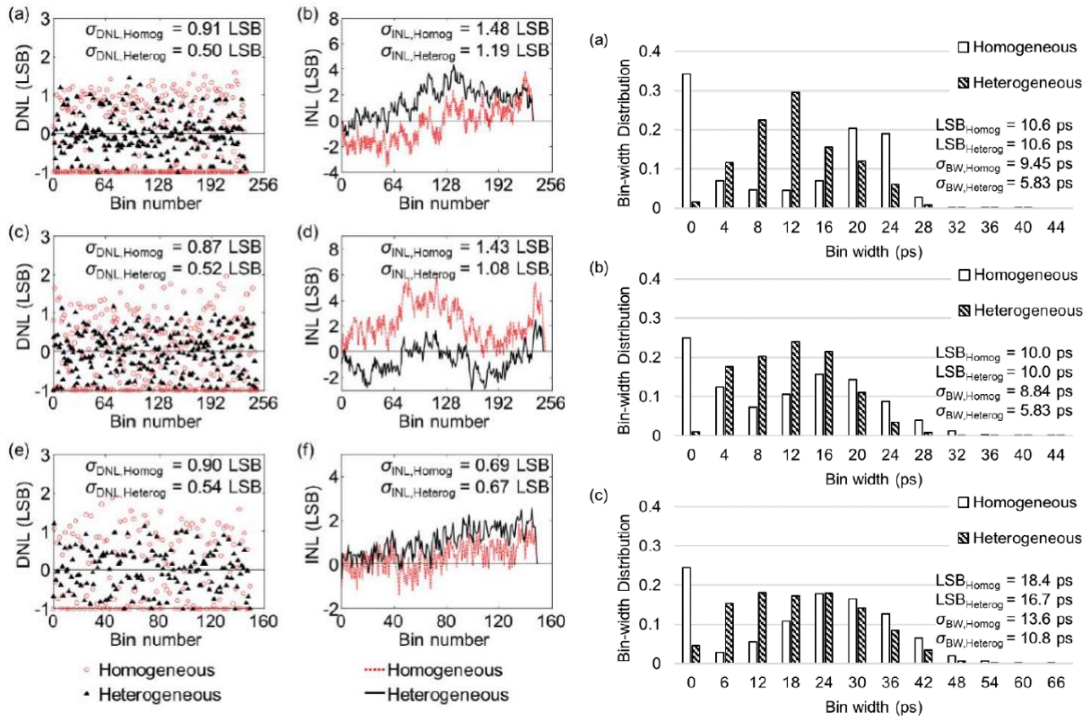
FIGURE 4.2 (LEFT) DNL, INL AND STANDARD DEVIATION OF THE HOMOGENEOUS AND HETEROGENEOUS TDCs. (A AND B) KINTEX-7. (C AND D) VIRTEX-6 (E AND F) SPARTAN-6. (RIGHT) BIN-WIDTH DISTRIBUTIONS AND STANDARD DEVIATION OF THE HOMOGENEOUS AND HETEROGENEOUS TDCs. (A) KINTEX-7. (B) VIRTEX-6. (C) SPARTAN-6 [151].

4.2.2 Multiple-event, direct histogram architecture

The multiple-event, direct histogram TDC was proposed in 2014 [7]. The direct histogram integrates the distributed histogram counters into TDCs directly, which provides two unique features compared with conventional architectures: the capability to measure multiple events by a single TDL per measurement, and a new bubble-allowed encoding procedure.

The conventional TDC architectures contain a set of encoders to convert the outputs of TDLs, namely thermometer codes (1111000), to typical binary codes, which indicate the positions of transitions. This conversion can only deal with one event each time. Therefore, the highest sampling period of the conventional TDL-TDCs equals the total propagation delay of used TDLs, and the 'bubbles' must be removed entirely before the encoding. The multiple-event

direct-histogram architecture, shown in **Figure 4.3**, applied an additional hit signal toggle module, a multi-edge detection encoder and a direct-histogram counter bank to achieve the two unique features.
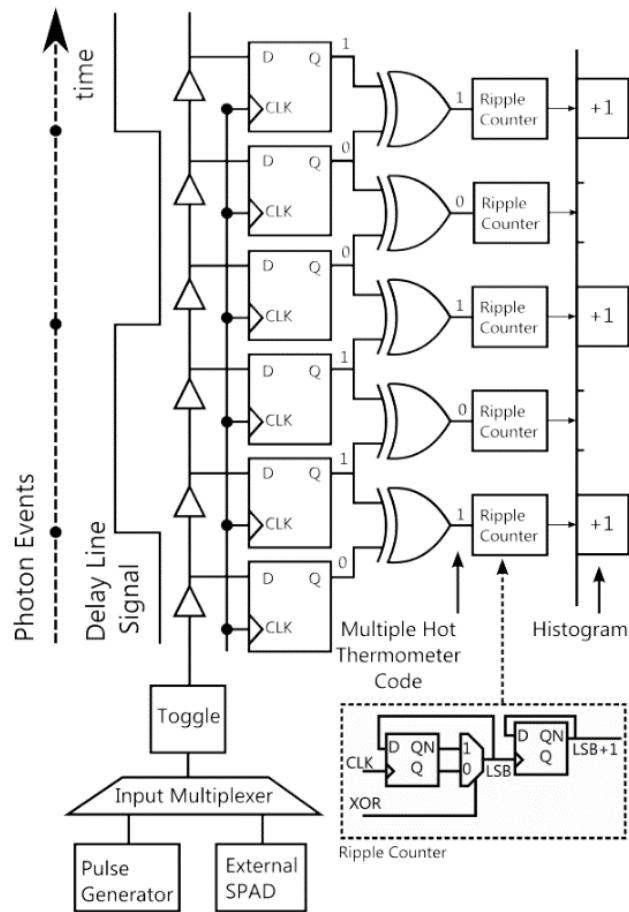
FIGURE 4.3 THE STRUCTURE OF THE MULTIPLE-EVENT DIRECT TO HISTOGRAM TDC [149]

a) HIT signal toggle module:

The HIT signal toggle module was used in Claudio's work [104] to increase the sampling rate, which is constructed by a D-FF and a feedback loop with an inverter. The toggle module is used for reshaping the signal propagated in TDLs and for turning it from signal-edge activated to dual-edge activated. To achieve this function, the signal 'START' or 'HIT' is connected with the clock port of the D-FF as a trigger signal. Once the D-FF is triggered, the outputs of the D-

FF are fed back to the D-port via the inverter and fed into the following TDL at the same time. By this process, the outputs are inverted once the D-FF is triggered and the signal transitions are identified by both rising and falling edges.

b) Multi-edge detection encoder

The multi-edge detection encoder is a bank of 2-input XOR gates. By sending two adjacent bits in a thermometer code to the XOR gates, the transition can be indicated when a '1' appears on the output of the XOR gate. Following this, each output bit of the XOR bank drives a histogram counter directly, instead of being encoded to fine binary codes. Through this structure, multiple transitions can be indicated at the same time and generate a multi-hot code by this encoder. The 'bubbles' are allowed and treated as normal transitions in this structure.

c) Direct-histogram counter bank

In conventional TCSPC systems, integrated memory space is used to build and store the histogram of timestamps. The fine binary codes of TDCs are used as addresses to fetch the counts from the memory module. Fetched counts are then increased and rewritten to their original address in the memory. The direct-histogram architecture uses a bank of counters as the memory module. Each counter is driven by the signal 'STOP' or the sampling clock, and is controlled by a corresponding bit in the multi-hot code. During measurements, bin counts are operated independently and stored in individual counters. Therefore, the maximum theoretical sampling rate of the direct-histogram architecture is related to the number of counters used. The histogram counters can be implemented either by ripple-counters or synchronous counters in FPGAs. Ripple-counters require lower logic resources than a synchronous counter with the same bit width. However, ripple-counters have longer latency. Synchronous counters can complete a counting operation in one clock cycle, and their addend can be modified. The original design uses ripple-counters [149]. In this design, synchronous counters are used for a histogram to apply the following fast bin-width calibration.

By combining these three unique modules, all TDC taps can be processed and counted

independently and concurrently. This architecture means that the sampling rate and deadtime overcome the limitations of the total propagation delay of TDLs. According to Dutton's work [149], the maximum theoretical and measured sampling rate achieved 61.7Gs/s and 6.17Gs/s respectively in a Virtex-5 FPGA. Furthermore, the various de-bubble methods for traditional TDCs which aggravate the nonlinearity are not used in the proposed design. The bubble bits are encoded to '1' bits in the multi-hot codes as normal transitions and indicate all of the possible locations of sampled events. In this way, the zero-width bins can be filled up with limited improvement in linearity, and the minimum DNL value was increased to -0.9LSB according to Dutton's work.

4.2.3 Multiple-sampling architecture

In this design, the multi-phase sampling architecture is applied to reduce the linearity impact from the clock skews. In a single-phase sampling architecture, the parts of events might be missed by the TDL [136] if the length of the TDL is less than the period of the sampling clock. To avoid this situation, increasing the length of TDLs and the frequency of the sampling clock are straightforward methods. However, both the highest operation clock frequency and the height of CRs are limited in FPGAs.

For the operation clock, various components and IP cores have their operation frequency ranges in FPGAs. In Virtex7 FPGAs, the maximum frequencies (with -1 speed grade) for BUFGs (global clock buffers) and BUFRs (regional clock buffers) are from 625MHz (1.6ns) and from 450MHz (2.22ns) respectively. For embedded Block RAMs, the maximum drive frequency for read/write operations is from 458.09MHz (2.18ns) to 601.32MHz (1.66ns), depending on device speed grades [115]. In practice, the maximum clock frequency for a sequential and combinatorial logic needs to be sufficient for a logic pathway with the longest delay. Otherwise, the logic will fail, leading to the risk of a system plagued with uncertainty or even a system crash.

The length and location of TDLs need to be appropriately controlled to avoid the significant nonlinearity around CR boundaries. In Virtex7 FPGAs, the height of CRs is 50 CLBs, which

contains 200 carry-out bits or taps. Considering that the average bin-width of plain TDLs in 7-serial FPGAs is around 10ps [128, 136, 151, 152], the total propagation delay of a TDL within a single CR is around 2ns. Therefore, a 500MHz clock is required to cover a TDL within one CR.

Considering that a shorter delay line can provide better linearity [153] and increase the design margin of different specifications, the triple-phase sampling architecture is used in this design. As shown in **Figure** 4.4**,** the three parallel 200-bin TDLs are sampled by the three clock signals with 0°, 120°, and 240° phase shifts, respectively. Each TDL covers one-third of the clock period. The diagram of the operation is shown in **Figure 4.5.** Theoretically, the minimum requested frequency of the sampling clock of TDLs in a single CR is reduced from 500MHz to around 166.67MHz in Virtex-7 FPGAs. However, the offsets among the three parallel TDLs reduce the effective TDL length as part of TDLs are overlapped. Therefore, the actual frequency of TDLs is slightly higher than the theoretical values, and these TDLs are located close to each other to minimise the offsets and overlap.
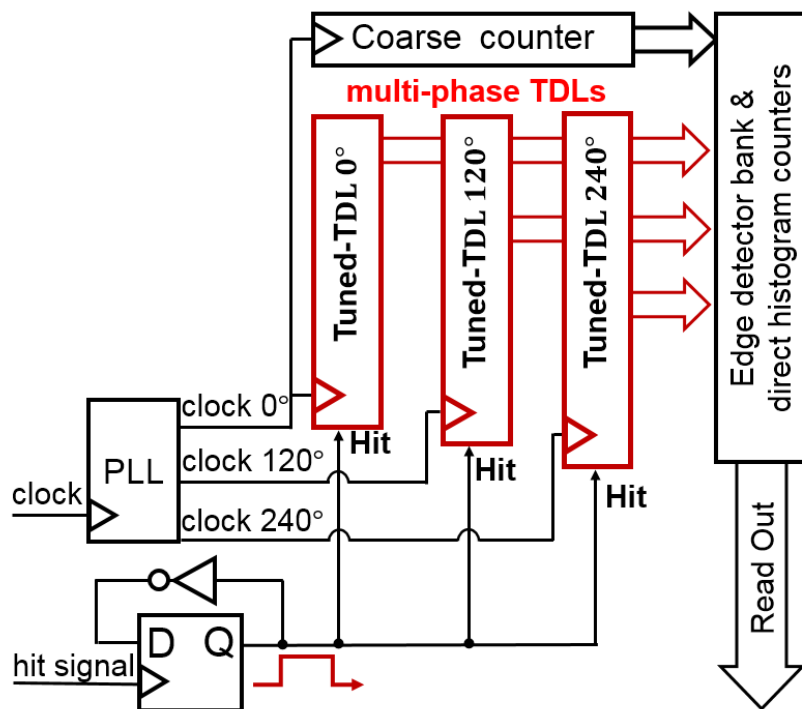
FIGURE 4.4 BLOCK DIAGRAM OF TRIPLE-PHASE SAMPLING ARCHITECTURE.
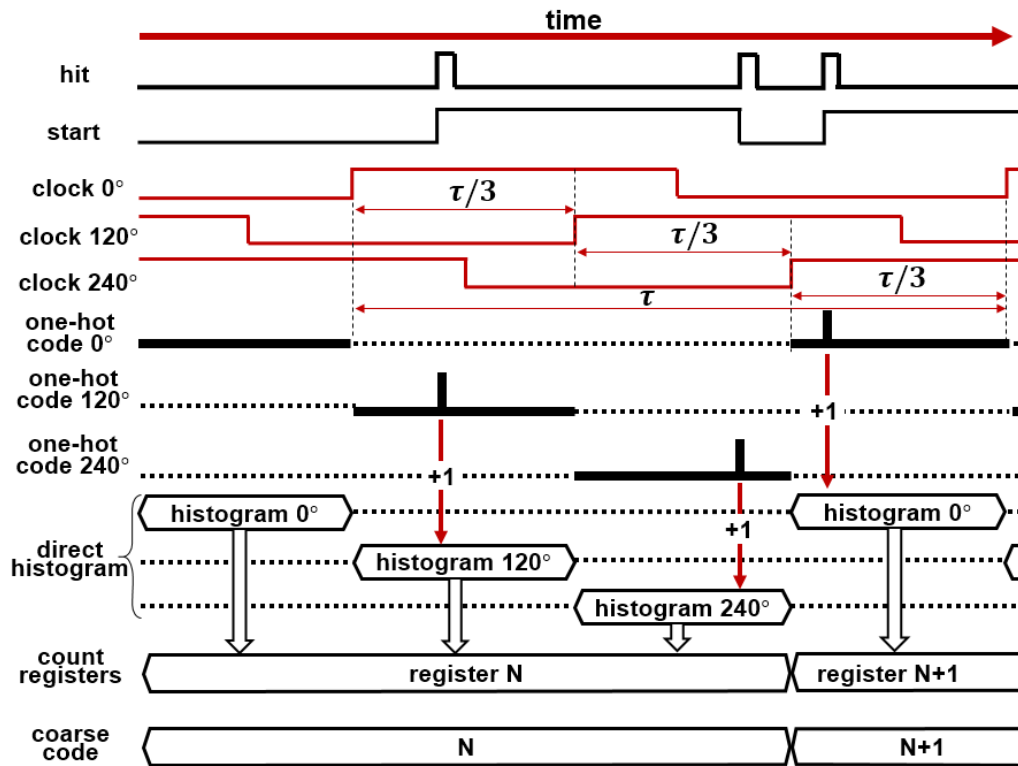
FIGURE 4.5 TIMING DIAGRAMS FOR THE PROPOSED TDL-TDC WITH TRIPLE-PHASE SAMPLING ARCHITECTURES. THE HIT SIGNALS ARE SAMPLED BY THREE CLOCK SIGNALS SEPARATELY AND RECORDED IN CORRESPONDING ONE-HOT CODES.

4.2.4 Calibration method for the nonuniformity of carry-chains

As discussed in Chapter 3, the calibration procedure is required to suppress the nonlinearity of FPGA-TDCs. In previous studies, the bin-by-bin calibration and the bin-width calibration were applied to suppress the static nonlinearity caused by the nonuniformity of carry-chains. To further improve linearity performance, this design invented a hardware-friendly bin-width calibration based on the direct-histogram architecture.

a) Bin-by-bin calibration

The bin-by-bin calibration method is designed to reduce the deviation of the uneven TDL bins in FPGA-TDCs [128, 140]. This method readdresses the output code to the centre of the bins,

which can be expressed as:

$$t[n] = \frac{W[n]}{2} + \sum_{k=0}^{n-1} W[k] \tag{4.1}$$

where $W[n]$ and $W[k]$ are the code bin-width of the *n-th* and *k-th* code bins, respectively. According to previous research [140], the RMS error, σ, is contributed by the bins with the lower, t_1, and upper, t_2, transit levels. When the output code of TDC is readdressed to a new point, t_c, within the bin, the RMS error can be expressed as below:

$$\sigma^2 = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (t - t_c)^2 \, dt$$
$$= \frac{(t_2 - t_c)^3 - (t_1 - t_c)^3}{3(t_2 - t_1)} \tag{4.2}$$

To simplify the above equation, it can be assumed that the $t_1 = 0$ and the $t_2 = w$, where the w is the width of bins. The equation can be rewritten as:

$$\sigma^2 = \frac{(w - t_c)^3 + t_c^3}{3w} = wt_c^2 - w^2 t_c + \frac{w^3}{3} \tag{4.3}$$

The RMS error is minimised when the $t_c = -\frac{-w^2}{2w} = \frac{1}{2}w$ and the RMS error can be expressed as below when $t_c = \frac{1}{2}(t_1 + t_2)$:

$$\sigma^2 = \frac{(t_2 - t_1)^2}{12} \tag{4.4}$$

Even if the RMS error can be reduced by calibrating output codes of TDCs to the centre of a bin, the RMS error is still considerable, as said RMS error is limited to larger than ~0.3LSB for ideal bin-width distribution. Moreover, the RMS error will increase to ~0.6LSB and ~0.87LSB with the widths of the bins at around 2LSB (DNL=1LSB) and 3LSB (DNL=2LSB), respectively. From the previously published FPGA-TDCs, it can be seen that the maximum DNL values are difficult to be restrained to lower than 1LSB. From **eq. 4.1** to **eq. 4.4**, it is clear that the ultra-wide bins will further increase the RMS error. As a result, it seems that the bin-

by-bin method is not sufficient for FPGA-TDCs with large DNL values. Furthermore, this method requires continuous updating of the calibration table with over 5℃ temperature drift, because it is mainly implemented by a pre-stored look-up-table [128]. This continuously updating process increases the complexity and the cost of the resources.

b) Bin-width calibration

For TCPSC systems or TDCs with histogram modules, a bin-width calibration is a practical approach to calibrate the uneven bin-width of TDLs [149]. The basic concept of this method is weighting the counts in histogram bins based on the width of corresponding bins. Theoretically, the calibration cannot process missing-codes and zero-width bins in TDLs. For TDCs with a small number of missing-codes and zero-width bins, the missing-codes can be ignored or merged with adjacent bins before applying the calibration. However, this method will lead to a noticeable impact on the temporal resolution in conventional FPGA-TDCs, since the missing-codes and zero-width bins take up a relatively large proportion, even more than 50% [62, 128, 142, 151]. Therefore, most of the previously-published FPGA-TDCs with poor DNL performance is not suitable for the bin-width calibration.

The direct-histogram TDC uses a group of calibration factors to calibrate the counts of histogram bins. The factors are calculated by referring to the ratio between the width of individual bins and the ideal bin (DNL+1). When performing the code density test, the counts stored in histogram bins are proportional to the bin-width in TDLs. The DNL values can be defined as:

$$DNL[k] = \frac{H[k] - H_{ideal}}{H_{ideal}} \tag{4.5}$$

where $H[k]$ is the actual count of k-th bin, and the ideal count, H_{ideal}, is:

$$H_{ideal} = \frac{\sum_{k=0}^{N-1} H[k]}{N} \tag{4.6}$$

where N is the number of bins in the tested TDC. According to the DNL values, the calibration factors of the k-th bin, $CF[k]$, can be calculated as:

$$CF[k] = \frac{1}{DNL[k]+1} \qquad (4.7)$$

The actual count in the k-th histogram bin can be used to calibrate $H_{cal}[k]$ as:

$$H_{cal}[k] = H[k] \times CF[k] \qquad (4.8)$$

The original CFs are floating-point numbers which contain enough fractional parts to guarantee the calibration effects. However, the digital circuits in FPGAs cannot process the floating-point numbers directly. The current advanced FPGAs always provide embedded dedicated digital signal processors (DSPs) or IP cores for the floating-point number calculation with high precision. However, the floating-point number calculation in FPGAs will generate longer operation latency and cost much more in terms of logic resources. **Table 4.1** presents a comparison of the resource costs of two finite impulse response (FIR) filters (floating-point and fixed-point version):

Table 4.1: resource costs and latency of floating and fixed-point calculation

	single-precision floating-point	fixed-point	improvement
Max frequency	500MHz	580MHZ	16% faster
Operation latency	91 cycles	12 cycles	7.6 times shorter
DSP48 cost	423	85	5 times less
LUT cost	23,106	1,973	11.7 times less

From the results, it can be seen that the floating-point based calculation significantly increases the system complexity and operation latency. Considering that bin counts need to be calculated independently, these increased costs will be multiplied for the bin-width calibration. Therefore, the fixed-point version of the calibration is utilised in the proposed design. The proposed design uses the synchronous counters with a fast calibration procedure to replace the ripple counters and embedded DSPs to minimise resources costs and data processing cycles.

The multiplication of **eq. 4.8** can be implemented by adding the $CF[K]$ into the count each iteration instead of incrementing by one each time, which can be expressed as below:

$$H_{cal}[k] = H[k] \times CF[k]$$

$$= \sum_{n=0}^{H[k]-1} CF[k] = \sum_{n=0}^{H[k]-1} \frac{1}{DNL[k]+1} \tag{4.9}$$

The floating-point version of the CF vectors needs to be converted to the fixed-point version before being used. The first step of the fixed-point calibration is confirming the position of the decimal point or the number of significant bits, M, in the fractional part. Following this, the original CF vectors are multiplied by 2^M and rounded to integers, CF_{fix}. Next comes changing the increment value or the addend of the k-th histogram counters from 1 to $CF_{fix}[k]$. Finally, the histogram results need to be divided by 2^M. Since bin counts are stored as the binary code, the division can be implemented by right-shifting M bits. These operations can be expressed by the equations below:

$$2^M H_{cal}[k] = \sum_{n=0}^{H[k]-1} CF_{fix}[k] = \sum_{n=0}^{H[k]-1} \frac{2^M}{DNL[k]+1} \tag{4.10}$$

$$2^M \cdot H_{cal} = \overline{H}_{cal}^*[k] = \overline{H}_{cal}[k] + \frac{2^M}{DNL[k]+1} \tag{4.11}$$

$$\overline{H}_{cal}^*[k][I-1:0] / 2^M \Rightarrow \overline{H}_{cal}^*[k][I+M-1:M] \tag{4.12}$$

where the bit-width of the histogram counter is I.

This method calibrates the histogram count directly during the measurement and histogramming, while the post-processing with additional operation latency and DSP modules is not needed. These features make this method more suitable for high sampling rate or throughput situations.

4.3 Experiments and performance evaluation

The presented low nonlinearity, multi-event direct-histogram TDC for TCSPC systems is implemented in a NetFPGA-SUME development board [154] with a Xilinx Virtex-7 XC7V690T FPGA chip, shown in **Figure 4.6**. To evaluate the proposed TDC, the code density tests, and the time-interval measurements were performed. The plain TDL and the tuned-TDLs

were one-to-one combined with the modified direct-histogram architectures (four combinations in total) for comparison. The effect of the bin-width calibration method was also tested. This section discusses and evaluates the performance of the proposed architecture in detail.
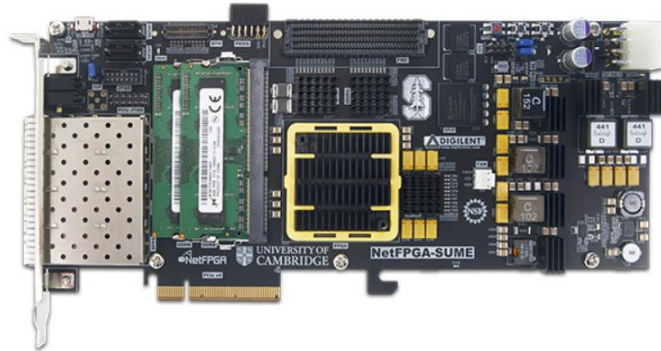
FIGURE 4.6, THE NETFPGA-SUME DEVELOPMENT BOARD WITH A XILINX VIRTEX-7 XC7V690T FPGA CHIP

4.3.1   Experiment setup and monitoring

The setup diagram of the code density test is shown in **Figure 4.7**. Two independent low-jitter crystal oscillators [155] (DSC1103 with <1ps RMS phase jitter) are used as the signal sources of two uncorrelated clock signals for the code density test. The two clock management tiles (CMTs) which contain a mixed-mode clock manager (MMCM) and a phase-locked loop (PLL) are used for clock management, de-skewing and jitter filtering. Two uncorrelated clock signals are fed into the TDC as the HIT signal and the sampling clock with random TIs.
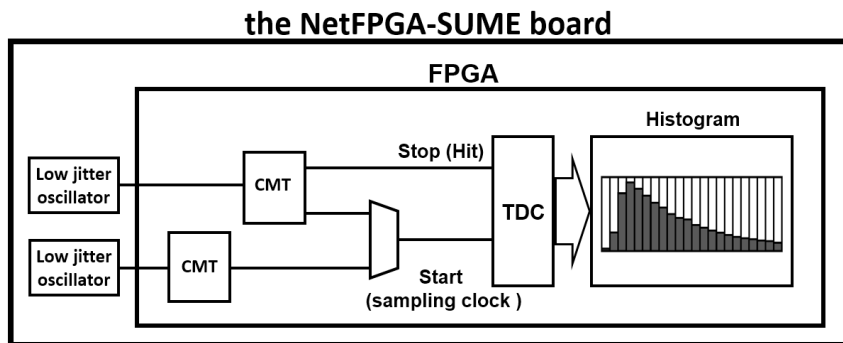
FIGURE 4.7, THE BLOCK DIAGRAM OF THE CODE DENSITY TEST SETUP

One of the preparations for this test is to ensure that the dynamic linearity is invariant so as to minimise its impacts. As a result, the system was placed in an environment with a stable room temperature. The operating voltage was controlled and maintained by the onboard power management chip. The System Monitor in the Xilinx FPGA was used to enhance the safety and reliability of the architecture through the real-time measuring and monitoring of the internal temperature and supply voltage during the operation. To ensure the accuracy of the measurement, numbers of on-chip voltage (maximum accuracy is +/-1%), temperature sensors (maximum accuracy is +/-4℃) and high-precision ADCs were used. After power-up and downloading of firmware, the internal temperature of the FPGA chip was raised and then stabilised at a certain small range. The chip temperature and supply voltages were sampled and sent to a software terminal continuously during the experiments. The code density tests and TI tests were performed after the temperature and supply voltage were stable. In this case, the internal temperature of FPGA was increased from 27.7°C to 31.2°C within the first 6 minutes. Following this, the temperature tended to be stable, since a heatsink with an electric fan had been mounted on the surface of the FPGA chip. The average temperature from 6 to 24 minutes was 31.54°C with σ=0.34°C. The measured temperature results are shown in **Figure 4.8**. For supply voltages, there are three main supply rails, including VCCINT, VCCAUX and VCCBRAM. Their voltages were monitored during measurements [156]. The measured details of the three supply power rails and temperature are summarised in **Table 4.2**.
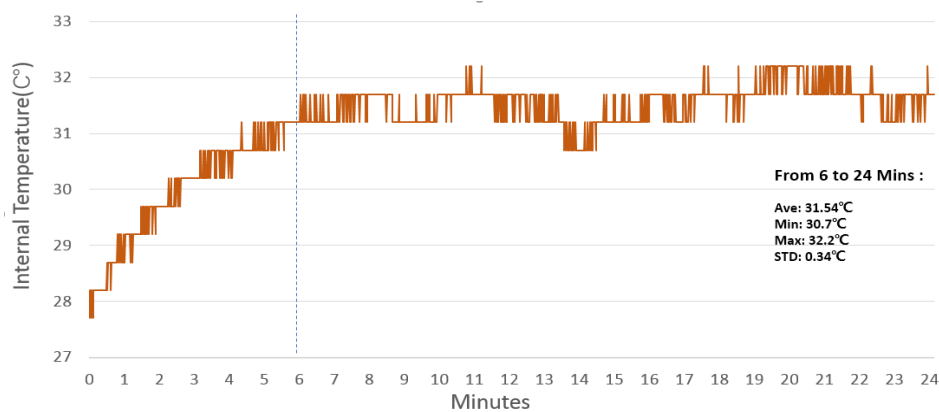
FIGURE 4.8 MEASURED INTERNAL TEMPERATURE RESULTS FROM 0-24 MINUTES AFTER POWER-UP AND THE DOWNLOADING OF THE FIRMWARE.

Table 4.2: measurement results of supply power and temperature

	VCCINT	VCCAUX	VCCBRAM	Temperature (6-24mins)
ave	0.948V	1.798V	0.949V	31.54°C
min	0.946V	1.796V	0.949V	30.7°C
max	0.949V	1.799V	0.952V	32.2°C
std	1.392mV	1.34mV	0.646mV	0.336°C

4.3.2   Full-length TDC test

A full-length TDC was implemented and tested in the target FPGA chip to investigate the impacts of the clock skews. The full-length TDC contains 500 CARRY4s and 2000 bins, which fully cover a column of CLBs and cross 10 CRs. The detailed structure is shown in **Figure 4.9**. Two uncorrelated clock signals for the code density test were configured with the proper frequency by two CMTs. Considering that the Xilinx's 28nm process FPGA devices can provide around the 10-11ps averaged bin-width of a typical TDL [136, 151], the total propagation delay of the full-length is approximately 22ns, and the sampling clock was set to 40MHz. Both a traditional and the proposed architecture (without calibration) were tested in the full-length TDC, respectively.
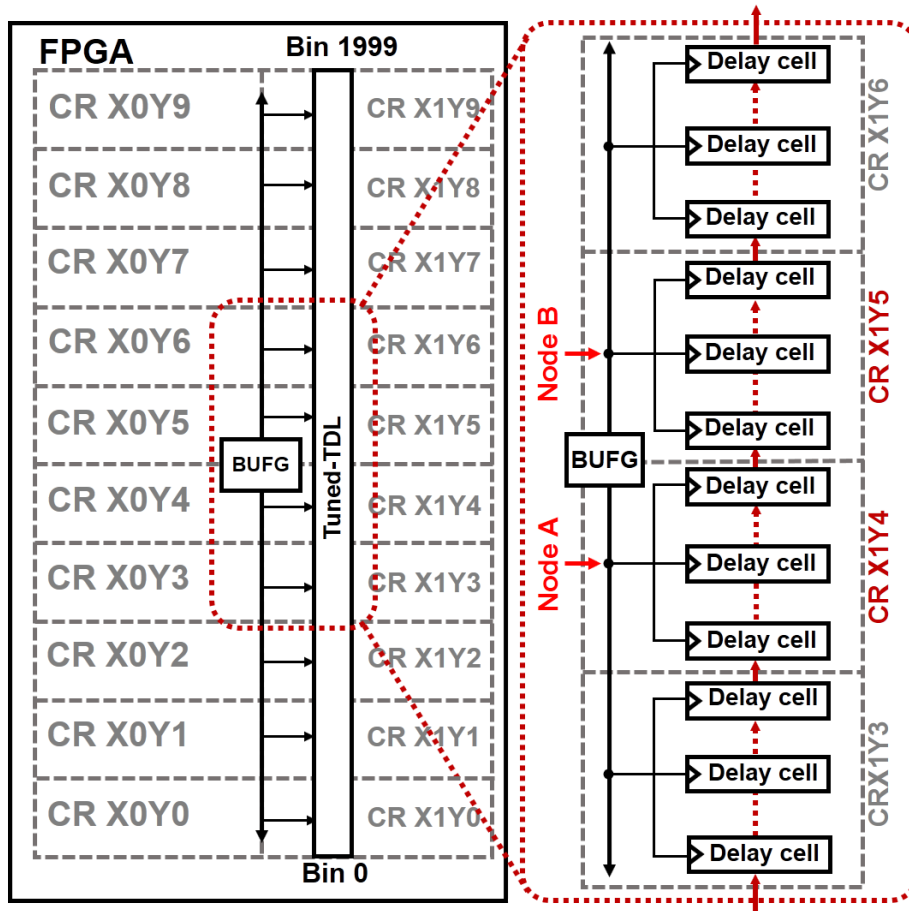
FIGURE 4.9 CLOCK ROUTES, CR, AND THE CLOCK SIGNAL CONNECTIONS OF A FULL LENGTH (2000 BINS) TDL.

The DNL test results of the traditional and proposed TDC architectures are shown in **Figure 4.10**. From the results, it can be seen that the proposed TDC architecture offers a lower nonlinearity. However, both of the two architectures contain numbers of the ultra-wide bins (DNL>2LSB) which are located at the boundaries of some CRs (at bins No.200, 400, 600, 800, 1200, 1400, 1600, and 1800). However, the ultra-wide bins did not appear at the boundary (bin 1000) between the two central CRs (CRX1Y4: from bin 800 to bin 999; CRX1Y5: from bin 1000 to bin 1199). This is because the clock routes of the two CRs are symmetrical. The wide-bins (DNL>1LSB) at bin 1100 (corresponding to Node B, at the middle point of CRX1Y5) and bin 900 (Node A, at the middle point of the CRX1Y6) are noticeable since minor clock skews exist between the two CARRY4 modules at the two sides of the CR middle point.
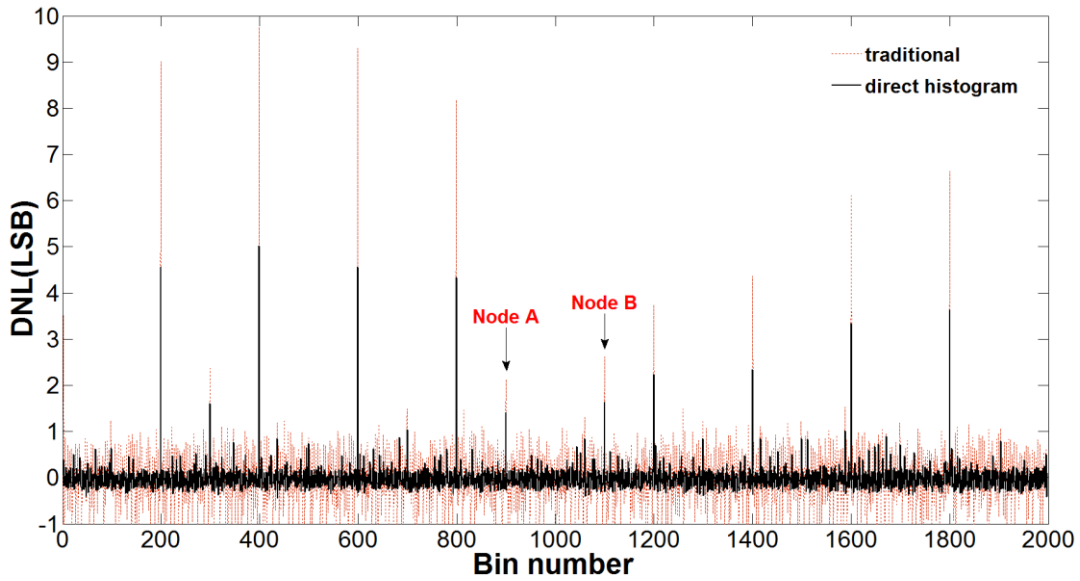
FIGURE 4.10 DNL PLOTS OF FULL-LENGTH (2000 BINS) TUNED-TDLs WITH THE TRADITIONAL AND THE DIRECT-HISTOGRAM ARCHITECTURES.

In order to minimise nonlinearity, this design set the length of a TDL to 200 bins, and the constrained location was from bin 900 to bin 1100. With different speed grades, the propagation delay of a 200-bin TDL in Virext-7 FPGAs is 2.0-2.2ns with 10-11ps average bin-width. As a result, the frequency of the sampling clock was set higher than 500Mhz for single phase architecture. By using three parallel TDLs following the proposed multi-phase method, the required frequency of the sampling clock signal was reduced from 500MHz to around 166.7MHz.

4.3.3 Linearity tests and comparisons among different methods and architectures

A series of tests and comparisons were performed with different architectures to evaluate the linearity improvement of the proposed TDC. The TDL carry pattern 'CCCC' was used in the traditional TDC and the original direct-histogram TDC. The carry pattern 'SCSC' was used in the tuned-TDL and the proposed direct-histogram architecture. All of the tested TDCs measured 5 million TI events during the code density test to ensure convergence.

The DNL, INL are two mainstream parameters to assess the linearity performance of TDCs and TCSPC systems. However, the assessment might be incomplete and biased if only the range or peak-to-peak values of DNL and INL values are evaluated. That is because the occasional worse bins might mislead the overall performance even the rest of the bins might have much better linearity. Therefore, other parameters, including the standard deviations of DNL and INL, the bin-width distribution, the equivalent bin-width and equivalent standard deviation, are also evaluated in this chapter.

a) DNL and INL

Figure 4.11 compares the code-density test results of the four different methods and architectures. **Figure 4.11(a)** and **(c)** compare the test results of a traditional TDL-TDC and an original direct-histogram TDC. By applying the direct-histogram architecture, the range of DNL was reduced from [-1, 4.34]LSB to [-0.96, 1.6]LSB, and the range of INL was reduced from [-6.85, 2.50]LSB to [-2.85, 1.61]LSB. By using the original direct-histogram architecture, the peak-to-peak values of DNL and INL, $DNL_{peak-peak}$ and $INL_{peak-peak}$, were reduced by more than half. The standard deviations of the DNL and the INL values, σ_{DNL} and σ_{INL}, were improved from 1.20LSB to 0.61LSB and from 1.85LSB to 0.92LSB, respectively.
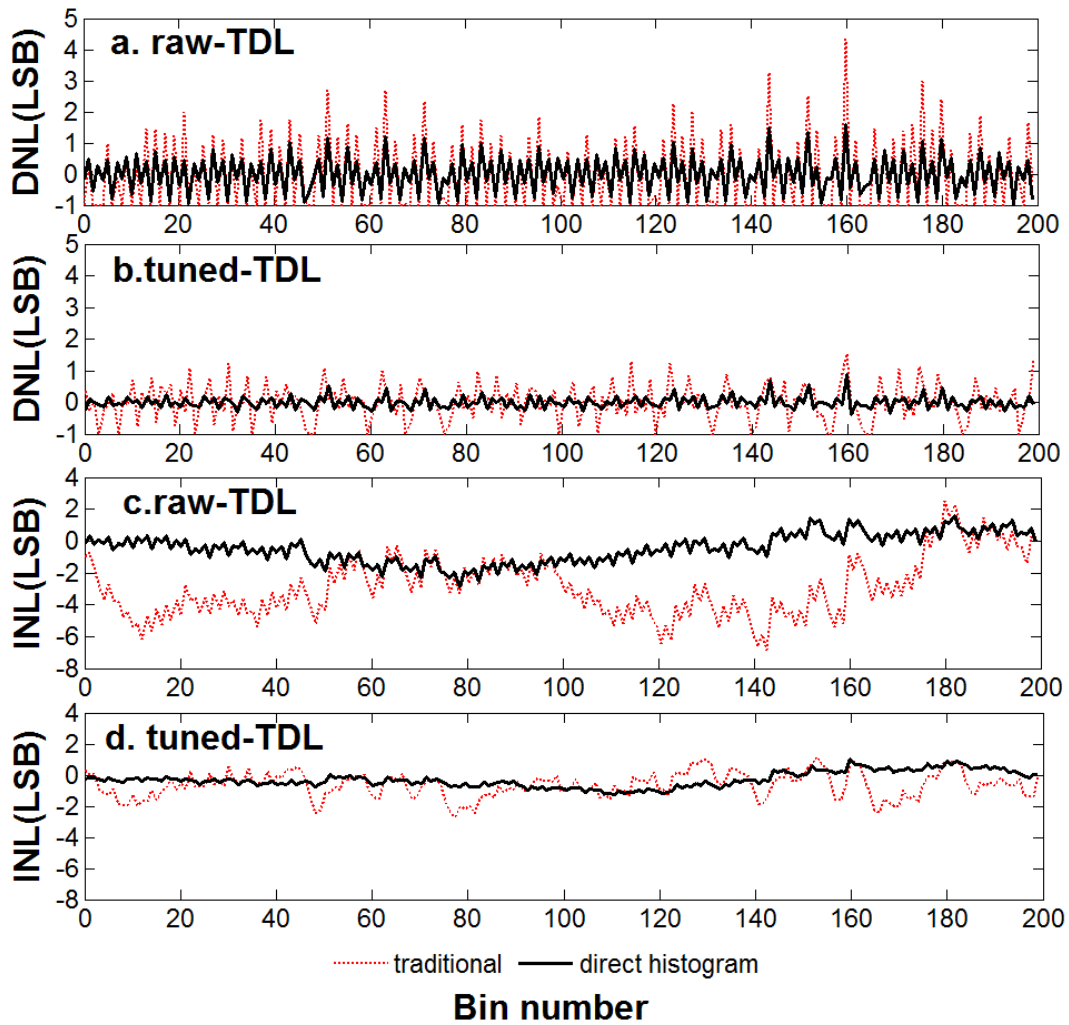
FIGURE 4.11 DNL PLOTS OF (A) RAW-TDL (B) TUNED-TDL. INL PLOTS OF (C) RAW-TDL AND (D) TUNED-TDL.

Figure 4.11(b) and **(d)** illustrate the performance improvements of the DNL and INL between a typical tuned-TDL TDC and the proposed design. By using the tuned-TDL alone, the DNL range was reduced to [-1, 1.53]LSB, and the INL range was reduced to [-2.66, 1.20]LSB. The standard deviations of the DNL and the INL values were improved to 0.58LSB and 0.81LSB, respectively.

From the results summarised in **Table 4.3** and **Figure 4.11**, it can be seen that a significant linearity improvement was achieved by combining the tuned-TDL and the direct-histogram

architectures. The DNL$_{min}$ of the proposed architecture was increased to -0.38LSB, far better than the threshold value of the missing codes. Compared with the traditional TDC, the $DNL_{peak\text{-}peak}$ and $INL_{peak\text{-}peak}$ were improved from 5.34LSB to 1.25LSB and from 9.35LSB to 2.25LSB respectively. The σ_{DNL} and the σ_{INL} were 0.20LSB and 0.50LSB, respectively.

Table 4.3: code density test results of different architectures

Unit: LSB	plain TDL		tuned-TDL	
	traditional	*direcet-histo*	*traditional*	*direcet-histo*
DNL	[-1, 4.34]	[-0.96, 1.60]	[-1, 1.53]	**[-0.38, 0.87]**
DNL$_{peak\text{-}peak}$	5.34	2.56	2.53	**1.25**
σ_{DNL}	1.20	0.61	0.58	**0.20**
INL	[-6.85, 2.50]	[-2.85, 1.61]	[-2.66, 1.20]	**[-1.23, 1.02]**
INL$_{peak\text{-}peak}$	9.35	4.47	3.86	**2.25**
σ_{INL}	1.85	0.92	0.81	**0.50**

b) Bin-width distribution

The bin-width distribution, derived from the DNL results and the averaged bin-width, demonstrates the uniformity of TDL bins, and can be used to evaluate the extent of zero-width bins, missing-codes and ultra-wide bins. A TDC with excellent linearity or bin uniformity should follow the Poisson distribution with limited deviation in its bin-width distribution.
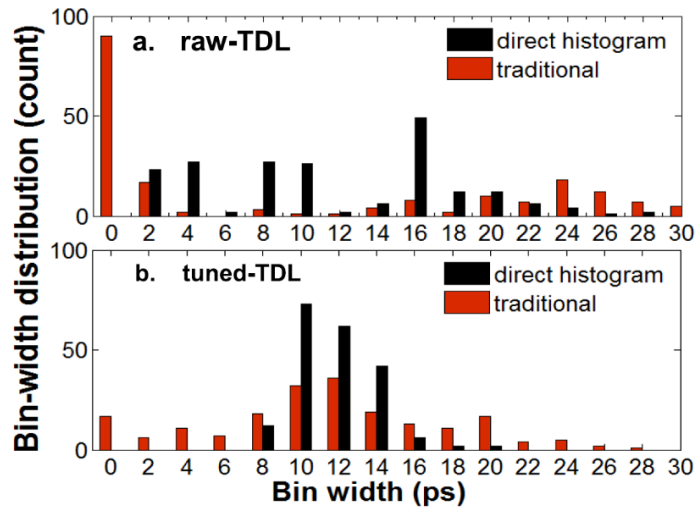
FIGURE 4.12   BIN-WIDTH DISTRIBUTIONS USING THE TRADITIONAL THERMOMETER-TO-

Figure 4.12 shows the bin-width distributions of the four architectures introduced above. In the traditional TDL-TDC, more than half of the bins (52%) are classified as the zero-width bins, while the missing-codes (**Figure 4.12(a)** with red bar), and bin-widths of 4.5% of bins are more than 2 times wider than the average bin-width. However, the proportion of the uniform bins (-0.5LSB <DNL <0.5LSB) was very low (only approximately 8.5%). These defects encumber the standard deviation of bin-width to 1.2LSB. Compared with the traditional TDL-TDC, the bin-width distributions were improved to a certain extent by using the tuned-TDL and direct-histogram individually. From **Figure 4.12(a)**, it can be seen that zero-width bins and ultra-wide bins are removed in direct-histogram architecture (black bar). However, missing-codes still exist with a proportion of 4%. The tuned-TDL method (**Figure 4.12(b)** with red bar) is able to provide better bin-width uniformity as 61.5% of bins were uniform bins, and no ultra-wide bins existed. However, the missing-codes (11%) and zero-width bins (8.5%) still cannot be removed entirely. The bin-width standard deviations of the direct-histogram architecture and the tuned-TDL method are 0.61LSB (6.40ps) and 0.58LSB (6.09ps) respectively.

By combining the tuned-TDL and direct-histogram architecture, the improvement in the uniformity of bin-width is more noticeable compared with the other three architectures. The proposed architecture achieved a much better bin-with distribution with the bin-width standard deviation of 0.2LSB (2.10ps) and the capacity of totally removing the missing-codes and ultra-wide bins. Moreover, the bin-width distribution demonstrates a shape which is close to the Poisson distribution, and 97.5% of the bins could be grouped into the uniform bins.

c) Equivalent bin-width and equivalent standard deviation

The equivalent bin-width, w_{eq}, and equivalent standard deviation, σ_{eq}, are the two new parameters which were presented in 2014 [19]. The resolutions of digitisers such as TDCs and ADCs can typically be considered as the primary parameters to evaluate their performances. The temporal resolution is calculated by averaging the width of all the bins in TDCs. However, the temporal resolution cannot describe the performance accurately, especially when severe

nonlinearity exists. This is because the deviation of bin-width is not described by temporal resolution. For example, a two-bin TDC with 1ps and 9ps bin-width has the same resolution as a TDC which has two 5ps bins. As a result, it may cause a biased interpretation if only the temporal resolution is focused on.

The w_{eq} and σ_{eq} consider temporal resolution and deviation of bin-width at the same time. These two parameters are calculated from the results of the code density test. During the code density test, a wider bin has a higher probability of capturing signal transitions, and the bin-widths are proportional to the counts of the corresponding bin. According to the definition of σ_{eq}, **eq. 4.13**, a narrower bin will provide better precision with a lower measurement error [157].

$$\sigma_{eq}^2 = \sum_i \left(\frac{w_i^3}{12W} \right) \tag{4.13}$$

where w_i is the bin-width of the i-th bin and W is the total width of a TDL which can be expressed as $W = \sum_i w_i$. The w_{eq} is further developed from the σ_{eq} and can be calculated as:

$$w_{eq} = \sigma_{eq} \sqrt{12} = \sqrt{\sum_i \left(\frac{w_i^3}{W} \right)} \tag{4.14}$$

Table 4.4: σ_{eq} and w_{eq} of different architectures

Unit: ps	plain TDL		tuned-TDL	
	traditional	*direcet-histo*	*traditional*	*direcet-histo*
w_{eq}	27.57	15.65	15.07	**11.15**
σ_{eq}	7.95	4.51	4.35	**3.22**

The w_{eq} and σ_{eq} of the four tested architectures are summarised in **Table.4.4**. The temporal resolutions of the four architectures are identical (around 10ps) since they were implemented in the Virtex-7 FPGA with the same configuration. However, the w_{eq} of these architectures are significantly different. The w_{eq} of the traditional architecture was decayed to 27.57ps, around 2.7 times of its temporal resolution. By applying the direct-histogram architecture and the tuned-TDL individually, the w_{eq} was improved to 15.65ps and 15.07ps, respectively. The proposed combination architecture improved the w_{eq} to 11.15ps with around 6% difference from its temporal resolution, and the σ_{eq} was improved from 7.95ps to 3.22ps.

d) Linearity improvement by using the hardware bin-width calibration

As described in Section 4.2.4, the bin-width calibration cannot handle the zero-width bins. A higher DNL_{min} value is expected because a narrower bin needs to be calibrated by a larger calibration factor, which will amplify measurement jitter. According to the test results shown in **Table 4.3**, the proposed architecture meets the requirement of the bin-width calibration. The code density tests of calibrated TDCs with different M values are performed. A larger M value leads to a better calibration effect and a lower precision loss at the cost of increasing the bit-width of the histogram counters.

Figure 4.13 presents the plots of the DNL and INL of the proposed TDC with four different M values ($M=0$: *uncalibrated*). **Table 4.5** summarises more details of the calibration. By setting $M=5$, the $DNL_{peak-peak}$ and the $INL_{peak-peak}$ values were reduced by more than 16-fold and 17-fold, respectively. The standard deviations, σ_{DNL} and σ_{INL}, were decreased by 20-fold and 25-fold, respectively. The w_{eq} was reduced from 11.15ps to 10.55ps. From **Figure 4.14 and Table 4.5**, it can be seen that it was difficult for the linearity performance to be further improved when the M value was larger than 5.
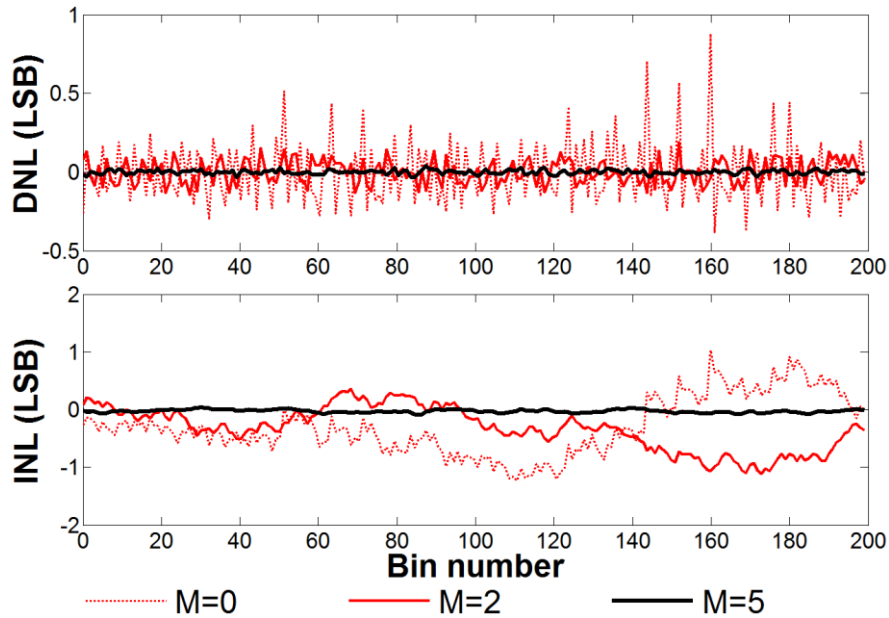
FIGURE 4.13 DNL AND INL CURVES OF A SINGLE TUNED-TDC WITH THE DIRECT-HISTOGRAM ARCHITECTURE AFTER BIN-WIDTH CALIBRATION WITH DIFFERENT M VALUES ($M = 0, 2, 5$).

Table 4.5: Results of calibrated TDC with various M values

	M=0	*M=1*	*M=2*	*M=3*	*M=4*	*M=5*	*M=6*	*M=7*
$DNL_{peak\text{-}peak}(LSB)$	1.25	0.62	0.34	0.20	0.11	**0.08**	0.08	0.07
$\sigma_{DNL}(LSB)$	0.20	0.14	0.08	0.04	0.02	**0.01**	0.01	0.01
$INL_{peak\text{-}peak}(LSB)$	2.25	2.20	1.47	0.74	0.30	**0.13**	0.16	0.14
$\sigma_{INL}(LSB)$	0.50	0.58	0.37	0.18	0.02	**0.02**	0.04	0.03
$w_{eq}(ps)$	11.15	10.81	10.59	10.52	10.55	**10.55**	10.55	10.55
$\sigma_{eq}(ps)$	3.22	3.12	3.06	3.04	3.05	**3.05**	3.05	3.05
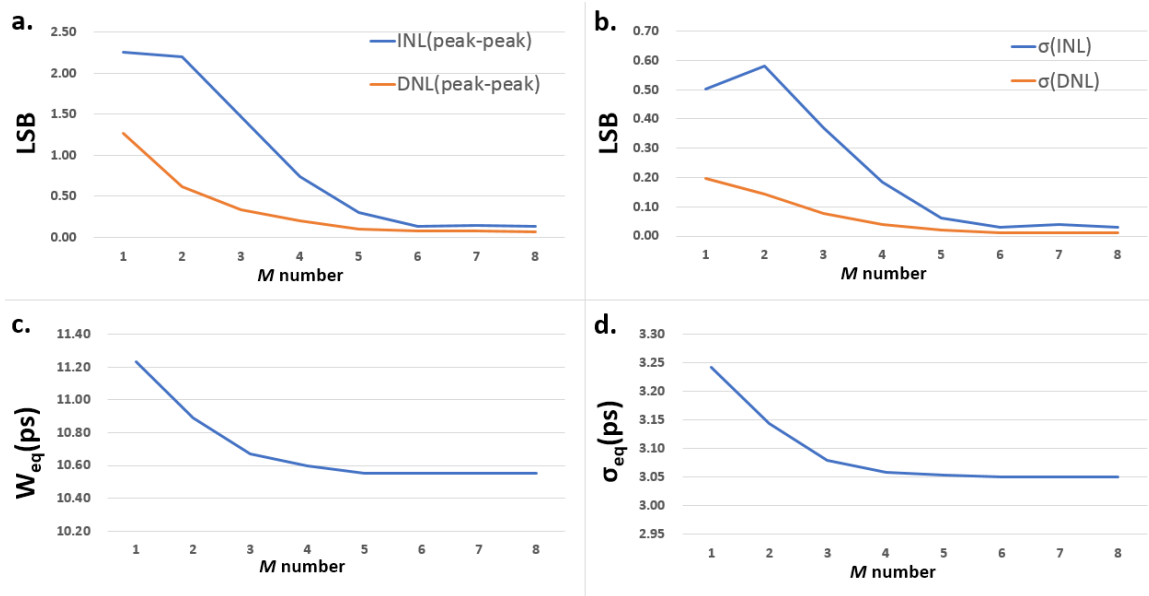
FIGURE 4.14 PLOTS OF LINEARITY PERFORMANCE OF PROPOSED TDC WITH DIFFERENT M VALUES. (A) THE PEAK-PEAK VALUES OF DNL AND INL, (B) THE STANDARD DEVIATION OF DNL AND INL, (C) THE EQUIVALENT BIN-WIDTH AND (D) THE EQUIVALENT STANDARD DEVIATION RESULTS.

4.3.4   Time interval (TI) measurement

The TI measurement repeatedly measures signal pairs with known TIs to evaluate offset and IRF values. In this design, programmable delay generators in FPGAs are used to produce the known signal pairs with adjustable TIs.

In Virtex-7 FPGAs, the programmable delay generators, IDELAYE2 modules, can be used to generate the known TIs. The IDELAYE2s are able to delay the external or internal signals via specified TIs, which are dynamically adjustable [158]. The IDELAYE2 modules are continuously calibrated by an IDELAYCTRL2 module based on an external low jitter reference clock to maintain the stability and resist the PVT deviation. The adjustable step delay of IDELAYE2 is 39±5ps per step when the reference clock is working at 400MHz [115]. The test system setup is shown in **Figure 4.15**. The signal pairs are generated by an onboard low jitter crystal oscillator and managed by CMTs for deskewing and jitter filtering. An IODELAY2 module delays one of the signals before being fed into TDCs. At the same time, the two signals are outputted via an FPGA Mezzanine Card (FMC) and SMA connectors and measured by a high-performance TCSPC module (PicoQuant PicoHarp 300, 4ps resolution and DNL<5%, peak<1% RMS) as the gold standard. Because the time intervals are generated in the same FPGA chip and sent to TDCs directly, it can be concluded that an advantage of this setup is that it minimises both additional jitters from the external signals generator and the noise from board circuits and interfaces.
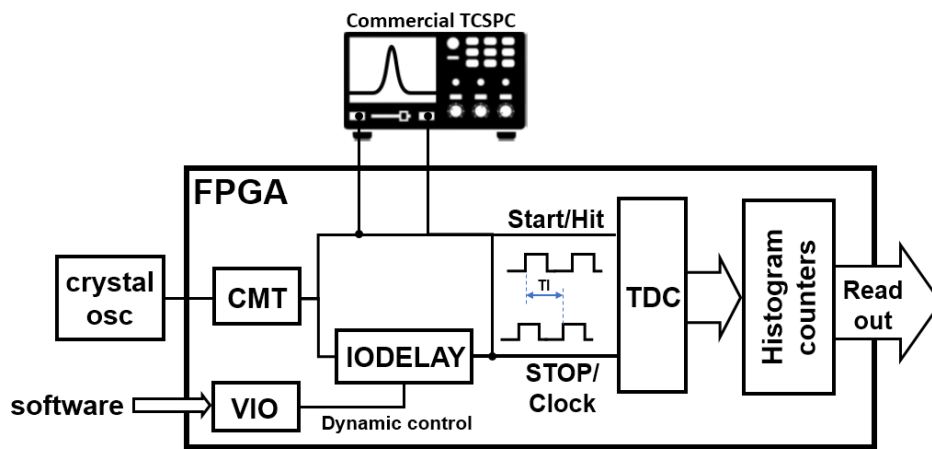
FIGURE 4.15 BLOCK DIAGRAM OF THE SETUP OF THE TIME INTERVAL TEST SYSTEM
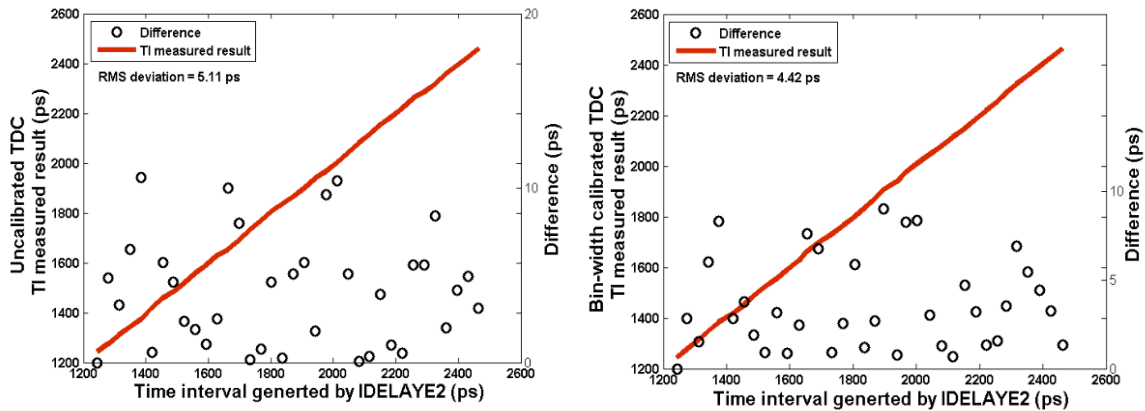
FIGURE 4.16   RESULTS OF TIME INTERVAL MEASUREMENTS OF AN UNCALIBRATED TDC
(LEFT) AND A CALIBRATED TDC (RIGHT)

During these measurements, the delay of the IODELAY2 was dynamically controlled by a software end via a virtual I/O (VIO) interface. The signal pairs with a fixed time interval were repeatedly measured (more than 100 000 samples) by both the proposed FPGA-TDC and the commercial TCSPC system. The time intervals were measured again after the delay step of IODELAY2 modules was increased. The time intervals were increased from 1244ps to 2464ps and measured by both uncalibrated and calibrated TDCs. **Figure 4.16** demonstrates the measurement results and differences compared with results provided by the commercial TCSPC. The actual temporal resolution was 10.5ps. The standard deviations of the measurement differences were 5.11 and 4.42ps for the uncalibrated and calibrated TDCs, respectively.

4.4   Laser ranging test

A typical laser ranging system was applied to evaluate the linearity performance of the proposed TDC. The system contains a 2x8 SPAD array, a picosecond pulse laser (PicoQuant LDH series, with 635nm excitation wavelength, 80MHz repetitive rate and a <120ps pulse width) and the proposed FPGA-TDC (with and without calibration for comparison).

The FPGA outputs the sampling clock of the proposed FPGA-TDC to trigger the laser and. The sampling clock is generated by an onboard low jitter crystal oscillator and managed by an FPGA

CMT. White cardboard with around 85% reflectivity is mounted vertically on a sliding rail as the target. The SPAD detector, which is mounted beside the pulse laser head, makes the photosensitive area face to the whiteboard to receive reflected photons. The output port of one SPAD pixel is connected to the FPGA-TDC as the signal 'HIT' via SMA connectors. The arrival time of the photons is measured by the FPGA-TDC and the commercial TCSPC system. Finally, the histogram data was transferred to software via a UART port for further analysis and processing. To ensure that the SPAD detector worked properly, a dedicated power supply board was used to power the SPAD board. The whole laser ranging test was performed in a dark environment at around 25℃ room temperature.

4.4.1   SPAD detector

The used SPAD was fabricated via the 180nm high-voltage CMOS process designed by the National Chiao Tung University [157, 159]. The SPAD has a breakdown voltage V_{bd} of 81.5V, while its DCR is approximately 500Hz. The SPAD works with a constant excess bias control circuit, which keeps the SPAD bias at a constant voltage for a stable performance. The SPAD and PCB board are shown in **Figure 4.17**.
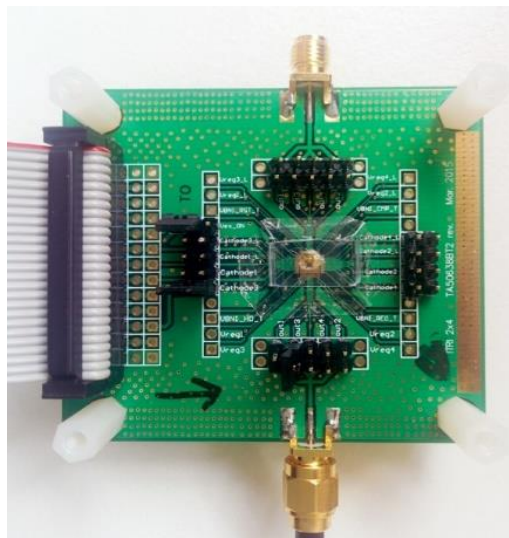
FIGURE 4.17 THE 2x8 SPAD DETECTOR AND PCB BOARD,

Figure 4.18 is the output signal waveform of the SPAD recorded by an oscilloscope. Two events were detected and identified by two 1.2V pulses with 21ns width. The rising edges of the pulses

identify the arrival time of the photons.

FIGURE 4.18 THE SPAD DETECTOR OUTPUT SIGNAL

4.4.2   Experiment results

A series of fixed distances with the same step was measured by the proposed FPGA-TDC (with and without calibration), a traditional FPGA-TDC and a commercial TCSPC system (PicoHarp 300, 4ps resolution) respectively. The measurement results of a fixed distance are shown in **Figure 4.19**. Compared with the traditional TDC, the proposed TDC achieved much better linearity performance. After the calibration, the proposed TDC achieved similar results to those produced by the commercial TCSPC system. **Figure 4.20 (left)** demonstrates the measurement results of linear incremental distances provided by the proposed FPGA-TDC with the calibration and the differences between TDC measured results and expected values. **Figure 4.20 (right)** presents the standard deviations of measurement versus the number of captured events in different TDCs. From the results, it is possible to see that the proposed FPGA-TDC with calibration can provide much lower deviations with the same number of samples.
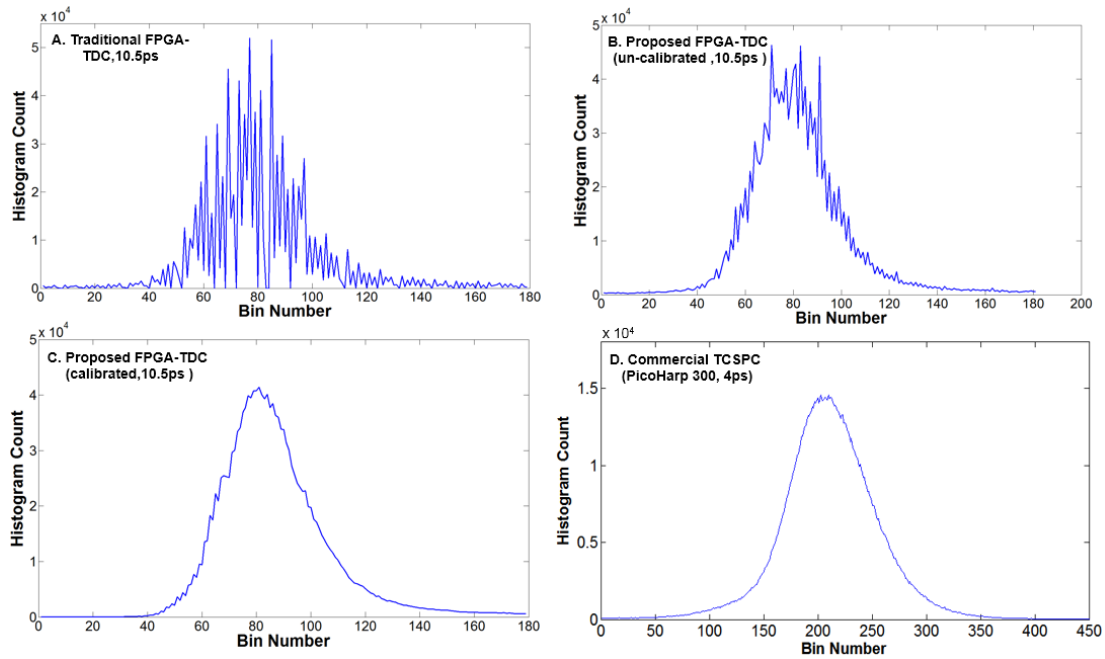
FIGURE 4.19 RANGING TEST RESULTS OF A FIXED DISTANCE FROM (A) A TRADITIONAL FPGA-TDC, (B) THE PROPOSED FPGA-TDC WITHOUT CALIBRATION, (C) THE PROPOSED FPGA-TDC AFTER CALIBRATION (D) A COMMERCIAL TCSPC (PICOHARP 300,4PS).
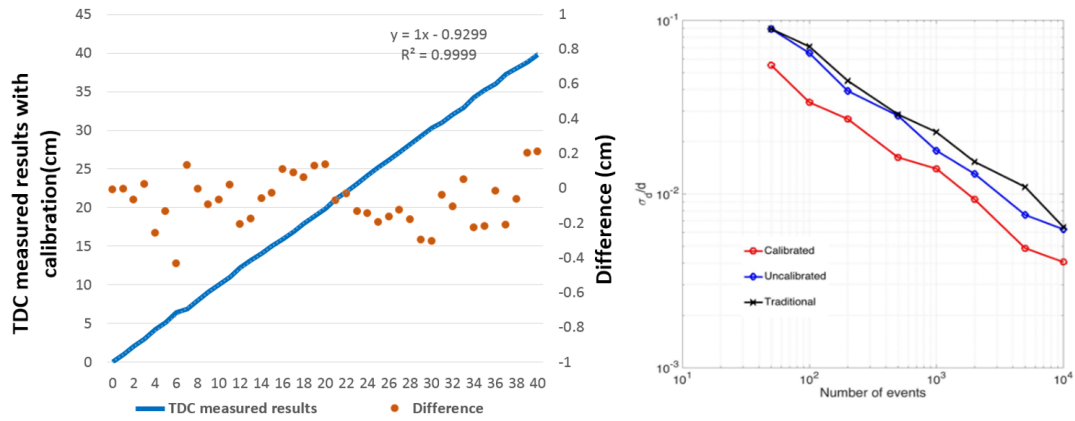
FIGURE 4.20 (LEFT) MEASUREMENT RESULTS AND THE DIFFERENCES BETWEEN THE MEASURED AND EXPECTED VALUES FOR THE PROPOSED FPGA-TDC WITH BIN-WIDTH CALIBRATION. (RIGHT) MEASURED STANDARD DEVIATIONS VERSUS THE NUMBER OF CAPTURED EVENTS OF A TRADITIONAL FPGA-TDC, THE PROPOSED FPGA TCSPC (WITH AND WITHOUT CALIBRATION)

4.5　　Hardware resource utilisation

The resource utilisation was estimated using the Vivado software according to the register-transfer level (RTL) design and synthesise result of the proposed TDC design. **Table 4.4** summarises the three primary logic resources which are used by the proposed TDC with the minimum MR, including look-up-tables (LUTs), slice registers (FFs) and CARRY4 modules. **Figure 4.21** is the place&route layout in the Virtex-7 FPGA.
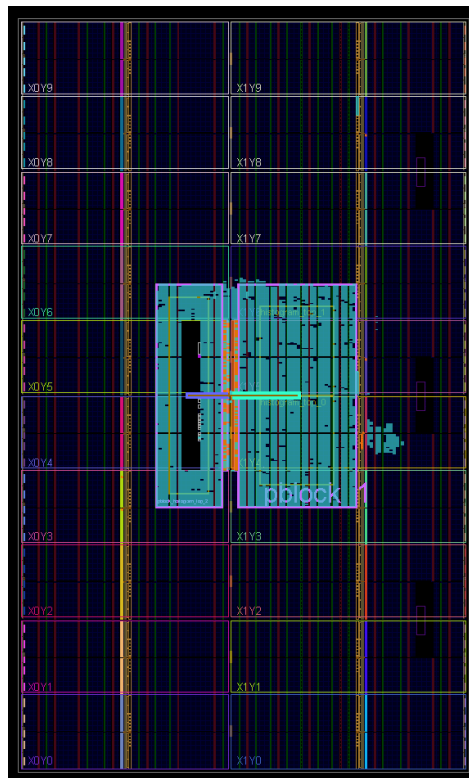
FIGURE 4.21 THE PLACE&ROUTE LAYOUT RESULT OF THE PRESENTED TDC AFTER THE MR EXTENSION

The total resource consumption includes the logical resources cost of the main body of the proposed TDC with 8.2ns MR, debugging and readout modules (ILA, VIO and UART). The total design costs are approximately: 8.2% LUTs, 5.5% slice registers and 6.13% CARRY4 modules. For applications demanding a longer MR, coarse counters can be utilised to achieve the interpolation architecture, and the number of histogram counters is fold increased based on the expected MR. The number of used LUTs and registers is increased to 30867 and 38880

respectively to extend their MR to 25.2ns; the total LUTs and registers costs are increased to 21.84% and 13.69%.

Table.4.6 the hardware resource utilisation of the TDC design

Resource type	Slice LUT	Slice register	CARRY4
Available	433200	866400	111000
single sampling phase			
TDL	200	400	200
Encoder	0	399	NA
histo-counters	11125	14397	NA
Triple sampling phase			
TDLs	600	1200	600
Encoders	0	1197	NA
histo-counters	33565	43209	NA
Other modules			
ILA	811	1262	NA
VIO	410	793	NA
Total cost	35521(8.2%)	47802(5.52%)	6858(6.13%)

4.6   Summary

This chapter proposed a 10.5ps, low nonlinearity FPGA-TDC for the TCSPC system combined with the tuned-TDL, the modified direct-histogram, the multi-phase architecture and a fast bin-with calibration. From the code density test results, it can be seen that the synergistic effects of the combination architecture significantly suppress the nonuniformity of carry-chains. Missing-codes and ultra-wide bins are entirely removed. The multi-phase architecture introduces extra design margins and flexibility in this design by reducing requirements of TDL length and clock timing. The direct-histogram architecture overcomes the sampling rate limits. A theoretical maximum sampling rate of 95.2Gsamples/s can be achieved with the temporal resolution of 10.5ps. Based on the modified direct-histogram architecture, a hardware-friendly bin-width calibration was further presented and evaluated. After the calibration, the $DNL_{peak-peak}$ and $INL_{peak-peak}$ were reduced to 0.08 and 0.13LSB, respectively. Compared with a traditional TDL,

the w_{eq} and σ_{eq} were improved from 27.57ps and 7.95ps to 10.55ps and 3.05ps respectively.

The proposed direct histogramming architecture has not been widely applied and tested before. Considering the advantages of the proposed TDC in the aspect of linearity and sampling rate, more tests and studies are suggested in the near future to evaluate the performances in different applications. However, compared with traditional architectures, the proposed TDC consumes a greater number of hardware resources. As a result, at this stage, the latter architecture design may not be suitable for multiple channel designs. Furthermore, the MR of the TDC is still limited, since the hardware resource costs for the direct-histogram counters will dramatically increase if the measurement range is extended.

Chapter 5: Multi-channel, low non-linearity time to digital converter based on 20nm and 28nm FPGAs

5.1 Motivation

The demand for multi-channel TDCs and TCSPC systems is growing rapidly, especially in areas such as LIDAR, ToF-PET, time-resolved spectroscopy and fast-FLIM applications. The TDCs of these applications normally have substantial requirements in terms of both high linearity performance and low consumption of hardware resource.

The FPGA-TDC proposed in Chapter 4 showed a significant improvement in linearity performance. However, the direct-histogram modules used in the first proposed TDC design consumed sizable hardware resources according to the resource utilisation report. This disadvantage limits the number of channels and the range of measurement, which in turn inhibits the direct-histogram architecture from multi-channel designs and long-range applications. For multi-channel applications, this study proposed and verified several novel methods to implement a resource-friendly multi-channel FPGA-TDC. Besides, the hope is to achieve a level of linearity performance similar to that of the first proposed FPGA-TDC.

5.2 System design and implementation

This TDC design aims to achieve an FPGA-based TDC/TCSPC systems with the features of high linearity, an extendable MR, and low resource consumption. In order to achieve these objectives, several novel methods and architectures are presented below:

● A sub-TDL averaging topology provides an ideal bubble-removal to correct the disorder

       problem from the carry-chain-based TDL structure.

- A new tap timing test method is proposed based on the sub-TDL structure to investigate the exact timing details of TDLs.

- A histogram compensation architecture and a mixed calibration are proposed and tested to limited hardware resource cost and calibrate long-term INL offset and nonlinearity.

In the present study, these methods and architectures were implemented in two different series of FPGA devices which have varied manufacturing processes and CLB structures in order to verify the universality of the proposed designs. Furthermore, two 96-channel TDCs/TCSPCs were built in these two FPGAs to demonstrate the former's potential for multi-channel TCSPC designs.

5.2.1 Sub-TDL averaging topology

The structure of the CLBs was evolved in the UltraScale series FPGA. As a result, this study applied both the Xilinx 28nm 7-series and 20nm UltraScale series FPGAs to verify the proposed designs. In the 7-series and earlier Xilinx FPGAs, each CLB module contains two independent Slices. The carry-chains of two Slices are in parallel without any intersections. In the UltraScale FPGAs, there is one Slice in each CLB module which consumes twice as much logic resource as the previous FPGA series. The simplified diagrams of the carry-chain in two types of Slice are shown in **Figure 5.1**.
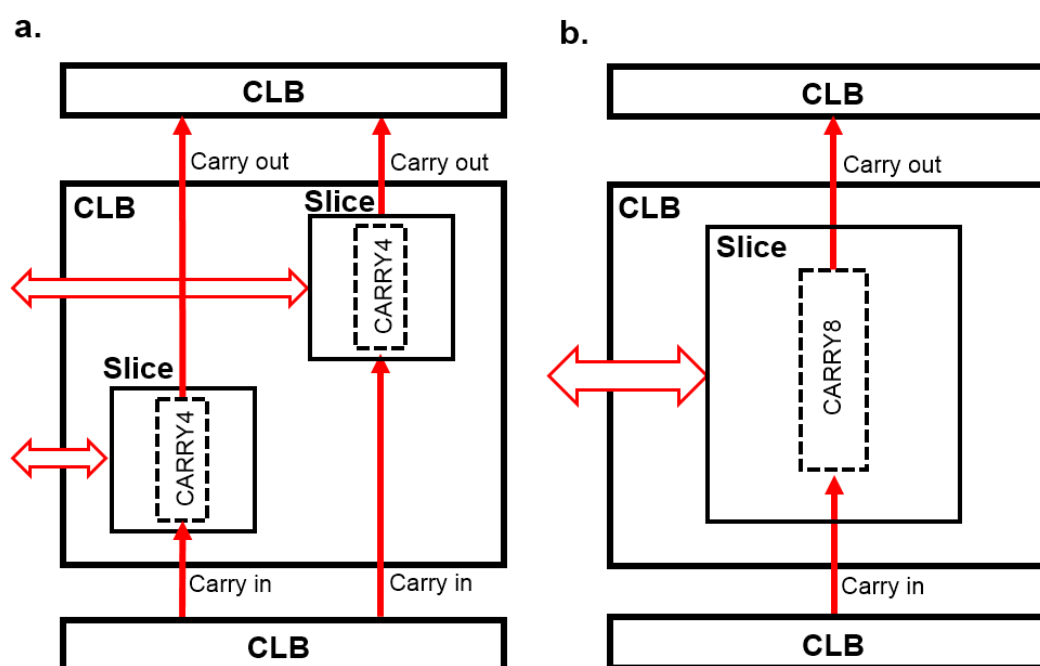
FIGURE 5.1, THE SIMPLE STRUCTURE DIAGRAM OF (A) 7-SERIES AND (B) ULTRASCALE FPGAS

In earlier FPGAs (before Xilinx UltraScale series), a CARRY4 module had four carry elements, and only one of two carry output types (type 'CO' or 'O') could be selected and registered within its local Slice [112, 113]. In UltraScale FPGAs, the new Slice structure contains 8 carry elements with 16 carry outputs in a carry-module (CARRY8). Different from CARRY4 modules, 16 D-FFs are included in each Slice and these 16 D-FFs are able to register all 16 carry outputs at the same time [114]. The details of carry-modules in both series FPGAs are shown in **Figure 5.2**.
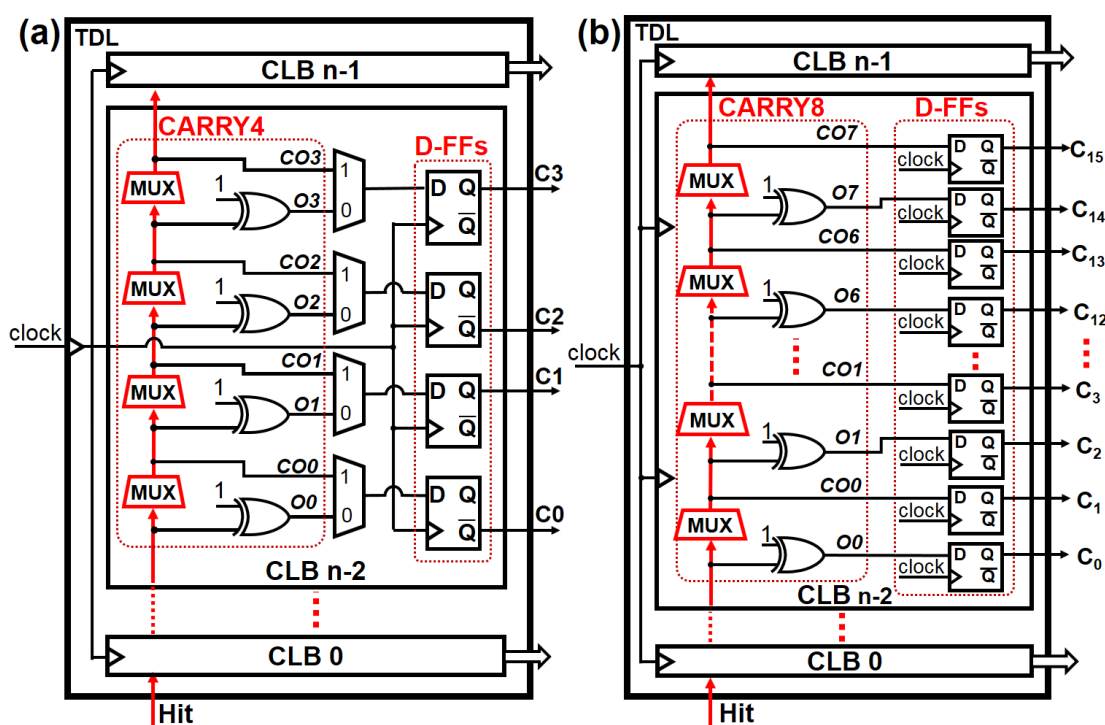
FIGURE 5.2 BLOCK DIAGRAM OF THE CARRY-CHAIN AND THE TDL IMPLEMENTED IN (A) VIRTEX-7 AND (B) ULTRASCALE FPGA

A traditional TDC assembles all of the carry-bits orderly to a single thermometer code. Bubble removal circuits and encoders follow the TDLs to convert the thermometer code to a one-hot code and finally encode the latter into a binary format (the fine-code). However, the disorder of the fast lookahead-carry architecture will cause problems such as bubbles, missing codes and poor linearity, as described in Chapter 3. In the UltraScale series FPGAs, the advanced technological process and the new structure of the Slice significantly reduce the average bin size. However, these non-linearity problems are more evident with a higher resolution.

The sub-TDL averaging topology is designed to solve the disorder and the 'bubble' problem. Different from the thermometer code conversion of traditional TDCs, the proposed topology segments output carry-bits and regroups them into multiple short thermometer codes (sub-thermometer codes) at first. This procedure is based on the location of carry-bits in CARRY4s or CARRY8s. For example, the first sub-thermometer codes consist of all of the first carry-bits of the CARRY4 or CARRY8 modules and follow their original order to assemble these carry-bits. Each sub-thermometer code is converted and encoded to individual binary code (sub-fine

code) with much fewer carry-bits. In this way, a complete TDL is segmented into several sub-TDLs equivalently. Finally, all sub-fine codes are summed together to form a complete fine-code which can be combined with a coarse counter in an interpolation architecture to extend the MR. The structure of the sub-TDL averaging topology for the 7-series and UltraScale devices is shown in **Figure 5.3**. In Xilinx 7-series FPGAs, a plain TDL is segmented into four sub-TDLs. In the UltraScale series FPGAs, a TDL can be segmented into 8 to 16 sub-TDLs based on the resolution requirement.
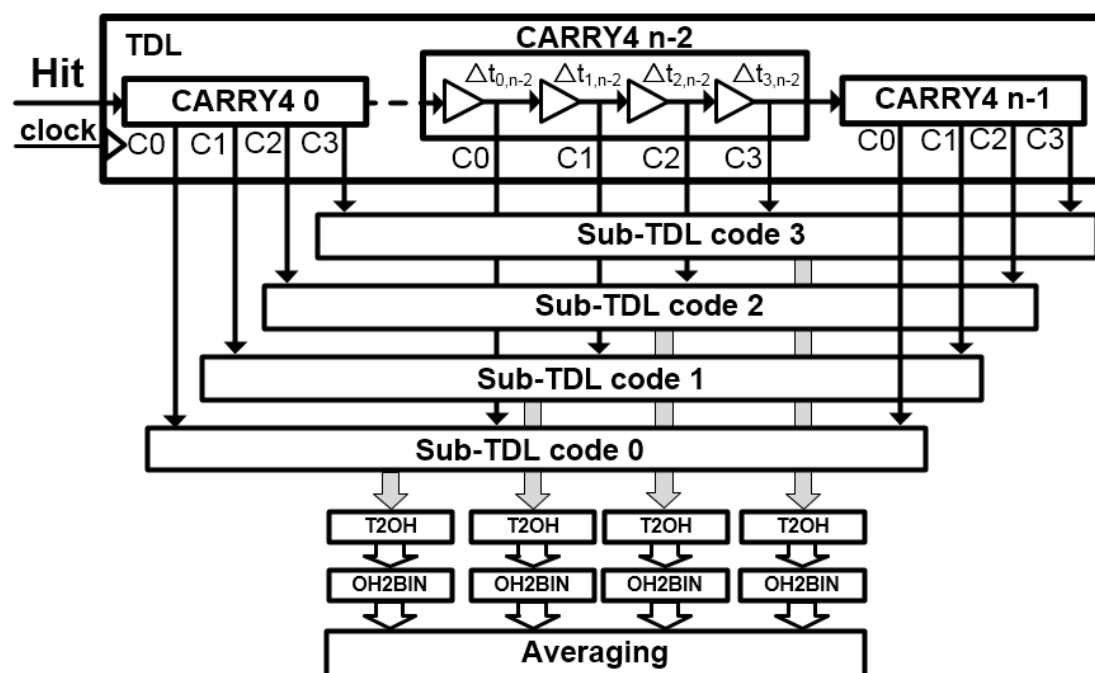
FIGURE 5.3 BLOCK DIAGRAM OF THE SUB-TDL TDC IMPLEMENTED IN A VIRTEX-7 FPGA.
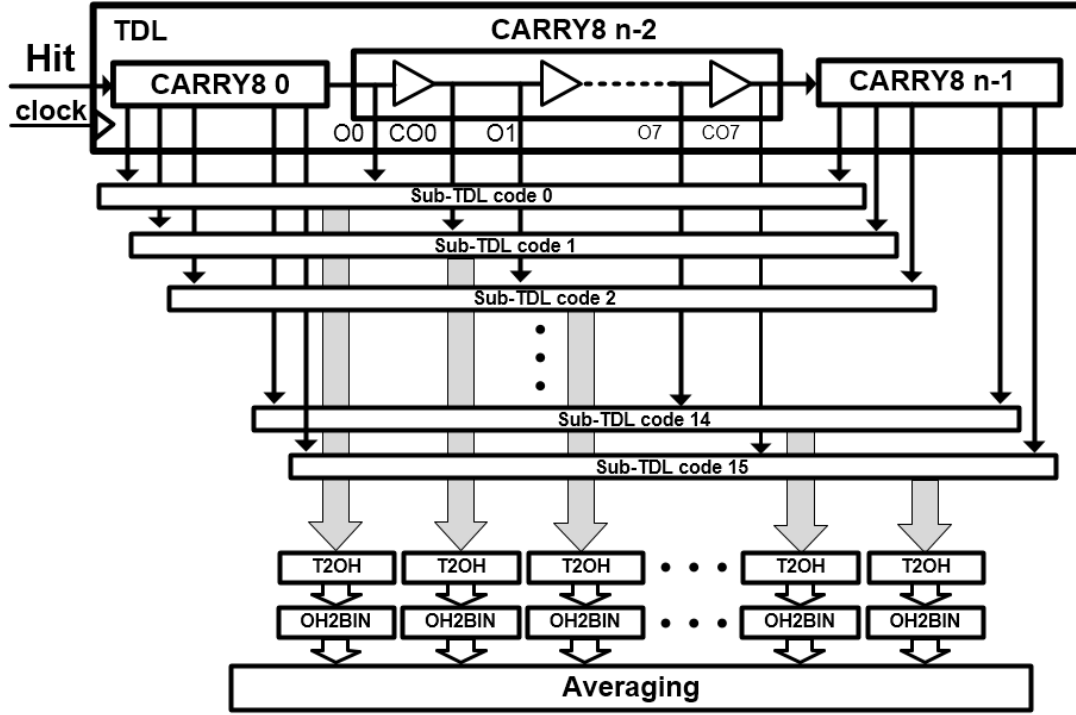
FIGURE 5.4 BLOCK DIAGRAM OF THE SUB-TDL TDC IMPLEMENTED IN AN ULTRASCALE FPGA.

An outstanding feature of the sub-TDL averaging topology is that the 'bubble' problem can be ideally solved without any additional nonlinearity, resource cost and encoding latency involved. Each sub-TDL has the same physical length and total propagation delay as the original TDL. At the same time, the number of taps in each sub-TDL is divided by the number of sub-TDLs. In this way, the bin-width of sub-TDLs is extended to the propagation delay of an entire carry-module, and the adjacent taps in a sub-TDL are located in separate carry-modules. For example, in series-7 devices, the bin-width of a plain TDL with $4n$ taps can be calculated as [73]:

$$LSB_{plain} = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{3}\Delta t_{j,i}}{4n} = \frac{4n \cdot \Delta t_{Ave}}{4n} = \Delta t_{Ave} \tag{5.1}$$

where n is the number of used CARRY4s, $\Delta t_{j,i}$ is the propagation delay of the j-th tap in the $i\text{-}th$ CARRY4 module, and the Δt_{Ave} is the average propagation delay of TDL taps. The delays from the backbone of the TDL to the corresponding FFs can be considered as a constant value which can be ignored in a simple model. Then, the bin-width of a sub-TDL can be calculated

as below:

$$LSB_{Sub} = \frac{\displaystyle\sum_{i=0}^{n-2}\sum_{j=0}^{3}\Delta t_{j,i}}{n} \approx \frac{4(n-1)\cdot\Delta t_{Ave}}{n} \tag{5.2}$$

As a result, the disorder problem caused by the lookahead-carry architecture cannot impact the monotonicity of sub-TDLs, and the bubble codes will not exist in sub-TDLs. Several code density tests were performed for individual sub-TDLs in both FPGAs. The tests system, which is shown in **Figure 5.5**, takes each sub-TDL as an independent TDC and utilises several individual memories to record the histogram of each sub-TDL. In these tests, no bubble removal or recognition methods were used, and a strict edge recognition pattern, '1111000', was utilised. Therefore, if the 'bubbles' codes are found in TDL outputs, this will cause a zero-width bin in the histogram.
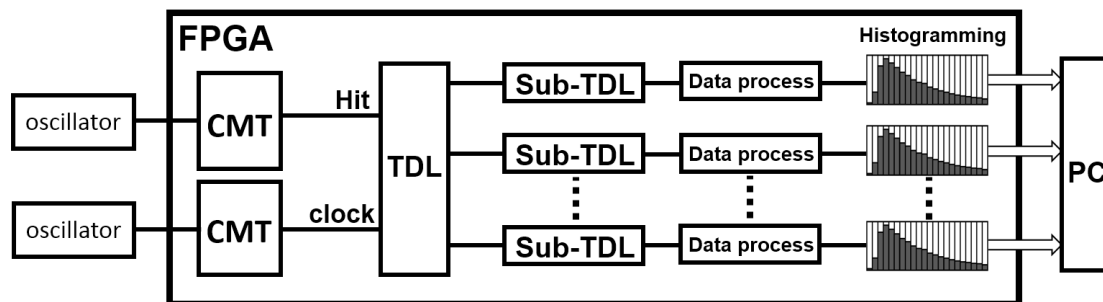
FIGURE 5.5 THE SYSTEM SETUP DIAGRAM OF THE CODE DENSITY TEST OF SUB-TDLS

Figure 5.6 shows the test results of individual sub-TDLs in both the Virtex-7 (4 sub-TDLs were tested) and the UltraScale (16 sub-TDLs were tested) FPGAs. According to the DNL plots, no missing-code and zero-width bin exist in all of the sub-TDLs in both FPGAs.
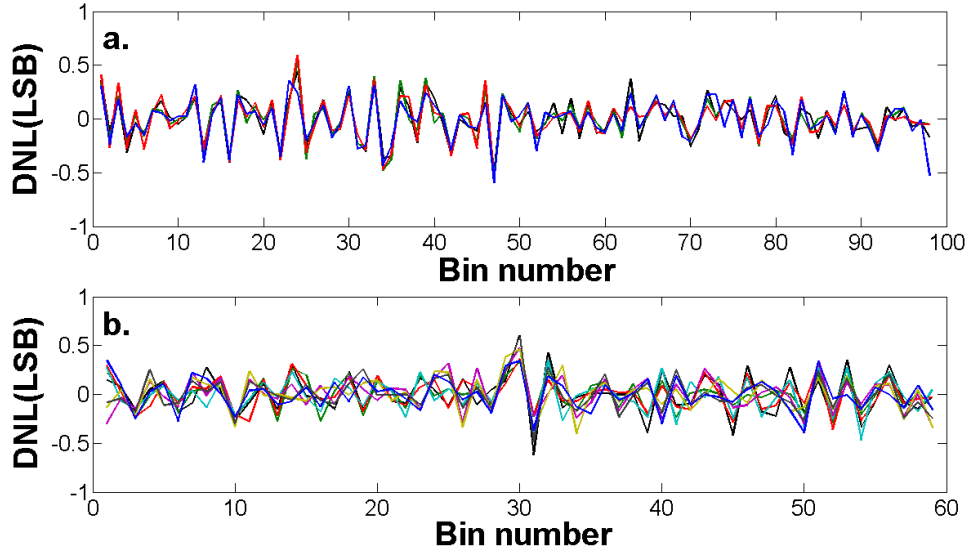
FIGURE 5.6 THE DNL AND INL PLOTS OF THE CODE DENSITY TEST OF SUB-TDLS

Due to the feature of bubble-free in sub-TDLs, the sub-fine codes converted from the outputs of sub-TDLs can be directly used to represent the propagation distances of HIT signals and the number of 'ones' in their thermometer codes. By summing all of the sub-fine codes, complete fine-codes are restructured, and equivalent multi-chain TDL averaging operations are performed. After the summing, the resolution of the entire TDC is calculated as below:

$$LSB_{Ave} = \frac{[4(n-1) \cdot \Delta t_{Ave}]}{4n} \approx \frac{LSB_{Sub}}{4} \tag{5.3}$$

when $n \gg 1$, there will be $LSB_{Ave} \approx LSB_{Plain}$. This operation can be equivalent to the 'ones-counting' method [5]. However, the proposed sub-TDL method will not introduce any additional resource cost or operation latency.

Three pseudo photon events (**a**, **b** and **c**) are shown in **Figure 5.7** to demonstrate the principle of the sub-TDL topology. The pseudo-events **a**, **b** and **c** fall into the 43rd, 44th and 45th bin, respectively. The disorder problem is revealed in the timing of four sub-TDLs, which causes the 'bubble' codes via conventional structure and encoding methods. By using the sub-TDL topology, the fine-codes of sub-TDLs are summed to form a final fine-code after encoding, and the measurement and encoding results of the three events are summarised in **Table 5.1**. From the findings it is possible to see that the sub-TDL averaging topology can obtain correct results and provide the capability to resist the misleading of the disorder problem.
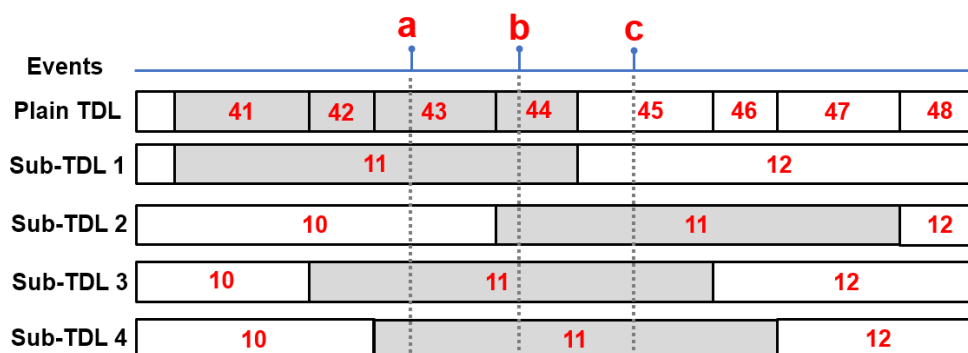
FIGURE 5.7 PRINCIPLE DEMONSTRATION OF THE SUB-TDL AVERAGING TOPOLOGY

Table 5.1 the measurement and encoding results of the three events

Photon Events	a	b	c
Location in a TDL	43	44	45
Fine-code of Sub-TDL 1	11	11	11
Fine-code of Sub-TDL 2	11	11	11
Fine-code of Sub-TDL 3	10	11	11
Fine-code of Sub-TDL 4	11	11	12
Final fine-code	43	44	45

5.2.2   Tap Timing Test

Base on the sub-TDL topology, the tap timing test was invented to provide a method with which to analyse and investigate the actual timing of each tap in TDLs. A traditional approach used to study the timing character of TDLs is the code density test. The code density test is a typical statistical analysis based on actual measurement results and is an accurate and widely-used method to estimate the bin-widths of TDLs. Nevertheless, this test method cannot fully reflect the natural status of the disorder problem nor the quantified timing of individual taps.

The tap timing test is another statistical test which feeds a mass number of random HIT signals into a TDC. Various conventional de-bubble methods and encoding circuits will transform the original output codes of TDLs. However, the sub-TDL topology will not modify the output codes of TDLs and can reflect the actual timing of taps correctly even 'bubble' codes exist. The

fine-codes of individual sub-TDLs are read out directly instead of being counted into histograms. Each measurement will generate multiple sub-fine codes, and differences between these sub-fine codes are statistically analysed further. Since all of the carry-chain modules in an FPGA chip have an identical structure, similar layout and approximate timing characteristics, the tap timing test can be developed into two main versions – a simplified version and an enumerate version. The simplified version can analyse the general timing information of carry-modules. In this test, the timing differences between adjacent taps versus the average bin-width are expressed as the equation below:

$$
\begin{cases}
D_n = \dfrac{\sum\limits_{m=0}^{m=L-1}(B_{n,m}-B_{n+1,m})}{L}, n = 0,1,...N-2, \\[4mm]
D_n = \dfrac{\sum\limits_{m=0}^{m=L-1}(B_{n,m}-B_{0,m+1})}{L}, n = N-1,
\end{cases}
\tag{5.4}
$$

where N is the number of used sub-TDLs in a plain TDL. The test measures L events and generates L groups of sub-fine codes. The $B_{n,m}$ is the sub-fine-code of the n-th sub-TDL and is generated by the m-th measurement. The simplified version of the tap timing test can be easily obtained. However, it cannot reflect the impacts of clock skews and process deviation.

The enumerate version of the tap timing test will calculate the timing of every tap in a TDL. The measured results need to be grouped according to the sub-fine code of a certain sub-TDL and multiple sets are thus generated. Developing from the simplified version, the equation of the enumerate version of the tap timing test is as below:

$$
\begin{cases}
D_n[i] = \dfrac{\sum\limits_{m=0}^{m=L[i]-1}(B_{n,m}[i]-B_{n+1,m}[i])}{L[i]}, n = 0,1,...N-2, \\[4mm]
D_n[i] = \dfrac{\sum\limits_{m=0}^{m=L[i]-1}(B_{n,m}[i]-B_{0,m+1}[i])}{L[i]}, n = N-1,
\end{cases}
\tag{5.5}
$$

where i is from 0 to I-1, and I is the number of the cascaded carry-chain module. In practice, i is the sub-fine code of a fixed sub-TDL. To obtain complete timing information of every tap, much more measurements need to be enumerated.

Figure 5.8 illustrates an example result of the simplified tap timing test of a TDL in the UltraScale FPGA. All of the taps in each CARRY8 (CO0 to CO7 and O0 to O7) are measured to study the timing characteristic thoroughly. From the results, it can be seen that the widest actual bin is approximately 2.3LSB in the single-sampling mode or 4.6LSB in the dual-sampling mode (from *CO1* to *CO5*). The narrowest bin is less than 0.1LSB (from *CO7* to *O4*). The timing bins of taps within the same CARRY8 (highlighted in red) indicate the actual status of the disorder problem and how said problem contributes to the nonlinearity of FPGA-TDCs. The number of sub-TDLs and taps in CARRY8 is selectable based on the requirements of temporal resolution and linearity. If partial taps are selected, the results of the tap timing test can be used as a reference to select taps with better timing uniformity. In this design, 8 out of 16 taps were used as working taps in the single sampling method.
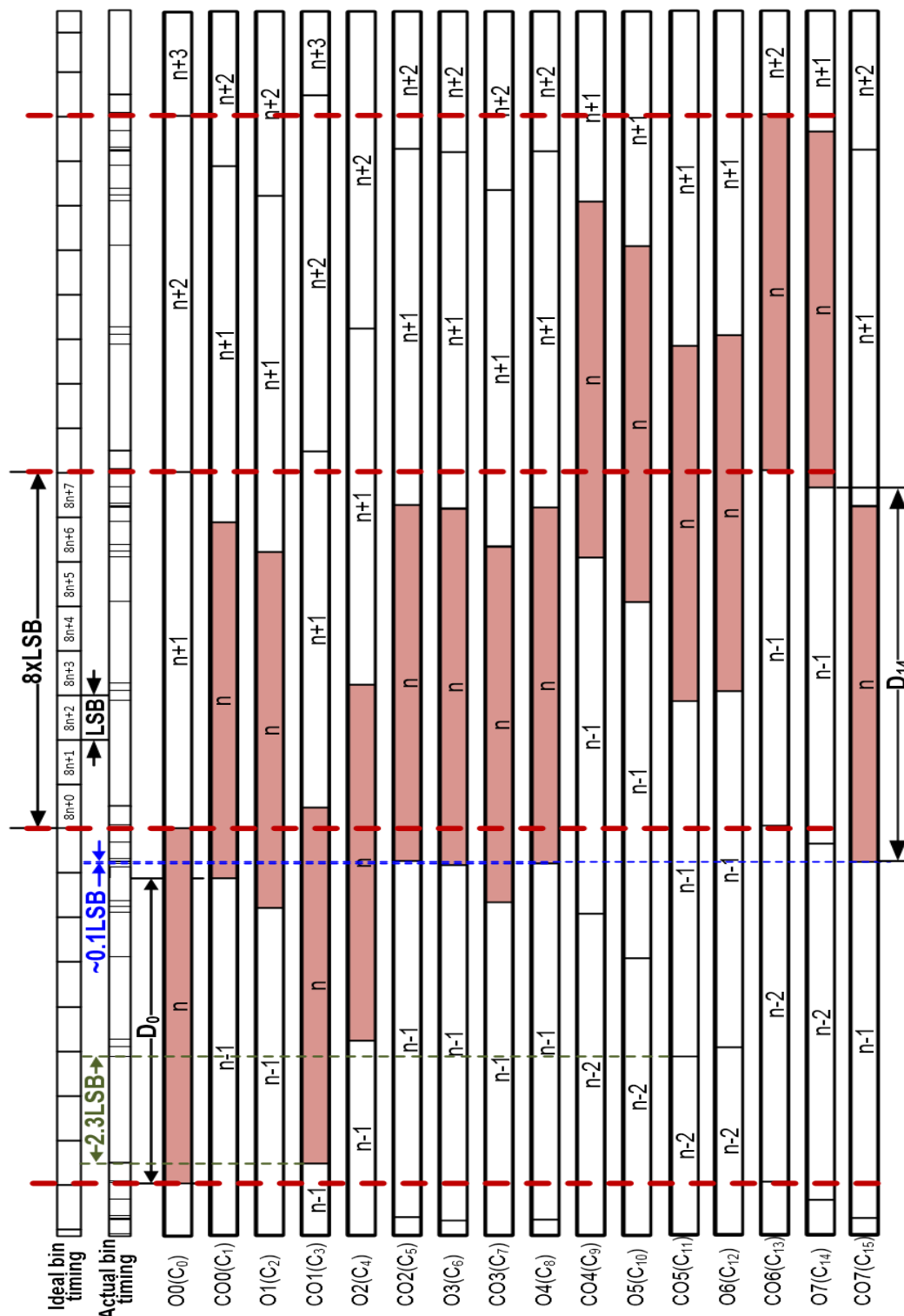
FIGURE 5.8 TIMING DIAGRAM BASED ON THE TAP TIMING TESTS OF THE 16 TAPS IN THE ULTRASCALE FPGA

5.2.3   Compensated histogram and mixed calibration method

In the first proposed FPGA-TDC design, the direct histogram counters had high logical resource consumption, which is not suitable for multiple-channel and long-range measurement designs. The compensated histogram architecture and mixed calibration method are designed for saving resources and improving the linearity at the same time.

The fine-codes of TDCs are used as the address of histogram bins. In traditional methods, each TDC channel generates one fine-code for each measurement. Therefore, the nonuniformity of a TDL is directly reflected in its corresponding histogram bin and generates missing-codes and ultra-wide bins. In order to solve this problem, the pseudorandom dither remapping method and a post-processing algorithm were presented in 2009 [104] and 2016 [13] respectively. However, the algorithm costs extra operational cycles and computing resources, especially for multi-channel architectures. Furthermore, additional analysis and evaluation are required, as the details of the algorithm and test results were not presented in [13].

The proposed histogram compensation is designed for fast and hardware-friendly offset and nonlinearity correction. In this design, the histogram is stored in block RAMs (BRAMs) to reduce logical resource consumption. Different from traditional histogramming methods, the histogram compensation method generates multiple fine-codes simultaneously for each measurement. The code density test for original TDLs was performed to obtain the natural timing character of TDLs at first. The equation to calculate the k-th code transition level, $T[k]$, is shown below:

$$T[k] = \sum_{n=0}^{k-1} W[n] = \sum_{n=0}^{k-1} \left\{ LSB \times (DNL[n]+1) \right\} \tag{5.6}$$

where $W[n]$ is the width of the n-th bin. Base on calculated code transition levels, **Figure 5.9** demonstrates the remapping theory of the proposed histogram compensation architecture. The compensation process is performed by utilising the main and the compensation calibration factor sets, BCF_m and BCF_c.
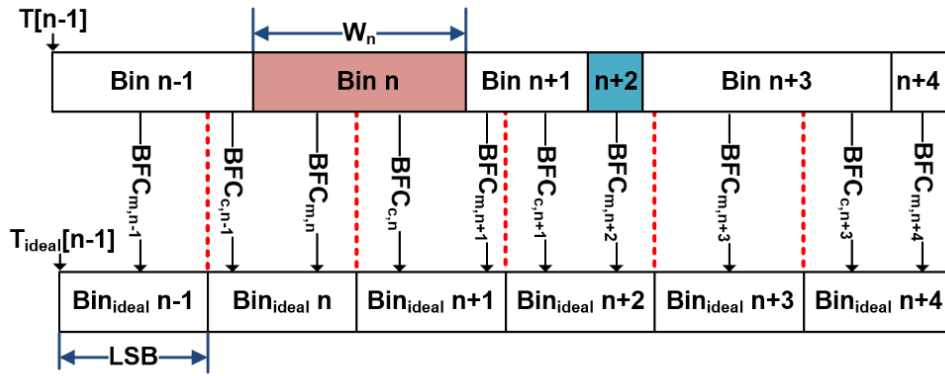
FIGURE 5.9   CONCEPT OF THE HISTOGRAM COMPENSATION METHOD.

The top row in **Figure 5.9** (from Bin[n-1] to Bin[n+4]) is the actual timing of a plain TDL fragment. The bottom row (from Bin_{ideal}[n-1] to Bin[n+4]) is the timing of a normalised TDL fragment with ideal linearity, and the widths of its bins are identical. The W_n is the width of the *n-th* bin. The T_{ideal}[n] and T[n] are the code transition level of the *n*-th ideal and actual histogram bin, respectively. In **Figure 5.9**, some bins are projected on different ideal bins, such as Bin *n,* and these are highlighted in red. Said kinds of bins are remapped into the two corresponding ideal bins by generating and using calibration factors, $BCF_{m,n}$ and $BCF_{c,n}$, as their fine codes during histogramming. For the bins which are projected onto a single ideal bin (such as Bin *n+2*, which is highlighted in blue in **Figure 5.9**), only the one calibration factor, $BCF_{m,n}$, is valid. The BCF_m and BCF_c sets are calculated by software normally, and the pseudocode of the calculation is shown below:

if (Tactual[k] < Tideal[k])

if (Tactual[k+1] < Tideal[k])

BCF$_m$ = K-1

BCF$_c$ = void

else if (Tideal[k]<Tactual[k+1])

BCF$_m$ = K-1

BCF$_c$ = K

else if …….

During the histogram compensation, the counts of histogram bins are modified and cause distortion. Therefore, the distorted counts of histogram bins need to be calibrated by the bin-

width calibration method. The bin-width calibration which is described in Chapter 4 needs to be modified to fit the histogram compensation architecture. The modified bin-width calibration contains multiple calibration factor sets, WCF_m and WCF_c, to correspond to the BCF_m and BCF_c sets. $BCF_{m,n}$, $BCF_{c,n}$, $WCF_{m,n}$ and $WCF_{c,n}$ are grouped into a mixed calibration factor set. A single-port BRAM or a distributed RAM can be used to store and hand out the mixed calibration factor sets as a calibration LUT. The fine-code of TDLs is used to address the mixed calibration factor sets which are stored in the calibration RAM.

One way to generate the bin-width calibration factors is to re-execute the code density test after the BFC_m and BFC_c have been reloaded, and all of the WCF_m and WCF_c have been reset to 2^M. After the second code density test, the WCF_m and WCF_c, corresponding to k-th fine-code can be obtained as below:

$$WCF_m[k] = \frac{2^M}{DNL(BFC_m[k]) + 1}$$

$$WCF_c[k] = \frac{2^M}{DNL(BFC_c[k]) + 1}$$

(5.7)

where M is the multiplication factor for the fixed-point calculation as described in Chapter 4. In this design, the calculation of the mixed calibration factors was performed by MATLAB. FPGA-embedded softcore or hardcore microprocessors are also available for on-the-fly calculation. The flow chart of the proposed histogram compensation procedure and the mixed calibration for TDCs in Virtex-7 FPGA are shown in **Figure 5.10**.
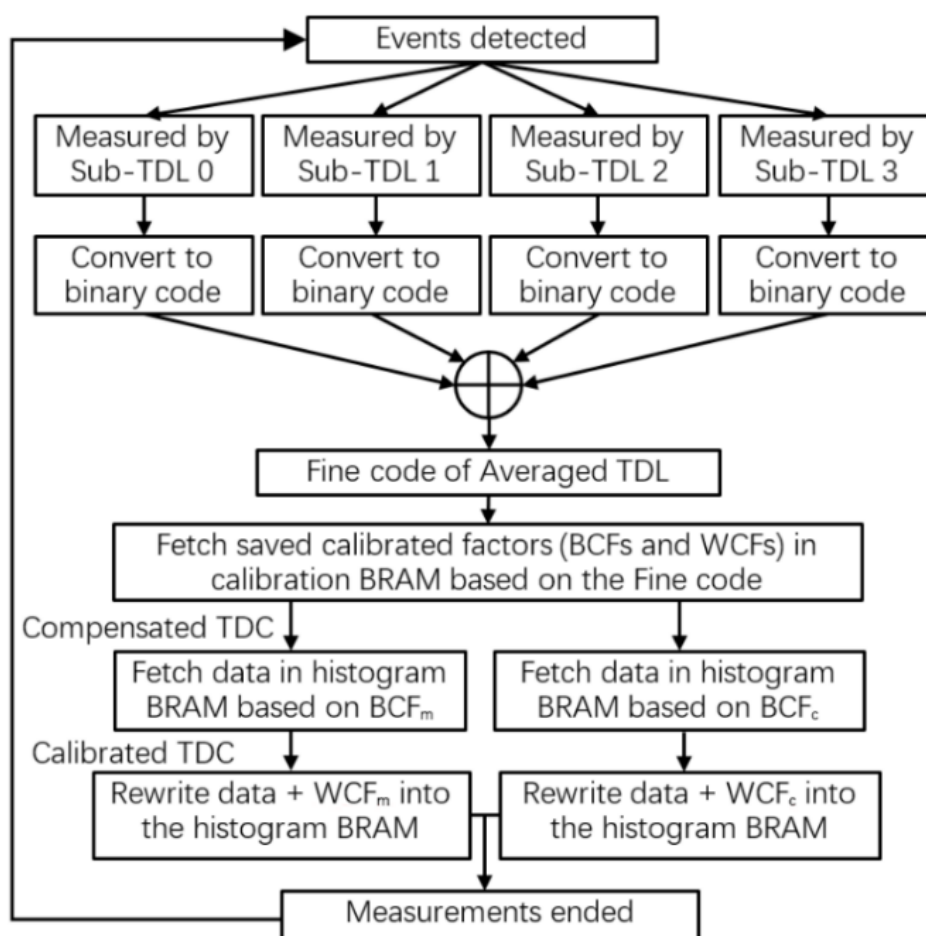
FIGURE 5.10 FLOW CHART OF THE TDC MEASURING EVENTS IN THE VIRTEX-7 FPGA.

In FPGAs, the histogram memory can be implemented by the true dual-port BRAMs, which provide two independent write/read ports to access a shared memory space at the same time. If only one BCF_c set is used, a single true dual-port BRAM can handle both the main and compensation histogramming procedures simultaneously, as shown in **Figure 5.10**. The original fine-codes are used to fetch the corresponding mixed calibration factor set from the calibration LUT and deliver it to the histogram memory. The BFCm and BFCc are used to address two bin counts in the histogram memory, respectively. Following this, the two bin counts will be increased by WCF_m and WCF_c respectively. Finally, the two increased bin counts will be overwritten into their original address in the histogramming memory. If the Fine&Coarse interpolation architecture is used, the BFC_m and BFC_c need to be multiplied by the coarse-codes. In this way, BRAM consumption can be minimised by using the true dual-

port BRAM. One disadvantage is that extra memory operation cycles or latency (2 to 4 clock cycles, depending on different read/write modes) for the operations of increment and overwriting are required. However, this latency can be ignored if the interpolation architecture is used, and the extended MR is longer than the extra operation cycle.
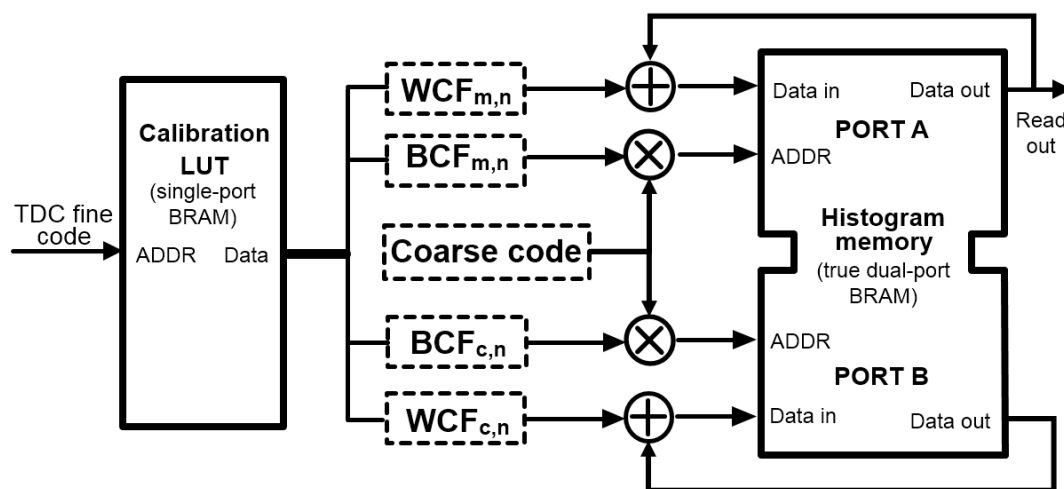
FIGURE 5.11 BLOCK DIAGRAM OF THE HISTOGRAM COMPENSATION WITH MIXED CALIBRATION WITH A SINGLE TRUE DUAL-PORT BRAM

For some applications which require a high sampling rate and a short MR, two or more true dual-port Block RAMs can be used in the parallel pipeline mode. As shown in **Figure 5.12**, the operations of increment and overwriting can be executed in an independent port. Thereby, the overwriting operation does not occupy the reading port, and the overall latency of histogramming can be minimised.
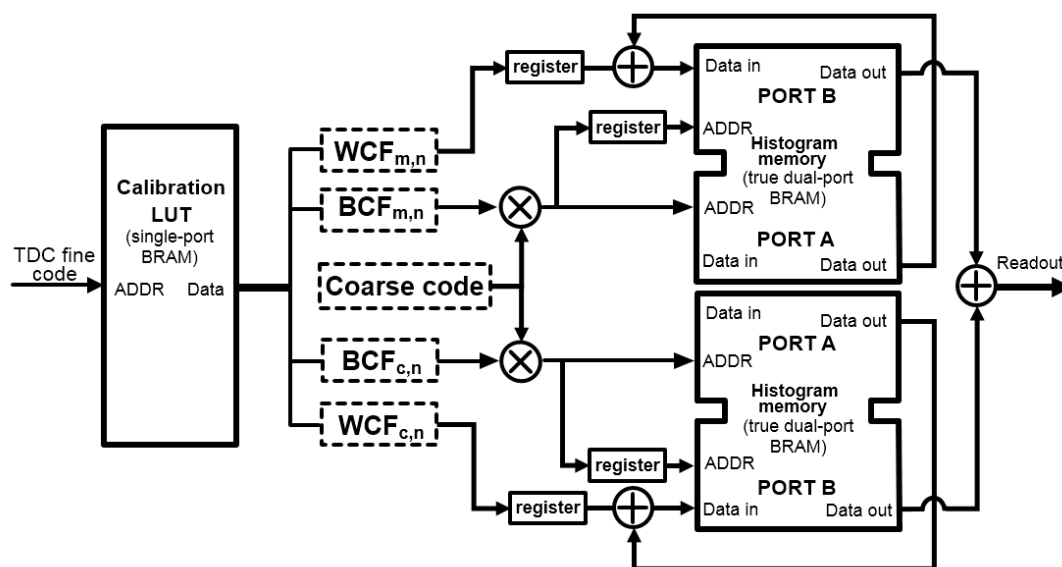
**FIGURE 5.12 BLOCK DIAGRAM OF THE HISTOGRAM COMPENSATION WITH MIXED
CALIBRATION WITH TWO SINGLE TRUE DUAL-PORT BRAM IN PIPELINE MODE**

5.2.4 Multi-channel FPGA-TDC configuration and hardware resource utilisation

The configuration of multi-channel FPGA-TDCs is required to find the balance among the
quantity of TDC channels, logical resource consumption, design density or routing congestion,
and linearity performance. To avoid impacts from large clock skews around the boundaries of
clock regions as described in Chapter 3, the location and length of TDLs need to be carefully
constrained based on the architecture of the FPGA chips used.

In this design, each CR in the UltraScale FPGA has 60 CARRY8 rows with around 2.4ns total
propagation delay. Considering that the propagation delay of TDLs should be longer than one
clock period, the frequency of the sampling clock should be higher than 416.66MHz in the
single-phase architecture and 208.33MHz in the dual-phase architecture. For the Virtex-7
FPGA, each CR has 50 CARRY4 rows with around 2.1ns total propagation delay. The
minimum frequency of the sampling clock is 476MHz in the single-phase architecture, which
oversteps the maximum operation frequency of BRAMs [115]. Therefore, the length of TDLs

is extended to 100 CARRY4 modules with 400 taps. In this way, the total propagation delay of TDLs is doubled to 4.2ns, and the minimum sampling frequency is reduced to 239MHz. Besides, the locations of TDLs are constrained in the two symmetrical middle CRs. This design implemented 96 TDC channels with single-phase architecture and 48 TDC channels with dual-phase architecture in both Virtex-7 and UltraScale FPGAs. The TDC layouts for two series FPGAs are shown in **Figure 5.13,** and the constrained TDLs are highlighted in orange.
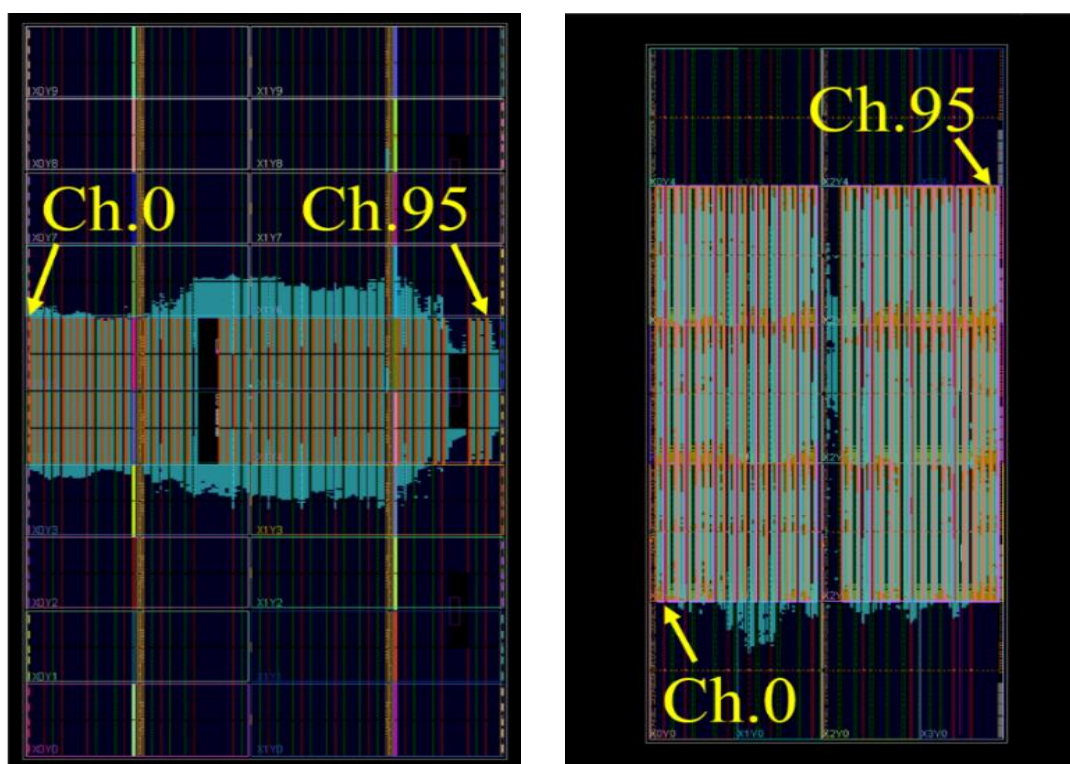
FIGURE 5.13 PLACE AND ROUTING RESULTS OF THE 96-CHANNEL TDCs IN VIRTEX-7(LEFT) AND ULTRASCALE (RIGHT) FPGAs.

During the implementation phase, the sufficiency of logic components and primitive resources (such as the FFs, LUT, carry-chain and BRAMs) is the primary factor when it comes to determining the number of channels. Besides these factors, the design density and the congestion of vertical and horizontal routing resources were also considered. The number of interconnecting routes is fixed within a certain area. The congestion problem will cause implementation failures if the design density is increased to a certain extent, even though the components and primitive resources are sufficient. To prevent the congestion problem, certain

space between adjacent TDC channels was guaranteed, and the vertical and horizontal routing utilisations were analysed by development software in the implementation phase.
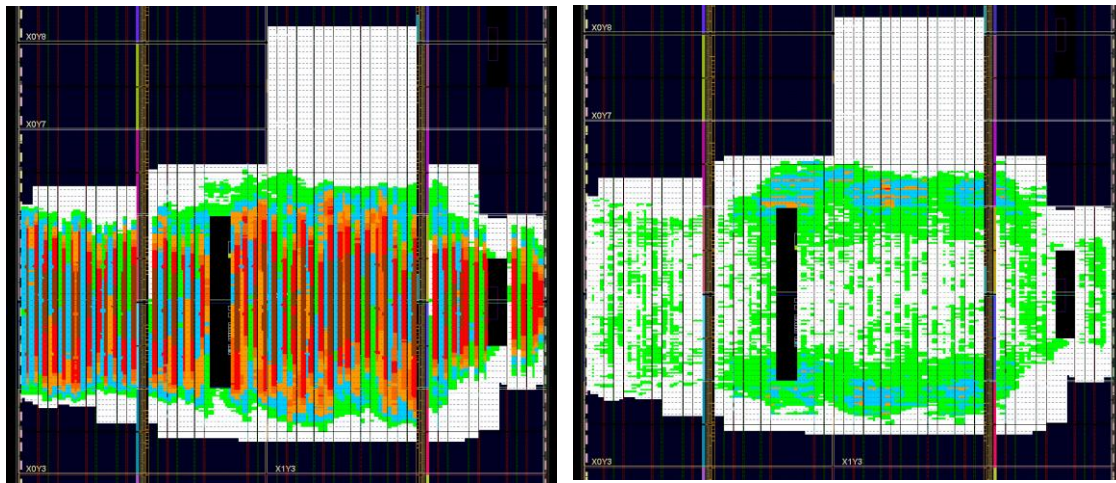
FIGURE 5.14 THE ROUTING UTILIZATION ANALYSATION AND VERTICAL(LEFT) AND HORIZONTAL(RIGHT) CONGESTION PLOTS OF THE 96-CHANNEL TDCS IN THE VIRTEX-7 FPGA
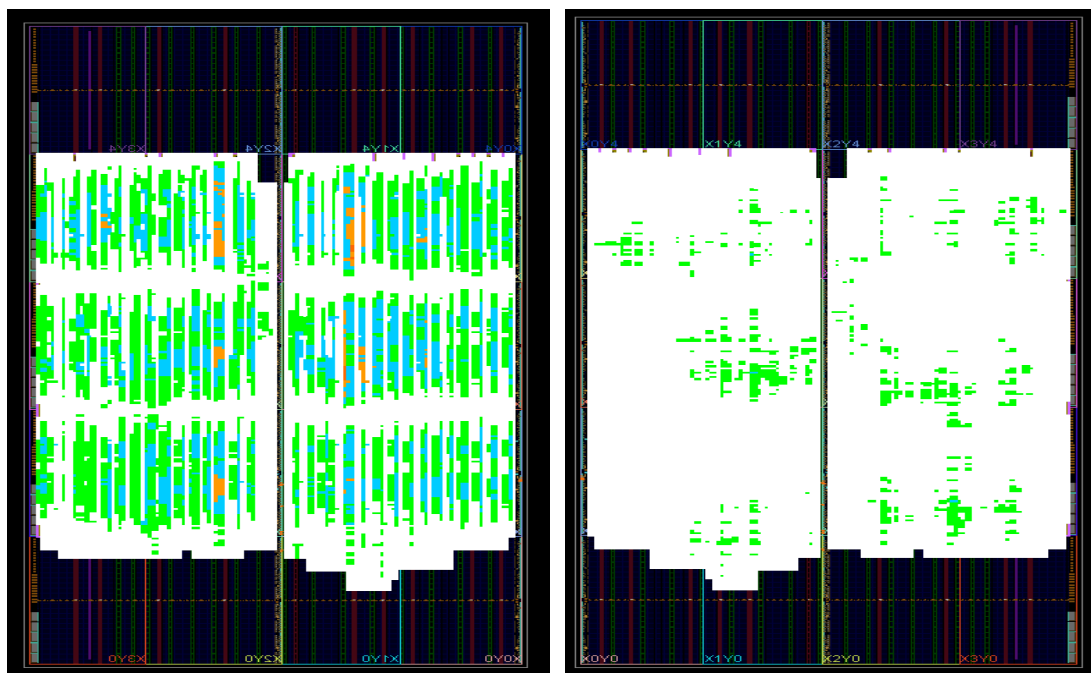
FIGURE 5.15 THE ROUTING UTILIZATION ANALYSATION AND VERTICAL(LEFT) AND HORIZONTAL(RIGHT) CONGESTION PLOTS OF THE 96-CHANNEL TDCS IN THE ULTRASCALE FPGA

Chapter 5: Multi-channel, low non-linearity time to digital converter based on 20nm and 28nm FPGAs

The routing utilisation analysis of multi-channel TDCs in the Virtex-7 and UltraScale is shown in **Figure 5.14**, **Figure 5.15 and Table 5.2**, respectively. The vertical and horizontal routing utilisation are described in different colours on the layout maps. Based on the analysis, the location constraint can be modified to achieve the balance between the routing congestion and the design density.

Table 5.2 Results of vertical and horizontal routing congestion analysis

Congestion range (%)	Colour	ratio of used CLB in Virtex-7(%)		ratio of used CLB in UltraScale (%)	
		Vertical	Horizontal	Vertical	Horizontal
0-20		28.4	28.24	41.38	61.54
20-40		5.8	34.55	15.69	33.92
40-60		15.0	30.83	26.89	4.51
60-80		16.8	5.5	14.87	0.02
80-100		9.0	0.84	1.11	0
100-150		5.7	0.03	0.05	0
150+		19.2	0.03	0	0

According to the congestion analysis in **Table 5.2**, the vertical routing resources are drained more quickly than the horizontal routing resources. For multi-channel TDCs in the Virtex-7 FPGA, around 25% of used CLBs have a >100% vertical congestion ratio. This phenomenon indicates that the circumjacent routing resources of a CLB will be occupied to compensate for the routing resources of adjacent CLBs. From the routing utilisation analysis in **Figure 5.14**, it can be seen that the occupied routing area diffuses away from the constrained area of TDLs. As a result, there is not much room left to increase design density within the four middle CRs. For the TDCs implemented in the UltraScale FPGA, more than 80% of CLBs consume less than 60% of vertical routing resources, and rare CLBs have a >100% congestion ratio. This result indicates that the design density of multi-channel TDCs has the potential to be enhanced. If large clock skews around the boundaries of CRs can be tolerated, there will be more room in the top and bottom areas of FPGAs to further increase the number of TDC channels.

Table 5.3 Logic resources utilisation

	Resource	Available	single channel		96-channel	
			Used	Utilisation %	Used	Utilisation %
Virtex-7	Slice	108300	712	0.65	24637	22.74
	LUTs	433200	1145	0.26	55790	12.87
	FFs	866400	1916	0.22	91968	10.61
	BRAM	1470	1.5	0.20	144	9.79
UltraScale	Slice	30300	80	0.26	7680	25.35
	LUTs	242400	703	0.29	68357	28.20
	FFs	484800	1195	0.24	114761	23.67
	BRAM	600	1.5	0.25	144	24

Table 5.3 summarises the hardware resource utilisation of a single, as well as the 96-channel TDCs, in both FPGA devices, respectively. Compared with the direct histogram architecture as described in Chapter 4, the proposed sub-TDL averaging topology and histogram compensation architecture have much lower hardware resource consumption, since the direct histogram counters were replaced by hardcore Block RAMs. From the results, it can be seen that the proposed TDC architecture provides the capability for multi-channel TDC designs with more than 200 channels.

5.3 Experiment results and discussion

This section introduces the experimental methods and discusses the test results of the proposed TDC design. The linearity performance of the TDC is thoroughly evaluated when all of the proposed methods are combined. The experimental setup and environment are similar to the first TDC design described in Section 4.3. The TDC design for 20nm UltraScale devices was implemented in a Kintex UltraScale XCKU040-2FFVA1156E FPGA on a KCU105 Evaluation Board [160], as shown in **Figure 5.16**. The clock signal sources for the code density tests were generated from a Silicon Labs Si5335A clock generator [161] and a Si570 programmable low-jitter 3.3V LVDS differential oscillator [162].
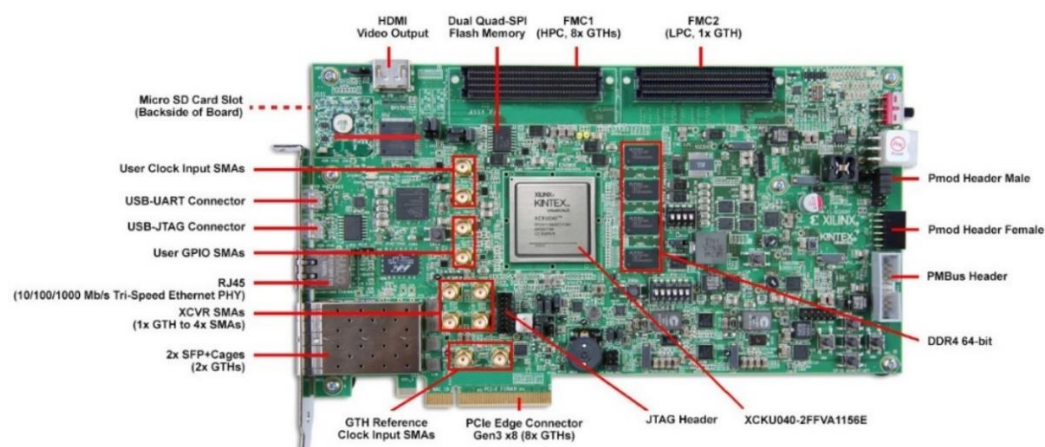
FIGURE 5.16 KCU105 EVALUATION BOARD WITH A KINTEX ULTRASCALE FPGA [160].

5.3.1 Result evaluation of sub-TDL averaging topology and tap timing test

To evaluate the improvement of the sub-TDL averaging topology and the tap timing test, tests to compare the traditional and the proposed TDCs were performed. In the Virtex-7 FPGA, the tuned-TDL and the sub-TDL average topology were combined. The plots of DNL and INL are shown in **Figure 5.17**. The $DNL_{peak-peak}$ is reduced from 4.78LSB to 2.73LSB, and the σ_{DNL} is reduced from 1.15LSB to 0.52LSB. The INL range is contracted from [-0.88, 5.90]LSB to [-2.54, 2.61]LSB, and the σ_{INL} is slightly reduced from 1.10LSB to 1.03LSB.

Chapter 5: Multi-channel, low non-linearity time to digital converter based on 20nm and 28nm
FPGAs
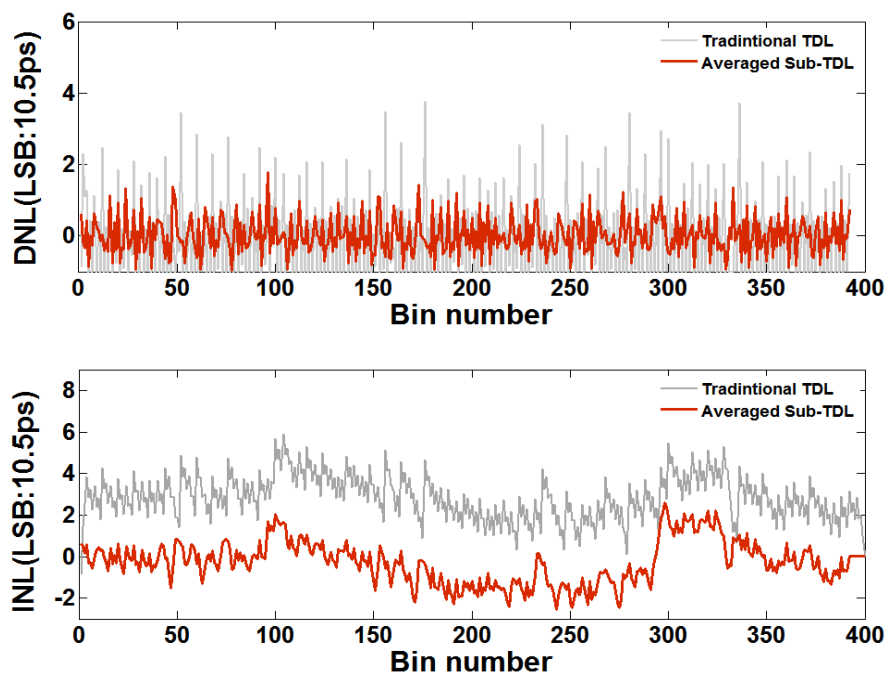
FIGURE 5.17 DNL RESULTS AND BIN-WIDTH DISTRIBUTION OF TRADITIONAL PLAIN TDC AND
THE TDL APPLY THE SUB-TDL AVERAGING TOPOLOGY IN A XILINX VIRTEX-7 FPGA DEVICE
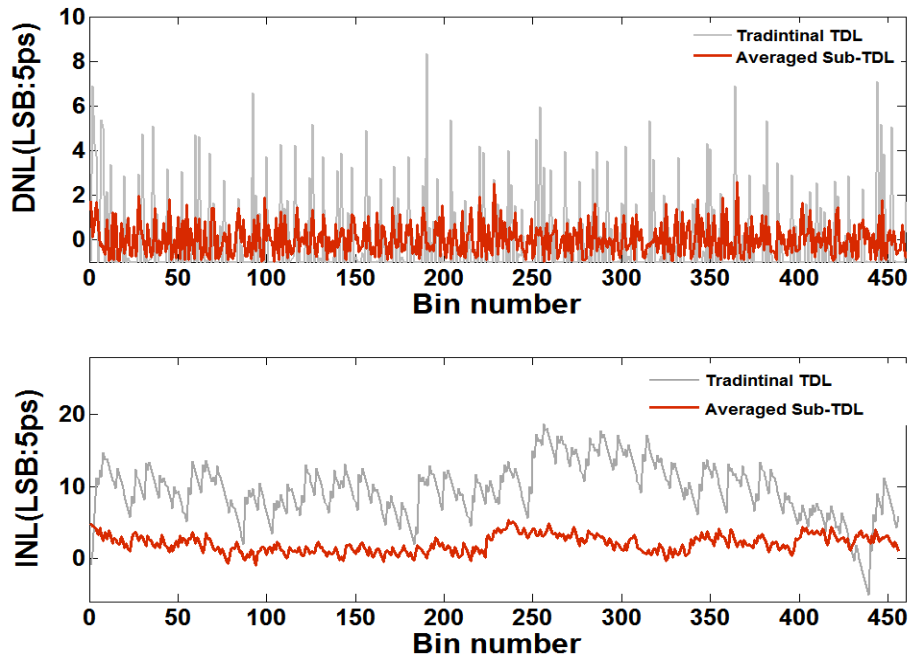
FIGURE 5.18 DNL RESULTS AND BIN-WIDTH DISTRIBUTION OF TRADITIONAL PLAIN TDC AND
THE TDL APPLY THE SUB-TDL AVERAGING TOPOLOGY AND TAP TIMING TEST IN A XILINX
ULTRASCALE FPGA DEVICE

For the UltraScale FPGA-TDC, 8 out of 16 taps in CARRY8 modules are selected based on the result of the tap timing test. **Figure 5.18** demonstrates the plots of the DNL and INL values. The $DNL_{peak-peak}$ is reduced from 9.09LSB to 3.53LSB, and the σ_{DNL} is reduced from 1.79LSB to 0.74LSB. The INL range is reduced from [-5.12, 18.76]LSB ($INL_{peak-peak}$ = 23.88LSB) with 3.94LSB standard deviation to [-0.90, 5.38]LSB ($INL_{peak-peak}$ = 6.28LSB) with 1.21LSB of σ_{INL}.
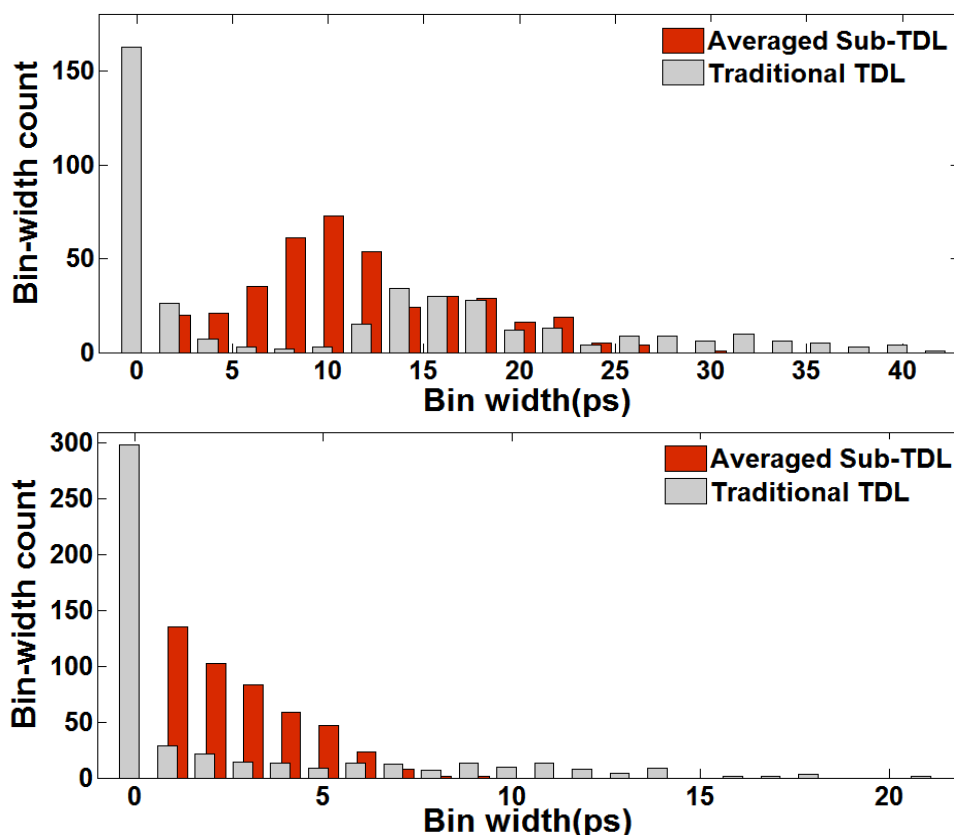
FIGURE 5.19 DNL RESULTS AND BIN-WIDTH DISTRIBUTION OF TRADITIONAL PLAIN TDC AND THE TDL APPLY THE SUB-TDL AVERAGING TOPOLOGY AND TAP TIMING TEST IN A XILINX ULTRASCALE FPGA DEVICE

The linearity improvement can also be demonstrated from the bin-width distribution shown in **Figure 5.19**. For the tested traditional TDC, the proportion of the zero-width bins is 40.7% in the Virtex-7 FPGA and 62.3% in the UltraScale FPGA. By applying the sub-TDL averaging topology in both FPGA-TDCs, the zero-width bins are removed, and the bin-width distribution is converged and optimised significantly.

The nonlinearity problems are more severe in the UltraScale FPGA-TDC compared with the Virtex-7 FPGA-TDC, as the average bin-width of the UltraScale FPGA-TDC is halved to around 5ps with more complicated carry-chains. The linearity performances are summarised in **Table 5**.4. From the results, it can be seen that the sub-TDL topology and tap selection based on the tap timing test provide more improvements in a more advanced UltraScale FPGA device.

TABLE 5.4. CODE DENSITY TEST RESULTS OF TDCS IN BOTH FPGA DEVICES

Unit: LSB	Virtex-7 (28nm)			UltraScale (20nm)		
	Raw-TDL	**Ave-TDL**	**Improve**	Raw-TDL	**Ave-TDL**	**Improve**
DNL	[−1, 3.78]	**[−0.95, 1.77]**		[-1, 8.35]	**[-0.96, 2.56]**	
DNL$_{pk-pk}$	4.78	**2.73**	**42.9%**	9.35	**3.53**	**62%**
σ$_{DNL}$	1.15	**0.52**	**54.8%**	1.85	**0.74**	**60%**
INL	[-0.88, 5.90]	**[-2.54, 2.61]**		[-5.12, 18.76]	**[-0.90,5.38]**	
INL$_{pk-pk}$	6.78	**5.14**	**24.1%**	23.88	**6.28**	**73.7%**
σ$_{INL}$	1.10	**1.03**	**6.4%**	3.94	**1.21**	**69.2%**

5.3.2 Result evaluation of the histogram compensation architecture

After the histogram compensation, the plots of DNL values and bin-width distributions of TDCs in both FPGA devices are shown in **Figure 5.20**. The missing codes and ultra-wide bins are readdressed to normalised bins by using the histogram compensation architecture. After the readdressing, the missing codes are compensated adequately, and the long-term drifting or offset (INL) are corrected.
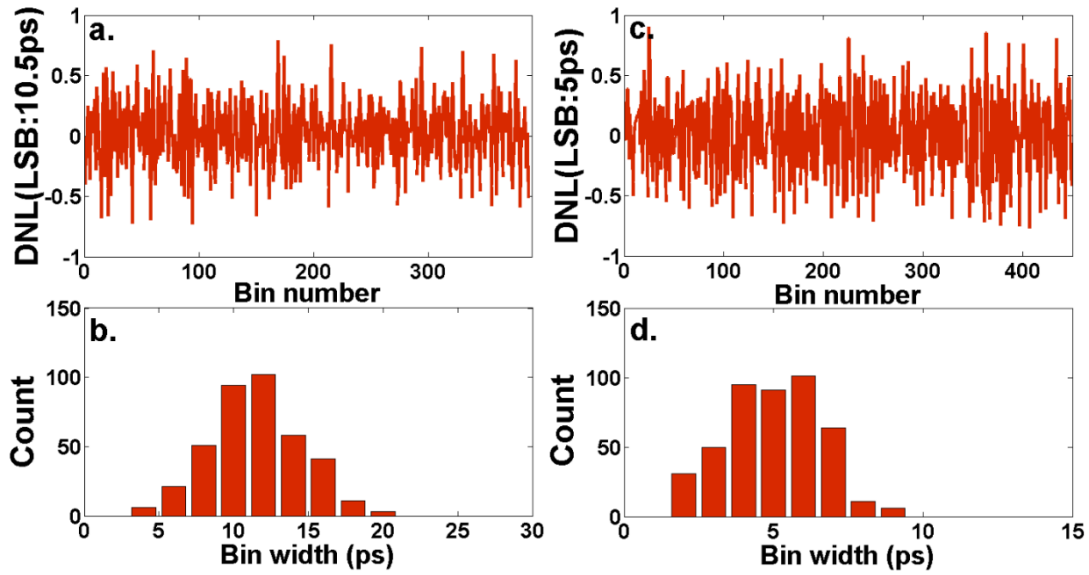
FIGURE 5.20 (A) DNL PLOT AND (B) BIN-WIDTH DISTRIBUTIONS OF THE COMPENSATED TDCS FOR VIRTEX-7. (C) DNL PLOT AND (D) BIN-WIDTH DISTRIBUTIONS OF THE COMPENSATED TDCS FOR ULTRASCALE FPGAS

From the DNL results, it can be seen that the $DNL_{peak-peak}$ values are improved from 2.73LSB to 1.52LSB in the Virtex-7 FPGA-TDC, and from 3.53LSB to 1.61LSB in the UltraScale FPGA-TDC. The maximum DNL value of both FPGA-TDCs is restrained lower than 1LSB, and achieving this value is difficult in traditional FPGA-TDCs. The σ_{DNL} is reduced from 0.52LSB to 0.29LSB, and from 0.74LSB to 0.35LSB in the Virtex-7 and UltraScale FPGA-TDCs respectively. From the bin-width distributions, it can be seen that missing codes and ultra-wide bins are eliminated in both FPGA-TDCs, and the shapes of bin-width distributions are much closer to the Poisson distribution.

5.3.3   Result evaluation of the mixed calibration

Figure 5.21 illustrates the comparisons between the uncalibrated and calibrated versions of the compensated TDCs in both FPGA-TDCs. The details of all linearity parameters of compensated and calibrated TDCs are shown in **Table 5.5**. Compared with the traditional FPGA-TDCs, the $DNL_{peak-peak}$ values of calibrated TDCs are decreased by around 36-fold and 40-fold, while the $INL_{peak-peak}$ values are decreased by around 34-fold and 37-fold in both FPGA devices.

Regarding the equivalent bin-width, w_{eq}, the results show that significant improvements are achieved by using the histogram-compensation. After the calibration, the w_{eq} values of TDCs are converged to their average bin-width.
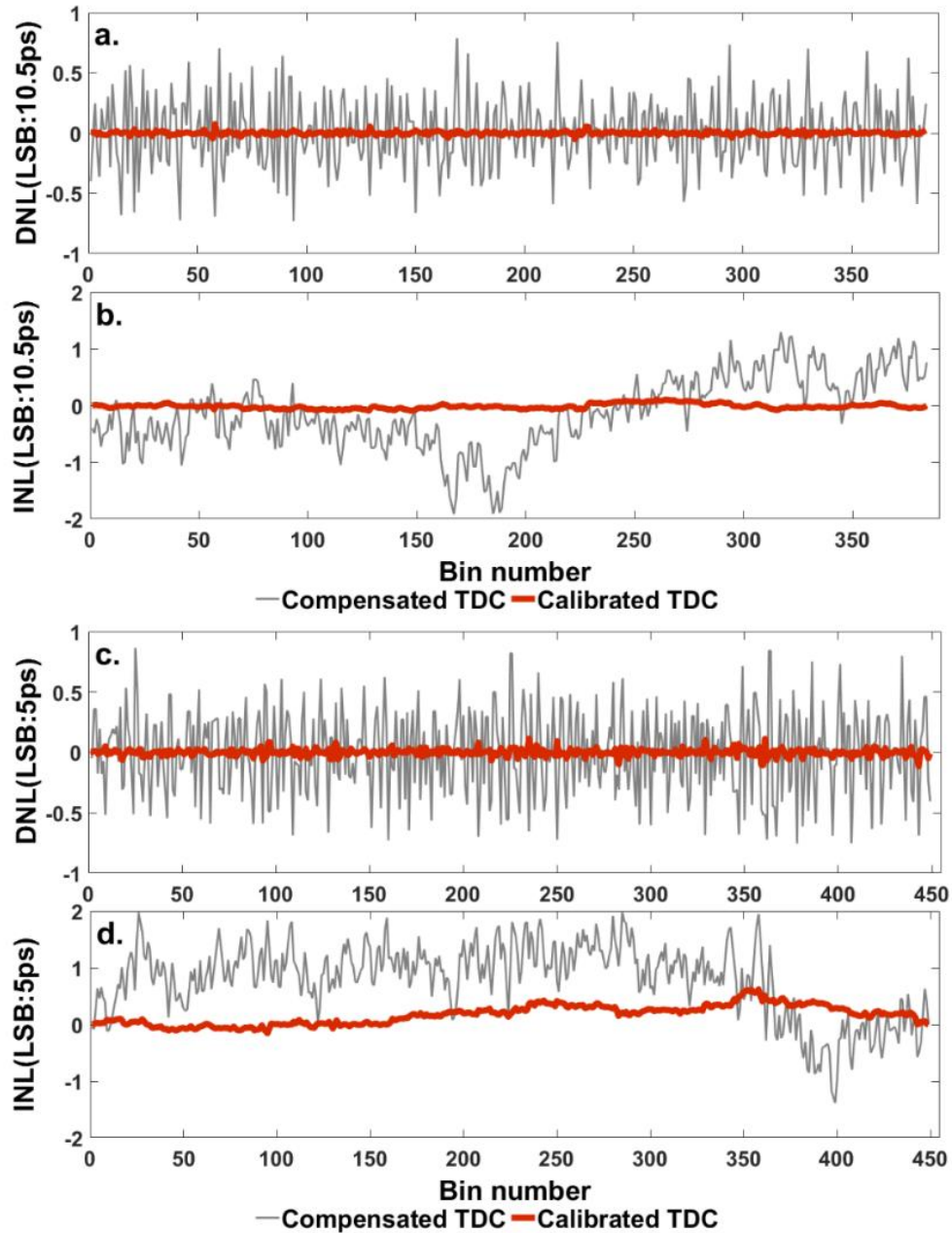
FIGURE 5.21 (A) DNL AND (B) INL PLOTS OF THE COMPENSATED AND CALIBRATED TDCS FOR THE VIRTEX-7 FPGA, AND (C) DNL AND (D) INL PLOTS OF THE COMPENSATED AND CALIBRATED TDCS FOR THE ULTRASCALE FPGA

Table 5.5 linearity parameters of traditional, compensated and calibrated TDCs

Unit: LSB	Virtex-7 (28nm)			UltraScale (20nm)		
	Traditional	Compensated	Calibrated	Traditional	Compensated	Calibrated
LSB(ps)		10.54			5.018	
DNL	[−1, 3.78]	[−0.73, 0.79]	**[−0.05, 0.08]**	[-1, 8.35]	[-0.75, 0.86]	**[-0.12, 0.11]**
$DNL_{pk\text{-}pk}$	4.78	1.52	**0.13**	9.35	1.61	**0.23**
σ_{DNL}	1.15	0.29	**0.01**	1.85	0.35	**0.03**
INL	[-0.88, 5.90]	[-1.91, 1.30]	**[-0.09, 0.11]**	[-5.12, 18.76]	[-1.38, 1.98]	**[-0.18, 0.46]**
$INL_{pk\text{-}pk}$	6.78	3.21	**0.20**	23.88	3.36	**0.65**
σ_{INL}	1.10	0.63	**0.04**	3.94	0.64	**0.16**
W_{eq} (ps)	26.75	11.73	**10.55**	25.09	5.85	**5.03**
σ_{eq} (ps)	7.72	3.39	**3.04**	7.24	1.69	**1.45**

5.3.4 Time interval measurement results

The measurement precision, the temporal resolution and the RMS resolution of TDCs can be obtained from the time interval measurement. The measurement setup is shown in **Figure 5.22**.
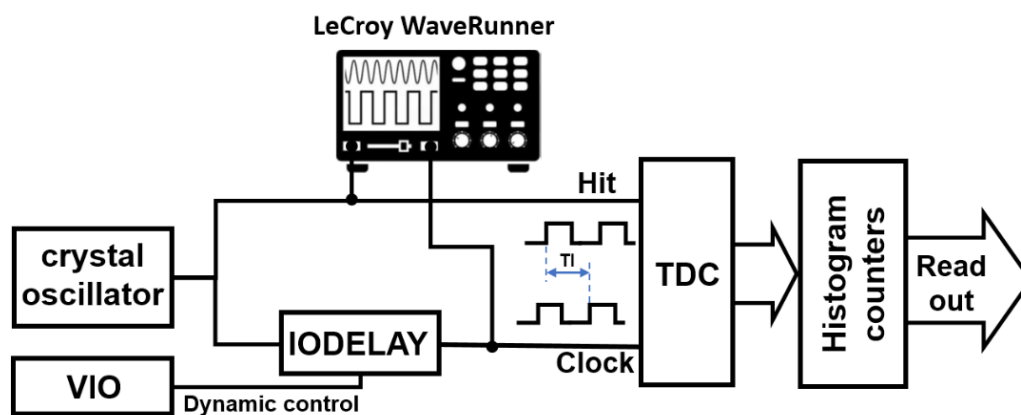
FIGURE 5.22 SETUP DIAGRAM OF THE TIME INTERVAL MEASUREMENT

The programmable delay generators, IDELAYE2 in the Virtex-7 FPGAs [158] and IDELAYE3 in the UltraScale FPGAs [163], are used to generate the known time intervals between the 'HIT' signals and the sampling clock signals. The time intervals are measured by the proposed TDCs (with the mixed calibration) and a high-end commercial oscilloscope (Teledyne LeCroy WaveRunner 640Zi) [164] at the same time. The measured signal pairs are connected to the

oscilloscope via two SMA connectors. The IODELAY modules are continuously calibrated by the low jitter reference clock via IDELAYCTRL modules to resist the drift and jitter of the voltage and temperature during measurements. In order to minimise the additional jitter, the IODELAYs are constrained to a fixed location, closest to the 'HIT' port of the TDCs.

The IODELAYs are dynamically controlled by the VIO and the ChipScope, while the slope tests are performed with an increasing step of 39ps and 4.6ps for the IDELAY2 and IDELAY3 modules respectively. A few IODELAY modules are cascaded to generate enough range of time intervals to cover the full length of the tested TDLs. Each test records 80,000 samples, and the time intervals are calculated based on the histogram distribution and the temporal resolution.
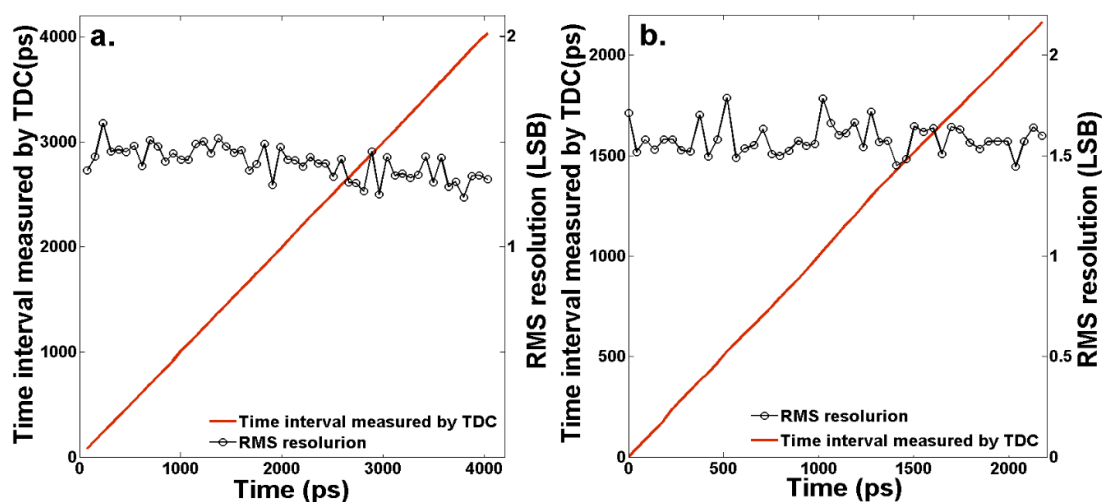
FIGURE 5.23 TIME INTERVAL MEASUREMENT RESULTS AND RMS RESOLUTIONS OF THE CALIBRATED TDCS FOR (A) VIRTEX-7 AND (B) ULTRASCALE FPGAS

The measurement results and the RMS resolution are shown in Figure 5.23. The average RMS resolution is 14.59ps with $\sigma = 0.84$ps for the Virtex-7 FPGA-TDC and 7.80ps with $\sigma = 0.45$ps for the UltraScale FPGA-TDC. The standard deviations of the time intervals measured by the oscilloscope are 14.86ps for the Virtex-7 FPGA and 8.55ps for the UltraScale FPGA. The standard deviations of the differences between the measured results (obtained by the TDC) and the expected results (obtained by the oscilloscope) are 4.04ps and 5.37ps for the Virtex-7 and the UltraScale FPGAs, respectively.

5.3.5 The uniformity of multiple channel TDCs

The linearity uniformity among the 96 TDC channels needs to be tested. Testing all of the 96 TDC channels is time- and resource-consuming. As a result, 16 out of 96 TDC channels in both Virtex-7 and UltraScale FPGAs are tested by the code density test. These selected 16 TDC channels are evenly located in the constrained area. From the DNL and INL values and their standard deviations in **Table 5.6**, it can be seen that the linearity performances of the TDC channels in different locations have achieved good uniformity.

Table 5.6. the linearity performance of 16 out of 96 TDC channels in both FPGAs

	channel	0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	ave
Virtex-7	DNL_{pk-pk}	0.17	0.20	0.14	0.15	0.17	0.12	0.22	0.12	0.12	0.15	0.18	0.14	0.15	0.13	0.18	0.15	**0.15**
	σ_{DNL}	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	**0.02**
	INL_{pk-pk}	0.32	0.32	0.36	0.35	0.38	0.37	0.32	0.34	0.38	0.33	0.45	0.27	0.29	0.29	0.45	0.43	**0.35**
	σ_{INL}	0.08	0.06	0.09	0.06	0.08	0.09	0.07	0.08	0.09	0.07	0.10	0.05	0.05	0.06	0.10	0.10	**0.08**
UltraScale	DNL_{pk-pk}	0.30	0.27	0.27	0.30	0.27	0.27	0.31	0.27	0.29	0.28	0.27	0.25	0.31	0.23	0.25	0.22	**0.27**
	σ_{DNL}	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	**0.04**
	INL_{pk-pk}	0.81	0.69	0.48	0.69	0.75	0.45	0.69	0.60	0.57	0.41	0.55	0.60	0.64	0.49	0.62	0.37	**0.59**
	σ_{INL}	0.18	0.15	0.10	0.18	0.17	0.11	0.17	0.13	0.13	0.08	0.10	0.12	0.16	0.10	0.12	0.07	**0.13**

5.4 Summary

In this chapter, several novel methods and architectures are presented, including the sub-TDL averaging TDL topology, the tap timing test, a hardware-friendly histogram compensation architecture, and a mixed calibration method. The 96 multi-channel TDC designs are demonstrated and evaluated in two different FPGA series.

The sub-TDL averaging topology can completely remove the bubbles and zero-width bins without extra resources and operation time costs and any linearity loss. This capability provides an ideal solution for the 'bubble' problem. Based on the sub-TDL averaging topology, the novel tap timing test can statistically quantify the actual tap timing of TDLs and elucidate the nature of the nonuniformity of TDLs. For the multi-channel implementation and long measurement

range, this chapter proposed the histogram compensation and mixed calibration methods to correct the measurement offset and calibrate the bin-width deviation directly with limited resource consumption. Said benefits and improvements can support the application of these new methods to other FPGA-TDC designs.

The proposed TDC designs integrate these methods and are implemented and tested in two different FPGA series which have varied architectures and performances. The temporal resolutions (LSB) of 10.5ps and 5.02ps with the RMS resolutions of 14.59ps and 7.80ps were verified in the Virtex-7 and the UltraScale FPGAs, respectively. The tests show that the second TDC improved the linearity to the same level as the first design and provided good applicability in different FPGA devices. By implementing 96-channel TDCs in both FPGAs, the proposed methods and architecture demonstrate their suitability for multi-channel designs and excellent uniformity on linearity. The proposed architecture also has the significant potential for applications which require parallel measurements, such as the fast 3D ranging or time-resolved imaging and Raman spectroscopy.

Chapter 6: An integrated 40nm CMOS 192 x 128 SPAD-TDC array sensor for time-correlated wide-field imaging

6.1   Introduction

Wide-field time-correlated imaging requires both spatial and temporal information on the detected photons simultaneously. Scanning and large-scale array are two implementation architectures for wide-field time-correlated applications [3, 61]. Scanning methods are normally combined with a single high-performance photon detector for one-dimensional scanning. A TCSPC channel is used to measure and collect the timestamps of detected photons. However, scanning methods usually suffer from several shortcomings: 1) the limited number of channels due to the large size and high price of detectors and commercial TCSPC systems; 2) limited sampling rate and imaging speed as a result of the scanning speed and dead-time of sensors and TCSPC systems; 3) the increasing complexity and cost of the entire TCSPC system due to the mechanical or optical scanning system. By applying the advanced CMOS technology, a large-scale SPAD and TDC array can be implemented within a single chip. A distinct advantage of large-scale arrays is that all pixels are independent and in parallel with each other to implement an independent TCSPC channel. The benefit of the relatively low dead-time of SPAD devices is that the overall sampling or acquisition rate of the entire array can be enhanced to the GHz level [13, 165]. However, SPAD arrays have low fill factors because the partial area of pixels is allocated to electronic circuits and in-pixel TDCs [12, 61, 67].

This chapter presents a large-scale multi-channel TCSPC system for wide-field FLIM applications. This system is based on a cutting-edge, large-scale sensor and TDC array [67, 68], an FPGA firmware, and software UI. The architecture and features of the SPAD array are presented in this chapter. The design and implementation of the firmware and software UI, as

well as the characterisation of the TCSPC system, are fully described. In oder to verify the applicability of the proposed system, this design further performs a basic FLIM experiment.

6.2 The 192x128 SPAD sensor with in-pixel TDC array

A large-scale photoelectric sensor array consists of 192x128 SPAD&TDC pixels and is manufactured by STMicroelectronics 40nm CMOS technology. It is used in the proposed TCSPC system. The overall layout and packaging of the sensor array are shown in **Figure 6.1**.
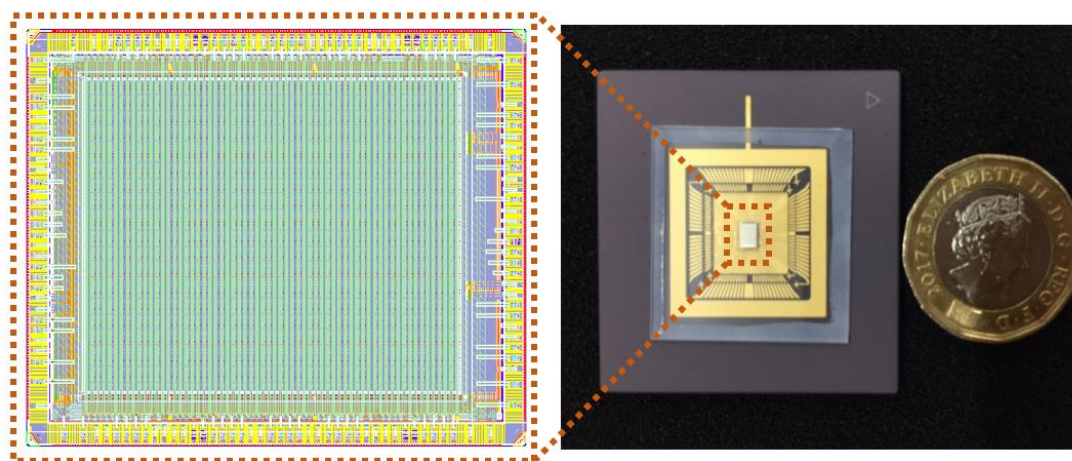
FIGURE 6.1 THE APPEARANCE AND OVERALL LAYOUT OF THE SENSOR ARRAY

The sensor can be operated in either the Photon Counting (PC) mode or the Time-Correlated (TC) mode to provide intensity and time-correlated imaging. In the PC mode, the number of detected photons during a certain exposure period is counted using a 9-bit in-pixel counter as light intensity. In the TC mode, the flight time of detected photons is measured by in-pixel RO-TDCs, and the digital timestamps are outputted. The features of the SPAD&TDC array are summarised below:

- 128x192 pixel SPAD and time correlate imaging array
- 18.396 μm x 9.198 μm pixel pitch
- 12.4% fill factor and up to 41% fill factor with the microlenses array
- provides PC and time-correlated modes and time gating mode

- In-pixel 13-bit RO-TDCs for parallel photon timestamping
- tunable TDC time resolution from 33ps to120ps for different applications
- 9-bit photon counter to provide a good intensity depth.
- 365 M pixels per second and up to 15 Kf/s operation and readout rate.
- Higher frame rates are accessible by dynamic disable or bypass a subset of pixels rows.
- Capability to provide and receive synchronisation signal to and from external lasers with different signal standard as the master and slave mode.
- On-board independently adjustable power supplies.
- Robust 180-pin CPGA optical package for easy device handling.

6.2.1   Pixel Architecture

Each pixel of the sensor consists of five blocks: a SPAD frontend, an RO-TDC, a ripple counter with control logic and readout circuits. The block diagram of the pixel architecture is shown in **Figure 6.2**. All pixels can be manually configured to the PC mode or the TC mode via a global signal, namely 'TCSPC'. Therefore, the sensor array is not able to work in a mixed-mode.
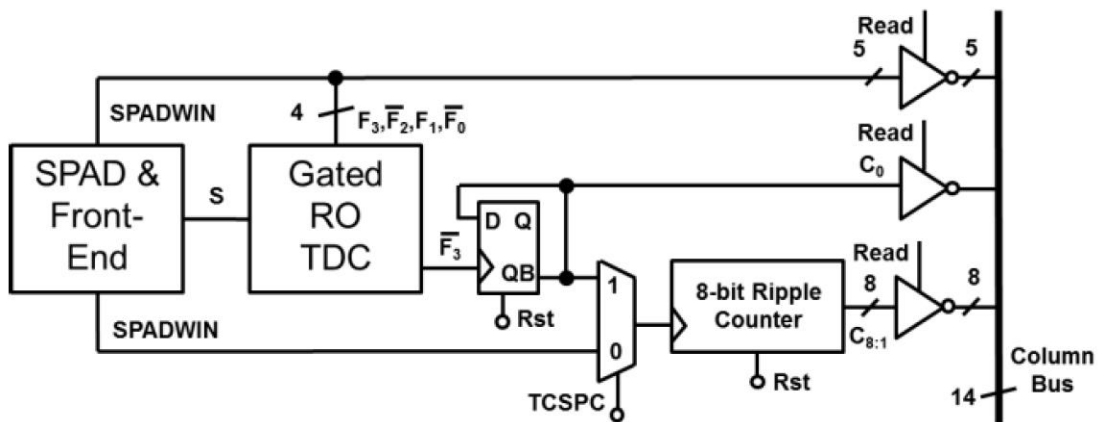
FIGURE 6.2 THE ARCHITECTURE DIAGRAM OF A SPAD PIXEL [67]

a)   SPAD frontend and logic control circuits

The schematics of the SPAD frontend and logic control circuits are shown in **Figure 6.3**. The SPAD is placed in the Geiger mode by applying its cathode with a voltage 'VHV'. The VHV is

set above the breakdown voltage of the SPAD (>12.9 V), and the excess amount is called the excess bias. The avalanche effect will be triggered in the SPAD fires when a photon is detected.
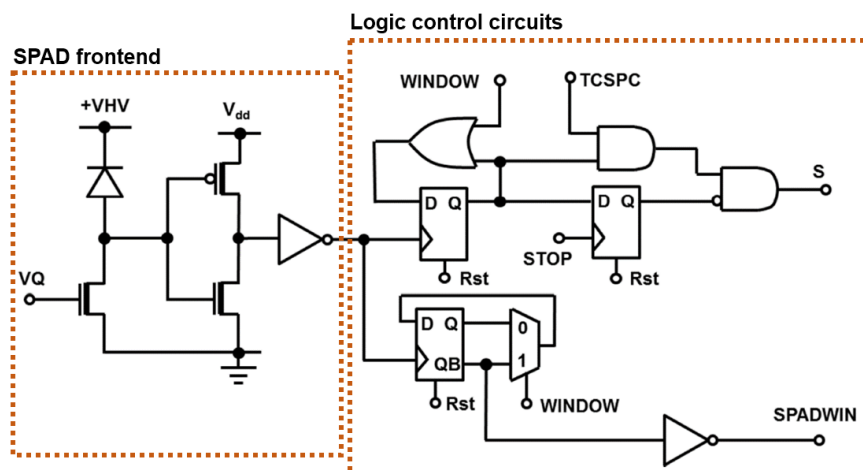
FIGURE 6.3 SCHEMATIC OF THE SPAD AND FRONTEND ELECTRONICS [67].

Once the avalanche effect is triggered at a SPAD frontend, a current flow from the VHV through the diode and the voltage at the SPAD anode will rise rapidly to the excess bias voltage. When the voltage of the diode is lower than the breakdown voltage, the current will cease flowing through to the diode. Following this, the voltage at the anode will be quickly pulled down to the ground level by a passive quench circuit, and the diode will be reset to the Geiger mode for the next detections. A global gate voltage, V_{quench}, controls the quench transistor, and a higher V_{quench} leads to a faster rate of quench and recovery. This procedure generates pulse signals, and the signal at the SPAD anode will further be shaped into an active-low pulse signal by the first inverter, which has a thick oxide gate. The second inverter is a standard logic cell which changes the polarity of the shaped SPAD signals and feeds them into the following logic control circuits. **Figure 6.4** demonstrates the waveform examples of three different nodes (the SPAD anode and two pulse shaping circuits) in the SPAD frontend when photons are detected.
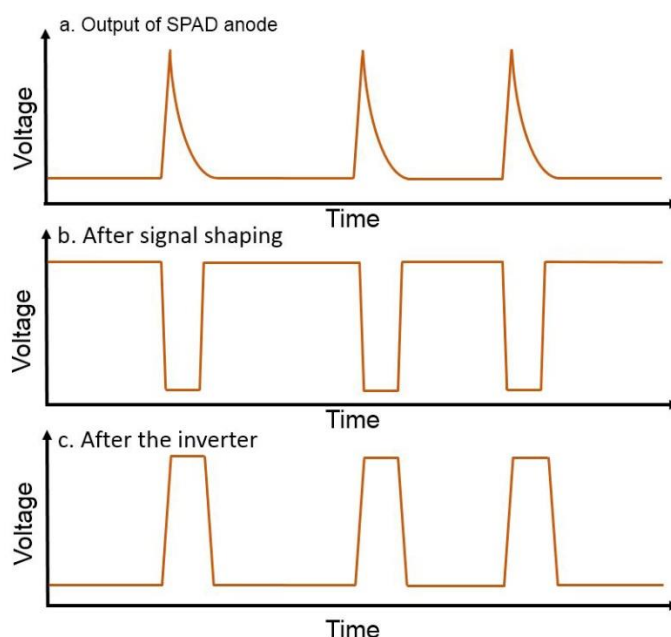
FIGURE 6.4 WAVEFORM EXAMPLES AT DIFFERENT NODES OF THE SPAD FRONTEND NODES

b) Logic control circuits

As shown in **Figure 6.3**, the logic control circuits are designed to provide the enable signals, 'S' and 'SPADWIN', for the RO-TDC and the ripple counter, respectively. Besides the pulse signals from the SPAD frontend, there are five control signals fed into the circuit, including 'TCSPC', 'STOP', 'WINDOW' and 'Rst'.

By setting the signal 'TCSPC' to logic high, the working mode of the sensor is switched to the TC mode, and the signal 'S' is activated for the RO-TDC. The signal 'S' is generated by a compact edge trigger circuit which consists of two D-FFs and three logic gates. The signal 'TCSPC' has a top priority to dominate the signal 'S', since it will be locked to '0' as long as the 'TCPSC' is logic low. When the current measurement is finished, the signal 'Rst' will reset the outputs of three D-FFs to '0'. The pulses from the SPAD frontend will be ignored if the 'WINDOW' is '0' and will pull up the output of the first D-FF (upper left) if the 'WINDOW' is '1'. In this way, the 'WINDOW' is used as a time gating signal to reduce the exposure period. By properly controlling the signal 'WINDOW', background noises which are uncorrelated with the laser excitation can be effectively suppressed in applications such as FLIM, Raman

spectrometer and Lidar. The trigger signal or reference clock of lasers, the signal 'STOP', is connected with the CLK port of the second D-FF (upper right). When the subsequent rising edge of the 'STOP' appears, the signal 'S' will be pulled down. Therefore, the pulse width of the signal 'S' is related to the time intervals between the rising edges of SPAD pulses and the signal 'STOP'. An example of the waveform in the TC mode is shown in **Figure 6.5**.
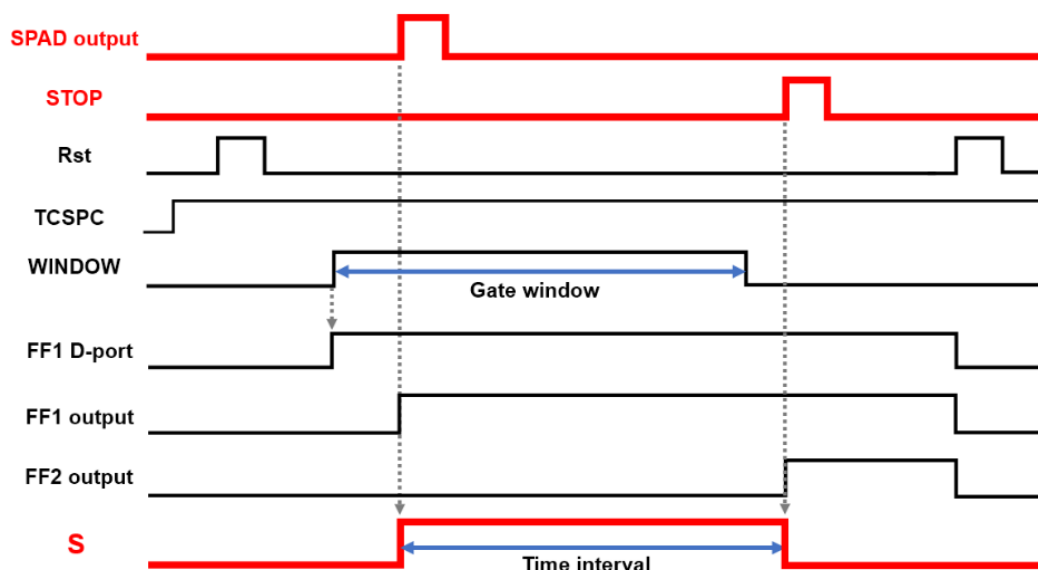
FIGURE 6.5 THE TIMING WAVEFORM OF SIGNAL 'S' IN TC MODE

If the signal 'TCSPC' is at logic-low, the sensor is working in the PC mode. As seen in **Figure 6.2**, the signal 'SPADWIN' will be connected to and drive the ripple counter via a multiplexer in the PC mode. As shown in **Figure 6.3**, a feedback circuit which consists of a D-FF (bottom left) and a multiplexer will generate pulses on the signal 'SPADWIN'. In the period during which the signal 'WINDOM' is at logic-high, the feedback circuit will be enabled as an oscillation circuit and will invert the signal 'SPADWIN' once the SPAD frontend generates pulses. A waveform example of the PC mode is shown in **Figure 6.6**.
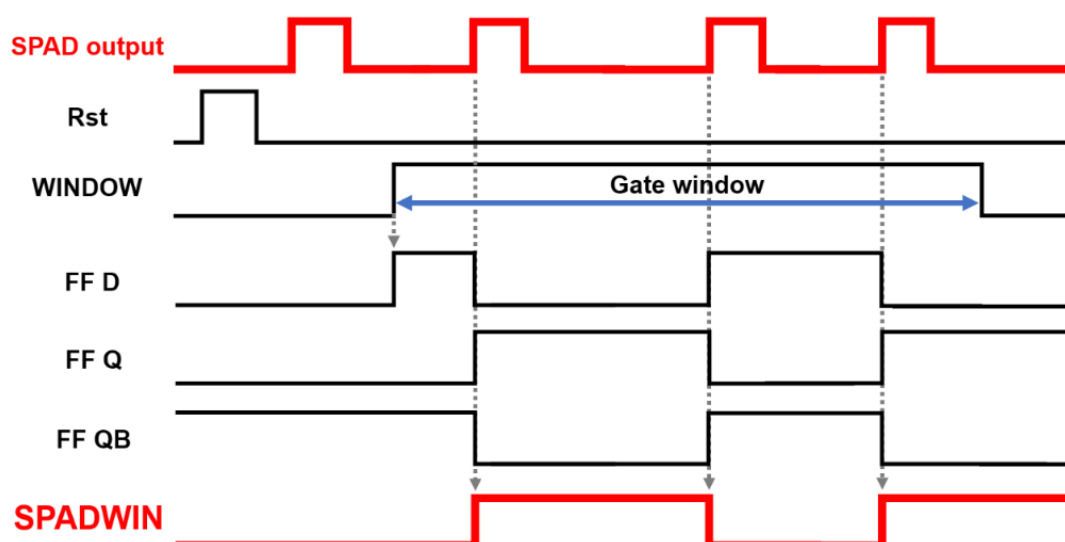
FIGURE 6.6 THE TIMING WAVEFORM OF SIGNAL 'SPADWIN' IN PC MODE

c) Gated Ring Oscillator

The RO architecture is utilised for the large-scale in-pixel TDC array to reduce the resource cost, power and area consumption. The basic structure of the RO is shown in **Figure 6.7**. The signal 'S' and its inverted version, 'S̄', are used to enable and disable the RO. The enabled oscillator is turned to an unstable state since the input and output of inverters have the same logic state, and the precondition of oscillation occurs. The oscillation propagates from a stage to another stage along with the ring structure of the oscillator until the signal 'STOP' is pulled down, and the signal 'S' freeze the signal status in the RO.
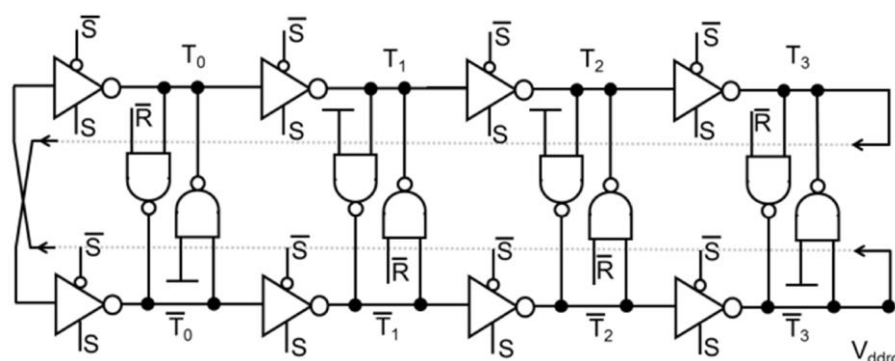
FIGURE 6.7 GATED RING OSCILLATOR [67]

The oscillator is supplied by an exclusive supply rail, Vdd_{RO}, which can be dynamically adjusted to change the frequency of the oscillator and the TDC resolution. The rising edge of the '*S*' starts the ring oscillator, which operates over a range of around 2-4 GHz, depending on the Vdd_{RO}. The intermediate nodes, T_0-T_3 and $\overline{T_0}$ -$\overline{T_3}$, are buffered and converted into a digital format (F_3, $\overline{F_2}$, F_1, $\overline{F_0}$) as the logic status outputs of the RO. By being combined with the ripple counter, the RO-TDC can act as a fine interpolator in the coarse&fine interpolation architecture.

d) Ripple counter

The ripple counter is multiplexed by the signal 'TCSPC' to act as either a photon counter in the PC mode or a coarse counter for the MR extension in the TC mode. In the PC mode, the 8-bit output of the ripple counter is used to form eight high significant bits of the final photon count, C[8:1], since the ripple counter is single edge sensitive. The SPADWIN signal is used as the least significant bit of the final photon count, C[0]. Therefore, the 9-bit photon counter and 512 count depth are achieved in the PC mode. During the TC mode, one of the output bits of the ring oscillator, $\overline{F}3$, is used to drive a high-speed flip-flop. The output of the flip-flop, which represents the cycles of the RO, is counted by the ripple counter as the coarse count of an RO-TDC. The coarse&fine interpolation TDC provides 8x512=4096 time bins.

e) Readout circuit

To read the photon counts in the PC mode and timestamps in the TC mode, each pixel generates a 14-bit data bus per frame, and a bank of tri-state inverters are used to enable and drive data signals, as shown in **Figure 6.2**. To handle the whole sensor array which contains 192x128=24576 SPAD pixels and generates 336Kb data per frame, a readout circuit along with a readout procedure is designed to transmit the data via a limited number of output ports in the sensor chip. The simplified structure diagram of the array is shown in **Figure 6.8**.
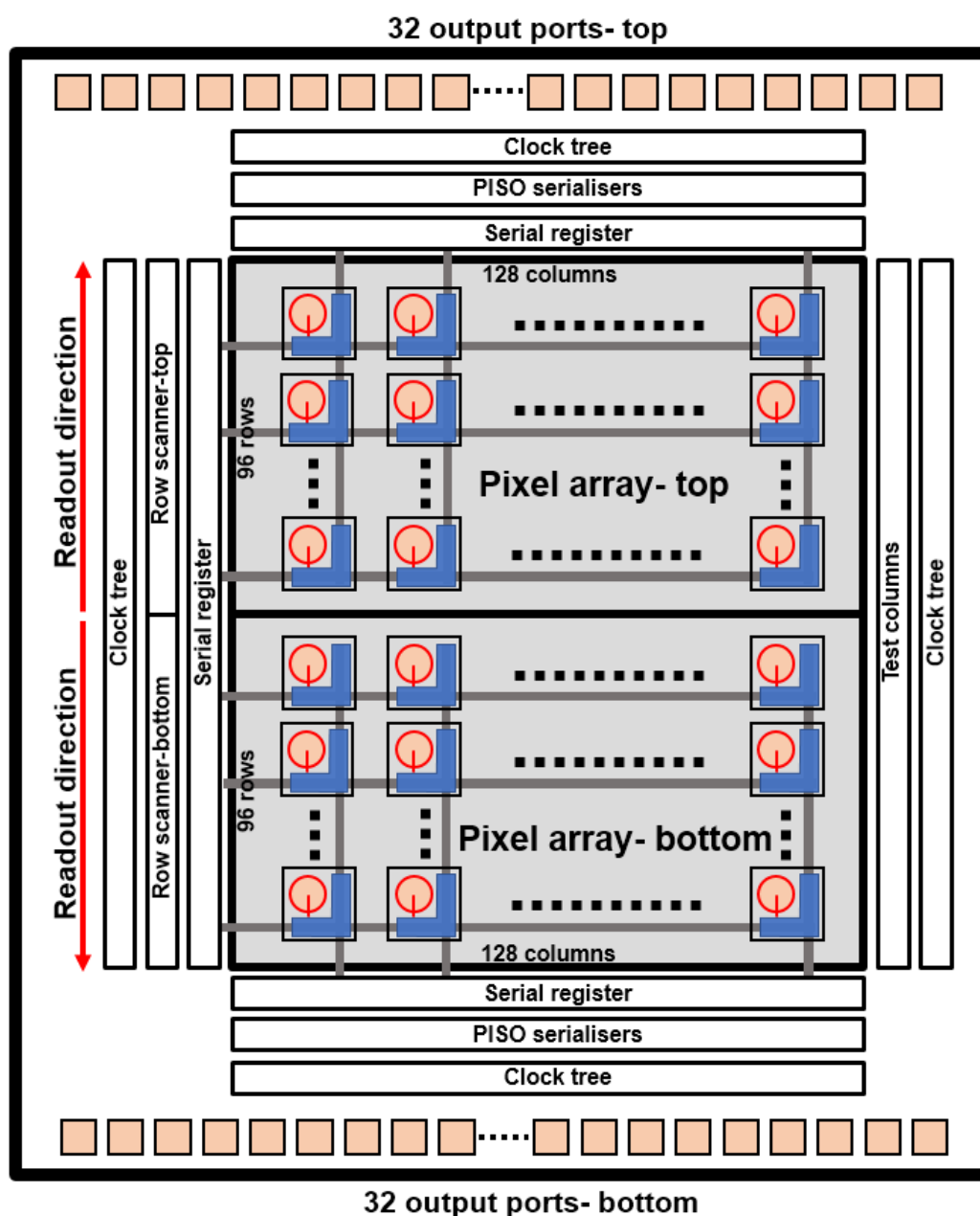
FIGURE 6.8 SIMPLIFIED STRUCTURE DIAGRAM OF THE 192X128 SPAD&TDC ARRAY

The chip package provides 32x2 output ports located symmetrically at the top and bottom side. Each output port is allocated to four half-columns of SPAD pixels and is responsible for transmitting the serial data of the four half-column pixels. Three clock signals (the Frame-clock, Line-clock and Data-clock) and several logic circuit modules, including parallel in serial out (PISO) serialisers, serial registers, and row scanners, collect the data and control the readout

procedure. A readout sequence is initiated when both rising edges of the Line-clock and Frame-clock arrive. The sequence generates a readout token and a reset token in each row of the scanner module. The read tokens are delivered from the central rows of the array to the edge rows at each rising edge of the Line-clock. The read tokens load the pixel data of all the pointed rows onto the 192-column bus. The reset tokens follow the read token to reset the data in the loaded pixel rows. A rising edge on the 'load' pulse loads the column data into the PISO serialisers where it transmits 1 bit serially at a certain time on the rising edge of the Data-clock.

6.3　　Hardware design of the imaging system

The structure of the proposed TCSPC system is shown in **Figure 6.9**. The hardware modules are mainly implemented in a PCB mainboard and an FPGA daughter board.
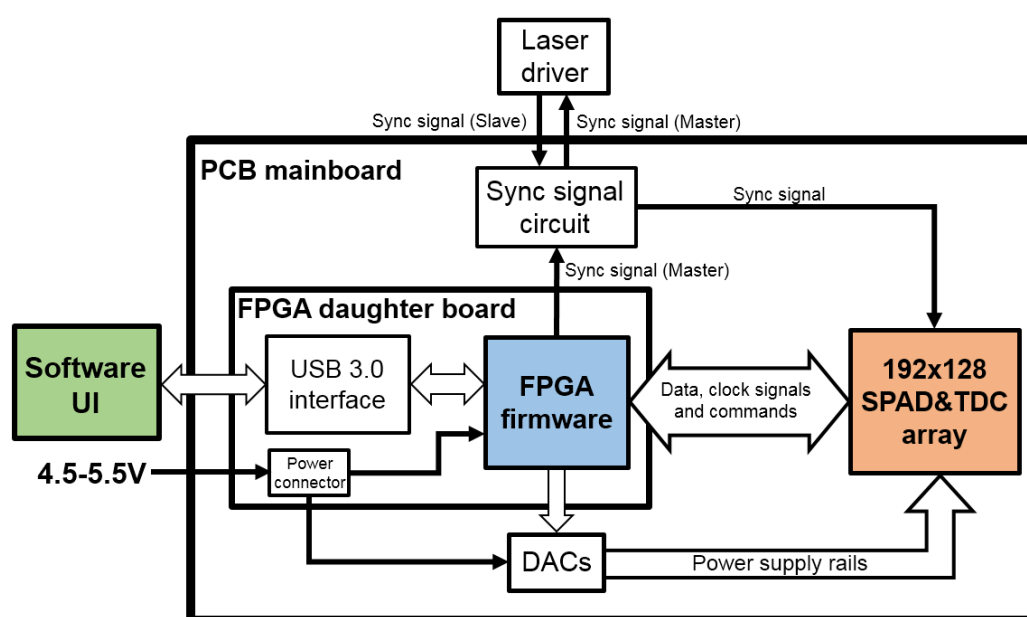
FIGURE 6.9　STRUCTURE DIAGRAM OF THE PROPOSED COMPACT TCSPC SYSTEM

6.3.1   PCB mainboard

The mainboard of this proposed system is a 148x103 mm, 6-layer PCB board. It is designed to power up the sensor, implement the signal connections between the sensor and the FPGA daughterboard, and build a signal transformation circuit for laser synchronisation. A pluggable IC Socket which is compatible with the standard 15x15 PGA (pin grid array) chip package is utilised on the mainboard to mount the sensor chip. Two 80-pin board connectors (BTE-040-01-F-D-A) manufactured by SAMTEC are used to connect the FPGA daughter board with the mainboard. The front side and backside of the mainboard are shown in **Figure 6.10**.
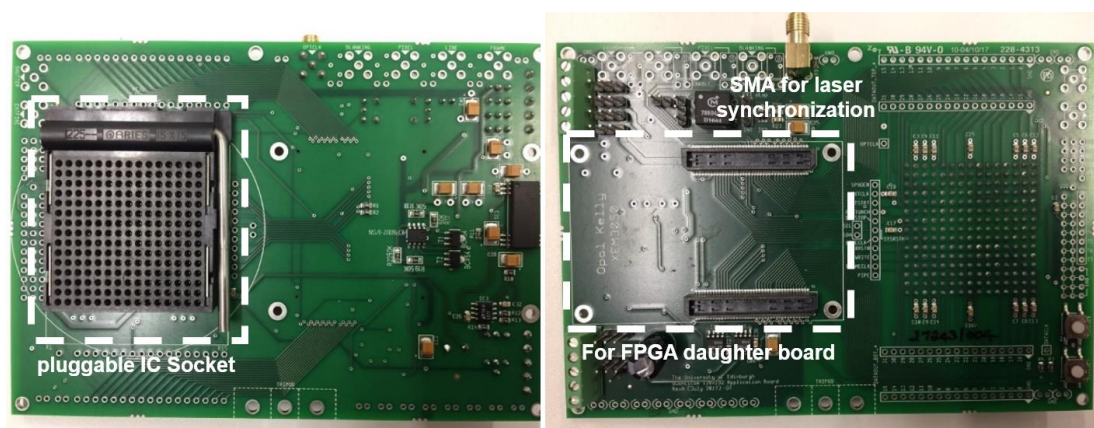
FIGURE 6.10 THE FRONT SIDE (LEFT) AND BACKSIDE OF THE MAINBOARD.

To avoid the SPAD sensor being affected by ambient lights from the LEDs on the FPGA daughterboard, the board connectors for the daughterboard are located at the backside of the mainboard. An SMA connector is mounted at the backside of the mainboard to introduce (in Slave mode) or provide (in Master mode) the trigger or synchronising signal from, or to, laser drivers. A sync signal circuit with a pulse transformer (MURATA POWER SOLUTIONS 78604/2C) [166] is added in the mainboard to make it compatible with different lasers or equipment which have various IO standards. The circuit schematic of the sync signal circuit is shown in **Figure 6.11**. By configuring jumpers manually, the system can be made compatible with TLL (2.5 to 3.3V), LVCMOS (0.6 to 1.2V and 1.2V to 3.3V) and NIM-logic (-0.3 to -0.8V) standards.
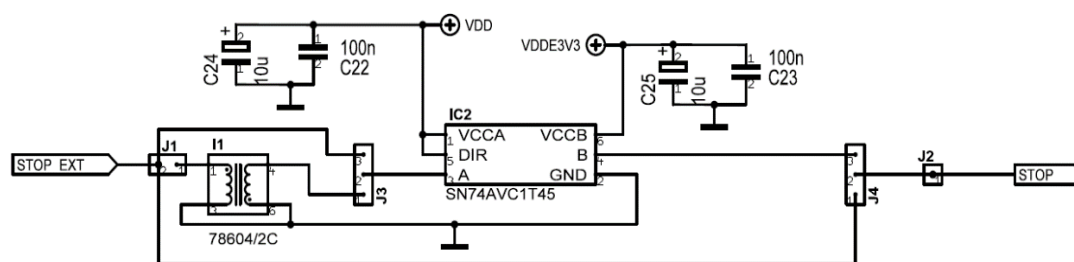
FIGURE 6.11 THE CIRCUIT SCHEMATIC OF THE SYNC SIGNAL CIRCUIT.

Four dedicated power rails are required to supply the sensor work properly. The maximum voltage of the four power rails (VHV) is lower than 16V. In order to control the voltage well, two dual-channel Digital to Analog Converter (DAC) IC chips, an operational amplifier (LT1077, LINEAR TECHNOLOGY) [167] and two low-power operational amplifiers (MCP6002, Microchip) [168] are integrated into the mainboard. A DAC control module is included in the FPGA firmware to configure the output voltages of the DAC chips dynamically.

6.3.2   FPGA daughter board

The FPGA daughterboard is a compact FPGA development board (XEM6310- LX150) [169] manufactured by Opal Kelly, as shown in **Figure 6.12**. The XEM6310 board contains a Xilinx Spartan-6 FPGA, a USB3.0 interface and several driver modules, as well as a 128MB DDR2 memory and a DC power port to supply the entire TCSPC system. Two expansion connectors are located at the backside of the XEM6310 board for connection with the mainboard.
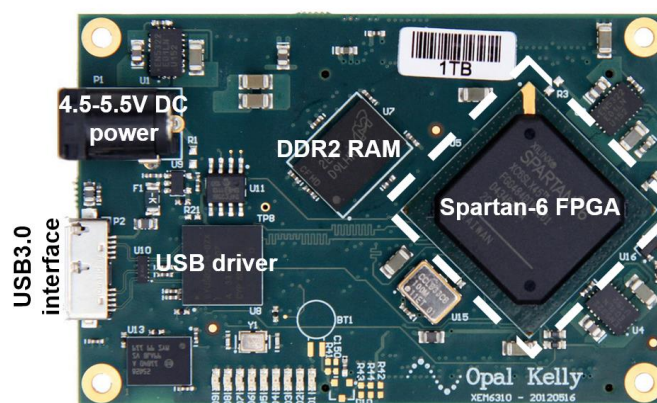
FIGURE 6.12 THE FRONT VIEW OF THE **FPGA** DAUGHTER BOARD.

The block diagram of the XEM6310 board is shown in **Figure 6.13**. The firmware is downloaded into the Spartan-6 FPGA via the USB interface, and it is designed to process and transmit the data and configure the sensor and DACs.
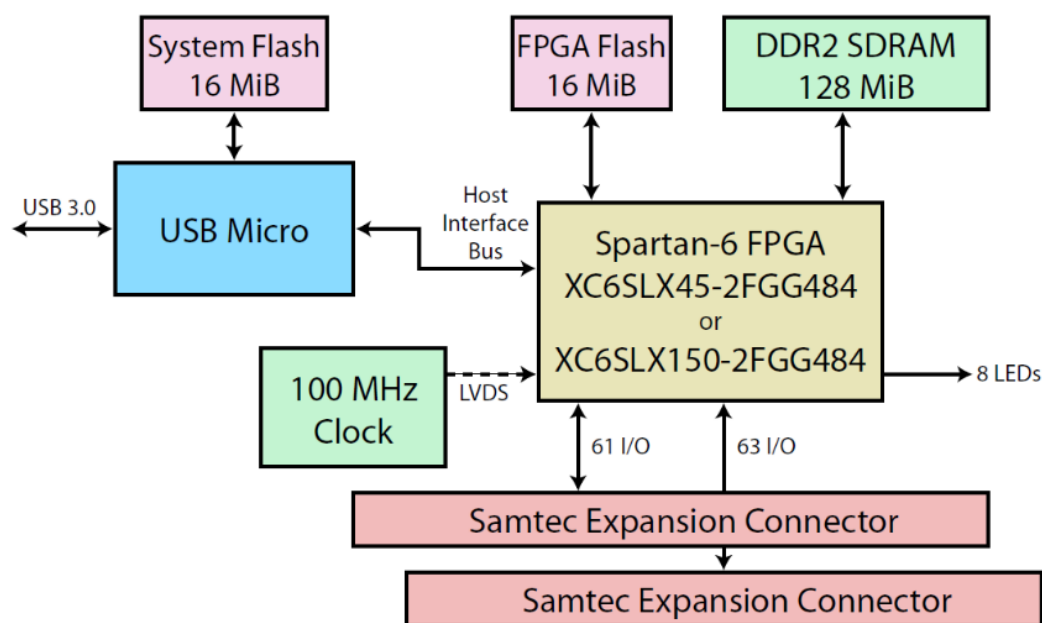
FIGURE 6.13 BLOCK DIAGRAM OF THE XEM6310 FPGA DAUGHTER BOARD [169].

6.4   Firmware Design

The Firmware plays critical roles in the TCSPC system, including:

- configuring and controlling the SPAD sensor and DAC chips
- collecting, encoding and processing the pixel data,
- communicating and transmitting the data with the software via USB 3.0 link,
- generating a synchronisation signal to lasers and the sensor in the master mode.

6.4.1 USB interface and endpoints

The XEM6310 FPGA board is available to use a series of pre-designed soft cores in FPGA firmware to perform various types of bidirectional data transmission with the software via the USB link. These soft cores instance various endpoints of the USB link in the FPGA firmware, which can be directly connected with the internal signals and modules of the firmware. A host interface module, 'OKHostInterface', must be instantiated as a hub and control centre of all used endpoints to connect the onboard USB driver chip.

The endpoints can be classified into four types, including Wire, Trigger, Pipe and BT Pipe. The block diagram of the USB interface and endpoints is shown in **Figure 6.14**. The Wire endpoints, 'OKWireIn/Out', can transfer an individual signal or bus between the software and the firmware asynchronously, such as LED signals or virtual switch signals. The Trigger endpoints, 'OKTriggerIn/Out', provide synchronous connections between the software and the firmware to trigger single events. The Pipe and BT Pipe endpoints, 'OKPipeIn/Out', are designed to transfer the bulk data stream between the software and the firmware.
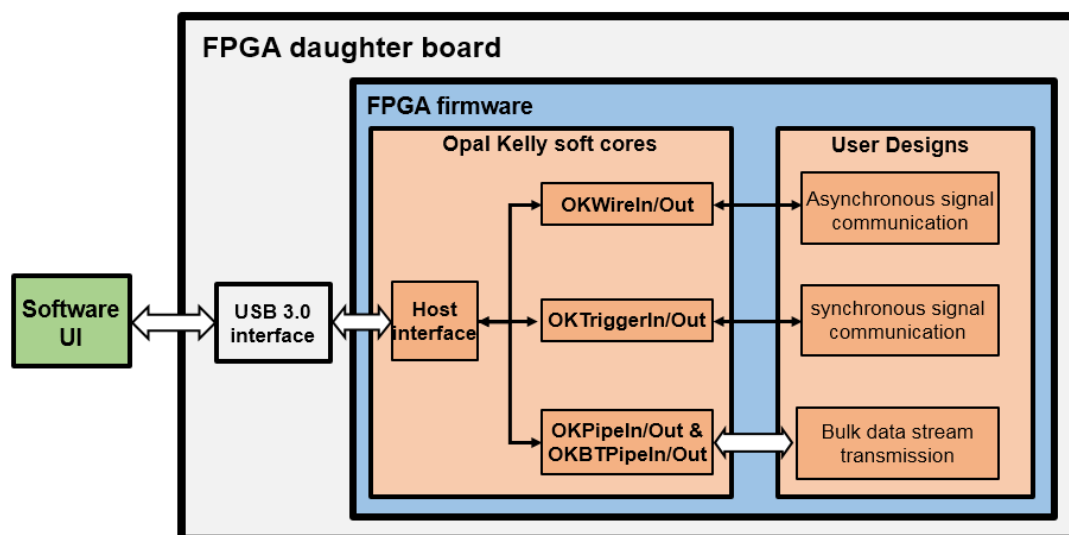
FIGURE 6.14 BLOCK DIAGRAM OF THE USB INTERFACE AND OPAL KELLY ENDPOINTS.

6.4.2   DAC control module for power supply rails

To power up the 192x128 SPAD&TDC array correctly in both the PC and TC modes, four power supply rails need to be configured independently. The 'Vdd' is the power supply of all digital and readout circuits in the sensor. The Vdd_{RO} supplies the power to the RO-TDCs and determines the frequency of the RO. The VHV places the SPAD frontend to work in the Geiger mode and affects the Photon Detection Efficiency (PDE) and the Dark Count Rate (DCR) of the SPAD. The V_{quench} controls the passive quench circuit and affects the dead-time of the SPAD frontend. The whole power supply system for the sensor is shown in **Figure 6.15**.
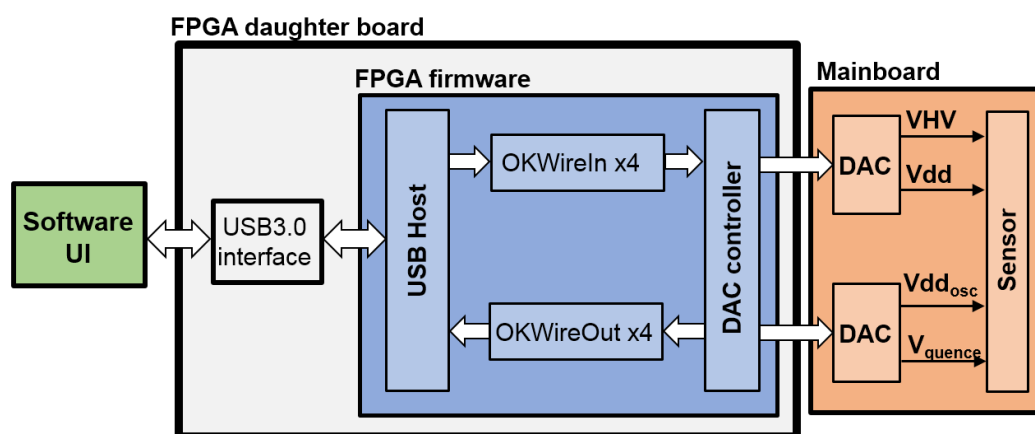
FIGURE 6.15 BLOCK DIAGRAM OF THE POWER SUPPLY SYSTEM DESIGN.

These four power rails are provided by either onboard DACs or external power sources. The rails from the external power sources tend to have better stability and precision for characterisation. However, onboard DACs as power sources will enhance the compactness of the system and reduce its complexity. The DAC control module in the firmware configures DACs by receiving the instructions from the software. The configuration values of the four power rails are transmitted via four 'OKWireIn' endpoints, respectively. The DAC controller encodes these configuration values and transmits them to the two DAC chips, following the Serial Peripheral Interface (SPI) protocol. This design utilises two LTC1446 dual-channel 12-bit DAC IC chips [170] provided by Micropower, and each TDC chip provides two independent outputs with a 12-bit resolution. The reference voltage of the DACs is provided by the FPGA daughter board via one of the expansion connectors. With 4.095V typical Full-Scale Voltage

(V_{FS}), the LSB of the DAC output can be calculated as [170]:

$$\text{LSB} = \frac{(V_{FS} - V_{os})}{(2^{12} - 1)} \approx \frac{4.095V}{4095} = 1mV$$

where V_{os} is the offset error whose typical values are $\pm 2mV$. After the configuration of DACs, the values of four power supply rails will be sent back to the software UI via four OKWireIn endpoints for verification.

6.4.3 Command serial interface module

In order to configure the operation mode and pixels of the sensor, serial commands need to be transmitted to the sensor via the firmware and dedicated command links. As **Figure 6.16** shows, the command serial interface module is designed to undertake receptions, decoding, encoding and transmission of the serial commands.
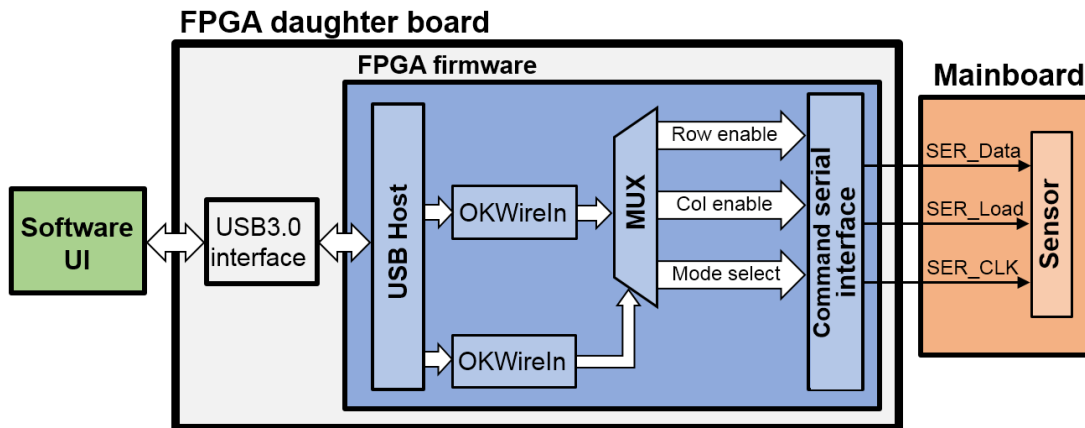
FIGURE 6.16 BLOCK DIAGRAM OF SERIAL COMMAND TRANSMISSION.

The serial command contains 323 bits, including 192 bits for row enable, 128 bits for column enable and 3 bits for work-mode configuration. To reduce the number of used endpoints, two endpoints are working in the multiplexing approach to receive the 323 bits command data step by step. The sensor is configured by three wire links (SER_Clk, SER_Load and SER_Data). The command serial interface module reshapes the 323-bit command data and controls the three wire signals to load the commands into a dedicated command register in the sensor.

6.4.4   Clock signals generation and data pipeline module

Three clock signals (Data-clock, Line-clock and Frame-clock) and a signal, 'LOAD', need to be generated by the firmware to drive the readout circuit and data transmit procedure in the sensor. The clock source is a 100MHz, low-jitter oscillator with LVDS outputs on the FPGA daughter board. Following this, several CMT primitives are instantiated to adjust the frequency and phase of the three clock signals. The signal 'Data-clock' drives each bit of the pixel data from the PISO serialisers into the output ports of the sensor. When all bits of the current pixel rows are transmitted, the signal 'Line-clock' drives the Row scanners to point at the next rows. After the data of the entire array has been transmitted, the signal 'Frame-clock' resets the row scanner and serial registers for the next frame imaging. Since each output port of the sensor is responsible for four half-column pixels and 14-bit data width, one Line-clock period contains 56 Data-clock cycles. Since the top and bottom parts of the array are read separately, 96 Line-clock cycles are required to read the complete frames, and 2 additional clocks are needed to generate the read and reset tokens.

The architecture of the Data-pipeline is shown in **Figure 6.17**. Two deserialiser groups (Deserialiser_TOP/BOTTOM) are built in the data-pipeline module, and each deserialiser group contains 32 deserialisers. Each deserialiser is responsible for the sampling of one sensor port and generates a 64-bit data bus with an 8-bit address identifier per Line-clock cycle. Therefore, 4096-bit data will be generated by 2 deserialiser groups per Line-clock cycle. For the following data buffering procedure, a multiplexer is used to segment the 4096-bit data into four 1024-bit data buses.
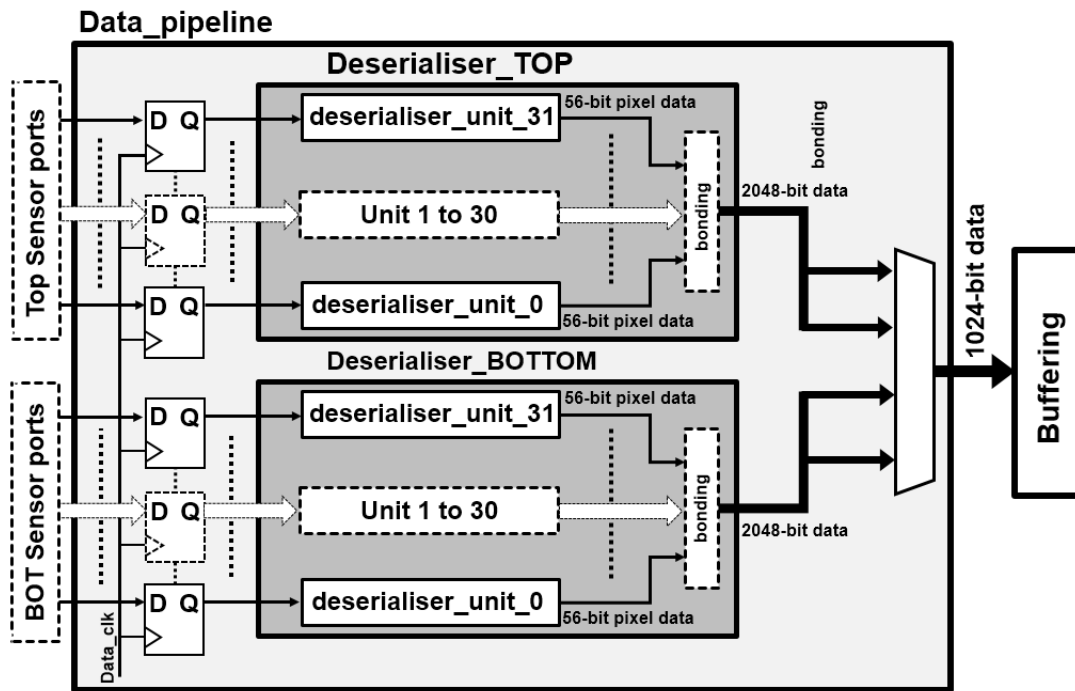
FIGURE 6.17 BLOCK DIAGRAM OF THE DATA PIPELINE MODULE

6.4.5 Data buffing and transmit module

The pixel data is transmitted to the software via a 32-bit endpoint, namely 'OKBTPipeOut'. Besides this, the endpoint is driven by a dedicated USB clock (100.8MHz) [169] which is uncorrelated with the system clock and the Data_clock. Therefore, a data buffing and transmission module is utilised to disassemble the 1024-bit data bus into multiple 32-bit data buses and transmit data across different clock domains to avoid metastability problems. In this work, FIFOs which are implemented by BRAMs [171] are used for the data buffering through three clock domains. Since FIFO has a limited ratio between the input width and output width, two FIFOs are utilised as a two-stage buffering. The first FIFO reduces the data width from 1024-bit to 128-bit, and the second FIFO reduces the width further to 32-bit. A handshake controller is used to coordinate the read/write operation between the two FIFOs. The 32-bit data bus outputted from the second FIFO can be directly fetched by the OKBTPipeOut and transmitted to the software via the USB link. The block diagram of the data buffing and transmit module is shown in **Figure 6.18.**

FPGA firmware
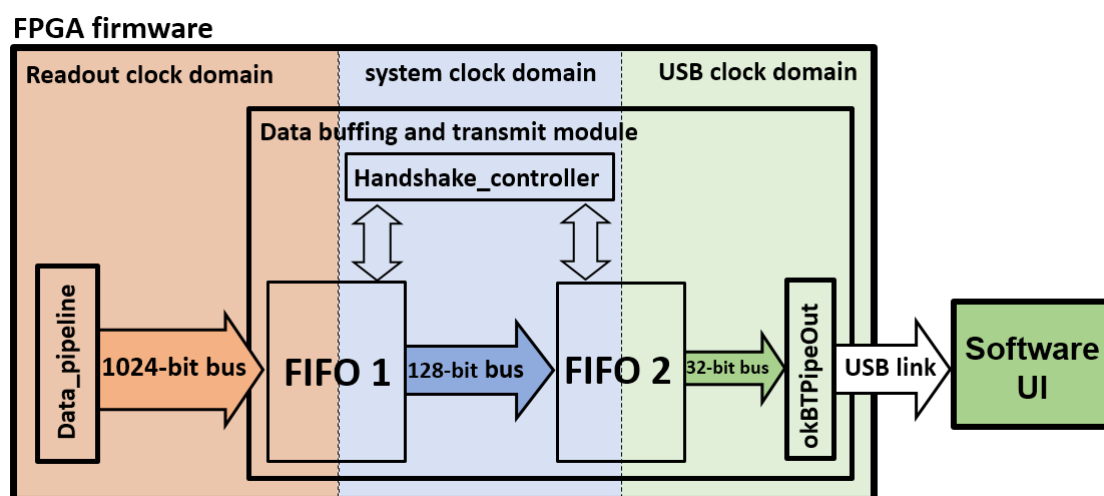
FIGURE 6.18 BLOCK DIAGRAM OF THE DATA BUFFERING AND TRANSMISSION.

6.4.6 Firmware histogramming

The histogramming module is designed to reduce data processing time and data throughput in the TC mode. The pixel data represents the timestamps of detected photons in the TC mode. Depending on the SNR level, hundreds to thousands of timestamps are required to form histograms in time-correlated applications. For applications with low light intensity, plenty of time is costed for the transmission and histogramming of the timestamps from the entire sensor array.

The histogramming can be performed in the FPGA chip instead of the software, whereby only histogram counts are required to be transmitted via the USB link to accelerate the histogramming process. Besides, these histogram counts can be reserved in the FPGA instead of overwriting the previous frame data when they are sampling the current frames. In this way, hardware-friendly algorithms can be directly implemented in the FPGA chip based on stored histogram counts.

Enough memory space is required to build the histograms of the entire array in FPGA. If the 16-bit histogram bins are used, the entire sensor array will cost $16 \times 2^{12} \times 192 \times 128 =$

$192\ MByte$. Considering that the maximum memory space of BRAMs in the FPGA is 603Kbyte [172], an onboard DDR2 RAM chip (Micron MT47H64M16HR, 128-MByte) [173] is instead used to store the histograms of the entire array. By reducing the TDC MR from 212 to 211, the requested memory space is halved to 96Mbyte, which can be stored in the DDR2 RAM.

A histogramming controller is designed to handle the DDR2 SDRAM for histogramming and data transmission. This design utilises a dedicated hardcore module, namely a Memory Controller Block (MCB), to manage the DDR2 SDRAM as a housekeeper, since driving the DDR2 RAM directly is extremely complicated. The MCB greatly simplifies the interaction procedure between the FPGA and the DDR2 RAM, and an MIG tool [174] further enhances the usability of the MCB, which can be instantiated and configured in a relatively easy way. The block diagram of the firmware histogramming is shown in **Figure 6.19**.
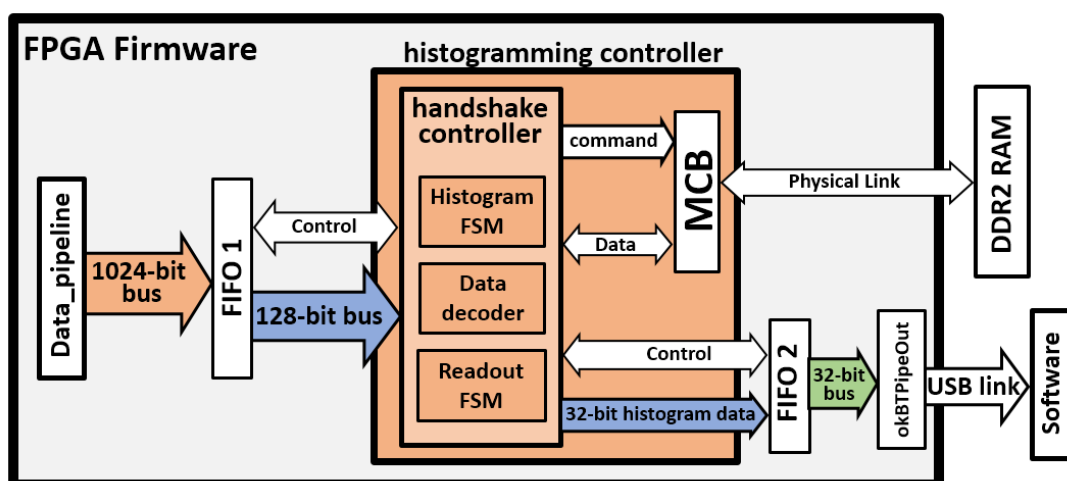
FIGURE 6.19 BLOCK DIAGRAM OF THE FIRMWARE HISTOGRAMMING MODULE

As shown in **Figure 6.19**, a handshake controller is designed to perform the histogramming procedure. A code decoder in the handshake controller decodes the 128-bit data from the first buffing FIFO. The histogramming Finite State Machine (FSM) issue commands the MCB to fetch counts addressed by the decoded timestamps. The histogramming FSM increases the fetched counts and writes the increased count to the same addresses in the DDR2 RAM. The

circulation of this process will be ceased once the measurement has been completed, and the readout FSM is responsible for transmitting the histogram from DDR2 SDRAM to the OKBTPipeOut endpoint via the second buffering FIFO.

6.5   Software Design

The software is implemented by a MATLAB Graphical User Interface (GUI) in order to complete the following tasks: download the firmware, configure and control the sensor array, receive and process the pixel data, and imaging. The Application Programmer's Interfaces (APIs) and Dynamic Link Library (DLL) files are provided by the Opal Kelly for different platforms, programming languages and third-party software. By including the DLLs and calling the APIs in the project, the MATLAB GUI can configure and communicate with the firmware via the USB link. The main UI is shown in **Figure 6.20**.
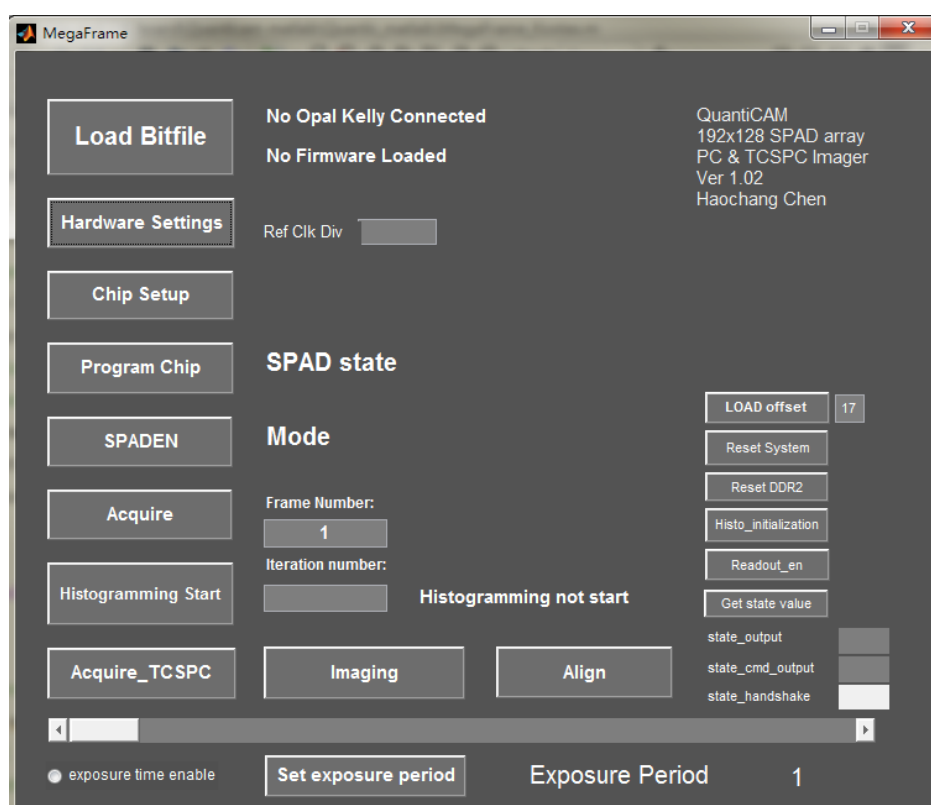
FIGURE 6.20 MAIN UI OF THE SOFTWARE.

A popup window, as shown in **Figure 6.21 (left)**, can be called by clicking the 'Hardware Setting' button to set the power supply voltages of the sensor. The four sliders modify the voltages of four power supply rails independently, and the voltage settings are transmitted to the firmware via four 'OKWireIn' endpoints. The 'QuanticSetup' is designed to switch the work mode and enable a certain row and column of pixels. The popup window 'QuanticSetup' is called by clicking the 'Chip Setup' button, which is shown in **Figure 6.21 (right).** By clicking the 'Acquire' button, a certain number of the frames or the histogram data are received from the firmware via the 'OKBTPipeOut' endpoint and saved in a local location for processing and further imaging.
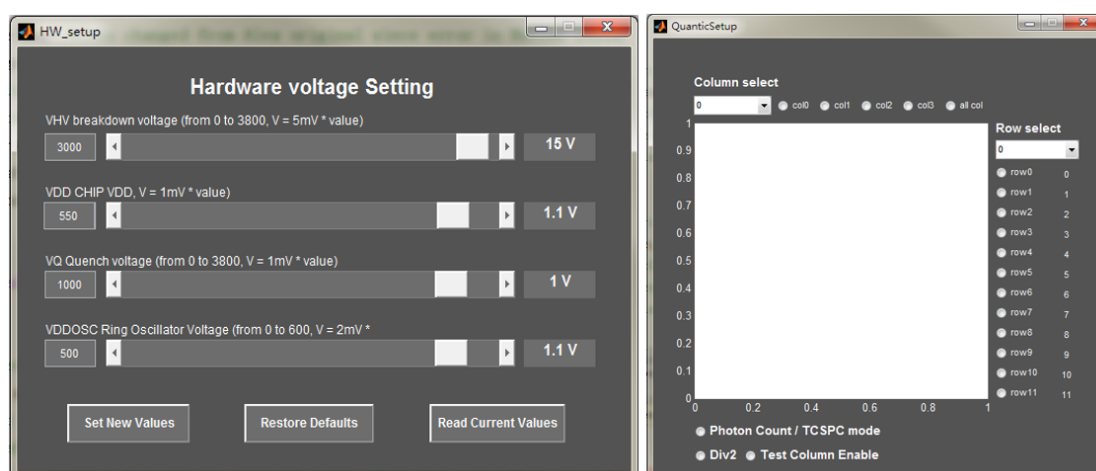
FIGURE 6.21 (LEFT)HARDWARE VOLTAGE SETTING UI AND (RIGHT)SENSOR ARRAY SETUP UI.

A pixel data decoding and imaging function is called by the 'Imaging' button in the main UI. **Figure 6.22** demonstrates single-frame images which are decoded from the received data in the PC and TC mode, respectively. In the PC mode, the sensor can be used as a camera for grey images. **Figure 6.22 (right)** shows the timestamps of the entire array under ambient lights. Since the ambient lights are uncorrelated with the synchronisation signal, the random timestamps are acquired in the TC mode.
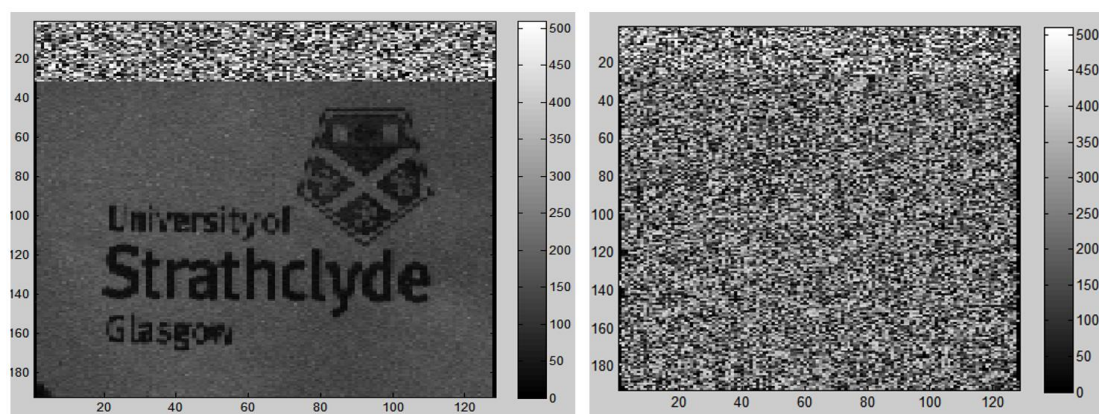
FIGURE 6.22 (LEFT) A GREY IMAGE ACQUIRED IN THE PC MODE, (RIGHT) A TIMESTAMP MAP OF THE ENTIRE ARRAY IN THE TC MODE.

6.6 Characterisation and basic test

The characterisation mainly focuses on the temporal performance in the TC mode. The following tests, including the code density test, IRF measurement and time interval test, are used for evaluations of TDC resolution and the TDC output offset.

6.6.1 Code density test

The DNL and INL of the RO-TDCs were acquired from the code density test, which was performed by exposing the sensor to ambient light. The FPGA presents the synchronisation signal 'STOP' to the RO-TDC array, and more than 300K frames of the timestamp were collected in the TC mode. The DNL/INL plot of a typical TDC pixel is shown in **Figure 6.23**. The DNL range is [-0.40, 0.49]LSB, with 0.14LSB of standard deviation. The INL value is [-1.677, 3.961]LSB, with 1.14LSB of standard deviation. As seen in the figure, the tested TDC achieved satisfactory DNL values ($<\pm 0.5$ LSB). However, the long-term deviation of the INL results needs to be corrected for long-range measurement. This is one typical drawback of RO-TDC for long-range measurements, since the error and jitter of the ring oscillators continue to accumulate.
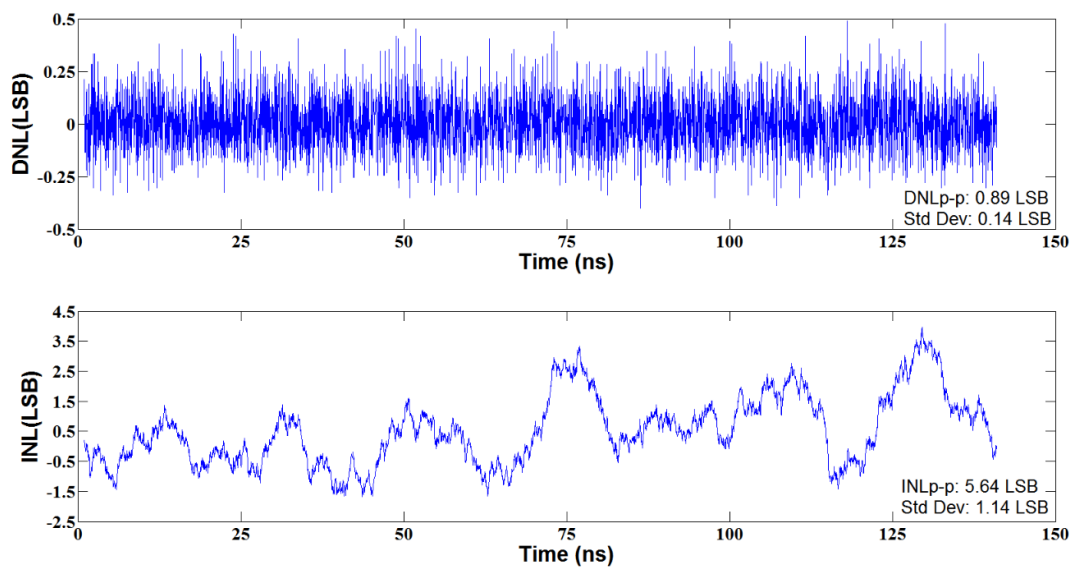
FIGURE 6.23 DNL AND INL PLOT OF A TYPICAL RO-TDC.

6.6.2 IRF measurement

To measure the IRF of the proposed TCSPC system, a pulse laser with 70ps FWHM of pulse width (Hamamatsu PLP-10 685nm) and a laser controller (Hamamatsu, C10196) are utilised. **Figure 6.24** demonstrates the measured IRF plot of one of the typical TCSPC channels with 1.5V excess bias. The all SPAD and TCSPC array are measured, and the FWHM map of IRF is shown in **Figure 6.25**. After removing the hot and bad pixels, the average FWHM is 219ps, with a standard deviation of 26.7ps. According to the previous report, the original jitter of the SPAD is around 170ps [175]. Therefore, the jitter of the STOP signal, the RO-TDC and the laser system contribute around 138ps of jitter.
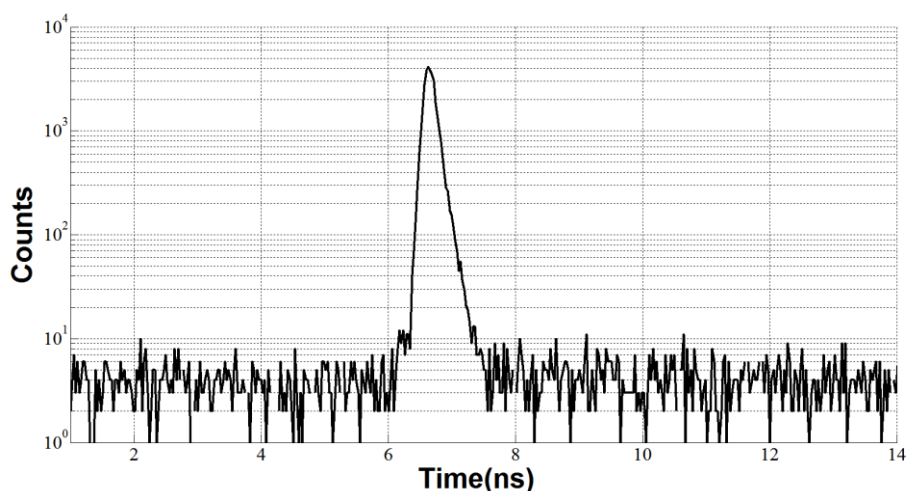
FIGURE 6.24 TYPICAL IRF OF A SINGLE TCSPC CHANNEL
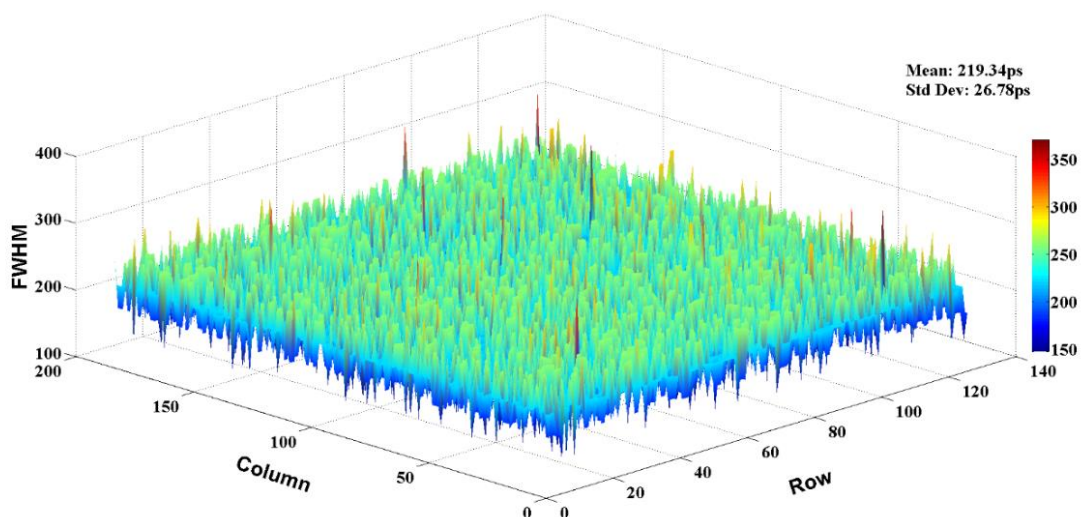
FIGURE 6.25   IRF FWHM MAP OF THE ENTIRE ARRAY

6.6.3   Time interval test

The RO-TDC resolution vs VDD_{RO} and the offset map of the entire array can be obtained from the time interval test. All RO-TDCs are sampled by the same synchronisation signal 'STOP'. By modifying the delay between the signal 'STOP' and the laser pulse, the temporal resolutions of RO-TDCs can be calculated using various Vdd_{RO} values. The measured results of a typical pixel are shown in **Figure 6.26**. From the figure, it can be seen that the resolution of TDC is changed from 120ps to 33ps when the Vdd_{RO} is increased from 0.7V to 1.2V. The full MR of a

TDCs is adjustable from around 135ns to 491ns.
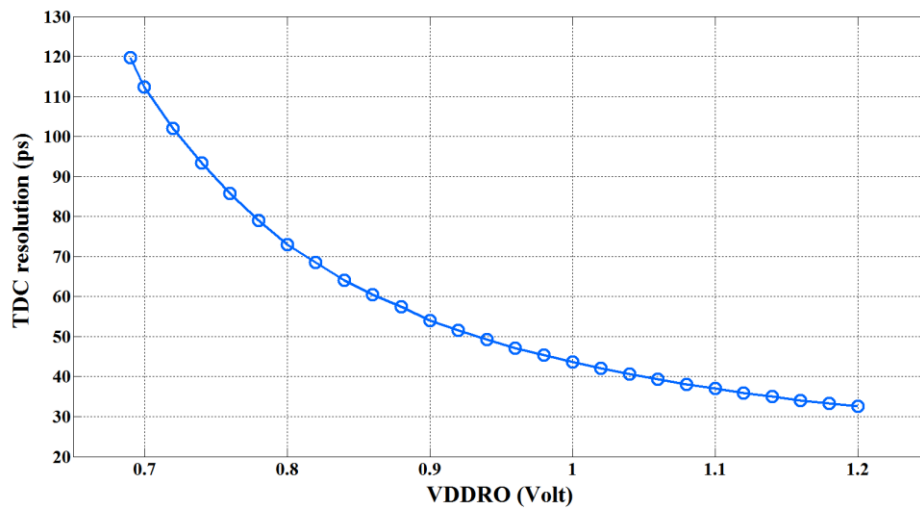
FIGURE 6.26   TEMPORAL RESOLUTION OF AN **RO-TDC** VS **VDD$_{RO}$.**

Figure 6.27 demonstrates the mean output codes of the entire RO-TDCs array in the measurements of a fixed time interval. From the results, it can be seen that there is an incremental drift and offset in the pixel rows. This is mainly because the signal 'STOP' is fed into the sensor chip at the left side and propagated to the right side horizontally. The TDC offset is static and therefore can be corrected by remapping it based on the TDC code map.
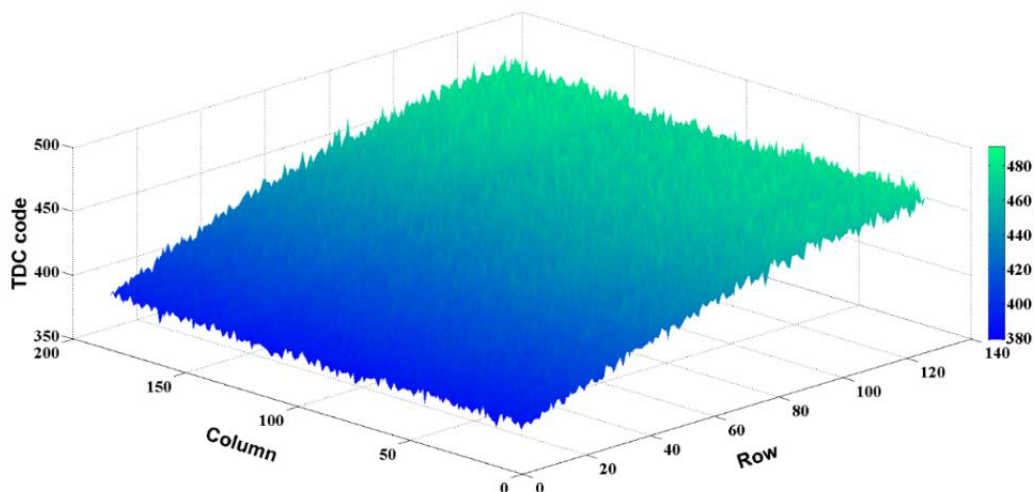
FIGURE 6.27   TDC CODE OFFSET MAP OF THE ENTIRE ARRAY.

This design also evaluates the improvement from the Cylindrical microlenses which cover the SPAD sensors. The uniformed distributed lights focus on the active areas of a pair of SPAD sensors and avoid the inactive areas, such as where RO-TDCs and readout circuits are located. By comparing the photon counts of the ordinary and microlenses versions, the equivalent fill factor increases by 28%, from 13% to 41%. All of the performance details are summarised in **Table 6.1.**

Table 6.1 the performance of the proposed TCSPC system and a previous work

parameter	This work	[18]	[19]
Process	40nm	150nm	180nm
array size	192x128	32x32	340x96
pixel pitch	18.4μm x 9.2 μm	44.64μm	25μm
Fill factor	13%(42%Microlens)	19.8%	70%
Median DCR	25Hz	600Hz	6Hz
TDC area	84.6um^2	402.7 um^2	31,000 um^2
TDC Resolution	33ps to 120ps	204.5ps	208ps
TDC MR	135ns to 491ns	53ns	852ns
Bit width of TDC	12-bit	8-bit	12-bit
DNLpk-pk (LSB)	0.9LSB	1.5LSB	0.52LSB
INLpk-pk(LSB)	5.64LSB	2.17	1.22LSB
IRF FWHM	219ps	-	-

6.6.4 Fluorescence lifetime measurement

A fluorescence lifetime measurement was also performed in this study to verify the TCSPC system. Fluorescein was measured as the standard fluorophore, since it has high quantum yields, photostability and monoexponentially lifetime decay with around a 4.1 ± 0.1ns lifetime in water solution [176, 177]. Considering that the absorption spectral range of fluorescein is between 465nm and 490nm approximately [178], a 470nm wavelength NanoLED laser with <200ps pulse width (HORIBA N-473L) was used as the excitation light source. Since the emission spectral range is between 490nm and 520nm approximately [178], a 505nm long-pass filter was used to block background noises. The VDD_{RO} is set at 1.08V, and the average bin-width of RO-TDCs is set at 37.9ps.

All of the available pixels in the sensor were counted as a single pixel in this measurement to reduce the acquisition time and increase the data processing speed. Before the experiment, the dark pixels caused by manufacturing defects and the hot pixels which have abnormally-high sensitivity were excluded from the data. The TDC code offsets were corrected to avoid the distortion of decay and lifetime estimation. The mean TDC code map of the entire array, which is shown in **Figure 6.27**, was used as the reference of the timing offset correction. The IRF test results of the integrated pixel with and without timing offset correction are displayed in **Figure 6.28**. From the results, it can be seen that the IRF FWHM of the integrated pixel is reduced from 76 bins (3017.2ps) to 6 bins (238.2ps) after the timing correction.
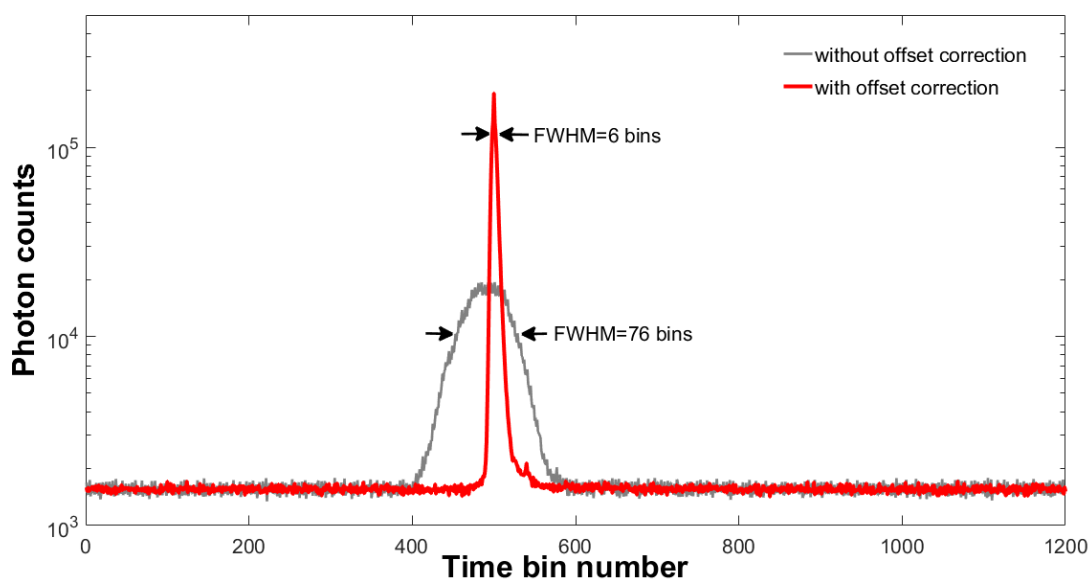
FIGURE 6.28 IRF MEASUREMENT RESULTS OF THE SUMMED PIXELS WITH AND WITHOUT OFFSET CORRECTION.

A hardware-friendly fast algorithm, the Centre-of-Mass Method (CMM), was used to estimate the fluorescence lifetime, τ. The CMM estimates the τ of single exponential decays, $\int(t) = A * exp\left(-t/\tau\right)$, by calculating the Centre-of-Mass (CM) and its deviation from actual τ as [31, 179]:

$$CM = \frac{\int_o^T tf(t)dt}{\int_0^T f(t)dt} = \tau - \frac{Te^{-T/\tau}}{1-e^{-T/\tau}} \cong \tau_{CMM} \tag{6.1}$$

where T is the width of the measurement window as $0 \le t \le T$. If the T>7τ, the deviation between CM and τ can be neglected. When the T<7τ, the deviation part in **Eq(6.1)** needs to be calculated, and the **Eq(6.1)** can be rewritten as:

$$\frac{\tau}{T} = \frac{\tau_{CMM}}{T} + \frac{e^{-T/\tau}}{1-e^{-T/\tau}} \tag{6.2}$$

For a TCSPC system, the CM presents the averaging timestamps of all captured photons in the discrete time domain [31]. The CM of a single exponential decay in a TCPSC system can be calculated as:

$$\tau_{CMM} = \left(\frac{\sum_{j=0}^{M-1} jN_j}{N_c} + \frac{1}{2} \right) \times h \tag{6.3}$$

where M and N_c are the total number of the bins and captured photons in the measurement window respectively, and the h is the average bin-width of the TCSPC. By using the recursive approximation method described in [31], a LUT can be created to map the τ_{CMM} to real τ directly. As a result, for a TCSPC system, the τ can be calculated as [31]:

$$\tau = \Omega\left(\frac{\tau_{CMM}}{T}\right) \times T = \Omega\left[\frac{1}{M}\left(\frac{\sum_{j=0}^{M-1} jN_j}{N_c} + \frac{1}{2} \right) \right] \times Mh$$

$$= \Omega\left[\left(\frac{\sum_{i=1}^{N_c} D_i}{N_c} + \frac{1}{2} \right) \right] \times h \tag{6.4}$$

where Ω is the LUT, and the D_i is the output of the TDC in the TCSPC system. By using an accumulator for $\sum_{i=1}^{N_c} D_i$, a counter for N_c and a memory for the LUT, Ω. This calculation can be easily implemented in FPGAs or other processors.

After removing the hot and dark pixels, **Figure 6.29** presents the fluorescence decay curve of the fluorescein measured by the proposed TCSPC system. The CMM calculated τ of the fluorescein in water solution is 3.93ns with 10K frames and 3.99ns with 400K frames. The standard deviations of τ of all valid pixels are 0.74ps and 0.34ps with 10k and 400K frames. The lifetime measurement result verifies that the proposed integrated TCSPC system can provide accurate lifetime values with a hardware-friendly algorithm.
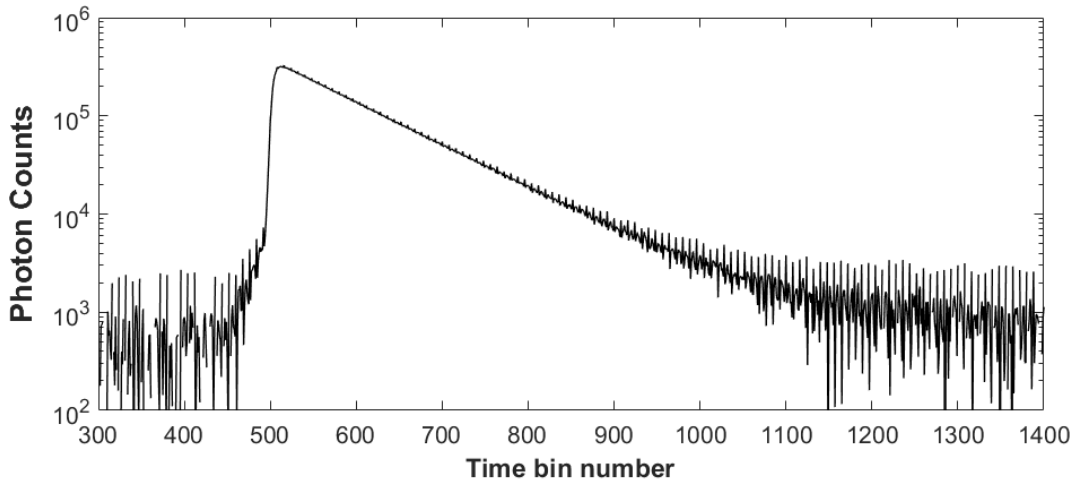
FIGURE 6.29. MEASURED DECAY CURVE OF FLUORESCEIN AFTER BAD&HOT PIXELS REMOVAL AND OFFSET CORRECTION.

6.7   Summary

This chapter presents the features, architecture, performance and application examples of the advanced large-scale TCSPC array based on the 192x128 SPAD and TDC array. With the advanced 40nm CMOS technique, a high-performance SPAD and a RO-TDC can be integrated into the same pixel to achieve the independent and parallel photon detection and temporal measurement.

Chapter 6: An integrated 40nm CMOS 192 x 128 SPAD-TDC array sensor for time-correlated wide-field imaging

By combining the dedicated hardware and FPGA firmware, the compact wide-field imager with 24.5k channels of TCSPCs has been implemented in the system. From the characterisation results, it can be seen that there is great potential for further developing of the hardware-friendly algorithms in the firmware for video-rate FLIM or ToF imaging. Compared with the conventional scanning imaging system, which is only based on a single channel detector and a TCSPC channel, the new system proposed in this chapter has more useful and advanced features which can be further developed for fast time-correlated applications. Furthermore, by using the Cylindrical microlenses, one of the main shortcomings of the large-scale SPAD array, fill factor, has also been greatly improved from 13% to 42%.

Chapter 7: Conclusions

7.1 Summary

This thesis successfully presented the design, implementation and evaluation of two new high-linearity FPGA-TDC designs and an integrated large-scale TCSPC array system. The statistical test results show that the two proposed FPGA-TDCs achieved significant improvements in linearity by proposing and utilising various novel architectures and methods. These proposed architectures and methods restrain the biggest drawback of the FPGA-based TDC/TCSPC designs, namely poor linearity, and promote the application values of FPGA-TDC/TCSPC in time-resolved applications. These improvements are based on the critical review of existing designs and methods and the analysis of the source and principle of non-linearity in FPGA-TDC. After characterising the proposed TDCs implemented in different mainstream FPGA devices, it was established that these architectures and methods successfully met the project aims, thus enhancing the level of linearity of FPGA-TDCs and meaning that the former can compete with ASIC-TDCs. Besides, an integrated large-scale TCSPC array based on a world-leading SPAD&TDC array and FPGA-based firmware was designed for fast time-correlated measurements and wide-field imaging applications.

As the core component of the TCSPC system, the analogue and digital TDC designs are reviewed in Chapter 2. In this chapter, I compared different methods and architectures of digital TDCs and discussed the features of two implementation platforms, namely ASIC and FPGA. It can be seen from the previous literature that although analogue TDCs still have advantages in resolution compared with current digital TDCs, these disadvantages, such as bulky size, high cost, delicate and complicated systems, all limit their applications. Given the design and manufacturing features of ASIC-TDCs, they are more suitable for general-purpose, mass-produced devices. However, FPGA-TDCs tend to be more suitable for scientific experiments, prototyping and high-end instruments. Through the development of semiconductor technology,

every aspect of the performance of FPGA-TDCs has been further improved. Moreover, FPGA-TDCs have irreplaceable advantages such as high integration, flexibility and low cost. However, the poor linearity of FPGA-TDCs is the main performance bottleneck, which has a conspicuous impact on its measurement accuracy and must be addressed.

Chapter 3 focuses on the non-linearity of FPGA-TDC and current solutions. Published methods can effectively monitor and correct dynamic non-linearity, which is the main cause of the linear drift of TDCs' temporal resolution. The static non-linearity of FPGA-TDC causes short-term jitter and accumulated offset, and for TDL-TDC architectures, the propagation nonuniformity of a TDL and the clock skews are the primary sources of the static non-linearity. The two non-linearity sources will cause bubble problems, missing-codes and ultra-width bins, which will aggravate the non-linearity performance of FPGA-TDCs. Especially, the bubble problems caused by the non-monotonicity of the TDL intensify other non-linearity problems and lead to encoding failures. The solutions from previous studies only offer limited improvements in linearity performances, which also cost additional resources and processing cycles.

A 10.5ps bin-width, low non-linearity, and direct histogram FPGA-TDC for the TCSPC system are presented in Chapter 4. The FPGA-TDC integrates the tuned-TDL into the direct-histogram architecture to create a novel combinational architecture for the first time, so as to improve the linearity performance with the features of the missing-codes free. In this new innovative design, the dual sampling phase method is modified to avoid the large clock skews. According to the code density test results, the combination architecture demonstrates a significant improvement in the linearity performance, as the values of $DNL_{peak-peak}$ and $INL_{peak-peak}$ have been reduced to 1.25LSB and 2.25LSB respectively, and the missing-codes have been completely removed. By applying the proposed hardware-friendly bin-width calibration, this study further enhances the linearity to $DNL_{peak-peak}$=0.08LSB and $INL_{peak-peak}$=0.13LSB. These linearity performances are even better than most of the ASIC-TDCs and commercial products. Besides this, reaping benefits from the direct-histogramming architecture, the theoretical maximum sampling rate can achieve 95.2G samples/s. However, this design also has a limitation; for example, it may not be suitable for long-range measurement and multiple channel applications because the direct-histogramming architecture has high resources consumption.

For multiple channel applications and long-range measurement, a high-linearity FPGA-TDC design with much lower logic resource consumption is presented in Chapter 5. Two 96-channel TDC/TCSPC arrays are implemented in two different FPGA series. This design invented several methods and architectures, including the sub-TDL averaging topology, tap timing test, compensation histogramming and mixed calibration. The sub-TDL averaging topology is an optimal solution for bubble problems compared with reported methods. It is equivalent to direct ones-counter encoding and provides ideal robustness for the non-monotonicity or mismatch of TDLs. Compared with the traditional ones-counter method, the proposed topology does not require any extra logic resource or processing latency. Furthermore, said topology can remove the zero-width bins and restrain the non-linearity to some extent. The tap timing test is an innovative statistical test method based on the sub-TDL architecture to analyse the actual temporal characteristic of each tap in a TDL. This test provides more details and elaborates on the timing relations among taps compared with the code density test. The compensation histogramming architecture is built by the BRAM in FPGAs to minimise the resource and area cost. By combining with the mixed calibration method, this architecture achieves offset correction and the bin-width calibration at the same time. By being integrated and characterised in two FPGA series which have different delay line architectures and specifications, these proposed methods demonstrate excellent compatibility with different devices. From the test results, it can be seen that the new design presents a satisfactory linearity improvement on both FPGA series. The $DNL_{peak-peak}$ and $INL_{peak-peak}$ are reduced to 0.15LSB and 0.35LSB with 10.54ps resolution and to 0.27LSB and 0.59LSB with 5.02ps resolution respectively. Furthermore, two 96-channel TDC/TCSPC arrays were implemented in both FPGA series, which demonstrated the feasibility of multi-channel applications and the excellent uniformity of linearity among different channels.

The design, implementation, characterisation and application of a large-scale multi-channel TCSPC array for wide-field imaging system is presented in Chapter 6. The system is based on a world-leading 192x128 SPAD sensor and TDC array, as well as a FPGA-based data processing, histogramming and transmission system. The sensor can work in either the photon counting or the time-correlated imaging mode for typical intensity imaging and time-resolved applications. By implementing the histogramming in the firmware, 192x128 independent TCSPC channels are feasible for fast wide-field and ToF imaging. By modifying the supply

voltage of in-pixel RO-TDCs, the temporal resolution of the TCSPCs is tunable from 33ps to 120ps. The averaging IRF of the entire TCSPC array is 219ps, with a standard deviation of 26.78ps. The code density test results show that a DNL<±0.5 LSB has been achieved. To verify the applicability of the proposed system in biomedical areas, a typical fluorescence lifetime measurement was performed by combining the multiple pixels as a single TCSPC channel. The measured fluorescence decay curve was calculated by the fast CMM algorithm to extract the fluorescence lifetime of fluorescein. The calculated lifetime value is 3.9ns (the expected value is 4ns) which verifies that the proposed TCSPC system is accurate and reliable for fluorescence lifetime measurements.

7.2 Future Work

If sufficient time and resource were available, there are several potential works which could be further developed based on the current studies. There are three main directions: 1) related applications based on the proposed FPGA-TDCs, 2) the development of new FPGA-TDC architectures for the temporal resolution and linearity, and 3) the system improvement of the large-scale TCSPC array and the implementation of fast hardware fluorescence lifetime algorithms.

The two proposed TDCs have high potentials for different ToF and time-resolved applications. The first proposed direct-histogram TDC has a prominent sampling rate performance and multiple-event measurement capability. For applications without spatial discrimination requirements, such as fast time-correlated flow cytometry, the first TDC can measure the photons from multiple detectors at the same time to reduce the acquisition period for low light applications. The second proposed FPGA-TDCs can serve as the solution for the applications which have multiple detectors and require the parallel measurement. For example, the proposed 5ps TDC can be integrated with high PDE (> 86%) superconducting nanowire single-photon detector arrays developed by the Single Quantum B.V. and Delft University of Technology [180] to cultivate multichannel picosecond sensor/stopwatch systems. Successfully delivering this work would not only bring about impacts on ToF applications but also promise to conquer the holy grail of 10-picosecond PET imaging [181] and millimetre resolution for early cancer

diagnosis.

The second direction is to improve TDC performance based on the current proposed TDCs. One method is to integrate the dual-sampling architecture and the WU method based on the sub-TDL averaging topology to increase the temporal resolution. It is expected that the theoretically-temporal resolution of a single chain of TDL-TDC will be reduced to around 1.25ps implemented in an UltraScale FPGA. Another method aims to reduce the accuracy loss and uncertainty of TDC in FPGA fundamentally. All of the reported methods for linearity and accuracy improvement can be considered as post-processing calibration procedures, and they have tried to correct and restrain the negative effects of the existing measurement errors and significant accuracy loss. Since the intrinsic structure of TDLs in FPGAs is fixed, the significant accuracy loss is generated once a TDL captures a HIT signal, and wider bins in TDL will generate larger accuracy loss and more measurement uncertainty. To solve the accuracy loss from wide bins, a novel interpolated-TDL architecture is designed and tested to segment the wide bins. The implementation and verification of the interpolated-TDL have started based on the proposed sub-TDL averaging topology. Furthermore, future studies can focus on designing a compact TDL closed loop in FPGAs, which could significantly reduce the length and resource costs of TDLs and increase the design density and the number of TDC channels.

In terms of the large-scale integrated TCSPC array, the firmware and software still have potentials to be further improved. Besides increasing the frequency of operation and sampling clocks to improve acquisition speed, faster peripheral memory devices such as DDR3 or DDR4 RAMs can be used to boost the firmware histogramming.

For the fast fluorescence lifetime measurement, the proposed TCSPC does not contain the estimated algorithm of the lifetime values, and additional software calculation is necessary. This method will bring about heavy pressure on the data handling capacity of both firmware and software. The 192x128 SPAD and TCSPC array will yield around 100MB of data to be transferred and processed at each measurement. Therefore, it is valuable to conduct more research on the firmware implementation of the centre-of-mass method (CMM) [31] and Adv. CMM [182] for single- and two-exponential decay lifetime estimation. After using these hardware-friendly algorithms, the data volume of each frame can be reduced to dozens of

kilobytes in FPGA. Finally, the presented large-scale integrated TCSPC array can be combined with various pulse light sources such as solid-state and semiconductor laser or micro-LEDs [183] for widespread time-resolved applications, including fluorescence lifetime wide-field imaging [68], endoscopy [30], ToF-PET/MRI [184], ToF-imaging [25], hyperspectral imaging [69] and time-resolved flow cytometry [36].

References

[1]   B. Guinot, "Solar time, legal time, time in use," *Metrologia,* vol. 48, no. 4, p. S181, 2011.

[2]   J. Kalisz, "Review of methods for time interval measurements with picosecond resolution," *Metrologia,* vol. 41, no. 1, pp. 17-32, 2003.

[3]   W. Becker, *Advanced time-correlated single photon counting applications*. Berlin: Springer International Publishing, 2005.

[4]   D. O'Connor, *Time-correlated single photon counting*. Orlando, Florida: Academic Press, 1984.

[5]   S. Cova, A. Lacaita, M. Ghioni, G. Ripamonti, and T. Louis, "20-ps timing resolution with single-photon avalanche diodes," *Review of scientific instruments,* vol. 60, no. 6, pp. 1104-1110, 1989.

[6]   S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Applied optics,* vol. 35, no. 12, pp. 1956-1976, 1996.

[7]   S. Cova, M. Bertolaccini, and C. Bussolati, "The measurement of luminescence waveforms by single-photon techniques," *physica status solidi (a),* vol. 18, no. 1, pp. 11-62, 1973.

[8]   S. Kinoshita, H. Ohta, and T. Kushida, "Subnanosecond fluorescence-lifetime measuring system using single photon counting method with mode-locked laser excitation," *Review of Scientific Instruments,* vol. 52, no. 4, pp. 572-575, 1981.

[9]   A. Sloman and M. Swords, "A fast and economical gated discriminator," *J. Phys. E: Sci. Instrum.,* vol. 11, no. 6, p. 521, 1978.

[10]   W. Becker, B. Su, and A. Bergmann, "Fast-acquisition multispectral FLIM by parallel TCSPC," in *Multiphoton Microscopy in the Biomedical Sciences IX*, 2009, vol. 7183: International Society for Optics and Photonics, p. 718305.

[11]   M. Micic, D. Hu, Y. D. Suh, G. Newton, M. Romine, and H. P. Lu, "Correlated atomic force microscopy and fluorescence lifetime imaging of live bacterial cells," *Colloids and Surfaces B: Biointerfaces,* vol. 34, no. 4, pp. 205-212, 2004.

[12]   J. Richardson *et al.*, "A 32× 32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging," in *Custom Integrated Circuits Conference, 2009. CICC'09. IEEE*, 2009: IEEE, pp. 77-80.

[13]   S. Burri, H. Homulle, C. Bruschini, and E. Charbon, "LinoSPAD: a time-resolved 256x1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second," in *Optical Sensing and Detection IV*, 2016, vol. 9899: International Society for Optics and Photonics, p. 98990D.

[14]   A. T. Erdogan, R. Walker, N. Finlayson, N. Krstajic, G. O. Williams, and R. K. Henderson, "A 16.5 Giga Events/s 1024× 8 SPAD Line Sensor with per-pixel Zoomable 50ps-6.4 ns/bin Histogramming TDC," in *2017 Symposium on VLSI Circuits*, 2017: IEEE, pp. C292-C293.

[15]   J. Kalisz, R. Pelka, and A. Poniecki, "Precision time counter for laser ranging to satellites," *Review of scientific instruments,* vol. 65, no. 3, pp. 736-741, 1994.

[16]   K. Maatta and J. Kostamovaara, "A high-precision time-to-digital converter for pulsed time-of-flight laser radar applications," *IEEE Trans. Instrum. Meas.*

, vol. 47, no. 2, pp. 521-536, 1998.

[17]   P. Chen, Liu, Shen-Luan, Wu, Jingshown, "A CMOS pulse-shrinking delay element for time interval measurement," *IEEE Trans. Circuits Syst. II, Analog. digital signal processing,* vol. 47, no. 9, pp. 954-958, 2000.

[18]   S. Tisa, A. Lotito, A. Giudice, and F. Zappa, "Monolithic time-to-digital converter with 20ps resolution," in *Solid-State Circuits Conference, 2003. ESSCIRC'03. Proceedings of the 29th European*, 2003: IEEE, pp. 465-468.

[19]   I. Nissinen and J. Kostamovaara, "On-chip voltage reference-based time-to-digital converter for pulsed time-of-flight laser radar measurements," *IEEE Trans. Instrum. Meas.,* vol. 58, no. 6, pp. 1938-1948, 2009.

[20]   Y.-H. Seo, J.-S. Kim, H.-J. Park, and J.-Y. Sim, "A 1.25 ps Resolution 8b Cyclic TDC in 0.13 umCMOS," *IEEE J. Solid-State Circuits.,* vol. 47, no. 3, pp. 736-743, 2012.

[21]   A. Hamza, S. Ibrahim, M. El-Nozahi, and M. Dessouky, "A low-power, 9-Bit, 1.2 ps resolution two-step time-to-digital converter in 65 nm CMOS," *2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS),* 2015.

[22]   A. I. Hussein, S. Vasadi, and J. Paramesh, "A 450 fs 65-nm CMOS millimeter-wave time-to-digital converter using statistical element selection for all-digital PLLs," *IEEE J. Solid-State Circuits.,* vol. 53, no. 2, pp. 357-374, 2018.

[23]   K. Funk, A. Woitecki, C. Franjic-Würtz, T. Gensch, F. Möhrlen, and S. Frings, "Modulation of chloride homeostasis by inflammatory mediators in dorsal root ganglion neurons," *Molecular pain,* vol. 4, no. 1, p. 32, 2008.

[24]   E. Scapparone, "The Time-of-Flight detector of the ALICE experiment," *J. Phys. G: Nucl. Part. Phys.,* vol. 34, no. 8, pp. S725-S728(4), 2007.

[25]   A. R. Ximenes, P. Padmanabhan, M.-J. Lee, Y. Yamashita, D. Yaung, and E. Charbon, "A 256×256 45/65nm 3D-stacked SPAD-based direct TOF image sensor for LiDAR applications with optical polar modulation for up to 18.6 dB interference suppression," in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, 2018: IEEE, pp. 96-98.

[26]   R. Ikuta *et al.*, "Wide-band quantum interface for visible-to-telecommunication wavelength conversion," *Nat. Commun.,* vol. 2, no. 537, p. 1544, 2011.

[27]   C. Wulker, A. Sitek, and S. Prevrhal, "Time-of-flight PET image reconstruction using origin ensembles," *Physics in Medicine & Biology,* vol. 60, no. 5, p. 1919, 2015.

[28]   M. Pavlovic, I. Huber, R. Konrad, and U. Busch, "Application of MALDI-TOF MS for the identification of food borne bacteria," *The open microbiology journal,* vol. 7, p. 135, 2013.

[29]   J. R. Lakowicz, *Principles of fluorescence spectroscopy*. Springer Science & Business

Media, 2013.

[30] G. O. Fruhwirth, S. Ameer-Beg, R. Cook, T. Watson, T. Ng, and F. Festy, "Fluorescence lifetime endoscopy using TCSPC for the measurement of FRET in live cells," *Optics express,* vol. 18, no. 11, pp. 11148-11158, 2010.

[31] D. D.-U. Li *et al.*, "Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm," *J. Biomed. Opt.,* vol. 16, no. 9, pp. 096012-096012-12, 2011.

[32] I. H. Munro *et al.*, "Toward the clinical application of time-domain fluorescence lifetime imaging," *J. Biomed. Opt.,* vol. 10, no. 5, p. 051403, 2005.

[33] K. Suhling, P. M. French, and D. Phillips, "Time-resolved fluorescence microscopy," *Photochemical & Photobiological Sciences,* vol. 4, no. 1, pp. 13-22, 2005.

[34] S. P. Poland *et al.*, "A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging," *Biomed. Opt. Express,* vol. 6, no. 2, pp. 277-296, 2015.

[35] J. L. Rinnenthal *et al.*, "Parallelized TCSPC for dynamic intravital fluorescence lifetime imaging: quantifying neuronal dysfunction in neuroinflammation," *PLoS One,* vol. 8, no. 4, p. e60100, 2013.

[36] W. Li, G. Vacca, M. Castillo, K. D. Houston, and J. P. Houston, "Fluorescence lifetime excitation cytometry by kinetic dithering," *Electrophoresis,* vol. 35, no. 12-13, pp. 1846-1854, 2014.

[37] G. Giraud *et al.*, "Fluorescence lifetime biosensing with DNA microarrays and a CMOS-SPAD imager," *Biomed. Opt. Express,* vol. 1, no. 5, pp. 1302-1308, 2010.

[38] M. Gersbach *et al.*, "A time-resolved, low-noise single-photon image sensor fabricated in deep-submicron CMOS technology," *IEEE J. Solid-State Circuits.,* vol. 47, no. 6, pp. 1394-1407, 2012.

[39] D. E. Schwartz, E. Charbon, and K. L. Shepard, "A single-photon avalanche diode array for fluorescence lifetime imaging microscopy," *IEEE J. Solid-State Circuits.,* vol. 43, no. 11, pp. 2546-2557, 2008.

[40] J. S. Karp, S. Surti, M. E. Daube-Witherspoon, and G. Muehllehner, "Benefit of time-of-flight in PET: experimental and clinical results," *J. Nucl. Med.,* vol. 49, no. 3, pp. 462-470, 2008.

[41] M. Ito, S. J. Hong, and J. S. Lee, "Positron emission tomography (PET) detectors with depth-of- interaction (DOI) capability," *Biomedical Engineering Letters,* vol. 1, no. 2, pp. 70-81, 2011.

[42] E. Venialgo *et al.*, "Towards a Full-Flexible and Fast-Prototyping TOF-PET Block Detector Based on TDC-on-FPGA," *IEEE Trans. Radiat. Plasma. Med. Sci.,* 2018.

[43] D. J. Kadrmas, M. E. Casey, M. Conti, B. W. Jakoby, C. Lois, and D. W. Townsend, "Impact of time-of-flight on PET tumor detection," *J. Nucl. Med,* vol. 50, no. 8, pp. 1315-23, 2009.

[44] A. S. Yousif and J. W. Haslett, "A fine resolution TDC architecture for next generation PET imaging," *IEEE Trans. Nucl. Sci.,* vol. 54, no. 5, pp. 1574-1582, 2007.

[45] S. Surti, A. Kuhn, M. E. Werner, A. E. Perkins, J. Kolthammer, and J. S. Karp, "Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities," *J. Nucl. Med,* vol. 48, no. 3, pp. 471-480, 2007.

[46]   L. Van Elmbt, S. Vandenberghe, S. Walrand, S. Pauwels, and F. Jamar, "Comparison of yttrium-90 quantitative imaging by TOF and non-TOF PET in a phantom of liver selective internal radiotherapy," *Physics in Medicine & Biology,* vol. 56, no. 21, p. 6759, 2011.

[47]   C. S. Levin, S. H. Maramraju, M. M. Khalighi, T. W. Deller, G. Delso, and F. Jansen, "Design features and mutual compatibility studies of the time-of-flight PET capable GE SIGNA PET/MR system," *IEEE transactions on medical imaging,* vol. 35, no. 8, pp. 1907-1914, 2016.

[48]   K. Vikman, H. Iitti, P. Matousek, M. Towrie, A. W. Parker, and T. Vuorinen, "Kerr gated resonance Raman spectroscopy in light fastness studies of ink jet prints," *Vibrational spectroscopy,* vol. 37, no. 1, pp. 123-131, 2005.

[49]   D. W. Shipp, F. Sinjab, and I. Notingher, "Raman spectroscopy: techniques and applications in the life sciences," *Advances in Optics and Photonics,* vol. 9, no. 2, pp. 315-428, 2017.

[50]   E. V. Efremov, J. B. Buijs, C. Gooijer, and F. Ariese, "Fluorescence rejection in resonance Raman spectroscopy using a picosecond-gated intensified charge-coupled device camera," *Applied spectroscopy,* vol. 61, no. 6, pp. 571-578, 2007.

[51]   I. Nissinen, J. Nissinen, P. Keränen, D. Stoppa, and J. Kostamovaara, "A 16 x 256 SPAD Line Detector With a 50-ps, 3-bit, 256-Channel Time-to-Digital Converter for Raman Spectroscopy," *IEEE Sensors Journal,* vol. 18, no. 9, pp. 3789-3798, 2018.

[52]   F. Bigongiari, R. Roncella, R. Saletti, and P. Terreni, "A 250-ps time-resolution CMOS multihit time-to-digital converter for nuclear physics experiments," *IEEE Trans. Nucl. Sci.,* vol. 46, no. 2, pp. 73-77, 1999.

[53]   A. Akindinov *et al.*, "Design aspects and prototype test of a very precise TDC system implemented for the Multigap RPC of the ALICE-TOF," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 533, no. 1, pp. 178-182, 2004.

[54]   Q. Shen *et al.*, "Time interval analyzer with FPGA-based TDC for free space quantum key distribution: Principle and validation with prototype setup," in *2012 18th IEEE-NPSS Real Time Conference*, 2012: IEEE, pp. 1-6.

[55]   P. Palojarvi, K. Maatta, and J. Kostamovaara, "Pulsed time-of-flight laser radar module with millimeter-level accuracy using full custom receiver and TDC ASICs," *IEEE Trans. Instrum. Meas.,* vol. 51, no. 5, pp. 1102-1108, 2002.

[56]   J. N. Lygouras, T. P. Pachidis, K. N. Tarchanidis, and V. S. Kodogiannis, "Adaptive High-Performance Velocity Evaluation Based on a High-Resolution Time-to-Digital Converter," *IEEE Trans. Instrum. Meas.,* vol. 57, no. 9, pp. 2035-2043, 2008.

[57]   J. F. Cavanaugh *et al.*, "The Mercury Laser Altimeter instrument for the MESSENGER mission," in *Space. Sci. Rev.*, vol. 131no. 1-4): Springer, 2007, pp. 451-479.

[58]   D. E. Smith *et al.*, "The lunar orbiter laser altimeter investigation on the lunar reconnaissance orbiter mission," *Space. Sci. Rev.,* vol. 150, no. 1-4, pp. 209-241, 2010.

[59]   J. A. Christian and S. Cryan, "A survey of LIDAR technology and its use in spacecraft relative navigation," in *AIAA Guidance, Navigation, and Control (GNC) Conference*, 2013, p. 4641.

[60]   A. Hamza, S. Ibrahim, M. El-Nozahi, and M. Dessouky, "A wideband 5 GHz digital PLL using a low-power two-step time-to-digital converter," in *Electronics, Circuits, and*

Systems (ICECS), 2015 IEEE International Conference on, 2015: IEEE, pp. 328-331.

[61]   L. M. Hirvonen and K. Suhling, "Wide-field TCSPC: methods and applications," *Meas. Sci. Technol.,* vol. 28, no. 1, p. 012003, 2017.

[62]   Z. Cheng, X. Zheng, M. J. Deen, and H. Peng, "Recent Developments and Design Challenges of High-Performance Ring Oscillator CMOS Time-to-Digital Converters," *IEEE Trans. Electron Devices,* vol. 63, no. 1, pp. 235-251, 2016.

[63]   Q. Yuan, B. Zhang, J. Wu, and M. E. Zaghloul, "A high resolution time-to-digital converter on FPGA for Time-Correlated Single Photon Counting," in *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2012: IEEE, pp. 900-903.

[64]   Y. Wang and C. Liu, "A 3.9 ps Time-interval RMS Precision Time-to-Digital Converter using a Dual-sampling Method in an UltraScale FPGA," *IEEE Trans. Nucl. Sci.,* vol. 63, no. 5, pp. 2617 - 2621, 2016.

[65]   H. Chen, Y. Zhang, and D. D.-U. Li, "A Low Nonlinearity, Missing-Code Free Time-to-Digital Converter Based on 28-nm FPGAs With Embedded Bin-Width Calibrations," *IEEE Trans. Instrum. Meas.,* vol. 66, no. 7, pp. 1912-1921, 2017.

[66]   H. Chen and D. D.-U. Li, "Multichannel, Low Nonlinearity Time-to-Digital Converters Based on 20 and 28 nm FPGAs," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 4, pp. 3265-3274, 2019.

[67]   R. K. Henderson *et al.*, "A 192× 128 time correlated single photon counting imager in 40nm CMOS technology," in *ESSCIRC 2018-IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, 2018: IEEE, pp. 54-57.

[68]   R. K. Henderson *et al.*, "A 192 x 128 Time Correlated SPAD Image Sensor in 40-nm CMOS Technology," *IEEE J. Solid-State Circuits.,* 2019.

[69]   A. D. Griffiths *et al.*, "Hyperspectral Imaging Under Low Illumination with a Single Photon Camera," in *2018 IEEE British and Irish Conference on Optics and Photonics (BICOP)*, 2018: IEEE, pp. 1-4.

[70]   P. Napolitano, A. Moschitta, and P. Carbone, "A survey on time interval measurement techniques and testing methods," in *Proc. IEEE Instrum. Meas. Techn. Conference (I2MTC),*, Austin, TX,USA, 2010: IEEE, pp. 181-186.

[71]   J. Doernberg, H.-S. Lee, and D. A. Hodges, "Full-speed testing of A/D converters," *IEEE J. Solid-State Circuits.,* vol. 19, no. 6, pp. 820-827, 1984.

[72]   N. L. a. A. Geraci, "Comparison of interpolation techniques for TDCs implementation in FPGA," presented at the 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015.

[73]   IEEE. IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters [Online] Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5692956

[74]   M. Bojan, T. Simone, A. V. Federica, T. Alberto, and Z. Franco, "A High-Linearity, 17 ps Precision Time-to-DigitalConverter Based on a Single-Stage Vernier DelayLoop Fine Interpolation," *IEEE Trans. Circuits Syst. I, Reg. Papers* vol. 60, no. 3, pp. 557-569, 2013.

[75]   J. P. Jansson, A. Mantyniemi, and J. Kostamovaara, "A CMOS time-to-digital converter with better than 10 ps single-shot precision," *IEEE J. Solid-State Circuits.,* vol. 41, no. 6, pp. 1286-1296, 2006.

[76]   J.-S. Kim, Y.-H. Seo, Y. Suh, H.-J. Park, and J.-Y. Sim, "A 300-MS/s, 1.76-ps-resolution, 10-b asynchronous pipelined time-to-digital converter with on-chip digital background calibration in 0.13-μm CMOS," *IEEE J. Solid-State Circuits.,* vol. 48, no. 2, pp. 516-526, 2012.

[77]   R. Lussana, F. Villa, A. Dalla Mora, D. Contini, A. Tosi, and F. Zappa, "Enhanced single-photon time-of-flight 3D ranging," *Optics express,* vol. 23, no. 19, pp. 24962-24973, 2015.

[78]   K. Park and J. Park, "Time-to-digital converter of very high pulse stretching ratio for digital storage oscilloscopes," *Review of scientific instruments,* vol. 70, no. 2, pp. 1568-1574, 1999.

[79]   A. Takahashi, M. Nishizawa, Y. Inagaki, M. Koishi, and K. Kinoshita, "New femtosecond streak camera with temporal resolution of 180 fs," in *Generation, amplification, and measurement of ultrashort laser pulses*, 1994, vol. 2116: International Society for Optics and Photonics, pp. 275-284.

[80]   T. E. Rahkonen and J. T. Kostamovaara, "The use of stabilized CMOS delay lines for the digitization of short time intervals," *IEEE J. Solid-State Circuits.,* vol. 28, no. 8, pp. 887-894, 1993.

[81]   J.-C. Lai and T.-Y. Hsu, "Cost-Effective Time-to-Digital Converter Using Time-Residue Feedback," *IEEE Trans Ind Electron.,* vol. 64, no. 6, pp. 4690 - 4700, 2017.

[82]   I. P. Dan, "Review of Sub-Nanosecond Time-Interval Measurements," *IEEE Trans. Nucl. Sci.,* vol. 20, no. 5, pp. 36-51, 1973.

[83]   N. Karpov, "Vernier method of measuring time intervals," *Meas. Sci. Technol.,* vol. 23, no. 9, pp. 817-820, 1980.

[84]   T.-i. Otsuji, "A picosecond-accuracy, 700-MHz range, Si bipolar time interval counter LSI," *IEEE J. Solid-State Circuits.,* vol. 28, no. 9, pp. 941-947, 1993.

[85]   L. Vercesi, A. Liscidini, and R. Castello, "Two-dimensions Vernier time-to-digital converter," *IEEE J. Solid-State Circuits.,* vol. 45, no. 8, pp. 1504-1512, 2010.

[86]   D. R. Hoppe, "Differential time interpolator," United States   Patent Appl. US4433919A, 1982.

[87]   M. J. Loinaz and B. A. Wooley, "A CMOS multichannel IC for pulse timing measurements with 1-mV sensitivity," *IEEE J. Solid-State Circuits.,* vol. 30, no. 12, pp. 1339-1349, 1995.

[88]   Y. Arai and M. Ikeno, "A time digitizer CMOS gate-array with a 250 ps time resolution," *IEEE J. Solid-State Circuits.,* vol. 31, no. 2, pp. 212-220, 1996.

[89]   P. Dudek, S. Szczepanski, and J. V. Hatfield, "A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line," *IEEE J. Solid-State Circuits.,* vol. 35, no. 2, pp. 240-247, 2000.

[90]   C. Ljuslin, J. Christiansen, A. Marchioro, and O. Klingsheim, "An integrated 16-channel CMOS time to digital converter," *IEEE Trans. Nucl. Sci.,* vol. 41, no. 4, pp. 1104-1108, 1994.

[91]   R. Pelk, J. Kalisz, and R. Szplet, "Nonlinearity correction of the integrated time-to-digital converter with direct coding," *IEEE Trans. Instrum. Meas.*
, vol. 46, no. 2, pp. 449-453, 1997.

[92]   J. Kalisz, R. Szplet, J. Pasierbinski, and A. Poniecki, "Field-programmable-gate-array-based time-to-digital converter with 200-ps resolution," *IEEE Trans. Instrum. Meas.,* vol. 46, no. 1, pp. 51-55, 1997.

[93] N. Paschalidis *et al.*, "An integrated time to digital converter for space instrumentation," in *Proc. 7th NASA Symp. VLSI Design*, 1998: Univ. New Mexico, pp. 5.4. 1-5.4. 8.

[94] E. Raisanen-Ruotsalainen, T. Rahkonen, and J. Kostamovaara, "A low-power CMOS time-to-digital converter," *IEEE J. Solid-State Circuits.*, vol. 30, no. 9, pp. 984-990, 1995.

[95] Y. Liu *et al.*, "A 6ps resolution pulse shrinking time-to-digital converter as phase detector in multi-mode transceiver," in *2008 IEEE Radio and Wireless Symposium*, 2008: IEEE, pp. 163-166.

[96] I. Nissinen, A. Mantyniemi, and J. Kostamovaara, "A CMOS time-to-digital converter based on a ring oscillator for a laser radar," in *ESSCIRC 2004-29th European Solid-State Circuits Conference (IEEE Cat. No. 03EX705)*, 2003: IEEE, pp. 469-472.

[97] K.-C. Choi, S.-W. Lee, B.-C. Lee, and W.-Y. Choi, "A time-to-digital converter based on a multiphase reference clock and a binary counter with a novel sampling error corrector," *IEEE Trans. Circuits Syst. II, Exp. Briefs,* vol. 59, no. 3, pp. 143-147, 2012.

[98] M. Z. Straayer and M. H. Perrott, "A multi-path gated ring oscillator TDC with first-order noise shaping," *IEEE J. Solid-State Circuits.*, vol. 44, no. 4, pp. 1089-1098, 2009.

[99] J. Chen, H. Yumei, and H. Zhiliang, "A multi-path gated ring oscillator based time-to-digital converter in 65 nm CMOS technology," *J. Semicond.*, vol. 34, no. 3, p. 035004, 2013.

[100] P. Lu, A. Liscidini, and P. Andreani, "A 3.6 mW, 90 nm CMOS gated-Vernier time-to-digital converter with an equivalent resolution of 3.2 ps," *IEEE J. Solid-State Circuits.*, vol. 47, no. 7, pp. 1626-1635, 2012.

[101] P. Lu, A. Liscidini, and P. Andreani, "A 2-D GRO vernier time-to-digital converter with large input range and small latency," *Analog Integrated Circuits and Signal Processing,* vol. 76, no. 2, pp. 195-206, 2013.

[102] R. Nutt, "Digital time intervalometer," *Review of scientific instruments,* vol. 39, no. 9, pp. 1342-1345, 1968.

[103] R. Szplet, J. Kalisz, and R. Szymanowski, "Interpolating time counter with 100 ps resolution on a single FPGA device," *IEEE Trans. Instrum. Meas.,* vol. 49, no. 4, pp. 879-883, 2000.

[104] C. Favi and E. Charbon, "A 17ps time-to-digital converter implemented in 65nm FPGA technology," in *Proc. FPGA' 09*, Monterey, California, USA, 2009: ACM, pp. 113-120.

[105] J. Wang, S. Liu, L. Zhao, X. Hu, and Q. An, "The 10-ps Multitime Measurements Averaging TDC Implemented in an FPGA," *IEEE Trans. Nucl. Sci.,* vol. 58, no. 4, pp. 2011-2018, 2011.

[106] C. Y. Yao, W. C. Hsia, and Y. J. Wen, "The Soft-Injection-Locked Ring Oscillator and Its Application in a Vernier-Based TDC," *IEEE Trans. Instrum. Meas.,* vol. 63, no. 8, pp. 2064-2071, 2014.

[107] J. Jansson, A. Mantyniemi, and J. Kostamovaara, "A delay line based CMOS time digitizer IC with 13 ps single-shot precision," in *2005 IEEE International Symposium on Circuits and Systems*, 2005: IEEE, pp. 4269-4272.

[108] J.-P. Jansson, A. Mantyniemi, and J. Kostamovaara, "Synchronization in a multilevel CMOS time-to-digital converter," *IEEE Trans. Circuits. Syst. I, Regul. Pap.,* vol. 56, no. 8, pp. 1622-1634, 2008.

[109] K. Kim, Y.-H. Kim, W. Yu, and S. Cho, "A 7 bit, 3.75 ps resolution two-step time-to-

digital converter in 65 nm CMOS using pulse-train time amplifier," *IEEE J. Solid-State Circuits.,* vol. 48, no. 4, pp. 1009-1017, 2013.

[110] A. Elshazly, S. Rao, B. Young, and P. K. Hanumolu, "A Noise-Shaping Time-to-Digital Converter Using Switched-Ring Oscillators-Analysis, Design, and Measurement Techniques," *J. Solid-State Circuits,* vol. 49, no. 5, pp. 1184-1197, 2014.

[111] Z. Cheng, M. J. Deen, and H. Peng, "A low-power gateable vernier ring oscillator time-to-digital converter for biomedical imaging applications," *IEEE Trans. Biomed. Circuits Syst.,* vol. 10, no. 2, pp. 445-454, 2016.

[112] Xilinx. Virtex-5 FPGA User Guide [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug190.pdf

[113] Xilinx. 7 Series FPGAs Configurable Logic Block [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug474_7Series_CLB.pdf

[114] Xilinx. UltraScale Architecture Configurable Logic Block [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug574-ultrascale-clb.pdf

[115] Xilinx. Virtex-7 T and XT FPGAs Data Sheet: DC and AC Switching Characteristics [Online]   Available: https://www.xilinx.com/support/documentation/data_sheets/ds183_Virtex_7_Data_Sheet.pdf

[116] Xilinx. Kintex UltraScale FPGAs Data Sheet: DC and AC Switching Characteristics [Online] Available: https://www.xilinx.com/support/documentation/data_sheets/ds892-kintex-ultrascale-data-sheet.pdf

[117] M. W. Fishburn, L. H. Menninga, C. Favi, and E. Charbon, "A 19.6 ps, FPGA-based TDC with multiple channels for open source applications," *IEEE Trans. Nucl. Sci.,* vol. 60, no. 3, pp. 2203-2208, 2013.

[118] Y. Wang, J. Kuang, C. Liu, and Q. Cao, "A 3.9 ps RMS Precision Time-to-Digital Converter Using Ones Counter Encoding Scheme in a Kintex-7 FPGA," *IEEE Trans. Nucl. Sci.,* vol. 64, no. 10, pp. 2713 - 2718, 2017.

[119] M. Daigneault and J. P. David, "A novel 10 ps resolution TDC architecture implemented in a 130nm process FPGA " in *Proc. 8th Int. NEWCAS Conf.*, Montreal, QC, Canada, 2010, pp. 281-284.

[120] R. Szplet, Z. Jachna, P. Kwiatkowski, and K. Rozyc, "A 2.9 ps equivalent resolution interpolating time counter based on multiple independent coding lines," *Meas. Sci. Technol.,* vol. 24, no. 3, pp. 35904-15, 2013.

[121] J. Wu and Z. Shi, "The 10-ps wave union TDC: Improving FPGA TDC resolution beyond its cell delay," in *Proc. IEEE Nuclear Science Symp. Conf. Rec.*, Dresden, Germany, 2008: IEEE, pp. 3440-3446.

[122] J. Qi, Z. Deng, H. Gong, and Y. Liu, "A 20ps resolution wave union FPGA TDC with on-chip real time correction," in *IEEE Nuclear Science Symposium conference record. Nuclear Science Symposium*, 2010, pp. 396-399.

[123] E. Bayer and M. Traxler, "A High-Resolution (<10 ps RMS) 48-Channel Time-to-Digital Converter (TDC) Implemented in a Field Programmable Gate Array (FPGA)," *IEEE Trans. Nucl. Sci.,* vol. 58, no. 4, pp. 1547-1552, 2011.

[124] R. Szplet and K. Klepacki, "An FPGA-integrated time-to-digital converter based on two-stage pulse shrinking," *IEEE Trans. Instrum. Meas.,* vol. 59, no. 6, pp. 1663-1670, 2010.

[125] M. Zhang, H. Wang, and Y. Liu, "A 7.4 ps FPGA-Based TDC with a 1024-Unit Measurement Matrix," *Sensors,* vol. 17, no. 4, p. 865, 2017.

[126] J. Imrek *et al.*, "FPGA Based TDC Using Virtex-4 ISERDES Blocks," *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE,* pp. 1413 - 1415, 2010.

[127] Xilinx. Virtex-5 FPGA Data Sheet: DC and Switching Characteristics [Online] Available: https://www.xilinx.com/support/documentation/data_sheets/ds202.pdf

[128] C. Liu and Y. Wang, "A 128-channel, 710 M samples/second, and less than 10 ps RMS resolution time-to-digital converter implemented in a Kintex-7 FPGA," *IEEE Trans. Nucl. Sci.,* vol. 62, no. 3, pp. 773-783, 2015.

[129] H. Menninga, C. Favi, M. W. Fishburn, and E. Charbon, "A multi-channel, 10ps resolution, FPGA-based TDC with 300MS/s throughput for open-source PET applications," in *2011 IEEE Nuclear Science Symposium Conference Record*, 2011: IEEE, pp. 1515-1522.

[130] Y. Wang, P. Kuang, and C. Liu, "A 256-channel multi-phase clock sampling-based time-to-digital converter implemented in a Kintex-7 FPGA," in *Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International*, 2016: IEEE, pp. 1-5.

[131] P. Chen, C.-C. Chen, J.-C. Zheng, and Y.-S. Shen, "A PVT insensitive vernier-based time-to-digital converter with extended input range and high accuracy," *IEEE Trans. Nucl. Sci.,* vol. 54, no. 2, pp. 294-302, 2007.

[132] W. Pan, G. Gong, and J. Li, "A 20-ps time-to-digital converter (TDC) implemented in field-programmable gate array (FPGA) with automatic temperature correction," *IEEE Trans. Nucl. Sci.,* vol. 61, no. 3, pp. 1468-1473, 2014.

[133] J. Song, Q. An, and S. Liu, "A high-resolution time-to-digital converter implemented in field-programmable-gate-arrays," *IEEE Trans. Nucl. Sci.,* vol. 53, no. 1, pp. 236-241, 2006.

[134] J. Wu, Z. Shi, and I. Y. Wang, "Firmware-only implementation of time-to-digital converter (TDC) in field-programmable gate array (FPGA)," presented at the Proc. IEEE Conf. Rec. NSS., Portland, Oregon, USA, 19-25 Oct. 2003, 2003.

[135] Xilinx. 7 Series FPGAs Clocking Resources User Guide [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug472_7Series_Clocking.pdf

[136] J. Y. Won, S. I. Kwon, H. S. Yoon, G. B. Ko, J.-W. Son, and J. S. Lee, "Dual-phase tapped-delay-line time-to-digital converter with on-the-fly calibration implemented in 40 nm FPGA," *IEEE Trans. Biomed. Circuits Syst.,* vol. 10, no. 1, pp. 231-242, 2016.

[137] J. Wang, S. Liu, Q. Shen, H. Li, and Q. An, "A Fully Fledged TDC Implemented in Field-Programmable Gate Arrays," *IEEE Trans. Nucl. Sci.,* vol. 57, no. 2, pp. 446-450, 2010.

[138] P. Chen, Y.-Y. Hsiao, Y.-S. Chung, W. X. Tsai, and J.-M. Lin, "A 2.5-ps Bin Size and 6.7-ps Resolution FPGA Time-to-Digital Converter Based on Delay Wrapping and Averaging," *IEEE Trans. Very Large Scale Integr. VLSI Syst.,* vol. 25, no. 1, pp. 114-124, 2017.

[139] R. Szplet, J. Kalisz, and Z. Jachna, "A 45 ps time digitizer with a two-phase clock and dual-edge two-stage interpolation in a field programmable gate array device," *Meas. Sci. Technol.,* vol. 20, no. 2, p. 025108, 2009.

[140] J. Wu, "Several key issues on implementing delay line based TDCs using FPGAs," *IEEE*

Trans. Nucl. Sci., vol. 57, no. 3, pp. 1543-1548, 2010.

[141] L. Zhao, X. Hu, S. Liu, and J. Wang, "The Design of a 16-Channel 15 ps TDC Implemented in a 65 nm FPGA," *IEEE Trans. Nucl. Sci.,* vol. 60, no. 5, pp. 3532-3536, 2013.

[142] Y. Wang and C. Liu, "A Nonlinearity Minimization-Oriented Resource-Saving Time-to-Digital Converter Implemented in a 28 nm Xilinx FPGA," *IEEE Trans. Nucl. Sci.,* vol. 62, no. 5, pp. 2003-2009, 2015.

[143] E. Sail and M. Vesterbacka, "A multiplexer based decoder for flash analog-to-digital converters," in *2004 IEEE Region 10 Conference TENCON 2004.*, 2004, vol. 500: IEEE, pp. 250-253.

[144] S. Kumar, M. Suman, and K. Baishnab, "A novel approach to thermometer-to-binary encoder of flash ADCs-bubble error correction circuit," in *2014 2nd International Conference on Devices, Circuits and Systems (ICDCS)*, 2014: IEEE, pp. 1-6.

[145] A. M. Amiri, M. Boukadoum, and A. Khouas, "A multihit time-to-digital converter architecture on FPGA," *IEEE Trans. Instrum. Meas.,* vol. 58, no. 3, pp. 530-540, 2009.

[146] S. S. Junnarkar, P. O'Connor, P. Vaska, and R. Fontaine, "FPGA-Based Self-Calibrating Time-to-Digital Converter for Time-of-Flight Experiments," *IEEE Trans. Nucl. Sci.,* vol. 56, no. 4, pp. 2374-2379, 2009.

[147] M.-A. Daigneault and J. P. David, "A high-resolution time-to-digital converter on FPGA using dynamic reconfiguration," *IEEE Trans. Instrum. Meas.,* vol. 60, no. 6, pp. 2070-2079, 2011.

[148] K. J. Hong, E. Kim, J. Y. Yeom, P. D. Olcott, and C. S. Levin, "FPGA-based time-to-digital converter for time-of-flight PET detector," in *Nuclear Science Symposium and Medical Imaging Conference*, 2012, pp. 2463-2465.

[149] N. Dutton *et al.*, "Multiple-Event Direct to Histogram TDC in 65nm FPGA Technology," in *Proc. IEEE PRIME*, Grenoble, France, 2014: IEEE, pp. 1-5.

[150] Q. Shen *et al.*, "A 1.7 ps equivalent bin size and 4.2 ps RMS FPGA TDC based on multichain measurements averaging method," *IEEE Trans. Nucl. Sci.,* vol. 62, no. 3, pp. 947-954, 2015.

[151] J. Y. Won and J. S. Lee, "Time-to-Digital Converter Using a Tuned-Delay Line Evaluated in 28-, 40-, and 45-nm FPGAs," *IEEE Trans. Instrum. Meas.,* vol. 65, no. 7, pp. 1678-1689, 2016.

[152] X. Qin, L. Wang, D. Liu, Y. Zhao, X. Rong, and J. Du, "A 1.15 ps Bin Size and 3.5 ps Single-Shot Preci-sion Time-to-Digital-Converter with On-Board Offset Correction in an FPGA," *IEEE Trans. Nucl. Sci.,* vol. 64, no. 12, pp. 2951 - 2957, 2017.

[153] S. Henzler, S. Koeppe, D. Lorenz, W. Kamp, R. Kuenemund, and D. Schmitt-Landsiedel, "A local passive time interpolation concept for variation-tolerant high-resolution time-to-digital conversion," *IEEE J. Solid-State Circuits.,* vol. 43, no. 7, pp. 1666-1676, 2008.

[154] Digilent. NetFPGA-SUME™ Reference Manual [Online] Available: https://reference.digilentinc.com/_media/sume:netfpga-sume_rm.pdf

[155] Microchip. DSC1103/23 Low-Jitter Precision LVDS Oscillator datasheet [Online] Available: https://www.xilinx.com/support/documentation/data_sheets/ds183_Virtex_7_Data_Shee

t.pdf

[156]   Xilinx. Power Methodology Guide [Online] Available: https://www.xilinx.com/support/documentation/sw_manuals/xilinx13_1/ug786_PowerMethodology.pdf

[157]   J.-Y. Wu, P.-K. Lu, and S.-D. Lin, "Two-dimensional photo-mapping on CMOS single-photon avalanche diodes," *Optics Express,* vol. 22, no. 13, pp. 16462-16471, 2014.

[158]   Xilinx. 7 Series FPGAs SelectIO Resources User Guide [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug471_7Series_SelectIO.pdf

[159]   P.-H. Chang, C.-M. Tsai, J.-Y. Wu, S.-D. Lin, and M.-C. Kuo, "Constant excess bias control for single-photon avalanche diode using real-time breakdown monitoring," *IEEE Electron Device Letters,* vol. 36, no. 8, pp. 859-861, 2015.

[160]   Xilinx. KCU105 Board User Guide [Online] Available: https://www.xilinx.com/support/documentation/boards_and_kits/kcu105/ug917-kcu105-eval-bd.pdf

[161]   S. Laboratories. Si5335A, 4 -Output, Any Frequency (< 350 MHz), Any Output, Clock Generator user guide [Online] Available: https://www.silabs.com/documents/public/data-sheets/Si5335.pdf

[162]   S. Laboratories. Si570 programmable low-jitter 3.3V LVDS differential oscillator [Online] Available: https://www.silabs.com/documents/public/data-sheets/si570.pdf

[163]   Xilinx. UltraScale Architecture SelectIO Resources User Guide [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug571-ultrascale-selectio.pdf

[164]   T. LeCroy. WaveRunner 6 Zi Oscilloscopes 400MHz - 4GHz datasheet [Online] Available: https://cdn.teledynelecroy.com/files/pdf/waverunner-6zi-datasheet.pdf

[165]   N. Krstajić *et al.*, "0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays," *Optics letters,* vol. 40, no. 18, pp. 4305-4308, 2015.

[166]   Murata-Power-Solutoin. 786 Series General Purpose Pulse Transformers datasheet [Online] Available: https://www.murata-ps.com/pub/data/magnetics/kmp_786.pdf

[167]   LINEAR-TECHNOLOGY, "LT1077 Micropower, Single Supply,Precision Op Amp datasheet."

[168]   Microchip. MCP6001/1R/1U/2/4 1 MHz, Low-Power Op Amp Datasheet [Online] Available: https://ww1.microchip.com/downloads/en/DeviceDoc/21733j.pdf

[169]   Opal-Kelly. XEM6310 integration board User's Manual [Online] Available: http://assets00.opalkelly.com/library/XEM6310-UM.pdf

[170]   LINEAR-TECHNOLOGY. LTC1446/LTC1446L   Dual 12-Bit Rail-to-Rail Micropower DACs in SO-8 [Online] Available: https://www.analog.com/media/en/technical-documentation/data-sheets/1446fa.pdf

[171]   Xilinx. Spartan-6 FPGA Block RAM Resources User Guide [Online] Available: https://www.xilinx.com/support/documentation/user_guides/ug383.pdf

[172]   Xilinx. Spartan-6 Family Overview Product Specification [Online] Available: https://www.xilinx.com/support/documentation/data_sheets/ds160.pdf

[173]   Micron. 1Gb: x4, x8, x16 DDR2 SDRAM Features [Online] Available: https://docs-emea.rs-online.com/webdocs/0e47/0900766b80e4746b.pdf

[174] Xilinx, "Spartan-6 FPGA Memory Interface Solutions User Guide," ed, 2011.

[175] S. Pellegrini *et al.*, "Industrialised SPAD in 40 nm Technology," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017: IEEE, pp. 16.5. 1-16.5. 4.

[176] J. A. Steinkamp and J. F. Keij, "Fluorescence intensity and lifetime measurement of free and particle-bound fluorophore in a sample stream by phase-sensitive flow cytometry," *Review of Scientific Instruments,* vol. 70, no. 12, pp. 4682-4688, 1999.

[177] D. Magde, G. E. Rojas, and P. G. Seybold, "Solvent dependence of the fluorescence lifetimes of xanthene dyes," *Photochemistry and Photobiology,* vol. 70, no. 5, pp. 737-744, 1999.

[178] R. Sjöback, J. Nygren, and M. Kubista, "Absorption and fluorescence properties of fluorescein," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 51, no. 6, pp. L7-L21, 1995.

[179] D.-U. Li, B. R. Rae, R. Andrews, J. Arlt, and R. K. Henderson, "Hardware implementation algorithm and error analysis of high-speed fluorescence lifetime sensing systems using center-of-mass method," *J. Biomed. Opt.,* vol. 15, no. 1, p. 017006, 2010.

[180] I. E. Zadeh *et al.*, "A single-photon detector with high efficiency and sub-10ps time resolution," *arXiv preprint arXiv:1801.06574,* 2018.

[181] P. Lecoq, "Pushing the limits in time-of-flight PET imaging," *IEEE Transactions on Radiation and Plasma Medical Sciences,* vol. 1, no. 6, pp. 473-485, 2017.

[182] D. D. Li *et al.*, "Advanced fluorescence lifetime imaging algorithms for CMOS single-photon sensor based multi-focal multi-photon microscopy," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013: IEEE, pp. 3036-3039.

[183] A. D. Griffiths *et al.*, "CMOS-integrated GaN LED array for discrete power level stepping in visible light communications," *Optics express,* vol. 25, no. 8, pp. A338-A345, 2017.

[184] L. H. C. Braga *et al.*, "An 8×16-pixel 92kSPAD time-resolved sensor with on-pixel 64ps 12b TDC and 100MS/s real-time energy histogramming in 0.13 μm CIS technology for PET/MRI applications," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013: IEEE, pp. 486-487.

Appendix

Journal Publications

[1]   H. Chen, Y. Zhang, and D. D.-U. Li, "A Low Nonlinearity, Missing-Code Free Time-to-Digital Converter Based on 28-nm FPGAs With Embedded Bin-Width Calibrations," IEEE Trans. Instrum. Meas., vol. 66, no. 7, pp. 1912-1921, 2017.

[2]   H. Chen and D. D.-U. Li, "Multichannel, Low Nonlinearity Time-to-Digital Converters Based on 20 and 28 nm FPGAs," IEEE Trans. Ind. Electron., vol. 66, no. 4, pp. 3265-3274, 2019.

[3]   R. K. Henderson, N. Johnston, F. M. Della Rocca, H. Chen, D. D.-U. Li, G. Hungerford, et al., "A 192 x 128 Time-Correlated SPAD Image Sensor in 40-nm CMOS Technology," IEEE J. Solid-State Circuits., pp. 1-10, 2019.

[4]   A. D. Griffiths, H. Chen, D. Li,, R. K. Henderson, J. Herrnsdorf, M. D. DAWSON, M. J. Strain, , " Multispectral time-of-flight imaging using light-emitting diodes," to be submitted to the Optics Express, vol. 27, no. 24, pp. 35485-35498, 2019.

Papers in Preparation

[5]   W. Xie, H. Chen, and D. Li, "Efficient design strategies towards 1 ps LSB time-to-digital conversion with maintained linearity in 20 nm FPGAs," submitted to IEEE Trans. Instrum. Meas. 2020.

Conference Submissions

[1]   R. K. Henderson, N. Johnston, H. Chen, D. D.-U. Li, G. Hungerford, R. Hirsch, et al., "A 192× 128 time-correlated single photon counting imager in 40nm CMOS technology," in ESSCIRC 2018-IEEE 44th European Solid-State Circuits Conference (ESSCIRC), 2018, pp. 54-57.

[2]   A. D. Griffiths, H. Chen, J. Herrnsdorf, D. Li, R. K. Henderson, M. J. Strain, et al., "Hyperspectral Imaging Under Low Illumination with a Single Photon Camera," in 2018

IEEE British and Irish Conference on Optics and Photonics (BICOP), 2018, pp. 1-4.

[3]   H. Chen and D. Li, "High-performance time-to-digital converters (TDC: high-precision stopwatches) and TDC arrays for time-resolved measurements," (C) Photonex Scotland: Advances in Photonics Techniques for Biomedical Sciences, Edinburgh, UK, 14th June 2018.

[4]   H. Chen and D. Li, "High-performance time-to-digital converters (TDC: high-precision stopwatches) and TDC arrays for time-resolved measurements," (C) 9th Annual SU2P Symposium, UK, Glasgow, 21st – 22nd May 2018.

[5]   H. Chen and D. Li, "Low nonlinearity, missing-code free time-to-digital converters based on 28nm FPGAs with embedded bin-width calibrations," (C) 2017 IEEE SENSORS, UK, Glasgow, 30th Oct-1st Nov 2017.

[6]   H. Chen and D. Li, "High linearity, low dead-time time-to-digital converters based on 28nm CMOS process suitable for multi-channel TCSPC applications," (C) 23rd PicoQuant Single Molecule Detection Workshop, Berlin, Germany, 13-15 September 2017.

[7]   H. Chen and D. Li, "Low nonlinearity, missing-code free time-to-digital converters based on 28nm FPGAs with embedded bin-width calibrations," (C) 12th FluoroFest International Workshop, Glasgow, UK, 24-26 April 2017.