Stephan Henzler

SPRINGER SERIES IN ADVANCED MICROELECTRONICS 29

Time-to-Digital Converters

Springer Series in Advanced Microelectronics 29

The Springer Series in Advanced Microelectronics provides systematic information on all the topics relevant for the design, processing, and manufacturing of microelectronic devices. The books, each prepared by leading researchers or engineers in their fields, cover the basic and advanced aspects of topics such as wafer processing, materials, device design, device technologies, circuit design, VLSI implementation, and subsystem technology. The series forms a bridge between physics and engineering and the volumes will appeal to practicing engineers as well as research scientists.

For other titles published in this series, go to
http://www.springer.com/series/4076

Stephan Henzler

Time-to-Digital Converters

Springer

Dr. Stephan Henzler
TU München
Lehrstuhl für Technische Elektronik
Arcisstr. 21
80290 München
Germany
henzler@tum.de

ISSN 1437-0387
ISBN 978-90-481-8627-3 e-ISBN 978-90-481-8628-0
DOI 10.1007/978-90-481-8628-0
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010921003

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Pour Julie

Contents

List of Symbols and Abbreviations

$A_{core}^{TDC-type}$	Core area of a TDC
ACF	Auto-correlation function
ADC	Analog-to-digital converter
ASDM	Asynchronous sigma delta modulator
BIST	Built-in self test
CML	Current mode logic
CMOS	Complementary metal oxide semiconductor
DAC	Digital-to-analog converter
DL	Delay-line
DL-TDC	Delay-line TDC
DLL	Delay-locked loop
DNL	Differential non-linearity
DR	Dynamic range
DTC	Digital-to-time converter
E_{gain}	TDC gain error
ELD	Early-late detector
E_{offset}	TDC offset error
ENOB	Effective number of bits
f_{clk}	Clock frequency
FCW	Frequency control word
FET	Field effect transistor
FOM	Figure of merit
FF	Flip-flop
GRO	Gated ring oscillator
INL	Integral non-linearity
k_{TDC}	TDC gain
k_{TDC}^{core}	TDC core gain
$k_{TDC}^{incremental}$	Incremental TDC gain
L	Transistor channel length
LC	Describes a resonator based on an inductance (L) and a capacitance (C)
LDO	Linear drop out (regulator)

LF	Loop factor
LO	Local oscillator
LPI	Local passive interpolation
LPV	Local process variations
LSB	Least significant bit
MOS	Metal oxide semiconductor
MS	Mixed-signal
MSB	Most significant bit
OPAMP	Operational amplifier
OSR	Oversampling ratio
$P_{core}^{TDC-type}$	Core power consumption of a TDC
PD	Phase detector
PFD	Phase-frequency detector
pdf	Probability density function
PLL	Phase-locked-loop
PPM	Pulse position modulation
PVT	Process, voltage, and temperature (variations)
PWM	Pulse width modulation
RDC	Resolution degradation coefficient
RF	Radio frequency
rms	Root-mean-square
RN	Asynchronous reset signal low-active
SCE	Short channel effects
SDM	Sigma delta modulator
SNDR	Signal-to-noise-and-distortion ratio
SNR	Signal-to-noise ratio
SOC	System-on-chip
SQNR	Signal-to-quantization-noise ratio
SSP	Single-shot precision
STA	Static timing analysis
std(.)	Standard deviation
T	Absolute temperature in Kelvin
TA	Time (interval) amplifier
T_{conv}	Conversion time of TDC
T_{dead}	Dead time, i.e. minimum time between two measurements
$T_{latency}$	Latency of TDC
T_{LSB}	Minimum time increment that can be resolved by a TDC
T_{max}	Maximum time interval that can be measured
T_{min}	Minimum time interval that can be measured
t_d	Delay of logic gate or circuit
TDC	Time-to-digital converter
VCO	Voltage controlled oscillator
V_{DD}	Positive supply potential of digital circuit
V_{LSB}	Minimum voltage increment that can be resolved by an ADC
V_{SS}	Negative supply potential of digital circuit

$V_T = \frac{kT}{q}$	Thermal voltage
V_{th}	Threshold voltage
V_{th}^0	Threshold voltage for $V_{DS} = 0$ and $V_{BS} = 0$
V_{tn}	NMOS threshold voltage
V_{tp}	PMOS threshold voltage
VLSI	Very large scale integration
W	Transistor channel width

Chapter 1
Foreword

Micro-electronics and so integrated circuit design are heavily driven by technology scaling. The main engine of scaling is an increased system performance at reduced manufacturing cost (per system). In most systems digital circuits dominate with respect to die area and functional complexity. Digital building blocks take full advantage of reduced device geometries in terms of area, power per functionality, and switching speed. On the other hand, analog circuits rely not on the fast transition speed between a few discrete states but fairly on the actual shape of the transistor characteristic. Technology scaling continuously degrades these characteristics with respect to analog performance parameters like output resistance or intrinsic gain. Below the 100 nm technology node the design of analog and mixed-signal circuits becomes perceptibly more difficult. This is particularly true for low supply voltages near to 1 V or below. The result is not only an increased design effort but also a growing power consumption. The area shrinks considerably less than predicted by the digital scaling factor. Obviously, both effects are contradictory to the original goal of scaling. However, digital circuits become faster, smaller, and less power hungry. The fast switching transitions reduce the susceptibility to noise, e.g. flicker noise in the transistors. There are also a few drawbacks like the generation of power supply noise or the lack of power supply rejection. Still, the advantages are overwhelming and suggest to implement as much system components as possible in the digital domain. As digital functionality nearly comes for free it usually makes no sense to optimize single gates or to elimination a few logic functions. In most cases the effort for these optimizations does not pay off. If however analog or mixed-signal performance can be increased by digital assist or enhancement techniques the effort should be spent. If digital circuit add-ons help to reduce analog power or improve robustness and reliability, the investment in these add-ons should be done. In the deep sub-micron regime a few thousand gates are not worth to be discussed. Very often the introduction of digital assist techniques or even all digital implementations cause a one time area and power overhead. The increased performance and new functionality should justify this expense. Moreover, the overhead will shrink in future technology generations so the full advantage of the increased performance and the robustness will pay off one or two generations later.

S. Henzler, *Time-to-Digital Converters*, Springer Series in Advanced Microelectronics 29,
DOI 10.1007/978-90-481-8628-0_1,

It is obvious that there are functions that can be done best with analog circuits. Hence, the message should not be to avoid analog circuits at all but to use digital techniques whenever they promise an advantage. The effort is justified very often.

By definition digital circuits (binary circuits to be more precise) cannot resolve any information in the amplitude domain. Instead they have a very high resolution in the time domain. The very first approach of trading voltage (current) resolution against time resolution is the sigma delta modulator: A coarse quantizer causes a considerable quantization error that is antagonized by oversampling with noise shaping. For high resolution at high bandwidths the quantizer often has more than two quantization levels so not the complete information is covered in the time domain. A generalization of this time domain concept encodes any information of the signal into the time domain. This can be done by providing time domain information at equidistant time instances: A pulse width modulated signal, for instance, consists of pulses that occur with a fixed rate and that contain the information within the width of the pulses. An asynchronous approach has not even a fixed sample rate but provides non-equidistant pulses. In this case both the sample information as well as the sampling instance is implicitly encoded by two time instances. Time-to-digital converters are highly precise stopwatches that convert time domain information into a digital representation. Hence, they represent the fundamental building blocks between the continuous time domain of time encoded data and the digital world. In principle TDCs only consist of sampled delay-lines. The challenges arise when continuous measurement, long time intervals, high resolution, linearity, low power consumption, and robustness come into the play. The architectures addressing one or more of these issues rapidly become more complex. This is the reason for this text on time-to-digital converters which is organized as follows: First an introduction to time domain signal processing is given in Chapter 2. The basic concepts of time-to-digital converters (TDC) are introduced and basic structures are discussed. Chapter 3 addresses the theory of TDCs. Performance figures are defined and compared to corresponding features of analog-to-digital converters. Noise and variability are discussed in detail. Their impact on the single-shot precision, the effective number of bits and the linearity is derived. Next, various TDC architectures for high dynamic range, linearity, bipolar measurement, matching, and low power consumption are discussed in Chapter 4. Many of these architectural concepts are based on research work by the author and are proposed and presented in this text for the very first time. High resolution below one gate delay is discussed in Chapter 5. Finally, representative TDC applications in the field of phase locked loops and analog-to-digital converters are discussed in Chapter 6.

The criterion for the selection of the material and for the focus of the text has been the advantage of the digital processing of time domain signals. The reason for this is the following: Time domain signal processing and so time-to-digital converters are proposed for high volume micro-electronics in order to circumvent analog impairments in nanometer-scale CMOS technologies. Hence, only TDC concepts are promising that have the ability to exploit the advantages of digital circuits. Time-to-digital converters that are based on sensitive analog approaches cannot take a real advantage of the time domain. If only such TDCs were available it would be

instructive to stay with the old analog/mixed-signal concepts that are based on the voltage domain but are architecturally mature, sophisticated, and proven for many years. A time domain signal representation, however, makes sense when the main building blocks are based on digital concepts. Anyhow, it is to be understood that time-to-digital converters are no digital circuits in the sense of synthesizeable semi-custom circuits. Yet some design automation is possible, but the careful knowledge and handling of the sensitive (continuous time) signals inside the TDC requires analog expert knowledge. This means that a TDC can exploit some advantages of digital circuits but still behaves like a mixed-signal circuit. Not least because of the fact that it processes continuous time signals.

With this book I want to give an overview and substantiated impression on time-to-digital converters. It is shown that the main challenges are not only the high time resolution but often lie in the architectures that focus on other performance figures such as high dynamic range and linearity. I have tried to motivate that time-to-digital converters have the potential to become a trend-setting technology for ultimately scaled CMOS technologies. TDCs are much more than delay-lines and can be used for many more applications than for simple phase detectors. I am confident that this new class of circuits, namely continuous time continuous time circuits, will take up an important role in upcoming RF and mixed-signal applications, especially in nanometer scale technologies. Perhaps they will even find a role in the world of integrated systems that will emerge after the transistor era.

The new concepts and investigations presented in this book mainly originate from my research work at Infineon Technologies AG during the last years. I highly appreciate that I got the great opportunity to work in this interesting field and for the permission to use unpublished material on time-to-digital converters in this book for the first time. I also want to thank Dominik Lorenz for his valuable contributions during his master thesis and his outstanding commitment during the start up phase of our TDC projects. Many thanks go to Professor Ulf Schlichtmann and Professor Yannos Manoli who take so much time in reviewing my research work and advising me in many situations. My deep gratefulness goes to my mentor Professor Doris Schmitt-Landsiedel, head of the Institute for Technical Electronics, Technische Universität München. She always encourages me to open my mind to new challenges, concepts, and ideas. Without this mindset I would not have the drive to excel the limits of current structures and concepts. I am very proud of having the possibility to teach together with her in the micro-electronics track of Technische Universität München where I try to teach not only technical stuff but also the fun of innovation.

Munich
November 2009

Stephan Henzler

Chapter 2
Time-to-Digital Converter Basics

Abstract On the basis of a generic mixed-signal system the scaling difficulties of analog and mixed-signal circuits based on a signal representation in the voltage domain are discussed for nanometer CMOS technologies. Therewith, the advantages of a signal representation in the time domain are emphasized. The primary approach to time-to-digital converters (analog TDCs) based on a two step approach translating the time interval into a voltage and this voltage into a digital value is explained. Analog impairments and resolution limitations are examined. Counter based time interval measurement and delay-line based TDCs (digital TDCs) are introduced and analyzed with respect to operating principle, basic implementation issues, and quantization error, i.e. resolution.

Key words: Basic Time-to-Digital Converter, Analog TDC, Digital Delay-Line TDC, Mixed-Signal Systems

2.1 Motivation – The Way to the Time Domain

Time-to-digital converters (TDC) – certainly most engineers link this expression with all-digital phase-locked loops (PLL) where a TDC serves as phase detector [34, 51–53]. Interestingly, TDCs have been used for more than 20 years in the field of particle and high-energy physics, where precise time-interval measurement is required [27, 28, 32, 54]. Other applications cover time-of-flight measurement, or measurement and instrumentation applications such as digital scopes and logic analyzers. Currently the micro-electronics community rediscovers time-to-digital converters and this is the justification for a complete book on this topic. While the all digital PLL is the first and most famous TDC application others emerge rapidly. TDC based analog-to-digital converter for instance shows that TDCs are not just phase detectors but useful in a much wider field [8, 48]. Of course the requirements for any micro-electronics application differ from the applications mentioned above with respect to acceptable price, reproducibility, and suitability for mass production. These issues will be discussed in detail throughout the book. At this point

S. Henzler, *Time-to-Digital Converters*, Springer Series in Advanced Microelectronics 29,
DOI 10.1007/978-90-481-8628-0_2,

the motivating question shall be discussed why TDCs suddenly become popular in mainstream micro-electronics: Modern VLSI technology is mainly driven by digital circuits. The reasons for this are the many advantages of digital compared to analog circuits: Atomic digital functions can be realized by very small and simple circuits. This results in a compact and cheap implementation of elementary logic functions and enables complex and flexible signal processing systems. A comparable complexity was not feasible with an analog implementation due to area and power consumption but also due to variability and signal integrity. Flexible means reconfigurable, adjustable or even programmable. Data can be stored easily in digital systems without any loss of information. The design of digital circuits is highly automated resulting in high design efficiency and productivity. However, the main advantage of digital signal processing is the inherent robustness of digital signals against any disturbances, i.e. noise and coupling, as well as the inherent robustness of digital circuits against process variations. Both signal integrity and variability in digital circuits are heavily discussed in technical literature during the last years. It is true that these are critical issues especially for large chips fabricated in ultra deep sub-micron technologies. But compared to analog realizations digital solutions are still by far more robust. As a consequence of all these advantages most digital signal processing systems are realized according to the generic structure depicted in Fig. 2.1. A small mixed-signal shell provides the interface between the digital core and the environment which is always analog.[1] While the mixed-signal interface is mainly responsible for the data conversion, the actual signal processing task is performed in the digital domain. In the input path some basic analog signal conditioning, e.g. coarse filters, amplifiers and mixers are followed by a sampler and a quantizer, i.e. an ADC. The sampler evaluates the input signal at discrete time instances and the quantizer maps the resulting continuous values to discrete ones. Both together result in a digital signal representation. In the output path the digital values provided by the signal processing core at discrete time instances are converted in voltages/currents and held constant until the next value occurs. An optional analog low pass filter removes the (sinc-filtered) mirror spectra and provides smooth output signals.

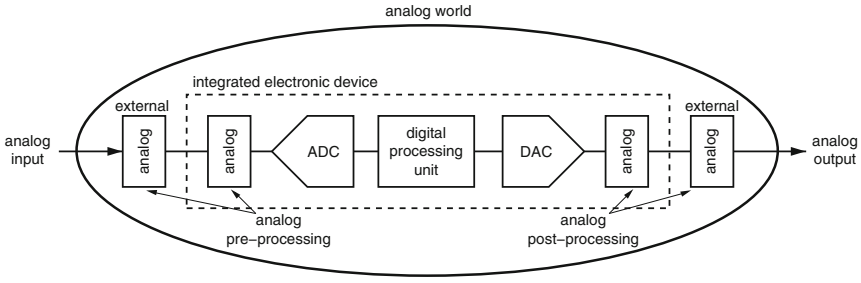

Fig. 2.1 Generic digital signal processing system comprising a digital core embedded in a mixed-signal shell for interfacing with the analog environment

[1] In fact even interfaces with digital functionality are analog. Digital signal transmission over high-speed serial links for instance poses lots of analog challenges, which become obvious on the lower abstraction levels.

But what has all of this to do with time-to-digital converters? What is the problem with this generic structure that has been successfully used for so many years? The difficulties arise from technology scaling in the ultra deep sub-micron regime: With each technology generation the intrinsic gain of a single MOS transistor namely the $\frac{g_m}{g_{ds}}$ decreases. This does not only result from parasitic short channel effects but is a fundamental result from MOS physics.[2]

In principle there are classical countermeasures to cope with decreasing transistor gain. Cascode transistors, for instance, increase the output resistance of basic amplifiers, current mirrors, and active loads dramatically, but cause at least one V_{Dsat} in the DC voltage budget. With the reduced supply voltages of scaled technologies there is often not enough voltage headroom to use cascodes. But even though this is possible, the speed of the circuit and the signal swing is reduced. The latter effect is particularly disadvantageous because of the reduced signal-to-noise ratio (SNR). In general technology scaling comes with a reduction in supply voltage and so reduced signal levels. As noise does not scale the SNR in the voltage domain is reduced in proportion to V_{DD}^2.

Altogether, the design of classical mixed-signal circuits becomes increasingly difficult. This rises two questions: First we have to check whether the mixed-signal content of modern signal processing systems is really reduced to a minimum. Second we have to answer why classical mixed-signal systems have that bad scaling behavior. In many systems mixed-signal circuits are just used for interfacing, so the first question can be approved. However, digital enhancement techniques within the mixed-signal building blocks, e.g. calibration techniques can help to improve the performance considerably. This means that the mixed-signal content of the generic system in Fig. 2.1 is minimized, but that the mixed-signal blocks themselves could be flavored with additional digital enhancement techniques.

The second question can be answered by considering the strengths of technology scaling. Although the main benefit of scaling in the deep sub-micron regime is the area reduction, there is still a speed improvement. This effect becomes smaller especially in low-power technologies, but continues to push digital performance. Despite increasing leakage currents [13] the power also scales to some extend. Good scaling behavior is thus achieved for all systems that take advantage of the fast digital switching speed. The reduction of the gate delay results in continuously improving temporal resolution which is contrary to the amplitude resolution [53]. Therewith, there is no principal restriction of mixed-signals blocks in aggressively scaled technologies but the problem is the signal representation in the voltage domain. An implementation of the same functionality in the time-domain would immediately take advantage of technology scaling again. The enabler for the time-domain processing of continuous signals is the time-to-digital converter. With this key building block a transformation of classical mixed-signal systems with a signal representation in the voltage domain is possible. This final rush for digitalization has just started.

[2] The channel length modulation coefficient $\lambda = \frac{1}{V_E L}$ is inverse proportional to the channel length L and the Early voltage V_E. The channel length scales considerably faster than V_E so the dependence of the drain current on the drain-to-source voltage increases with technology scaling [50].

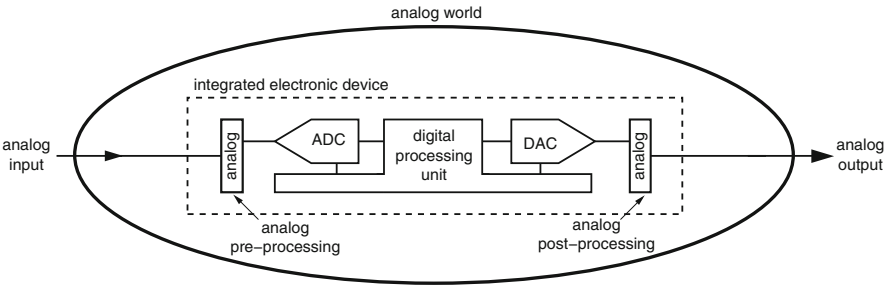

Fig. 2.2 Future digital signal processing system where the mixed-signal shell is heavily flavored with digital enhancement techniques that help to cope with the impairments of analog circuits in the deep sub-micron CMOS technologies

During the next years the amount of classical mixed-signal circuitry will be further reduced. The remaining mixed-signal components will be heavily supported by digital enhancement techniques and digital post-processing of imperfect data (ref. Fig. 2.2). Non-idealities due to analog impairments will be corrected by digital signal processing. For sure this is no digitalization in a VHDL or Verilog sense but a conversion of the continuous voltage domain into the continuous time domain. A substantiated expert knowledge in transistor level circuit design will still be the essential basis to cope with the increasing challenges in deep sub-micron CMOS technologies.

2.2 Analog Time-to-Digital Converters – The First Generation

The traditional approach to time-to-digital conversion is first to convert the time interval into a voltage. In a second step this voltage is digitized by a conventional analog-to-digital converter (ADC). A basic block diagram is given in Fig. 2.3. The start and the stop event are used to form a pulse with a width corresponding to the time interval to be measured. An analog integrator transforms this pulse into a voltage which is then fed to the ADC. A fundamental trade-off between the dynamic range DR, i.e. the maximum time interval to be measured and the maximum number of bits N the ADC can accomplish is revealed by the following equation:

$$DR = 2^N \cdot T_{LSB} \tag{2.1}$$

The minimum time interval that can be resolved is given by T_{LSB}. As the maximum resolution of the ADC is limited by analog constraints a long measurement interval means low resolution and vice versa. A high resolution measurement of long time intervals requires a two stage approach, i.e. a coarse quantization of the long time interval and a fine quantization of the remainder. Another approach is to integrate up and down periodically. The number of periods gives a coarse quantization and the

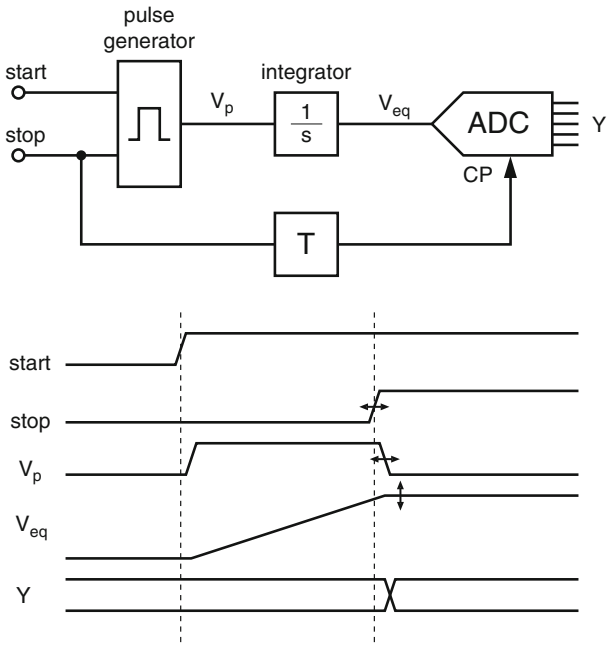

Fig. 2.3 Block and signal diagram of basic analog time-to-digital converter, e.g. [32, 42]

analog voltage level can be converted by the ADC to achieve a high resolution. We will see later in this book that the limitation given by eq. 2.1 does not exist for digital TDCs which is a great advantage over ADCs. There are also multi-stage approaches for digital TDCs (ref. Section 4.5) but the motivation behind this is the reduction of area and power.

Although the conversion principle in Fig. 2.3 is quite simple there are several analog issues that degrade the TDC performance: All building blocks, i.e. the pulse generator, the integrator, and the ADC have to meet the full linearity demands of the overall TDC. A basic integrator implementation that allows for high-speed, i.e. short measurement intervals, is a current source which is connected to an integration capacitance during the measurement interval. Due to the finite output resistance of the current source the linearity is weak. To overcome this problem an active RC-integrator may be used. As the potential at the virtual ground is nearly constant the linearity is improved. However, the finite bandwidth of the operational amplifier (opamp) limits the speed and so the minimum time interval considerably.

Absolute time measurement requires the knowledge of the current and the capacitance value. For integrated TDC implementations this is not feasible without calibration. A more elaborate TDC approach which does not require the knowledge of absolute device values is shown in Fig. 2.4. An exemplary signal diagram is given in Fig. 2.5. Again the pulse defined by the start and the stop signals is integrated. On the arrival of the stop signal a second integrator starts integrating but with a

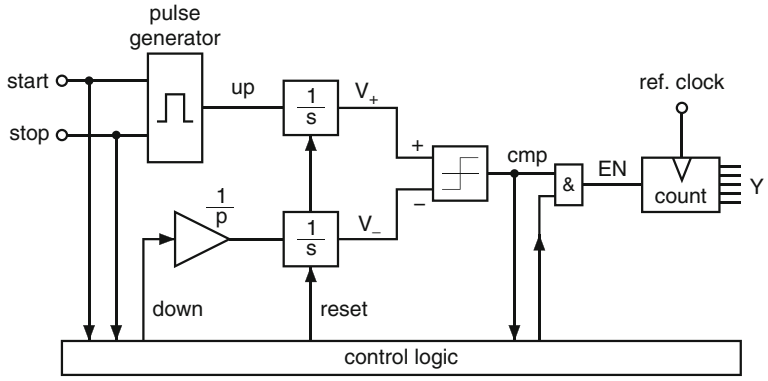

Fig. 2.4 Blockdiagram of analog TDC based on dual-slope analog-to-time-interpolation [43]

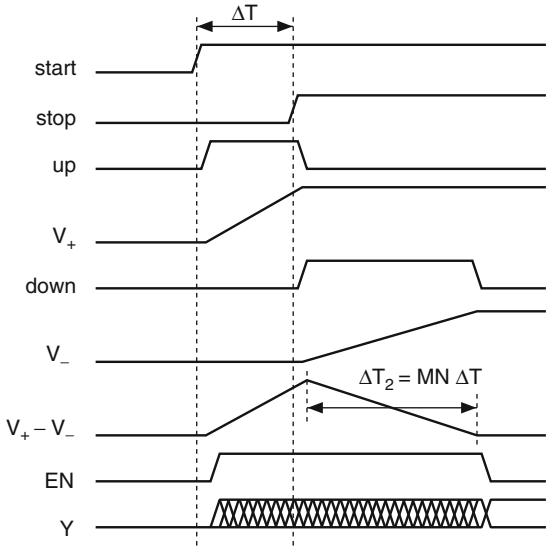

Fig. 2.5 Basic signal diagram of analog time-to-digital converter based on dual-slope analog-to-time interpolation. By means of pulse stretching the initial time interval ΔT is converted into a digital representation

reduced integration constant $\frac{1}{p}$. A comparator detects when the output of the second integrator is equal to the first one. This happens $(1+p)\Delta T$ seconds after the start event. The initial time interval ΔT is stretched by the factor $(1+p)$. Therefore this dual-slope approach is also known as time amplification. If p is large enough the enlarged time interval may be quantized by a simple digital counter.

In principle the dual-slope TDC is a reversed dual-slope analog-to-digital converter. In a dual-slope ADC an unknown analog signal is integrated for a well defined time. In the dual-slope TDC a well defined voltage namely the pulse height is

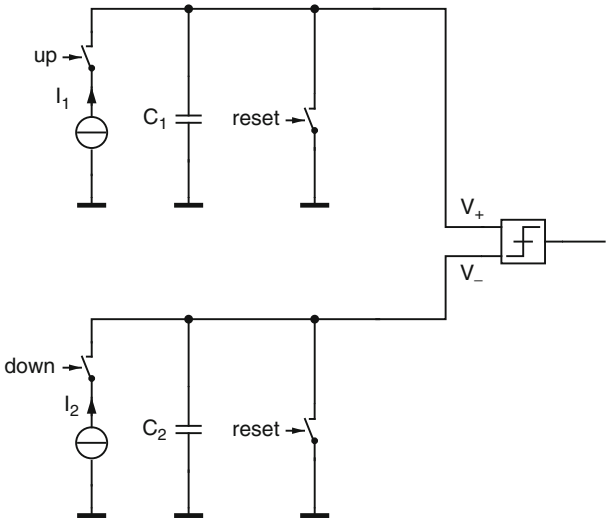

Fig. 2.6 Schematic circuit diagram of a basic dual-slope TDC [43]

integrated for the unknown time interval ΔT. In both cases the output voltage of the integrator after the first phase is integrated down by a predefined rate until a reference level is reached and a basic counter measures the required time. A major advantage of the dual-slope with respect to the basic single-slope ADC is the fact that absolute device values cancel out. The same should hold for the dual-slope TDC: To demonstrate this, the implementation in Fig. 2.6 is considered. At the beginning both capacitances C_1 and C_2 are discharged by the reset devices. During the time interval ΔT a first current source is connected to C_1. Assuming a perfect current source, the voltage V_+ increases linearly and reaches a final value given by $V_+ = \frac{I_1}{C_1}\Delta T$. On the arrival of the stop signal C_1 becomes floating and C_2 is connected to a second current source I_2. The voltage V_- across C_2 is given by $V_- = \frac{I_2}{C_2}\Delta T_2$ where ΔT_2 is the time elapsed since the stop event has occurred. The comparator connected to both capacitors detects when the two voltages V_+ and V_- are equal. This happens a time interval

$$\Delta T + \Delta T_2 = \left(1 + \frac{C_2}{C_1}\frac{I_1}{I_2}\right)\Delta T \tag{2.2}$$

after the start event. With a capacitance ratio $M := \frac{C_2}{C_1}$ and a current ratio $N^{-1} = \frac{I_2}{I_1}$ the initial time interval is stretched (amplified) by a factor $(1 + M \cdot N)$. As both factors M and N are ratios of physical quantities namely capacitances and currents the absolute values are not relevant. This makes the dual-slope approach very robust against process variations and enables an integration without the need for calibration. The fact that both ratios are multiplied is also advantageous for integration as large stretching factors can be realized even without unreasonably large capacitance or current ratios.

2.3 Fully Digital TDCs – The Second Generation

If the motivation for a TDC was just to implement a precise timer, an analog approach like one of those discussed in the previous section would be enough. However, the strategy to provide TDCs as generic mixed-signal building blocks for various applications rises questions about the suitability in ultimately scaled CMOS technologies. In Section 2.1 the advantages of time domain signal processing and the superior scaling properties of TDCs have been emphasized. Obviously this does not hold for any analog TDC which converts time domain information first into the analog and then to the digital domain. Such TDCs consist mainly of an ADC so have all the impairments of analog circuits in deep sub-micron technologies. The advantages of the time domain can be exploited only if there is no analog conversion step in the time-to-digital conversion. Only if the TDC is clearly dominated by digital circuitry the scaling and robustness arguments hold. Therefore, digital conversion techniques are investigated in the remainder of this book: The simplest technique to quantize a time interval is to count the cycles of a reference clock fitting into the respective measurement interval. As shown in Fig. 2.7 the measurement interval defined by the start and stop signal is completely asynchronous to the reference clock signal. This causes a measurement error ΔT_{start} at the beginning and ΔT_{stop} at the end of the time interval. The measurement interval ΔT can be expressed as

$$\begin{aligned} \Delta T &= N \cdot T_{CP} + (T_{CP} - \Delta T_{stop}) - (T_{CP} - \Delta T_{start}) \\ &= N \cdot T_{CP} - \Delta T_{stop} + \Delta T_{start} \\ &= N \cdot T_{CP} + \varepsilon_T \end{aligned} \tag{2.3}$$

$$\begin{aligned} \Delta T_{start} &\in [0; T_{CP}] \\ \Delta T_{stop} &\in [0; T_{CP}] \\ \varepsilon_T &= \Delta T_{start} - \Delta T_{stop} \in [-T_{CP}; T_{CP}] \end{aligned} \tag{2.4}$$

where N is the counter value and T_{CP} the reference clock period. ΔT_{start} and ΔT_{stop} are the time intervals between the start and the stop signal, respectively, and the next rising edge of the clock signal. The quantization error of the ΔT measurement is between $-T_{CP}$ and $+T_{CP}$, i.e. is limited to twice the period of the clock signal.

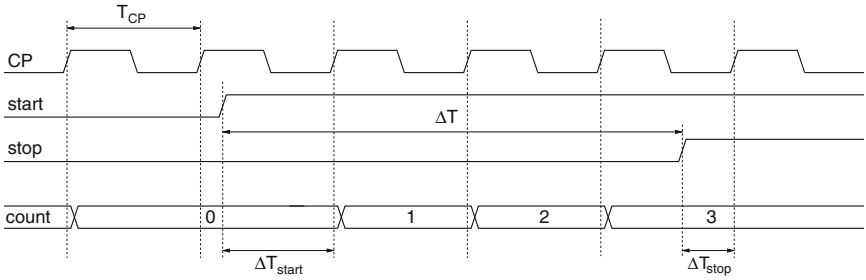

Fig. 2.7 Principle of counter based TDC

The measurement accuracy can be increased by a higher clock frequency. However, the higher the clock frequency the higher the power consumption for the generation and the processing of the clock signal. Above a certain clock frequency (corresponding to 3–4 FO2 inverter delays[3]) CMOS based oscillators are not available anymore, so even more expensive CML or LC oscillators are required for the clock generation. Timing restrictions in the counter as well as in the controller circuit pose another limit on the maximum frequency. In a 65 nm technology for instance the maximum frequency is limited to 5–10 GHz, i.e. a maximum measurement accuracy of $2 \cdot 100$–200 ps may be achieved.

An even higher resolution is achieved by subdividing one clock period asynchronously into smaller time intervals. The engine that performs this subdivision is actually what we call a digital time-to-digital converter. Hence, the resolution is the criterion that distinguishes a counter from a TDC. The measurement interval quantized with a TDC can be described by

$$\Delta T = NT_{CP} - (T_{CP} - \Delta T_{start}) + (T_{CP} - \Delta T_{stop}) \tag{2.5}$$

$$\Delta T_{start} = N_1 \frac{T_{CP}}{k} - \varepsilon_1 \tag{2.6}$$

$$\Delta T_{stop} = N_2 \frac{T_{CP}}{k} - \varepsilon_2 \qquad \varepsilon_1, \varepsilon_2 \in \left[0; T_{LSB} = \frac{T_{CP}}{k}\right] \tag{2.7}$$

$$\Delta T = NT_{CP} + N_1 \frac{T_{CP}}{k} - \varepsilon_1 - N_2 \frac{T_{CP}}{k} + \varepsilon_2 \tag{2.8}$$

$$\varepsilon_T = \varepsilon_2 - \varepsilon_1 \in \left[-\frac{T_{CP}}{k}; \frac{T_{CP}}{k}\right] \tag{2.9}$$

where the resolution $T_{LSB} = \frac{T_{CP}}{k}$ is increased by the factor k. The interpolation factor k describes in how many sub-intervals the reference clock cycle is partitioned, N_1 and N_2 indicate the position of the start and the stop event within such a reference clock cycle. The subsequent chapters describe the basic concepts of digital time-to-digital converters and explain how a higher resolution can be achieved even without a higher reference clock frequency.

2.4 Basic Digital Delay-Line Based TDC

To increase the measurement resolution beyond the maximum feasible clock frequency each counter clock cycle has to be sub-divided asynchronously by a time-to-digital converter. Figure 2.8 illustrates that the counter value then provides a coarse quantization of the measurement interval and the TDC a fine sub-quantization [37]. The subdivision of the reference clock interval, also known as reference clock

[3] FO2 refers to the fan-out of two inverter delay, i.e. to the delay of an inverter that is loaded with twice its input capacitance.

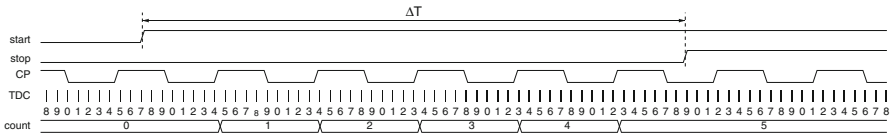

Fig. 2.8 Principle of a counter based TDC where the resolution is increased by subdividing one clock interval asynchronously into smaller time intervals [37]

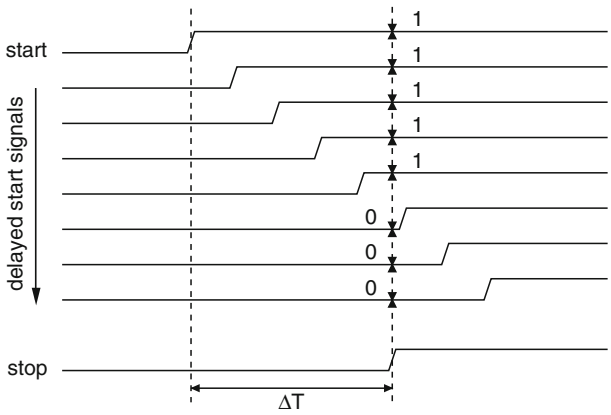

Fig. 2.9 Operating principle of a time-to-digital converter. Delayed versions of the start signal are sampled on the rising edge of the stop signal, resulting in a thermometer code at the outputs of the sampling elements (flip-flops or comparators)

interpolation, is done by using multiple phases of the reference clock. For a small interpolation factor, i.e. medium resolution, the phases may be directly derived from the oscillator which generates the reference clock. A ring oscillator consisting of k delay stages for instance generates k equally spaced versions of the clock signal. An even higher resolution is achieved by delaying the original reference clock in a chain of digital delay elements. The resolution then depends on the delay of the delay elements in the chain. Figure 2.9 illustrates the operating principle of a TDC based on a digital delay-line. The reference clock which is in a more general sense an arbitrary start signal is delayed along the delay-line. On the arrival of the stop signal the delayed versions start_i of the start signal are sampled in parallel. Either latches or flip-flops can be used as sampling elements. The sampling process freezes the state of the delay-line at the instance where the stop signal occurs. This results in a thermometer code because all delay stages which have been already passed by the start signal give a HIGH value at the outputs of the sampling elements, all delay stages which have not been passed by the start signal yet give a LOW value. The position of the HIGH-LOW transition in this thermometer code indicates how far the start signal could propagate during the time interval spanned by the start and the stop signal. Hence this transition is a measure for the time interval. The number N

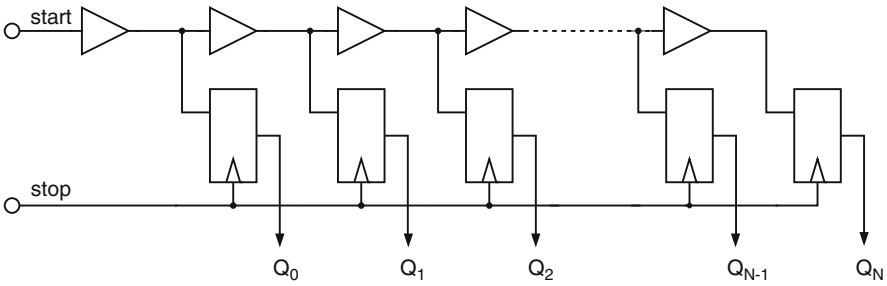

Fig. 2.10 Implementation of a basic delay-line based time-to-digital converter (DL-TDC)

of all sampling elements with a HIGH output is related to the measurement interval ΔT according to

$$N = \left\lfloor \frac{\Delta T}{T_{LSB}} \right\rfloor \tag{2.10}$$

where T_{LSB} is the delay of a single delay element in the delay-line. The time interval ΔT can be calculated from the number of HIGH outputs by

$$\Delta T = NT_{LSB} + \varepsilon \tag{2.11}$$

where ε describes the quantization error that arises as a delay element has been either passed by the start signal yet or not. Any intermediate state is not possible. The quantizer characteristic of a TDC and its non-idealities are discussed in Chapter 3 extensively.

An implementation of the basic delay-line TDC is shown in Fig. 2.10. The start signal ripples along a buffer chain that produces the delayed signals start_i. Flip-flops are connected to the outputs of the delay elements and sample the state of the delay-line on the rising edge of the stop signal. The stop signal drives a high number of flip-flops so a buffer-tree (not shown) is required. Any skew in this buffer-tree directly contributes to the non-linearity of the TDC characteristics. For a correct thermometer code the skew between adjacent branches in this tree has to be smaller than T_{LSB} which makes the design challenging.

2.4.1 Inverter Based Time-to-Digital Converter

The resolution of the delay-line based TDC discussed in the previous section is limited by the delay of the buffers. The resolution can be doubled by replacing the buffers by CMOS inverters. This, however, rises some implementation challenges which will be discussed next: The use of inverters means that both the rising and the falling signal transitions are used for measurement. Hence the thermometer code

at the outputs of the sampling elements becomes a pseudo thermometer code with alternating ones and zeros:

111111111111111 ⌋⌈ 000000000000000	buffer TDC
010101010101010 ⌋⌈ 010101010101010	inverter TDC

The length of the measurement interval is indicated not by a HIGH-LOW transition but by a phase change of the alternation HIGH-LOW sequence.

In principle the delays for a rising and a falling transition of a CMOS inverter are different and only partially correlated. There is some correlation due to common process steps during manufacturing of NMOS and PMOS devices. Examples for such process steps are the formation of the gate oxide and the gate lithography. However, there are also completely independent process steps such as the ion implantation for threshold voltage adjustment. This leads to systematic non-linearity of the converter characteristic. For nominal process conditions the rise and the fall delay can be made equal but any process variation imbalances the delays again. An even stronger effect has the asymmetric setup time of basic sampling elements such as master slave latch pairs. Hence, a high-resolution TDC based on a single inverter delay chain seems to be not feasible if process variations become significant. Figure 2.11 shows a robust inverter based TDC. Fully symmetrical differential flip-flops such as sense amplifier based flip-flops are used as sampling elements. Two delay chains propagate the start as well as the inverted start signal and provide differential data to the flip-flops. The inverting characteristics of the CMOS inverters is compensated by twisting the input signals of the flip-flops in each second stage. This compensates completely for the potential asymmetric setup time of the flip-flops and asymmetric rise and fall delays of the inverters. Local process variations may cause a faster signal propagation in one delay-line compared to the other. Coupling elements such as cross coupled inverter pairs in between two corresponding

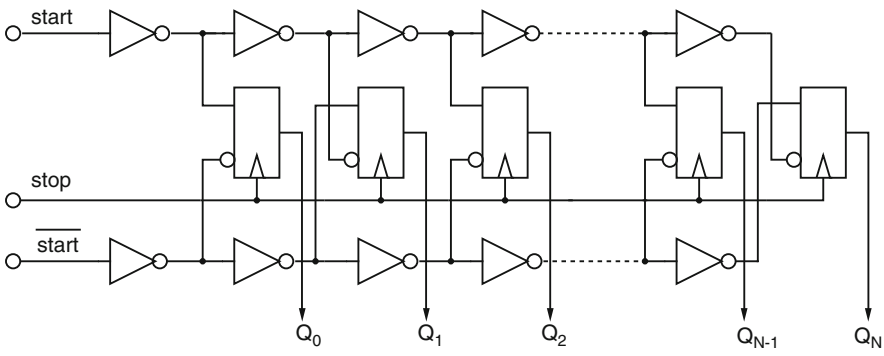

Fig. 2.11 Time-to-digital converter based on inverters instead of buffers. The resolution is doubled with respect to the implementation given in Fig. 2.10

Table 2.1 Performance summary of buffer and inverter delay-line TDC. The performance figures are introduced in the following chapter. However, for comparability with the results presented in Chapter 4 this table is already presented here

	Buffer Delay-Line TDC	Inverter Delay-Line TDC
Principle	Start signal propagates in linear chain of buffers. On the arrival of a stop signal the state of the delay-line is sampled by flip-flops/comparators	Differential start signal propagates in two coupled chains of inverters. On the arrival of a stop signal the state of the delay-line is sampled by (differential) flip-flops/comparators
Resolution T_{LSB}	$t_d^{buffer} = 2t_d^{inv}$	t_d^{inv}
Number of Stages N	$\frac{T_{max}}{T_{LSB}} = \frac{T_{max}}{t_d^{buffer}} = \frac{T_{max}}{2t_d^{inv}}$	$\frac{T_{max}}{T_{LSB}} = \frac{T_{max}}{t_d^{inv}}$
Core Area $A_{core}^{Vernier}$	$\frac{T_{max}}{T_{LSB}}\left(2A^{inv} + A^{FF}\right)$	$\frac{T_{max}}{T_{LSB}}\left(2.2A^{inv} + A^{FF}\right)$
Average Power $\langle P_{vernier}^{core} \rangle$	$f_{meas}\frac{T_{max}}{T_{LSB}}\left(2E_{rise}^{inv} + 2E_{fall}^{inv} + E^{FF}\right)$	$f_{meas}\frac{T_{max}}{T_{LSB}}\left(2.2E_{rise}^{inv} + 2.2E_{fall}^{inv} + E^{FF}\right)$ The factor 2.2 accounts for the overhead required to couple the two delay-lines (differential)
Conversion Time T_{conv}	T	T
Latency $T_{latency}$	0	0
Loop Structure	loop possible	loop possible
PROS	• Simple • Fully digital • Low power • Low latency • Easy control and embedding	• Same as for buffer TDC • Doubled resolution
CONS	• Low resolution • Resolution limited by technology	• Alignment of delay-lines • Resolution limited by technology • Doubled number of comparators for same dynamic range

delay elements reduces this drifting of the signals. However, there is still a very regular and symmetrical layout required for the two delay-lines. This is probably the most challenging task for an actual TDC implementation (Table 2.1).

2.5 Synchronous Versus Asynchronous Time Interval Measurement

Synchronous time interval measurement based on a coarse pre-quantization by a reference clock and a fine quantization with a TDC (ref. Fig. 2.8) has the fundamental drawback that the reference clock jitter deteriorates the measurement accuracy. This can be seen by rewriting eqs. 2.5–2.8 with a jitter term:

$$\Delta T = NT_{CP} - (T_{CP} - \Delta T_{start}) + (T_{CP} - \Delta T_{stop}) \tag{2.12}$$

$$\Delta T_{start} = N_1 \frac{T_{CP}}{k} - \varepsilon_1 + t_{jitter,1} \tag{2.13}$$

$$\Delta T_{stop} = N_2 \frac{T_{CP}}{k} - \varepsilon_2 + t_{jitter,2} \qquad \varepsilon_1, \varepsilon_2 \in \left[0; T_{LSB} = \frac{T_{CP}}{k}\right] \tag{2.14}$$

$$\Delta T = NT_{CP} + N_1 \frac{T_{CP}}{k} - \varepsilon_1 + t_{jitter,1} - N_2 \frac{T_{CP}}{k} + \varepsilon_2 - t_{jitter,2} \tag{2.15}$$

The jitter is a mean free not necessarily gaussian random variable characterized by its rms value σ_{jitter}. It can be seen that two partly independent jitter contributions add to the measurement uncertainty statistically. For short and medium time intervals it may be advantageous to abdicate the reference clock and to use a longer TDC. Therewith the measurement error caused from reference clock jitter vanishes. The start and stop signals for the TDC are directly extracted from the measurement pulse and not referred to any clock (asynchronous time interval measurement). The TDC has to be able to measure the complete time interval not just a reference clock period, i.e. the delay chain has to be long enough or an advanced TDC architecture like the looped TDC (ref. Section 4.2) has to be used. Of course there is also noise in the TDC which translates into a deviation of the switching instances of the delay elements from the ideal time instances and a variation of the sampling instance. The resulting measurement uncertainty increases with $\sqrt{T}$ where T is the length of the measurement interval. Hence for long measurement intervals the intrinsic timing uncertainty of the TDC may be larger than the reference clock jitter. For an actual design and a required measurement time the intrinsic TDC noise and the quality of a possible reference clock generator must be compared. Thereupon it can be decided whether a synchronous or asynchronous approach is preferable. Beside the noise the availability of the clock signal is another criterion as an additional clock PLL contributes considerably to the overall power consumption and die area.

Chapter 3
Theory of TDC Operation

Abstract This chapter addresses theoretical aspects of time-to-digital converters. First the basic shape of a TDC input–output characteristic is explained. The quantization error that arises from the mapping of a continuous into a discrete signal domain is revisited. Linear and non-linear imperfections of a TDC, namely offset and gain error as well as differential and integral non-linearities, are explained. Dynamic performance figures are motivated on the basis of analog-to-digital converters. The difference between ADC and TDC measurement is discussed. The interrelation between the single tone experiment of ADCs and the single shot experiment for TDCs is derived. An effective number of bits is defined for TDCs based on a single shot measurement. Therewith, an ADC compliant figure of merit is defined for TDCs. Local (process) variations are a critical issue for high resolution time-to-digital converters. Variations hit TDCs in the buffer tree of the stop signal, the delay-line, and the comparators. The susceptibility of each of these components is analyzed individually. Finally, the particular variation effects are combined.

Key words: TDC Characteristic, TDC Offset Error, TDC Gain Error, TDC Non-Linearity, Single-Shot-Precision, TDC Effective Number of Bits, TDC Process Variations

3.1 Basic Performance Figures

The static input–output behavior of a TDC is given by a quantizer characteristic such as that depicted in Fig. 3.1. The input time interval to be measured is plotted on the x-axis and the corresponding digital output word is represented by the y-axis. The term quantizer characteristic means that continuous time intervals at the TDC input are mapped to discrete output values. Consequently, there is a range of time intervals that is mapped to the same output value. The (ideal)[1] width of this time

[1] In actual TDC designs the step width varies from step to step. This non-linearity is described later in this chapter. A more precise definition for T_{LSB} will be given in eq. 3.4.

S. Henzler, *Time-to-Digital Converters*, Springer Series in Advanced Microelectronics 29,
DOI 10.1007/978-90-481-8628-0_3,

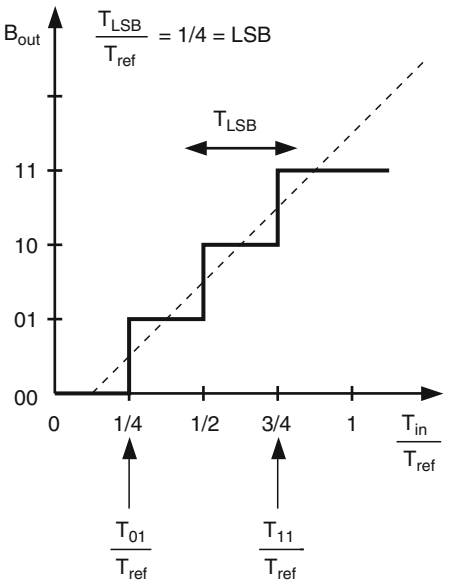

Fig. 3.1 Ideal input–output characteristic of time-to-digital converter

interval is the absolute resolution T_{LSB} and the corresponding output increment is called $1\,LSB$. The time-to-digital conversion is not invertible and the input–output behavior

$$T_{in} = B_{out} T_{LSB} + \varepsilon \qquad 0 \leq \varepsilon < T_{LSB} \tag{3.1}$$

can be described only by means of the so called quantization error ε. In the ideal case the unknown quantization error behaves like an equally distributed, uncorrelated random signal and contributes only to the noise floor of the measurement. This is only true, however, if the input signal changes fast enough and strong enough so that the quantization error appears to be de-correlated from the signal. For periodic operation, this is not the case and the quantization error which is actually not noise but deterministically dependent on the signal gives rise to harmonic distortion .

The time intervals where the steps occur are labeled by T_B where B is the quantization level after the step, i.e. $T_{0\ldots01}$ indicates the position of the very first and $B_{1\ldots11}$ of the last step. These step positions will be used throughout this chapter to define the various non-idealities such as offset, gain error, or non-linearity.

In principle, a TDC characteristic is identical with the characteristic of an analog-to-digital converter (ADC), except the fact that a continuous time and not a continuous voltage is measured. In ADCs, this voltage is referred to a reference voltage. The reference voltage V_{ref} is the voltage against which the input voltage is compared and indicates usually the maximum input level that can be measured without saturation effects. Usually the reference voltage is externally provided and the ADC derives the quantization voltage V_{LSB} from this reference voltage according to

$$V_{LSB} = \frac{V_{ref}}{2^N} \tag{3.2}$$

where N is the number of bits. In a TDC, however, not a reference time interval T_{ref} but a quantization time T_{LSB} is given. A corresponding reference time can be defined by

$$T_{ref} = 2^N \cdot T_{LSB} \tag{3.3}$$

The problem with this definition is the fact that, theoretically, a TDC can measure an arbitrarily long time interval. In the looped TDC that is discussed in Section 4.2, for instance, the timing event to be measured circulates in the TDC delay-line and a loop counter determines the number of full loop cycles. The number of bits in eq. 3.3 then depends on the number of bits in the loop counter and the reference time T_{ref} has no useful meaning. Yet, the loop structure means that the converter characteristic is constructed periodically by the characteristic of a single loop cycle. The time increment corresponding to this periodicity is then a much better definition for the reference time. The number of bits corresponding to this reference time is called the core resolution[2] N_c, whereas the overall resolution is named N. With this T_{ref} definition the definition of the quantization time interval according to

$$T_{LSB} = \frac{T_{ref}}{2^{N_c}} \tag{3.4}$$

makes sense again. It is important to understand that the resolution according to this formula is the ideal resolution, i.e. the best value that can be achieved with the underlying converter architecture. In reality impairments such as non-linearity and noise degrade the effective resolution that is actually measured. This will be discussed after the introduction of the various imperfections.

A further difference compared to ADCs is the fact that the steps in a TDC characteristic usually occur at the end of a quantization interval whereas for ADC characteristic the steps often lie in the middle of an interval. This is a minor difference that changes the definition of the TDC performance figures slightly.

3.2 Quantization Error Revisited

Converting a continuous into a discrete signal results in a quantization error ε according to eq. 3.1. In contrast to ADCs where the quantization error is usually symmetrical around zero, i.e. $-\frac{1}{2}V_{LSB} \leq \varepsilon < \frac{1}{2}V_{LSB}$, the quantization error of a TDC is not mean free ($0 \leq \varepsilon < T_{LSB}$). An equally distributed quantization error has the mean value

$$\langle \varepsilon \rangle = \frac{1}{T_{LSB}} \int_0^{T_{LSB}} \varepsilon d\varepsilon = \frac{1}{2} T_{LSB} \tag{3.5}$$

The quantization noise power

$$\langle \varepsilon^2 \rangle = \frac{1}{T_{LSB}} \int_0^{T_{LSB}} \varepsilon^2 d\varepsilon = \frac{1}{3} T_{LSB}^2 \tag{3.6}$$

[2] In a controlled TDC the reference time according to this definition corresponds usually to the time reference T_R of the control loop (ref. Section 4.4).

can be separated into a constant and a mean free variable component according to

$$\varepsilon = \langle\varepsilon\rangle + \eta \tag{3.7}$$

$$\langle\varepsilon^2\rangle = \langle\varepsilon\rangle^2 + \langle\eta^2\rangle \tag{3.8}$$

with

$$\langle\varepsilon\rangle^2 = \frac{1}{4}T_{LSB}^2 \tag{3.9}$$

$$\langle\eta^2\rangle = \frac{1}{12}T_{LSB}^2 \tag{3.10}$$

In the frequency domain the constant component $\langle\varepsilon\rangle$ yields a DC spectral line. The variable component results in spectral power with $f > 0$. This component of the quantization error is often assumed as white, i.e. its power spectral density is flat. This is justified for non-periodic signals that change sufficiently fast. If this condition is fulfilled, the power spectral density can be described by

$$psd\,(\varepsilon) = \frac{1}{4}T_{LSB}^2\delta(f) + \frac{T_{LSB}^2}{12f_s}\left[\sigma\left(f+\frac{f_s}{2}\right) - \sigma\left(f-\frac{f_s}{2}\right)\right] \tag{3.11}$$

For a sinusoidal signal $x(t) = A\sin(2\pi ft)$ with 2^{M-1} quantization steps for the signal amplitude[3] the ideal signal-to-(quantization)-noise ratio (S(Q)NR) is given by

$$SNR = 6.02\text{ dB}\cdot M + 1.76\text{ dB} \tag{3.12}$$

The DC component of the quantization error has been neglected. As for the ADC each additional bit in the amplitude domain increases the signal-to-noise ratio by 6.02 dB. This is the ideal signal-to-noise ratio as only the quantization but no physical noise is considered. The actual SNR is always lower than this theoretical value.

3.2.1 Linear Imperfections of TDC Characteristic

In an ideal TDC the first step occurs at the position $T_{00\ldots01} = T_{LSB}$. If this step and so the complete converter characteristic is shifted along the time axis, the converter is said to have an offset error E_{offset}, defined by

$$E_{offset} = \frac{T_{00\ldots01} - T_{LSB}}{T_{LSB}} \tag{3.13}$$

i.e., the offset error (ref. Fig. 3.2) is the deviation of the first step position from its ideal value T_{LSB} in terms of T_{LSB}. This is not necessarily the same as the output

[3] This definition sounds a little bit artificial. However, unlike the input voltage of ADCs the maximum input time interval of TDCs is not limited so the reference to a sinusoidal input signal with maximum amplitude is not possible.

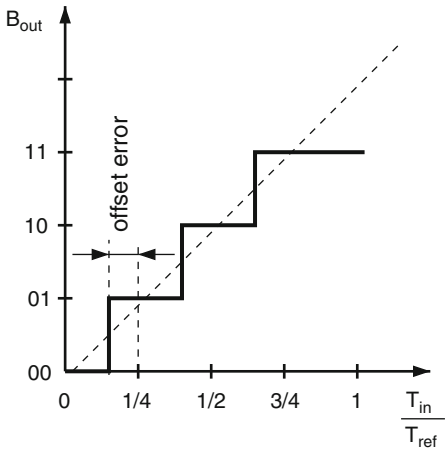

Fig. 3.2 Input–output characteristic of a TDC with offset error

value $B_{out}(T = 0)$ that is measured for a zero time interval. The reason for a mismatch between these two figures is a potential non-linearity as discussed in the next section. In contrast to an ADC, the input time interval is not bound in principle, so the definition of the TDC gain is somehow ambiguous. Generally, the gain k_{TDC} of a TDC is the steepness of the input–output characteristic, or mathematically,

$$k_{TDC} = \frac{\Delta B}{\Delta T} \tag{3.14}$$

Depending on the definition of the deltas, several gain definitions are possible: Using the whole converter characteristic yields the overall gain k_{TDC} of the TDC. In looped TDCs, i.e. for converter characteristics with a periodicity, it makes sense to refer to this periodicity and to call the respective gain the core gain k_{TDC}^{core}. Differences of these two gain definitions arise if the periodic parts do not fit together perfectly. If non-linearity is noticeable it makes sense to define the gain as the slope of a regression line. This is possible for both gain definitions given above. If only a small fraction of the TDC characteristic is used during operation, an incremental gain definition k_{TDC}^{incr} around an operating point is suited best. It depends on the application which gain definition is the most appropriate one. An ideal TDC has a gain of

$$k_{TDC} = \frac{1}{T_{LSB}} \tag{3.15}$$

i.e. a T_{LSB} increment of the input time interval reflects in an output increment of one LSB. Any variation of the gain may be quantified by the gain error E_{gain}, which is the deviation of the last step position from its ideal value in terms of LSB after the offset error has been removed (ref. Fig. 3.3):

$$E_{gain} = \frac{1}{T_{LSB}} \left(T_{11\ldots11} - T_{00\ldots01}\right) - \left(2^N - 2\right) \tag{3.16}$$

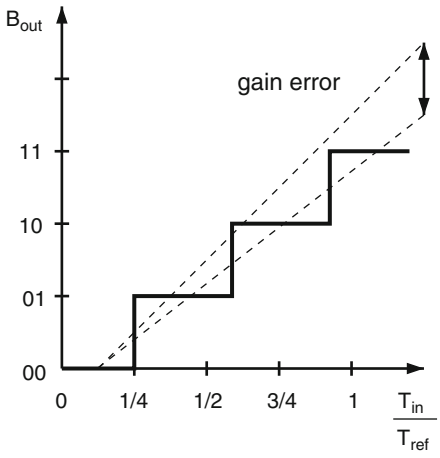

Fig. 3.3 Input–output characteristic of a TDC with gain error

Again, similar definitions may be used if only a certain part of the converter characteristic is of interest. The actual TDC gain can be expressed in terms of the gain error according to

$$k_{TDC} = \frac{2^N - 2}{(2^N - 2)\,T_{LSB} + E_{gain} T_{LSB}} \approx \frac{1}{T_{LSB}} \left(1 - \frac{E_{gain}}{2^N - 2}\right) \tag{3.17}$$

Finally, it shall be mentioned that the offset and the gain error are so called linear imperfections as they can be modeled by an additive or multiplicative term to the input time interval, respectively, so do not cause non-linear distortion.[4]

3.3 Non-Linear Imperfections of TDC Characteristic

Non-linear imperfections are all deviations of a TDC characteristic from its expected shape that lead to non-linear distortion. The integral non-linearity (INL) is a macroscopic description of the bending of a converter characteristic. It is defined as the deviation of the step position from its ideal value normalized to one T_{LSB}. The ideal value is defined by a straight line that can be either a line connecting the first and the last step or a best fit line. The former is a pragmatic approach, whereas the latter causes a little bit more computation effort but yields better insight especially for converters with the main non-linearity at the beginning and/or the end of the characteristic. The INL is defined for each step. However, sometimes a single number is given as INL value. This is then the maximum or root-mean-square (rms)

[4] Quantization is always a very non-linear operation. In this context, the term linear means that the output is linearly dependent on the input when the quantization error is neglected.

value over all steps. The differential non-linearity (DNL) describes the deviation of each step from its ideal value, namely T_{LSB}, so gives a more microscopic view on the non-linearity. Both the INL and the DNL are usually normalized to one T_{LSB}.

The terms 'integral' and 'differential' may suggest that the INL is obtained by summing up the DNL. However, if the T_{LSB} is the quantization time interval of the ideal converter, the DNL contains both gain and non-linearity error. This can be easily seen by considering a case where each step has a width of let's say $0.5 \cdot T_{LSB}$. The error with respect to a reference line is zero, i.e. there is no integral non-linearity, but the slope of this line differs by a factor of two from the ideal characteristic.

3.4 Dynamic Performance and Effective Resolution

The performance figures discussed so far result from so called static measurements. This means usually that the measurements are done slowly, without active peripheral circuits. For converter characteristics multiple measurements are averaged. This averaging means that any noise (physical noise, power supply noise, substrate noise) and any dynamic coupling effects are removed from the measurement results. In real operation where any single measurement counts and where many measurements are done subsequently noise cannot be neglected and reduces the effective resolution. In this section dynamic performance figures of conventional analog-to-digital converters are explained, transferred to time-to-digital converters, and compared.

The classic dynamic measurement of ADCs is the stimulation with a sinusoidal signal and the evaluation of the output spectrum. Therefore, output data is stored for multiple cycles of the input signal and transferred into the frequency domain via Fourier transformation. The resulting spectrum is a so called single tone spectrum. It contains a peak at the signal frequency, smaller peaks at multiples of the signal frequency caused by non-linear distortion, and a noise floor. The spectral components at multiples of the signal frequency are called harmonics. The noise floor describes power components at all other frequencies. It results from both quantization as well as physical noise. The signal-to-noise ratio is defined as the ratio between the power of the fundamental over the total noise power. Usually it is given in dB, i.e.

$$SNR = 10\ \mathrm{dB} \cdot lg\left(\frac{P_{signal}}{P_{noise}}\right) \tag{3.18}$$

The harmonic components are neglected during the calculation of the noise power. The signal-to-noise-and-distortion ratio (*SNDR*) defined by

$$SNDR = 10\ \mathrm{dB} \cdot lg\left(\frac{P_{signal}}{P_{noise} + P_{harmonics}}\right) \tag{3.19}$$

takes also the harmonics into account. Thus, the *SNDR* is always smaller than the *SNR*. Obviously, *SNR* and *SNDR* are functions of the signal amplitude. This dependence is usually demonstrated by a diagram where the signal amplitude is plotted

along the x-axis (logarithmic scale) and the *SNR* and *SNDR*, respectively, along the y-axis. For small amplitudes the *SNR* and *SNDR* curve are identical. This is due to the fact that harmonic components scale with A^n, where A is the signal amplitude and n the order of the respective harmonic. For larger signals the harmonics grow quickly, and the *SNDR* curve bends below the *SNR* curve. For a certain amplitude, the so called peak amplitude, the *SNR* reaches a maximum value. If the amplitude is increased beyond the peak amplitude both *SNR* and *SNDR* collapse rapidly. The peak *SNDR* is the value used for the computation of the effective resolution of the converter. According to eq. 3.12 the effective number of bits (*ENOB*) can be defined by

$$ENOB := \frac{max\,(SNDR) - 1.76\ \text{dB}}{6.02\ \text{dB}} \tag{3.20}$$

ENOB is the resolution of a hypothetical noise free converter that has the same *SNDR* as the converter under consideration. The dynamic range is defined as the input voltage range between the amplitude where the *SNR* is zero and full scale. Usually it is given in dB. It is important to understand that often this signal range cannot be used as the *SNR* already collapses below full-scale. For a more detailed description of dynamic ADC figures the reader is referred to classical mixed-signal literature, e.g. [45]. An excellent overview of data converter figures is given in [47].

In principle the same characterization strategy can be applied to TDCs. The problem, however, is to generate a sinusoidal sequence of time intervals with an accuracy better than the TDC to be measured. Thus, the classic dynamic measurement of a TDC is not a single tone experiment but the so called single shot experiment. In this experiment a fixed time interval T is applied repeatedly to the time-to-digital converter. Without noise each measurement would yield the same result. Noise, however, causes variations of the measurement values. The standard deviation of these measurement values is called single-shot precision (SSP). It describes how reproducible a TDC measurement is in the presence of noise. The single-shot precision usually depends on the time interval, i.e. it is actually not a single value but a function $SSP(T)$ of the time interval T. The reason for this becomes obvious when the topology of a TDC is considered: The delay along the delay-line is the accumulated gate delay of the delay elements. Each delay element has a certain delay variation, so contributes to a timing uncertainty. The longer the time interval the more delay elements are passed and contribute to the overall timing uncertainty. The sampling elements experience a noise induced threshold shift. The timing uncertainty related to this effect is independent on the position inside the delay-line and so on the time interval. If the comparators were neglected and if the delay variations of the delay elements were uncorrelated the timing uncertainty would increase along the delay-line in proportion to $\sqrt{T}$. However, in the advanced TDC architectures that will be discussed in Chapter 4 this assumption is often not true. In the looped-TDC, for instance, the delay elements are reused several times, i.e. there is a correlation of the individual noise contributions and the single shot precision scales nearly linearly.

The dominant noise effect in MOS transistors is flicker noise. Flicker noise describes a charge trapping effect at the interface between the channel region and the gate insulator. It results in a temporary shift of the transistor threshold voltage which

in turn results in an altered gate delay. The process is relatively slow and has a power spectral density that is proportional to $\frac{1}{f}$ (at least for low frequencies). Hence, the measurement values during the single-shot experiment do not vary completely randomly as the single-shot definition might suggest. In fact there is a correlation that is revealed when the auto-correlation function (ACF) and the power spectral density are considered. In many applications the near term variation of the measurement results is more relevant than the long term variation. Hence the ACF provides useful additional information about TDC noise.

To understand the dependencies for the measurement of different time intervals a double-shot experiment can be used. In this experiment two time intervals T_1 and T_2 are measured alternately. For both series a single-shot precision and an auto-correlation can be given. Additionally, a cross-correlation can be computed that indicates whether noise affects the measurements of both time intervals completely independently or in a correlated fashion. In a linear delay-line for instance the delay variations of the first delay elements affects the arrival time at all subsequent nodes. Hence, there is a correlation of the arrival times for all measurements during which the common delay elements are passed. The smaller the difference $|T_1 - T_2|$ the more common delay-elements contribute to the measurement and the stronger the correlation. In looped TDCs (ref. Section 4.2) the number of delay elements is small, and all delay elements are passed by the start signal repeatedly. Hence the correlation is much stronger. The cross correlation function reveals this dependency so may help to identify structural details of an unknown TDC circuit.

The definition of a TDC dynamic range is not as straight forward as for ADCs. This is because of the fact that the SNR cannot be measured easily, so the signal amplitude where the SNR is zero cannot be identified. Hence, in this book the dynamic range is defined as the maximum time interval that can be measured without saturation. It can be given as absolute time or with respect to T_{LSB} in dB. Obviously, the relative specification of the dynamic range is only useful when the resolution T_{LSB} is also given. The lack of a simple SNR measurement also complicates the definition of the effective number of bits. A definition that does not take the correlation into account is derived in the following section.

3.4.1 Basic ENOB Definition

As discussed in the previous section the single tone experiment and so the calculation of the signal-to-noise ratio and the effective number of bits is not straight forward for TDCs.[5] The standard methodology for dynamic TDC characterization is single shot experiment. For the sake of simplicity it is assumed during the following calculation that the single shot precision does not depend on the measurement time, i.e. $SSP(T) = SSP = const$. Without noise and for equally distributed time

[5] Such a measurement methodology would be highly desirable but, to the best knowledge of the author, it is not known till the publication date of this book.

intervals the quantization error is assumed to be equally distributed in the interval[6] $\left[-\frac{1}{2}T_{LSB};\frac{1}{2}T_{LSB}\right]$. Noise causes an effective variation of the time interval to be measured. The measurement result is then represented by neighboring steps of the converter characteristic so the quantization error is increased. The delay deviation τ of a nominal time interval T is modeled by a Gaussian distribution[7]:

$$p_\tau(\tau) = \frac{1}{\sqrt{2\pi}\sigma_\tau} exp\left(-\frac{\tau^2}{2\sigma_\tau^2}\right) \tag{3.21}$$

If k is the nominal measurement result for a noise free measurement of an interval T, $T - kT_{LSB}$ describes the position within a quantization interval. The joint probability density function for a nominal position $T - kT_{LSB}$ inside a quantization interval in conjunction with a noise induced time interval deviation τ is

$$p_{T-kT_{LSB},\tau}(T - kT_{LSB}, \tau) = \frac{1}{T_{LSB}} \cdot p_\tau(\tau) = \frac{1}{T_{LSB}} \cdot \frac{1}{\sqrt{2\pi}\sigma_\tau} exp\left(-\frac{\tau^2}{2\sigma_\tau^2}\right) \tag{3.22}$$

The quantization error depends on the quantization step, i.e. on the actual measurement result for the noisy signal. It can be described by

$$\varepsilon(T - kT_{LSB}, \tau) = \begin{cases} \cdots & \cdots \\ T - kT_{LSB} + 2T_{LSB} & -\frac{5}{2}T_{LSB} \le T - kT_{LSB} + \tau < -\frac{3}{2}T_{LSB} \\ T - kT_{LSB} + T_{LSB} & -\frac{3}{2}T_{LSB} \le T - kT_{LSB} + \tau < -\frac{1}{2}T_{LSB} \\ T - kT_{LSB} & -\frac{1}{2}T_{LSB} \le T - kT_{LSB} + \tau < \frac{1}{2}T_{LSB} \\ T - kT_{LSB} - T_{LSB} & \frac{1}{2}T_{LSB} \le T - kT_{LSB} + \tau < \frac{3}{2}T_{LSB} \\ T - kT_{LSB} - 2T_{LSB} & \frac{3}{2}T_{LSB} \le T - kT_{LSB} + \tau < \frac{5}{2}T_{LSB} \\ \cdots & \cdots \end{cases} \tag{3.23}$$

Therewith, the quantization noise power under physical noise conditions can be computed according to

$$\left\langle \varepsilon^2 \right\rangle = \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \int_{-\infty}^{\infty} \varepsilon^2(T - kT_{LSB}, \tau) \cdot p_{T-kT_{LSB},\tau}(T - kT_{LSB}, \tau)\, d\tau\, d(T - kT_{LSB}) \tag{3.24}$$

The integration can be partitioned into the respective quantization intervals:

$$\left\langle \varepsilon^2 \right\rangle = \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \sum_{n=-\infty}^{\infty} \int_{\left(n-\frac{1}{2}\right)T_{LSB}-T'}^{\left(n+\frac{1}{2}\right)T_{LSB}-T'} \left[T' - nT_{LSB}\right]^2 \frac{1}{T_{LSB}} \frac{1}{\sqrt{2\pi}\sigma_\tau} exp\left(-\frac{\tau^2}{2\sigma_\tau^2}\right) d\tau\, dT'$$

[6] For most applications the DC error is irrelevant thus for the sake of simplicity a symmetrical quantization error has been chosen during this calculation.

[7] In some respects a Gaussian timing uncertainty is problematic. It is chosen here anyhow as it can be easily handled during analytic calculations. The limitations of this model are discussed in Section 3.7.

$$= \frac{1}{\sqrt{2\pi}\sigma_\tau T_{LSB}} \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \sum_{n=-\infty}^{\infty} \left[T' - nT_{LSB}\right]^2 \int_{\left(n-\frac{1}{2}\right)T_{LSB}-T'}^{\left(n+\frac{1}{2}\right)T_{LSB}-T'} exp\left(-\frac{\tau^2}{2\sigma_\tau^2}\right) d\tau\, dT'$$

$$= \frac{1}{T_{LSB}} \sum_{n=-\infty}^{\infty} \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \left[T' - nT_{LSB}\right]^2 \left[\frac{1}{2} erf\left(\frac{\left(n+\frac{1}{2}\right)T_{LSB} - T'}{\sqrt{2}\sigma_\tau}\right) - \frac{1}{2} erf\left(\frac{\left(n-\frac{1}{2}\right)T_{LSB} - T'}{\sqrt{2}\sigma_\tau}\right)\right] dT' \tag{3.25}$$

where the substitution $T' = T - kT_{LSB}$ has been used. With a quantization interval dependent probability density function $p_n(T')$ defined by

$$p_n(T') = \frac{1}{2} erf\left(\frac{\left(n+\frac{1}{2}\right)T_{LSB} - T'}{\sqrt{2}\sigma_\tau}\right) - \frac{1}{2} erf\left(\frac{\left(n-\frac{1}{2}\right)T_{LSB} - T'}{\sqrt{2}\sigma_\tau}\right) \tag{3.26}$$

the quantization noise power in the presence of physical noise can be expressed by

$$\left\langle \varepsilon^2 \right\rangle = \frac{1}{T_{LSB}} \sum_{n=-\infty}^{\infty} \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \left[T' - nT_{LSB}\right]^2 p_n(T') dT' \tag{3.27}$$

For a nominal quantization error T' the probability density function $p_n(T')$ describes the probability that the measurement result $k+n$ is obtained. Figure 3.4 shows the pdf of the quantization error for various values of the timing uncertainty σ_τ. If σ_τ converges to zero, the quantization error is confined to the interval $\left[-\frac{1}{2}T_{LSB}; \frac{1}{2}T_{LSB}\right]$. For increasing timing uncertainty the quantization measurement result is more and more pushed to neighboring quantization intervals so the probability for larger quantization errors increases.

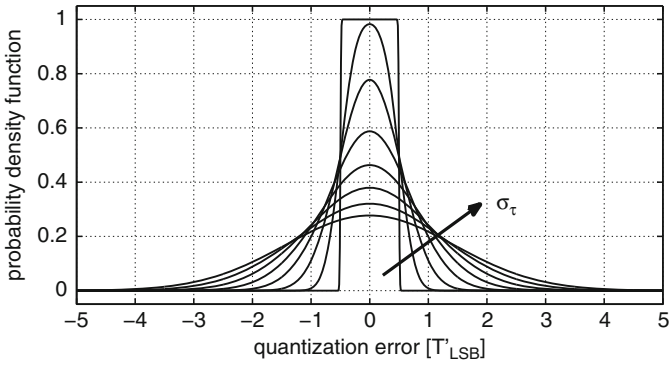

Fig. 3.4 Probability density function of quantization error in presence of physical noise for increasing timing uncertainty σ_τ

Without physical noise the quantization noise power is given by eq. 3.10. This ideal value is separated from the noise power given in eq. 3.27 according to

$$\left\langle \varepsilon^2 \right\rangle = \frac{T_{LSB}^2}{12} \frac{12}{T_{LSB}^3} \sum_{n=-\infty}^{\infty} \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \left[T' - nT_{LSB}\right]^2 p_n(T')dT' = \frac{(T_{LSB} \cdot RDC)^2}{12} \tag{3.28}$$

Therewith, the resolution degradation coefficient *RDC*

$$RDC := \sqrt{\frac{12}{T_{LSB}^3} \sum_{n=-\infty}^{\infty} \int_{-\frac{1}{2}T_{LSB}}^{\frac{1}{2}T_{LSB}} \left[T' - nT_{LSB}\right]^2 p_n(T')dT'} \tag{3.29}$$

describes the increase of T_{LSB} due to physical noise. The effective number of bits of the TDC can then be defined as

$$ENOB_{TDC} := log_2\left(\frac{DR}{T_{LSB} \cdot RDC}\right) = M - log_2\left(RDC\right) \tag{3.30}$$

where M is the ideal resolution for the noise free case. With this *ENOB* definition the classical ADC figures of merit for area and power can be defined:

$$FOM_P = \frac{\langle P \rangle}{f \cdot 2^{ENOB}} \tag{3.31}$$

$$FOM_A = \frac{\langle A \rangle}{f \cdot 2^{ENOB}} \tag{3.32}$$

Finally, the formalism derived above can be used to compute the single-shot precision. For any position T' within a quantization interval, the average measurement result is

$$\left\langle T_q(T') \right\rangle = \sum_{n=-\infty}^{\infty} n p_n(T') \tag{3.33}$$

and its standard deviation, i.e. the single-shot precision is given by

$$SSP(T') = \sum_{n=-\infty}^{\infty} \left[n - \left\langle T_q(T') \right\rangle\right]^2 p_n(T') \tag{3.34}$$

The single-shot precision according to this formula is plotted in Fig. 3.5 once for a nominal time interval that is positioned in the middle of a quantization interval and once for a position at the border. For a timing uncertainty larger than half a LSB there is no difference anymore and the single-shot precision increases quite linearly with the timing uncertainty. This means that the single-shot precision describes the timing uncertainty within the TDC pretty well.

It shall be mentioned that all derivations of this section do not take any correlations between the measurement values into account. Hence, the computed ENOB is

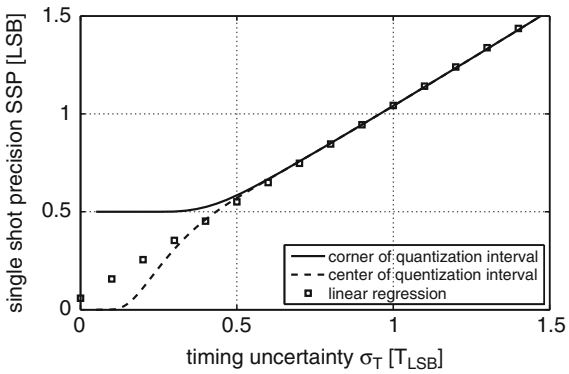

Fig. 3.5 Single-shot precision in dependence of the timing uncertainty inside the delay-line

a lower bound. Applications that can benefit from a correlation might achieve better values. Furthermore the formalism above can be extended for each single step of the TDC characteristic. Therewith, non-linear distortion is also considered by this ENOB definition.

3.5 Timing Figures

Obviously, the measurement of a time interval takes some time. The period between the start event and the availability of the measurement result is described by the conversion time T_{conv}. If the stop event is taken as reference the delay is called latency $T_{latency}$. Most TDCs need some time after a measurement before a new one can be started. This is described by the dead time T_{dead}. The minimum and maximum time interval that can be measured are described by T_{min} and T_{max}.

3.6 Noise Shaping in Time-to-Digital Converters

So far, only TDCs have been investigated where the start signal initiates a timing event that then propagates along the delay-line. Now it shall be assumed that a timing event is continuously propagating along a delay-line and that a measurement is instantaneously started when the previous measurement ends. From a discrete standpoint this means that both the start as well as the stop event are quantized. There are multiple scenarios where this may happen. One example is a looped TDC (ref. Section 4.2) with a timing event that is not injected on the arrival of a start signal but continuously circulating and which is always sampled on the arrival of a stop signal. Another example is a very long delay-line that may contain more than one measurement interval. Finally, the gated ring oscillator described in section 5.6

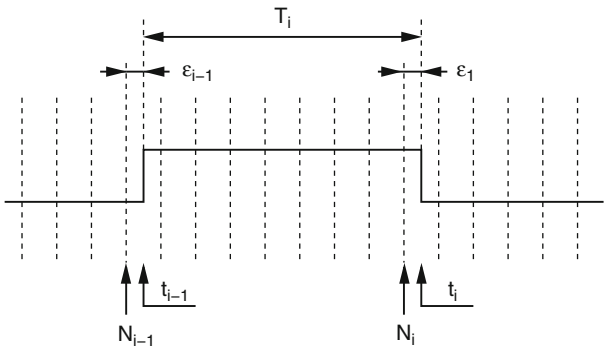

Fig. 3.6 Time interval measurement with quantized start and stop events

falls into this class of TDCs. In this context Fig. 3.6 shows a signal that carries pulse width encoded information. Let's assume that one measurement, namely the measurement of a low phase ends at the time t_{i-1}. The next measurement, namely the measurement of the subsequent high phase, starts at this time instance t_{i-1}. It ends at the time t_i that is also the start point for the next measurement and so on. Due to the discretization in the TDC not the times t_{i-1} and t_i but the quantized values $N_{i-1}T_{LSB}$ and N_iT_{LSB} are measured:

$$t_{i-1} = N_{i-1}T_{LSB} + \varepsilon_{i-1} \tag{3.35}$$

$$t_i = N_iT_{LSB} + \varepsilon_i \tag{3.36}$$

The measurement interval is thus represented by

$$(N_i - N_{i-1})\,T_{LSB} = T_i - (\varepsilon_i - \varepsilon_{i-1}) \tag{3.37}$$

i.e. the measured interval is equal to the actual time interval T_i plus a quantization error

$$\varepsilon = -(\varepsilon_i - \varepsilon_{i-1}) \tag{3.38}$$

This quantization error is the difference of two quantization errors that arise from delay-line sampling. Compared to a conventional TDC the quantization error has statistically doubled but is first order noise shaped. The noise shaping reduces low frequency noise components and shifts the quantization noise power towards higher frequencies. For some amount of oversampling the noise power falling into the signal band is thus reduced with respect to conventional approaches.

Physical noise affects both the propagation of timing events in the delay-line as well as the sampling process. The question is whether this noise component is also subject to noise shaping. To investigate this, a time interval with quantized start and stop signal is depicted in Fig. 3.7. The black curve describes the signal that would be measured without physical noise. The gray curve describes the same signals with all timing events shifted due to noise. The original switching instances t_{i-1} and t_i can be described in dependence on the measurement values N'_{i-1} and the N'_i,

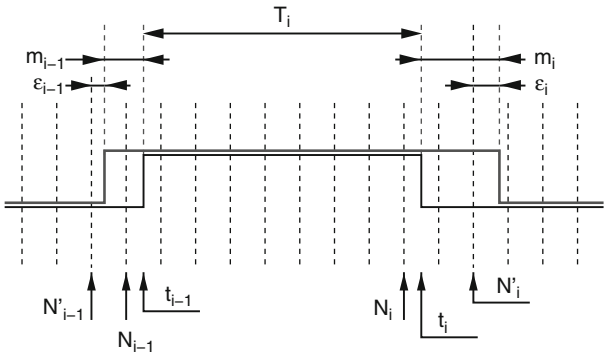

Fig. 3.7 Time interval measurement with quantized start and stop event in the presence of physical noise

the quantization errors ε_{i-1} and ε_i, and the noise induced delay shifts m_{i-1} and m_i according to

$$t_{i-1} = N'_{i-1} T_{LSB} + \varepsilon_{i-1} - m_{i-1} \tag{3.39}$$

$$t_i = N'_i T_{LSB} + \varepsilon_i - m_i \tag{3.40}$$

The measured time interval

$$\left(N'_i - N'_{i-1}\right) T_{LSB} = T_i - (\varepsilon_i - \varepsilon_{i-1}) + (m_i - m_{i-1}) \tag{3.41}$$

shows that both the quantization noise as well as the physical noise are first order noise shaped.

This feature, however, is only advantageous for low frequency components, i.e. correlated noise. If the noise component m_i is given by m_{i-1} plus another noise component the latter one is directly visible at the output without noise shaping. Yet, this happens in the delay line of a TDC: Assume that each delay element i contributes to a certain noise error η_i according to

$$t_{d,i} = T_{LSB} + \eta_i \tag{3.42}$$

These noise errors accumulate along the delay-line statistically. In a certain stage n the cumulative time shift is

$$m_n = \sum_{i=1}^{n} \eta_i \tag{3.43}$$

Consequently, the noise induced measurement error becomes

$$m_i - m_{i-1} = \sum_{i=1}^{N_i} \eta_i - \sum_{i=1}^{N_{i-1}} \eta_i = \sum_{i=N_{i-1}}^{N_i} \eta_i \tag{3.44}$$

This means that all noise that is picked up between the time instances spanning the measurement interval directly contributes to the measurement uncertainty. Thus, noise shaping is advantageous for quantization noise as well as correlated noise such as the measurement uncertainty arising from noise in the buffer tree of the stop signal. Noise in the delay chain is not subject to noise shaping. It depends on the measurement time which noise component dominates, i.e. whether a noise shaping TDC is advantageous or not.

3.7 Process Variations in TDCs

Process and environmental variations influence the behavior of each integrated circuit thus also the performance of time-to-digital converters. Global process variations affect the respective devices on the die in the same way. Some of them jointly affect NMOS and PMOS devices, others have only an impact on a single transistor type. Variations that affect both transistor types are for example variations in the oxide thickness or variations in the exposure time that leads to length and width variations. Variations of the dose of ion implantation for threshold voltage adjustment for instance affect only NMOS or PMOS devices respectively. Environmental operating parameters like the supply voltage or the temperature have also a global impact on the performance, at least for small circuit blocks like time-to-digital converters. On the other hand local process variations (LPV) affect each single device individually. The main reason for local variations are random dopant fluctuation and line edge roughness. Both phenomena are noticeable especially in heavily scaled technologies where quantization effects become visible: If the threshold voltage of a device is set by a small number of dopant atoms, the absence or presence of a single atom has a noticeable effect on the overall device performance. The reason for line edge roughness is the finite grain size in the transistors' poly-silicon gates. For large devices the relative impact on the transistor length and width is negligible but very small devices experience relatively large variations. In fact there is also something in between global and local variations, namely process or environmental variations that change across the die. These gradients become relevant on very large chips but may be neglected for small macros such as TDCs.

In a time-to-digital converter global variations alter the gate delays in the delay-line and the buffer tree of the stop signal. In the comparators the blackout time, i.e. the sampling instance is shifted. Functionally this affects the TDC offset, the gain, and the resolution. As long as the quantization error is low enough, even for slow process conditions, the latter effect is of minor importance. The offset error appears as a DC error that is not critical in most applications. The TDC gain, however, describes the change of the output word per change of the input time interval, i.e. global process variations can be modeled as a gain element cascaded with the TDC. This has a critical impact, e.g. on control loops because the TDC gain influences the loop dynamic. In TDC based ADCs the gain variation of the TDC affects the absolute output value. Consequently the TDC gain has to be either controlled or

measured for further calibration. To control the gain of a TDC the delay along the complete delay-line has to be measured and compared to time reference. Depending on the result the delay-elements can be configured either for smaller or larger delay until the delay matches the reference. Beyond the fact that a tunable delay-element is more complex than a basic delay-cell and hence more area and power intensive the, resolution is reduced because of the tuning headroom. Moreover, fine-grained delay tuning is a step back on the way towards the digital domain. Implementation details of such a delay-locked loop (DLL) gain calibration are discussed in Section 4.4.

Another possibility to cope with TDC gain variations is to use digital calibration. Therefore, the TDC is dynamically characterized by measuring a known reference time interval T_R. In a linear TDC the input time interval T and the output word B are related according to

$$B = \lfloor kT \rfloor + B_{offset} \tag{3.45}$$

where k is the TDC gain and B_{offset} the TDC output voltage for $T = 0$. This equation reveals that a single reference measurement is not enough as two unknowns (k and B_{offset}) have to be determined. Depending on the architecture of the TDC under investigation several calibration strategies are possible. One is to use two reference time intervals. Another strategy is to first measure the offset error by applying the same signal to the start and the stop terminal of the TDC. The gain is then determined in a second step by measuring a reference time interval. For process variations a one time TDC characterization e.g. during production test would be sufficient. However, digital delays are strongly dependent on the operating conditions, namely voltage and temperature, so the calibration has to be repeated regularly, e.g. during start-up or idle periods. It is interesting that some applications like the all-digital PLL (ADPLL) already work with normalized measurement values: In the ADPLL architecture proposed in [34, 51, 53] for instance, the TDC is used to interpolate one local oscillator (LO) period. This means that not the absolute time interval but the fraction of this time interval with respect to another time interval (the LO period in this case) is of interest. This is plausible as a PLL works on the phase error between the LO and the reference signal and not on the absolute skew between these signals. If the offset error B_{offset} is acquired in a first measurement, e.g. by applying the same signal to the TDC, the relative time interval measurement is described by

$$\frac{T_1}{T_2} = \frac{B_1 - B_{offset}}{B_2 - B_{offset}} \tag{3.46}$$

(neglecting quantization) and the TDC gain cancels out. This means that relative time interval measurements are somehow self-calibrated and a TDC characterization (except for the offset error) is not required. This holds particularly for TDC architectures that are inherently offset compensated, such as the bipolar TDC described in Section 4.1 or the continuously running TDC described in Section 3.6.

Local process variations hit the TDC at three points: At the buffer tree of the stop signal, the delay elements in the delay-line and the sampling elements. In contrast to synchronous logic where local variations are averaged along combinatorial

paths, the variation of every single cell (delay element, sampling element) directly affects the TDC converter characteristic and so the linearity. Hence, time-to-digital converters are particularly susceptible to local variations and their effect has to be investigated carefully.

3.7.1 Impact of Local Variations in Buffer Tree

The basic idea of a TDC is to sample the outputs of all delay elements at the same time. As the stop signal is distributed by a buffer tree the arrival time of the stop signal at the comparators may depend on the particular branch of the buffer tree. One reason for this arrival time uncertainty is the deterministic skew in a buffer tree which is not perfectly balanced. Another reason are local process variations that cause skew even in a perfectly balanced tree. The example in Fig. 3.8 shows two groups of delay elements together with their sampling flip-flops. A hypothetical skew T is assumed that can be either positive or negative. The resulting TDC characteristics are illustrated in Fig. 3.9. If the second group of delay elements is sampled later than the first group, i.e. if $T > 0$, the step position T_{100} representing the interface as well as all later step positions are shifted left by the time T. For a skew larger than one T_{LSB} this results not only in a DNL error but also in missing codes. If the second group is sampled earlier, i.e. if $T < 0$, all step positions related to the second group are shifted to the right. This causes an increased step width at the handover point of the two groups. In a real TDC the buffer tree usually has multiple levels of hierarchy and multiple branches. This means that the arrival time uncertainty of the respective branches is partially correlated and the prediction of variation effects becomes a little bit more complicated. The good news is, however, that the buffers in the tree use large devices, so local variations are relatively small. A design countermeasure to avoid skew is the use of a clock mesh[8] on the last hierarchy level(s) of the buffer tree. This increases the routing complexity and the power consumption but makes the skew essentially zero.

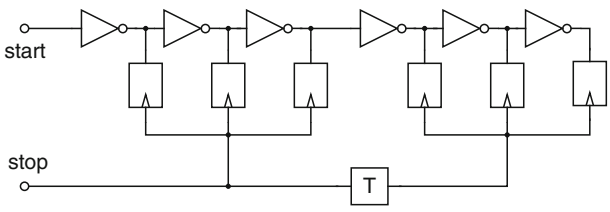

Fig. 3.8 Skew in buffer tree, each buffer level can cause skew in its associated branches

[8] The end points of all branches of a certain hierarchy level are shorted.

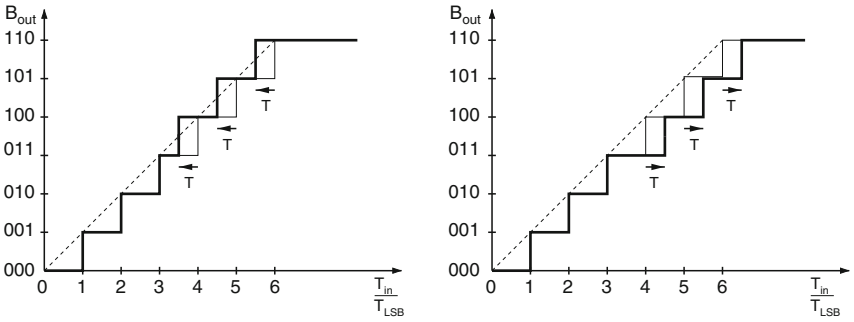

Fig. 3.9 Impact of clock skew on exemplary TDC depicted in Fig. 3.8 for the right branch sampled later (left) and earlier (right) than the left branch.

3.7.2 Impact of Local Process Variations on Delay-Line

Local process variations cause a modification of each gate delay along the delay-line that can be modeled by

$$T_{d,i} = T_{LSB} + \varepsilon_i \tag{3.47}$$

where ε_i is the delay error that can be positive or negative. In simulations and analytic calculations, ε_i is often modeled as a mean free Gaussian process. Although this model is easy to implement, it has some limitations arising from the tails in the pdf: If ε_i became smaller than $-T_{LSB}$, which was possible with a certain probability, the gate delay $t_{d,i}$ would become negative. This is not compliant with the causal switching sequence in the delay-line. Another explanation that ε_i cannot be Gaussian distributed is the fact that the basic device parameters affected by variations (such as the threshold voltage) are modeled by Gaussian processes. The gate delay, however, is not a linear function of these parameters, so its pdf cannot be Gaussian.[9] For the following discussion there are no special requirements for the distribution of ε_i except that it is mean free.

Assuming an ideal buffer tree and ideal comparators results in the switching times , i.e. the step positions of the TDC characteristic, given by

$$t_n = n \cdot T_{LSB} + \sum_{i=1}^{n} \varepsilon_i \tag{3.48}$$

If the buffer delays independently vary with a standard deviation $std(\varepsilon_i) = std(\varepsilon)$ the arrival time uncertainty after the nth delay element is characterized by the square root law

$$std\,(t_n) = std\,(\varepsilon) \cdot \sqrt{n} \tag{3.49}$$

[9] Anyhow, for the sake of simplicity Gaussian distributions are widely used for analytic investigations. This is usually acceptable but one has to be aware of the limitations.

The differential non-linearity can be calculated according to

$$DNL_n = \frac{t_{n+1} - t_n - T_{LSB}}{T_{LSB}} = \frac{\varepsilon_{n+1}}{T_{LSB}} \tag{3.50}$$

and is directly given by the delay variation of the respective delay element in terms of T_{LSB}. Along the delay-line variations sum up and contribute to the gain error E_{gain} and the integral non-linearity. The arrival time at the end of a delay-line consisting of N delay elements is

$$t_N = N \cdot T_{LSB} + \sum_{i=1}^{N} \varepsilon_i \tag{3.51}$$

Therewith the gain error

$$E_{gain} = \frac{1}{T_{LSB}} [t_N - t_1] - [N-1] = \sum_{i=2}^{N} \varepsilon_i \tag{3.52}$$

can be calculated. The reference step positions t'_n for the calculation of the integral non-linearity (INL) are

$$t'_n = t_1 + \frac{n-1}{N-1}(t_N - t_1) \tag{3.53}$$

The INL is then given by

$$\begin{aligned} INL_n &= \frac{1}{T_{LSB}} \left(t_n - t'_n\right) \\ &= \frac{1}{T_{LSB}} \left(\sum_{i=2}^{n} \varepsilon_i \left[1 - \frac{n-1}{N-1}\right] - \frac{n-1}{N-1} \sum_{i=n+1}^{N} \varepsilon_i \right) \end{aligned} \tag{3.54}$$

To understand the meaning of this formula, it is instructive to calculate the standard deviation of the integral non-linearity:

$$std\,(INL_n) = \frac{std(\varepsilon)}{T_{LSB}} \sqrt{(n-1)\frac{N-n}{N-1}} \tag{3.55}$$

Its maximum

$$\max_n (std(INL_n)) = \frac{std(\varepsilon)}{2T_{LSB}} \sqrt{N-1} \tag{3.56}$$

occurs in the middle of the delay-line, more precisely at the position

$$n = \frac{1}{2}(N+1) \tag{3.57}$$

This means that the arrival time uncertainty grows with the length of the delay line according to a square root of N law (eq. 3.49). Figure 3.10 shows this increasing uncertainty along the delay-line. The gain error resulting from local process variations

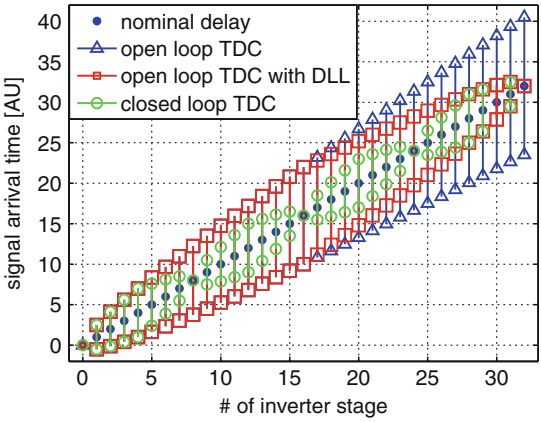

Fig. 3.10 Arrival time uncertainty in linear TDC before and after calibration

can be calibrated together with the gain error caused by global process variations. After this calibration the signal arrival time at the input of the first delay element and at the output of the last delay element is fixed. This gain compensation reduces the maximum arrival time uncertainty by a factor $\sqrt{2}$. The remaining variation contributes only to the INL described by eqs. 3.54, 3.55, and 3.56 which grows in proportion to $\sqrt{N}$. Hence, with or without calibration a good linearity can be achieved only if the length of the delay-line is as short as possible. This gives rise to the looped TDC that is discussed in Section 4.2. Another important finding is the fact, that the arrival time uncertainties are strongly correlated since the switching along the chain is causal. This means that the delay variation of any delay element contributes to the arrival time uncertainty of each delay element placed after it.

3.7.3 Impact of Local Process Variations on the Comparators

The susceptibility of the comparators on local process variations strongly depends on their actual implementation. Qualitatively, LPV shift the trigger threshold of each comparator so can be modeled by a voltage source in series to the respective data input. A threshold shift of ΔV is translated into a shift of the sample time ΔT according to

$$\Delta T = \frac{1}{m}\Delta V \tag{3.58}$$

where m is the slope of the comparator's input signal, i.e. of the signals in the delay-line. Hence, the impact of LPVs in the comparators can be reduced by decreasing the rise-time of the signals in the delay-line. This partly compensates for the effect of increasing local variations in scaled CMOS technologies as the gate delays and the signal transition times are scaled as well.

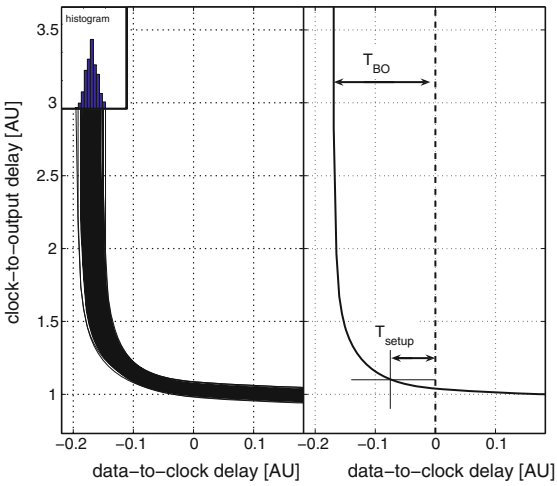

Fig. 3.11 Meta-stability curve of comparator with or without local process variations

Figure 3.11 illustrates the sampling process in a comparator/flip-flop with and without local process variations. The right graph shows the meta-stability curve without variations. The setup time is given according to the conventional 10% criterion. In synchronous digital circuits the setup time describes the minimum data-to-clock delay that is possible in order to guarantee a certain clock-to-output delay. In time-to-digital converters, however, not the delay of the comparator is of interest but rather the question whether the comparator switches or not. The data-to-clock delay for which the comparator barely samples a input signal is the so called blackout time of the comparator. The comparator delay under blackout conditions is much higher than the delay for relaxed setup conditions. If local variations in the comparator are discussed, the variation of the blackout time has to be considered. This is shown in the left graph of Fig. 3.11 where multiple realizations of the meta-stability curve are overlaid.

For densely spaced signals with overlapping switching intervals variations in the comparators may cause bubbles in the thermometer code. Large bubble induced errors can be avoided by a digital bubble correction logic inserted between the comparators and the thermometer-to-binary decoder.

3.7.4 Combined Impact of Local Variations on TDCs

In this section the effects of all types of local variations shall be considered together. The derivation is done for a basic delay-line TDC. Later it is adjusted to the various sub-gate delay resolution concepts (in the respective sections). For the sake of simplicity it is assumed that there is no latency along the stop tree. The stop signal

is considered as a global signal, so its variation causes an offset error but no local non-linearity. The arrival time uncertainty of the stop signal is described by ε_s. The blackout time is set to zero, but its local variations are considered. The start signal shall be injected into the delay-line at the time $t = 0$. The stop signal occurs at the time $t = T$. The arrival times of the start signal in two consecutive delay stages is given by

$$t_n = nT_{LSB} + \sum_{i=1}^{n} \varepsilon_i \tag{3.59}$$

$$t_{n+1} = (n+1)T_{LSB} + \sum_{i=1}^{n+1} \varepsilon_i \tag{3.60}$$

The stop signal arrives at the time

$$t_s = T + \varepsilon_s \tag{3.61}$$

and the blackout times of the respective flip-flops/comparators vary according to

$$t_n^{BO} = \beta_n \tag{3.62}$$

$$t_{n+1}^{BO} = \beta_{n+1} \tag{3.63}$$

In the nth stage a logic high signal is sampled if

$$t_s - t_n > t_n^{BO} \tag{3.64}$$

$$\Leftrightarrow T > t_{s,n} := nT_{LSB} + \sum_{i=1}^{n} \varepsilon_i + \beta_n - \varepsilon_s \tag{3.65}$$

A similar equation holds for $(n+1)$th stage, namely

$$t_s - t_{n+1} > t_{n+1}^{BO} \tag{3.66}$$

$$\Leftrightarrow T > t_{s,n+1} := (n+1)T_{LSB} + \sum_{i=1}^{n+1} \varepsilon_i + \beta_{n+1} - \varepsilon_s \tag{3.67}$$

These two equations can be used to calculate the differential non-linearity of the nth stage:

$$DNL_n = \frac{1}{T_{LSB}} (t_{s,n+1} - t_{s,n} - T_{LSB}) \tag{3.68}$$

$$= \frac{1}{T_{LSB}} (\varepsilon_{n+1} + \beta_{n+1} - \beta_n) \tag{3.69}$$

If this equation is compared to eq. 3.50 where only variations in the delay-line have been considered, it becomes obvious that now also the blackout time variations of both sampling flip-flops come into play. As a global stop signal is used the variation

in the stop tree cancel out.[10] In summary, the actual DNL varies with a standard deviation given by

$$\mathrm{std}(DNL_n) = \frac{1}{T_{LSB}} \sqrt{\mathrm{std}^2(\varepsilon) + 2\mathrm{std}^2(\beta)} \tag{3.70}$$

Finally, the probability for a basic bubble shall be computed. This is the event that in the nth stage a logic low signal is sampled, whereas a logic high signal is sampled in the $(n+1)$th stage. The probability of this event is maximized if the $(n+1)$th stage barely samples a logic high signal, i.e. if $T = t_{s,n+1}$. This condition can be inserted into the complementary relation of eq. 3.65. The resulting relation for the occurrence of a bubble error is thus

$$\varepsilon_{n+1} + \beta_{n+1} - \beta_n < -T_{LSB} \tag{3.71}$$

which corresponds to a negative DNL error larger than minus one LSB. The consequence of this bubble depends on the system and a potential bubble correction logic. A reasonable criterion for a robust TDC may be the demand for a differential non-linearity that is smaller than one LSB with a three sigma confidence, i.e.

$$3 \cdot std(DNL) < 1 \tag{3.72}$$

This criterion will be also used for the TDCs with sub-gate delay resolution discussed in Chapter 5.

[10] This is true for a clock mesh. The calculation with individual stop signals is done accordingly.

Chapter 4
Advanced TDC Design Issues

Abstract In this chapter architectural aspects of time-to-digital converters are discussed in detail. The design challenges are a high dynamic range, small offset and gain error, high linearity, a small die area, and finally a low power consumption. The time resolution is independently addressed in Chapter 5. Several architectures are proposed that particularly focus on at least one of the design challenges. A bipolar TDC allows for signed measurement without offset error. A loop architecture enables long measurement times with reasonable area consumption. A linear loop extension improves the linearity impairments of the basic loop architecture. For long measurement times a hierarchical TDC system can further reduce the power consumption. For complex measurement tasks that would require multiple time-to-digital converters a multi-event TDC can be used to reduce the number of converters and to improve matching. An on-chip measurement and characterization engine allows for low cost TDC characterization. Finally, a time domain quantizer is discussed that maps a discrete valued continuous time signal to a discrete time raster.

Key words: TDC Architectures, Bipolar TDC, Looped TDC, Linearly Extended TDC, Hierarchical TDC, Two-Stage TDC, Multi-Event TDC, Time-Domain Quantizer

4.1 Bipolar Time-to-Digital Converter

In principle a TDC has an asymmetrical structure: The start signal propagates in the delay-line and the stop signal triggers the comparators. This means that the start signal must arrive prior to the stop signal, i.e. only positive time intervals can be measured. In control applications with an integrating element in the control loop, the steady state time difference between an input signal and a feedback signal is essentially zero. Moreover, the time difference may become negative which means that the order of the two measurement signals is reversed. In a type II PLL, for

S. Henzler, *Time-to-Digital Converters*, Springer Series in Advanced Microelectronics 29,
DOI 10.1007/978-90-481-8628-0_4,

instance, the steady state phase difference is zero and the phase detector (i.e. the TDC in an all-digital PLL) has to measure small positive and small negative phase errors (time differences).

If the stop signal arrives earlier than the start signal the output of a conventional TDC is equal to zero as no signal has been injected into the delay-line yet. The conventional approach to allow for negative time intervals is to skew the stop signal. This means that the stop signal arrives later than the start signal even for a zero or slightly negative time interval. The main drawback of this technique is a large unknown offset error in the TDC characteristics that is highly susceptible to process variations and changes in the operating conditions (supply voltage, temperature). Furthermore, the skew element causes not only additional area and power but increases also the arrival time uncertainty and so the TDC's single-shot precision due to the intrinsic noise in the additional delay elements. A more advanced approach is depicted in Fig. 4.1. Two conventional TDCs are connected to the measurement signals x and y in such a way that the forward TDC measures the time interval $x \rightarrow y$ and the reverse TDC the time interval $y \rightarrow x$. Neither of the TDCs has to measure negative time intervals. The bipolar output signal is obtained by subtracting the output signal $TDC_{x \rightarrow y}$ of the forward TDC and the output signal $TDC_{y \rightarrow x}$ of the reverse TDC. As shown in Fig. 4.3 this yields a bipolar TDC characteristic without offset error (assuming two identical TDCs and negligible local process variations). Usually a TDC allows for slightly negative time intervals due to the latency along the buffer tree of the stop signal. This enables a reasonable overlap region between the measurement intervals of the two TDCs. In this overlap region both TDC outputs $TDC_{x \rightarrow y}$ and $TDC_{y \rightarrow x}$ vary with the time interval. Thus the TDC gain is doubled with respect to the gain obtained for large positive or negative time intervals. For applications where this non-linearity is an issue the implementation example of Fig. 4.2 can be used. As soon as one TDC delivers a zero output the

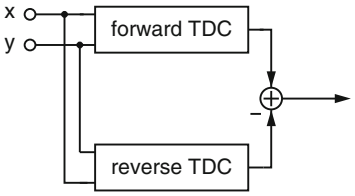

Fig. 4.1 Basic bipolar TDC

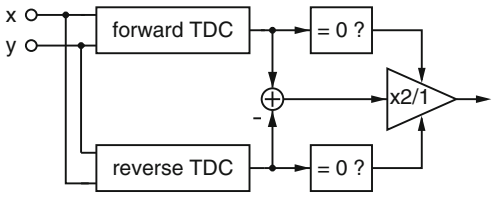

Fig. 4.2 Bipolar TDC with non-linearity correction

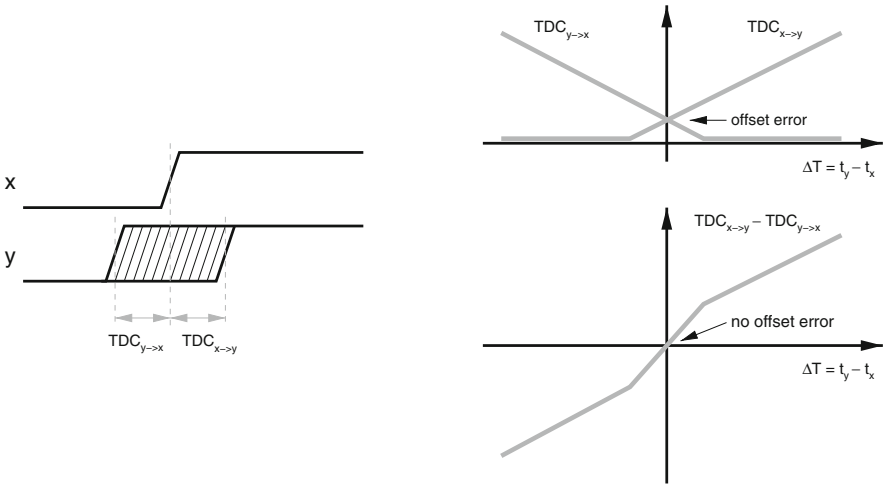

Fig. 4.3 Operating principle of a bipolar time-to-digital converter

overall output signal is doubled. This is a trivial correction that can be done in the digital domain with low overhead.

Each of the two sub-TDCs has to measure only half the time interval of the conventional approach with skew element. Hence the overall length of the delay-lines and so the area consumption is not increased. The subtraction of the two measurement results is done in the digital domain and causes negligible area and power overhead.

4.2 Looped Time-to-Digital Converter

The TDCs discussed so far are so called linear TDCs as they consist of one or more feed-forward delay-lines without any feedback. The length of these delay-lines and so the area of the TDC grow with the maximum time interval to be measured. Therefore, very long time intervals require large area. This can be avoided by looped TDC architectures (ref. Fig. 4.4) where a short delay-line is bent into a loop and traversed several times by the start event. As the start signal is fed into the delay-line again, when it has reached its end, the expression "reference recycling" is sometimes used to describe looped TDCs. A counter determines how often the delay-line has been passed by the start event before the TDC is stopped. The counter value B_{cnt} is then a coarse quantization of the time interval and the one-zero transition in the thermometer code provided by the (short) delay-line describes the position within one counter interval, i.e. the fine resolution B_{TDC}. The overall measurement value B can be calculated according to

$$B = M \cdot B_{cnt} + B_{TDC} \tag{4.1}$$

where M is the number of delay elements within the delay-line.

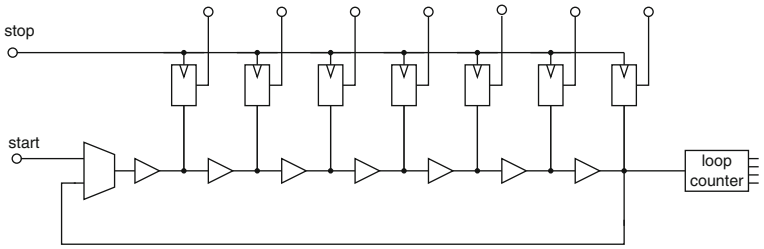

Fig. 4.4 Principle of looped TDC architecture

The start event is injected into the loop via a input multiplexer that passes either the external start signal or the feedback signal to the first delay element of the line. A control unit organizes the measurement as follows: First the multiplexer connects the input to the delay-line to enable the injection of the start event. The start event can be either a rising or falling signal transition or a pulse. As both is possible the expression "timing event" is used throughout this chapter.[1] When the start signal has entered the delay-line completely the multiplexer is toggled and waits for the start event emerging at the end of the line. The start event circulates in the loop until the measurement is stopped. On the arrival of the stop event the delay-line and the counter are sampled and the TDC is prepared for the next measurement. Therefore, the multiplexer is connected to the input, the old timing event is removed from the delay-line and the comparators are prepared for a new comparison (e.g. pre-charged). The most critical part during this control cycle is the detection of the stop event. The reason for this is the fact that the timing event in the loop and the stop signal are completely asynchronous. Hence, the control logic must detect the relative position of the timing event in the loop with respect to the stop event in order to generate appropriate control signals. The detection of the stop event and the generation of the control signals for the counter requires some time. To work correctly, the counter itself needs a signal with a pulse width not smaller than a certain minimum width at its clock input. These two observations mean that the counter may be disabled only if the timing event circulating in the loop has not yet reached the counter and if it is still sufficiently far away. If the counter cannot be disabled a further counting event happens and a digital correction is required later on. Figure 4.5 shows a technique to stop the counter reliably. Assume that the timing event is a pulse circulating in the loop. This pulse is used to sample the stop signal twice: Once in flip-flop A and a second time in flip-flop B. If the stop signal occurs while the pulse is in the right part of the delay-loop there is enough time to turn off the counter. This is the case if the stop signal is first sampled by flip-flop B and later by flip-flop A. In the other case, i.e. if the stop signal occurs while the pulse is in the left part of the loop, it is not possible to turn off the counter anymore. Hence, the

[1] In the general case the delay along the delay-line is different for rising and falling transitions. In the sampling elements, i.e. the comparators, there is also an asymmetry for sampling of a high or a low signal. Hence a pulse is preferable compared to signal transitions, especially in technologies with strong process variations.

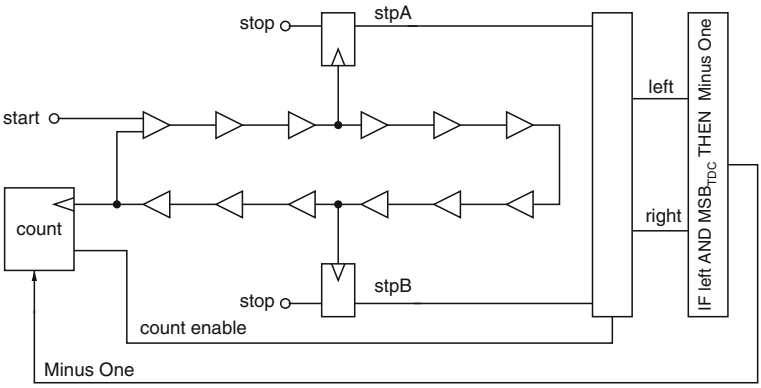

Fig. 4.5 Stop signal detection in looped TDC

counter value is incremented and is turned off not until the pulse has reached flip-flop A. It depends on the thermometer code whether the last increment was justified or not: If the most significant bit of B_{TDC} is not set, i.e. if the stop occurs while the pulse is in the upper left region, a full loop cycle has been completed and the counter value is correct. If the stop signal occurs while the pulse is in the lower left region, however, the loop cycle has not been completed and the counter value must be decremented by one. It shall be mentioned that a LSB error in the counter has to be avoided under all circumstances as a LSB of the counter represent an error of $N \cdot LSB$ where N is the number of elements in the loop. For a delay loop with 128 delay elements, for instance, 7 bits come out of the delay-line and all higher bits come from the counter. Consequently, a failure in the counter means a false 8th bit, which is a huge error.

While the circuit of Fig. 4.5 allows for reliably stopping the loop counter, the sampling flip-flops themselves are critical due to the asynchronous start and stop events: Suppose a circuit where the timing event in the loop is used to sample the stop signal. When the stop signal occurs, it is sampled by the flip-flop and the control logic turns the counter off as described above. If the time interval is increased slightly the stop event moves nearer to the timing event that clocks the flip-flop. As there is no constraining possible between the timing event in the delay-line and the stop signal a violation of the setup condition of the flip-flop is unavoidable. Hence, for certain measurement intervals the sampling flip-flop runs into its meta-stable state and its delay becomes long and unpredictable. This can lead to the following situation: A time interval is measured repeatedly and the stop signal is sampled properly. Then the time interval is increased slowly. The stop signal moves nearer and nearer to the sampling signal and the delay of the sampling flip-flop grows with an increasing rate. For some time intervals one of the sampling flip-flops runs into meta-stability and the control logic fails. Some applications are immune against occasional upsets. Others rely on each single measurement value and require an even more complex control logic that detects and eliminates meta-stability.

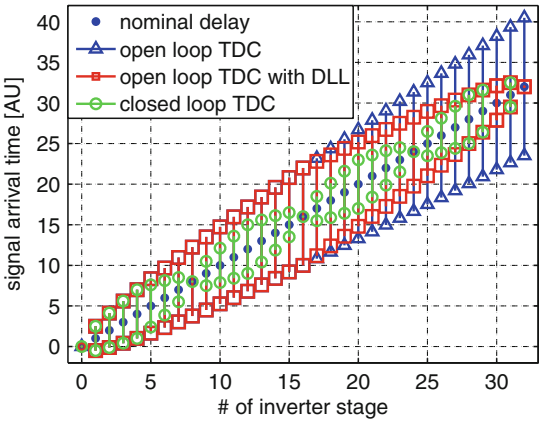

Fig. 4.6 Arrival time uncertainty in linear TDC (open loop), linear TDC with calibration, and looped TDC

Area consumption is just one reason to go for looped TDCs. Another one is the growing arrival time uncertainty in delay-lines with strong local process variations (ref. Section 3.7.2). Figure 4.6 shows various scenarios: Without local process variations the signal arrival time at a certain position within the delay-line grows linearly with the number N of the respective element. In the presence of local variations there is a arrival time uncertainty that grows according to the $\sqrt{N}$–law derived in eq. 3.49. Using gain compensation either by digital calibration or a controlled delay-line (ref. Section 4.4) forces the end-point of the delay-line to its ideal position. The $\sqrt{N}$-curve is thus bend and the maximum error is reduced by a factor of $\sqrt{2}$. A much bigger advantage yields the reduction of the number of delay elements: In short delay-lines only a few local variations can accumulate and the high dynamic range is achieved by a looped architecture. In the ideal case the arrival time uncertainty is reduced by a factor of $\sqrt{LF}$ where the length of the delay-line is reduced by the loop factor LF. However, the loop structure itself causes some linearity errors: The reason for this is the difficulty to bend a delay-line into a loop geometry without causing layout asymmetries. Even if half of the delay-line goes into one direction and the other half into the opposite direction the wiring at the bending points is long enough to cause visible non-linearity. This is particularly challenging in TDCs with sub-gate delay resolution in the order of some pico-seconds. The other source of non-linearity is the input multiplexer that has another delay than all the other delay elements in the loop. By replacing all delay elements by multiplexers with one input connected to the loop and the other one used as a dummy input this problem can be circumvented. However, area and power consumption are increased and the resolution is degraded. In practice there is a trade-off between non-linearity caused by a long delay-line and the structural non-linearity from the loop itself. The latter one can be reduced by a careful layout and reasonable sizing of the multiplexer. An advanced architecture that measures and eliminates this non-linearity is proposed in the following section.

In principle a looped TDC allows for arbitrary long time intervals to be measured. In practice, however, there is always a certain asymmetry in the delay-line, especially in the case of a looped architecture. This gives rise to parasitic pulse shrinking (or growing) that shortens the circulating timing event until it vanishes completely. The reason for this pulse shrinking is explained in Section 5.4 using the example of the pulse-shrinking TDC. Pulse-shrinking is an unwanted effect in all TDCs (except for the pulse-shrinking TDC that is based on intentional pulse-shrinking) and limits the maximum measurement interval. The looped TDC with linear extension can also solve this problem and is discussed next.

4.3 Linearly Extended TDC Loop

As discussed in the previous section the looped time-to-digital converter would be a perfect architecture for high dynamic range and good linearity if there was no asymmetry caused by layout and the coupling multiplexer. It is the intension of time-to-digital converters to measure time intervals with a picosecond resolution. The linearly extended TDC [18] uses this accurate measurement capability to actually characterize the delay asymmetry in the feedback loop itself and to correct it later on in the digital domain. In Fig. 4.7 the basic concept is illustrated schematically. The TDC consists of two linear delay chains namely the main TDC and the extender TDC. The main TDC is functionally configured in a loop structure. However, this loop structure does not reflect in a loop structure in the layout. The output signal is coupled from the output to the input of the main TDC without paying attention to the delay along this feedback path. When the measurement is stopped while the timing event passes the feedback loop, not the timing event itself but a redundant timing event (namely a copy of the original timing event) that propagates in the linear extension, i.e. the extender TDC, is used. Therewith, the measurement accuracy is decoupled from the delay in the feedback path and so from any asymmetries. As no carefully designed layout is required anymore automatic generation and place and route of highly accurate TDCs becomes possible.

The feedback does not disturb the linear signal propagation between the main TDC and the extender TDC because a small decoupling inverter is used. All other

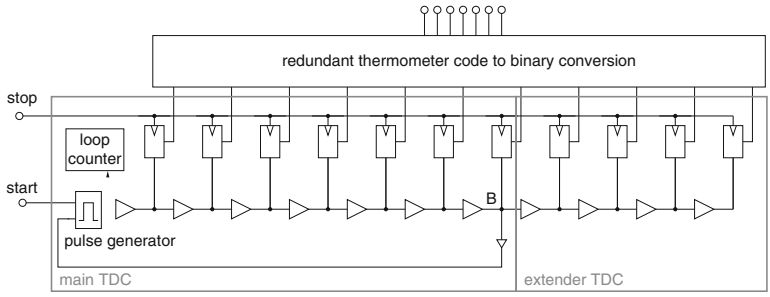

Fig. 4.7 Principle of linearly extended TDC loop

delay cells also possess such an inverter to provide various delayed versions of the timing event for control purposes.

The insensitivity to the feedback delay allows for the insertion of additional gates without degrading the TDC linearity. The insertion of a pulse generator at the input of the first delay line is particularly advantageous because it generates a fresh pulse with a well defined width not only at the beginning of the measurement but also after each loop cycle. For linearity reasons conventional TDCs cannot use such a pulse generator because it causes a serious discontinuity in the signal delay. A pulse is only generated at the beginning of the measurement and then circulates through the loop until a stop signal occurs. Due to parasitic pulse shrinking the pulse width diminishes. After some cycles the pulse has vanished completely and the maximum measurement interval is reached. In the linearly extended TDC with pulse generator the pulse width is restored after each cycle so in principle arbitrary measurement intervals are possible. The only factor that limits the maximum measurement interval is the range of the loop counter which can be easily adjusted.[2]

4.3.1 Operation and Calibration of Linearly Extended TDC

The purpose of the extender TDC is to continue the measurement linearly while the original timing event passes the feedback loop. Therefore, its measurement range can be smaller than the range of the main TDC. Three different cases can be distinguish after stopping the TDC:

1. A timing event is detected in the main TDC but not in the extender TDC. In this case the timing event has already run out of the extender TDC and the quantized measurement time interval can be calculated from the main TDC according to

$$t_d = N \cdot t_L + N_1 \cdot T_{LSB} \tag{4.2}$$

   where t_L is the delay of one complete loop cycle including the feedback, N is the number of loop cycles, and N_1 is the measurement value from the main TDC.
2. A timing event is detected in the extender TDC but not in the main TDC. Now, the original timing event is somewhere in the feedback loop and cannot be used for non-linearity reasons. In the extender TDC there is a linearly propagating redundant timing event so the measurement result is

$$t_d = N \cdot t_L + (max(N_1) + N_2)\, T_{LSB} \tag{4.3}$$

   where N_2 is the measurement value from the extender TDC and $max(N_1)$ is the length of the main TDC, i.e. the interval measured by the main and the extender TDC corresponds to the complete delay chain of the main TDC plus some delay stages in the extender TDC.

[2] Physical noise accumulates during the TDC measurement, i.e. a functional limitation of the maximum measurement interval is the noise specification of the overall system using the TDC.

3. A timing event is detected in both the main and the extender TDC. In this case the original timing event has already passed the loop but the redundant timing event is still seen by the extender TDC. The measurement value is then computed either as in the first or as in the second case. However, this case can be used to calculate the loop delay t_L including the discontinuous feedback. The required algorithm is described next.

In case 3, both TDCs provide measurement values, namely N_1 and N_2. If there was no delay in the feedback path, N_1 would be equal to N_2. As the feedback path and the pulse generator cause an unknown delay which is not a multiple of the TDC's unit delay T_{LSB}, N_1 and N_2 differ and the loop delay is given by $N_2 - N_1$. As both TDCs provide quantized measurement values this difference is actually not the exact loop delay t_L but a rough estimation. Averaging over many of these differences acquired for all measurements where a timing event is detected in both TDCs yields the exact loop delay t_L. This can be seen when the quantization errors are included in the computation: Starting from the branching point B (ref. Fig. 4.7) where the feedback path and the path into the extender TDC spread, the time interval T' represented by the output values of the TDCs can be written as

$$\begin{aligned} T' = t_{feedback} + (N_1 + \varepsilon_1)\, T_{LSB} &= (N_2 + \varepsilon_2)\, T_{LSB} \\ 0 \le \varepsilon_1 &< 1 \\ 0 \le \varepsilon_2 &< 1 \end{aligned} \tag{4.4}$$

N_1 and N_2 are the TDC output values and ε_1 and ε_2 are the quantization errors. In general the feedback delay $t_{feedback}$ is not a multiple of the quantization time T_{LSB}, i.e. the unit delay in the TDCs. Figure 4.8 illustrates the quantization errors of the two TDC measurements for a non integer loop delay with a fractional part d. Considering one step where both TDCs produce a measurement value, the quantization errors can be written as

$$\begin{aligned} \varepsilon_2 &= \frac{t}{T_{LSB}} \qquad 0 \le t < T_{LSB} \\ \varepsilon_1 &= \begin{cases} \frac{t}{T_{LSB}} + (1-d) & 0 \le t < d \cdot T_{LSB} \\ \frac{t}{T_{LSB}} - d & d \cdot T_{LSB} \le t < T_{LSB} \end{cases} \end{aligned} \tag{4.5}$$

where t describes the position within the quantization interval. The feedback delay can be calculated according to

$$t_{feedback} = (N_2 - N_1)\, T_{LSB} + (\varepsilon_2 - \varepsilon_1)\, T_{LSB} = (N_2 - N_1)\, T_{LSB} + \varepsilon \cdot T_{LSB} \tag{4.6}$$

with a quantization error

$$\varepsilon = \varepsilon_2 - \varepsilon_1 = \begin{cases} -(1-d) & 0 \le t < d \cdot T_{LSB} \\ d & d \cdot T_{LSB} \le t < T_{LSB} \end{cases} \tag{4.7}$$

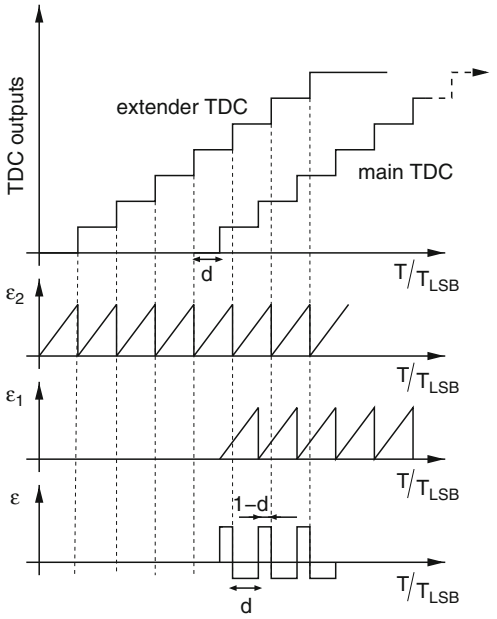

Fig. 4.8 Quantization error for the computation of the feedback delay

The mean value of this quantization error

$$< \varepsilon >= -d \cdot T_{LSB} \cdot (1-d) + (1-d)\, T_{LSB} \cdot d = 0 \tag{4.8}$$

vanishes, i.e. the exact feedback delay $t_{feedback}$ can be calculated according to

$$t_{feedback} = \langle N_2 - N_1 \rangle\, T_{LSB} \tag{4.9}$$

The loop delay is then given by

$$t_L = max(N_1) \cdot T_{LSB} + t_{feedback} = (max(N_1) + \langle N_2 - N_1 \rangle) \cdot T_{LSB} \tag{4.10}$$

If this calibration process runs in the background during the regular TDC operation the quantization error is sufficiently randomized, i.e. the averaging is feasible. Theoretically, the computed feedback delay and so the linearity are perfect, even if the delay varies during operation due to altering voltage and temperature conditions. In practice the achieved accuracy depends on the number of bits used for the computation. The effort for the background calibration is in the order of 1k gates.

Figure 4.9 shows exemplary histograms of the TDC output values for equally distributed input time intervals. The upper histogram refers to a conventional looped TDC with asymmetry in the feedback path due to the input multiplexer and layout asymmetries. The lower histogram shows the output data of the same TDC but now with active linearly extended loop. The implemented additional accuracy for the digital background calibration is two bits. The original histogram shows an increased probability for measurement intervals corresponding to multiples of the loop delay.

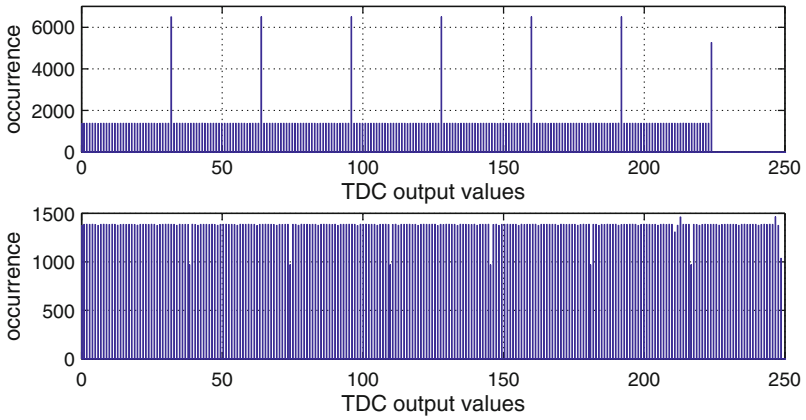

Fig. 4.9 Exemplary histograms of output data of a conventional looped TDC (upper histogram) and a looped TDC with linear extension (lower histogram) for equally distributed input time intervals

This is due to the fact that the timing event takes more than T_{LSB} to pass the feedback and the input multiplexer. Hence more measurement events fall into the corresponding bins. The background calibration removes this increased probability, by using the linear extension for the measurement of the problematic intervals. This reflects not only in a flat histogram, but also in a stretching along the x-axis. The reason for the stretching is the fact that the feedback path now accounts with its actual delay and not with just T_{LSB}.

4.4 Delay-Locked-Loop Based TDC

Time-to-digital converters are based on digital delay elements which, unfortunately, depend on process variations, temperature, and supply voltage (PVT variations).

Hence, a basic TDC can provide only a qualitative measurement in terms of one inverter delay, for instance. For absolute time interval measurement a calibration step is required before the measurement. If an absolute time interval measurement without calibration is of interest, the DLL (delay-locked loop) regulated time-to-digital converter can be used. This approach is particularly advantageous if a given reference time interval has to be subdivided into a given number of sub-intervals (ref. Section 4.8). The basic concept is illustrated in Fig. 4.10. A periodic start signal is injected into the delay-line. After each delay element a skewed copy of the original start signal is available. The signals at the input and at the end of the delay-line are fed to a phase detector that compares the phase of these two periodic signals. For the sake of simplicity it shall be assumed that approximately one period of the periodic start signal fits into the delay line. If the rising edge of the signal tapped from the end of the delay-line exactly occurs when the input signal rises, the delay along the line corresponds to the period of the start signal. If the rising edge at

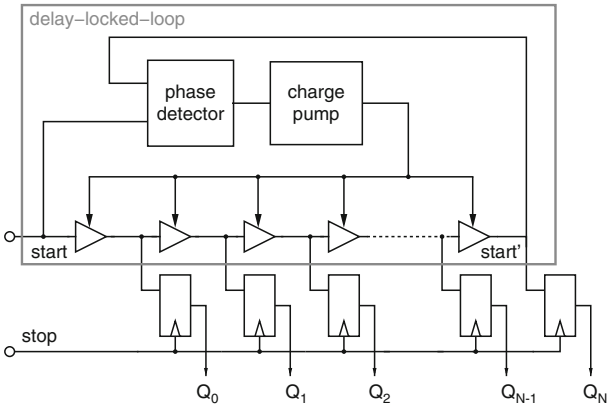

Fig. 4.10 Block diagram of time-to-digital converter embedded in DLL

the output node start' occurs prior to the rising edge of the start signal, the delay along the delay-line is too small. In the other case the delay is too large. Ideally, the phase detector provides a signal that is proportional to the time difference between the two rising edges. This information is used to control the delay of the individual delay elements. As the output signal contains not only skew information but also some higher frequency components a loop filter (low-pass filter) is required after the phase detector. For a vanishing phase difference in steady state the loop filter needs an integrating component. This integrating component is usually implemented as a charge pump (for analog implementations). Obviously, the dynamic of this closed-loop configuration can be adjusted by the poles and zeros of the loop filter.

In principle the DLL can be designed for more than one reference period in the delay-line. Usually, this has no advantage. However, a risk of control loops such as the DLL is a phenomenon called false locking: A basic phase detector only compares the arrival times of its input signals but has no idea about potential signal transitions within the delay line. Hence it can happen that the delay-line is intended to contain a single period but in fact contains two reference periods. To avoid false locking a more complex phase comparator, namely a phase-frequency detector, can be used. Alternatively a watchdog circuit can monitor the signals within the delay-line.

The delay of each delay element can be tuned by its supply voltage. The gain of this control mechanism (voltage-to-delay transfer function) is very high. However, the implementation is challenging as the complete supply current of the delay-line is drawn from the controlled node. This means that the charge pump has to be designed for large and highly varying load currents. If the supply voltage is used for delay tuning it may be advantageous to use a LDO instead of a charge pump for the supply voltage generation. This requires enough voltage headroom for the drop-out device. Another approach uses series transistors to starve the supply current of the delay elements. As shown in Fig. 4.11 such devices can be inserted either at the power or at the ground side or even at both sides. The advantage is that now the gates of

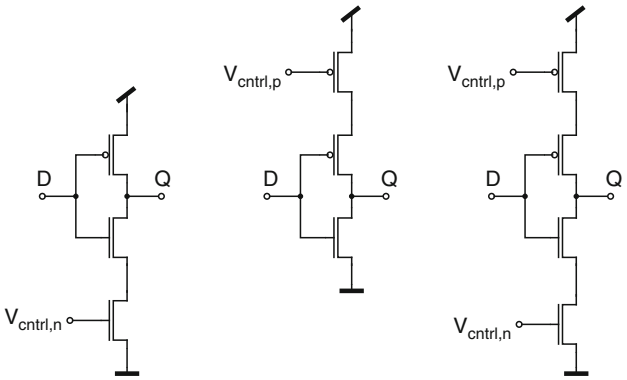

Fig. 4.11 Current starved inverters as tunable delay elements: Footer device (left), header device (middle), or both (right)

the series transistors are connected to the controlled node voltage. Thus, the output of the charge pump is only capacitively loaded which simplifies the control loop considerably and also reduces the overall power consumption. If a very wide tuning range of the DLL is required a coarse setting can be done by connecting more or less delay elements in parallel or by connecting more or less capacitance to the outputs of the delay elements. Fine tuning may be achieved with the techniques described above.

Finally, it has to be understood that the delay tuning always reduces the resolution of the TDC. This is simply because of the fact that linear operation of the control loop requires some tuning headroom towards both direction, i.e. towards slower and faster delay. Both analog or digital implementations of the control loop are possible. Currently, analog implementations are predominant. A system constraint of the DLL regulated TDC is the fact that the start signal has to be periodic. This requirement is very well fulfilled in digital PLLs for instance. If the start signal is not periodic (at least in average) the DLL approach is only possible if a periodic reference signal is provided to the system during regular calibration intervals.

4.5 Hierarchical Time-to-Digital Converter

The TDCs discussed so far measure the complete time interval with the full resolution. This causes a high power consumption especially for the high-resolution TDCs discussed in Chapter 5. Consequently, the measurement of long time intervals becomes unattractive. For long intervals the power can be significantly reduced by a hierarchical approach such as that depicted in Fig. 4.12. A coarse time-to-digital converter quantizes the input time interval coarsely. This can be done with a relatively short, so area and power inexpensive, delay-line. To avoid a large quantization error the residue is injected into a second TDC for fine quantization. This TDC

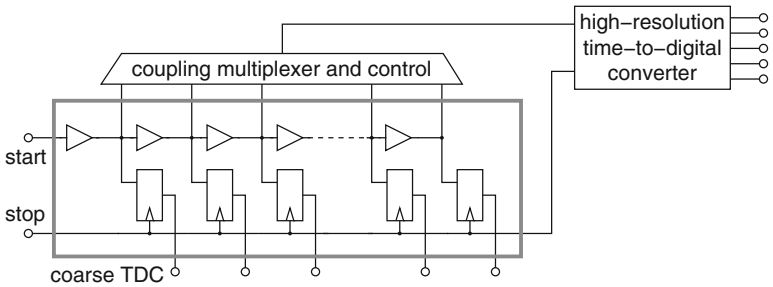

Fig. 4.12 Principle of the Hierarchical TDC [44] based on a coarse time quantizer, a controllable coupling stage followed by a fine time quantizer

(fine TDC) has a much higher resolution, but a small dynamic range, which in the ideal case is equal to the quantization interval of the first TDC. It is obvious that this architecture can reduce both area and power: The first stage due to the coarse quantization and the second stage due to the small dynamic range. Both stages are coupled by a large multiplexer logic. This multiplexer is the reason why the theoretically simple and advantageous hierarchical TDC architecture is very challenging to implement. For the injection of the residue time interval into the fine TDC the one-zero-transition in the coarse TDC has to be detected first (or at least the position has to be estimated). This causes not only control overhead but means also some latency that has to be compensated by additional delay elements in the stop path between the first and the second TDC stage. Additional delay elements always mean additional variability, additional power, and additional noise. Hence, the overall latency and so the number of compensation elements must be kept as small as possible. For slowly varying signals, the results from previous measurements can be used to determine the tap of the coarse TDC. This reduces the timing requirements for the control logic and the overall latency.

From a delay perspective, a multiplexer is not symmetrical with respect to its inputs. Even if the topology is completely symmetrical, for instance by using a wired-NOR structure, the layout and the wiring between the first TDC stage and the multiplexer unavoidably introduce some asymmetry. The wiring is particularly difficult as each input signal originates from another delay element, i.e. from another geometrical location. These effects cause a systematic non-linearity that makes the hierarchical approach unattractive for practical applications. To take advantage of the hierarchical TDC architecture anyhow, the number of coupling paths between the first and the second stage has to be reduced as much as possible. In the ideal case only a single coupling path exists. The upper schematic in Fig. 4.13 demonstrates this approach. The first stage consists of a very short delay-line that is coupled to the second converter stage at a single position. To achieve a high dynamic range in the coarse stage, the delay elements are configured in a conventional loop structure. The circulating timing event is injected periodically into the fine TDC. On the arrival of the stop signal both the coarse and the fine delay-lines are sampled. The leftmost transition in the fine TDC represents the actual measurement. If more than

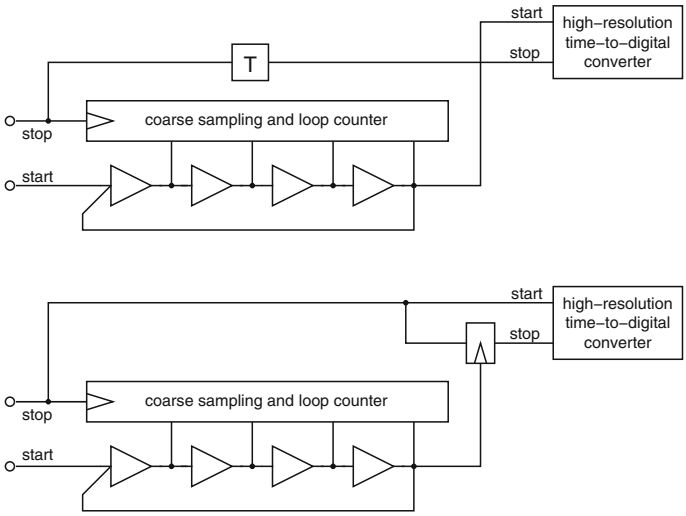

Fig. 4.13 Hierarchical TDC where the interface between the first and the second stage is reduced to a single signal. Straight forward approach (top), activity optimized approach (bottom)

one transition fit into the fine TDC the resulting period information can be used to calibrate the first and the second TDC stage against each other. This is required as the relative delays of the delay elements in the first and the second stage are not known. While the linearity of this interface optimized TDC is very good the power consumption is still high. This is due to the fact that in each cycle of the coarse TDC a new timing event is injected into the fine TDC. These timing events cause switching activity and so dynamic power consumption.

An improved hierarchical architecture is illustrated by the lower schematic of Fig. 4.13. In this approach not the periodically cycling timing event of the coarse stage but the global stop signal serves as the start signal for the fine TDC. Hence, only a single timing event per measurement propagates along the fine delay-line and the power consumption is reduced considerably. The timing event circulating inside the coarse TDC stops the measurement when it traverses the interface node for the first time after the arrival of the global stop signal. This causes similar challenges as the disabling of the counter in the looped TDC. The stop signal must be detected and the following timing event in the coarse loop must be passed to the fine TDC. Due to the asynchronous timing between the circulating signal and the stop signal latency and meta-stability in the control logic are serious design challenges.

The fine TDC needs a dynamic range that corresponds to the period of the oscillation in the coarse TDC.[3] As the latter one is designed for low power and so low resolution the period is usually quite high. To avoid an excessive area consumption in the fine TDC a compromise between the architectures in Figs. 4.12 and 4.13 can be found. In Fig. 4.14 an exemplary TDC is given that uses a loop of four coarse

[3] Latency in the control circuit requires an even higher dynamic range.

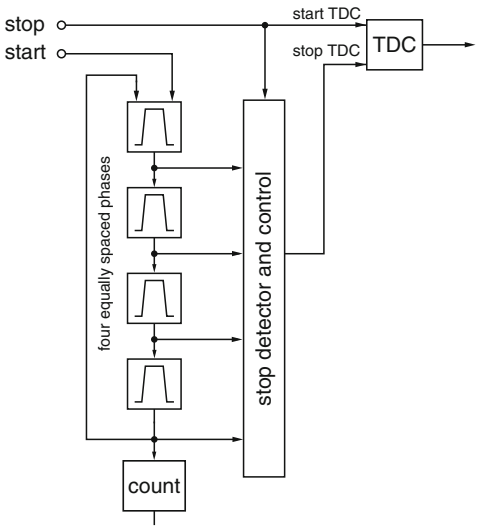

Fig. 4.14 Implementation example of a hierarchical TDC with a coarse looped TDC in the first stage and a high-resolution TDC in a second stage

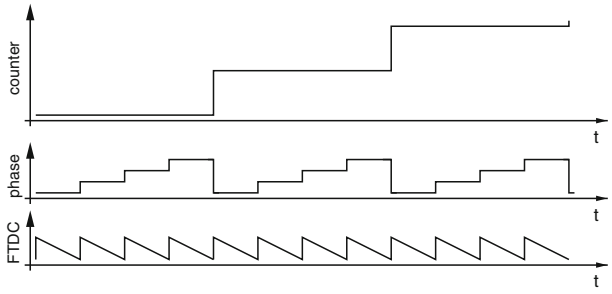

Fig. 4.15 Signal diagram of proposed hierarchical TDC of Fig. 4.14

delay elements. Each output is connected to the fine TDC via a 4:1 multiplexer. The number of delay elements and so the fan-in of the multiplexer is small enough to enable symmetrical layout. The use of four phases reduces the dynamic range of the fine TDC to one fourth of the coarse oscillation period. A complex meta-stability immune control logic monitors the global stop signal and selects the most appropriate phase of the coarse stage. The second stage comprises a linear high-resolution TDC. As depicted in Fig. 4.15 the output word of the hierarchical TDC is calculated according to

$$B_{out} = B_{cnt} + \frac{1}{4}B_{phase} + \left(\frac{1}{4} - \frac{1}{4}\kappa B_{FTDC}\right) \tag{4.11}$$

B_{cnt} is the value of the loop counter, B_{phase} the number of the phase that has actually stopped the fine TDC, and B_{FTDC} is the measurement result of the fine TDC. The relative calibration of the first and the second stage reflects in the coefficient κ (internal calibration).

4.6 Multi-Event Time-to-Digital Converter

Most applications do not require the measurement of isolated timing events but consecutive measurement of multiple events or even overlapping events. The drawback of most of the advanced TDC architectures described so far is the dead time after a measurement. This prevents the system from starting a new measurement simultaneously to the stopping of the previous measurement. Figure 4.16 describes an exemplary application where the timing information contained in a pulse width modulated signal (PWM signal) shall be extracted. As indicated below the figure there are two ways of extracting the full information contained in the signal: The first possibility is to start a first measurement at the time t_1, a second measurement at the time t_2, and to stop both measurements at the time t_3:

$$\text{Measurement 1: } t_1 \rightarrow t_3$$
$$\text{Measurement 2: } t_2 \rightarrow t_3$$

This means that in the interval $[t_2;t_3]$ two measurement run simultaneously. In the next PWM cycle the same measurements have to be performed, i.e. both of the measurements currently considered end at the time t_3 and a new one has to be started immediately.

For the second measurement strategy a first measurement is started at the time t_1 and stopped at the time t_2. A second measurement runs in between t_2 and t_3:

$$\text{Measurement 1: } t_1 \rightarrow t_2$$
$$\text{Measurement 2: } t_2 \rightarrow t_3$$

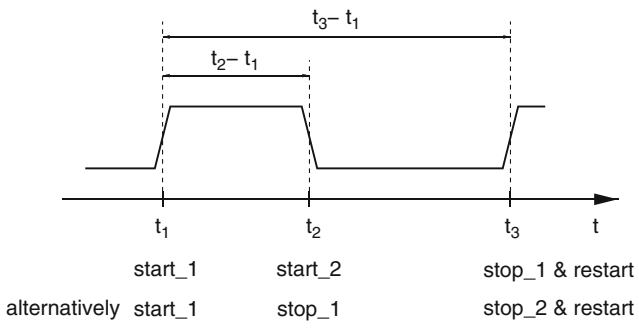

Fig. 4.16 Application scenario with multiple events measured simultaneously or consecutively

In this case there are no simultaneous measurements but each measurement is started at the same time instance where the previous measurement ends.[4] If the utilized time-to-digital converters have a dead time, two TDCs have to be used in a time interleaved manner. The first approach requires even three devices. In principle, these concepts can be easily implemented but result in a considerable area consumption. The major disadvantage, however, becomes visible if a mismatch between the various TDCs exists: Time domain multiplexing means that multiple TDCs are used periodically. If these TDCs have slightly different properties, e.g. a slightly different gain, a periodic error arises in the measurement result. Any periodic error leads to spurious tones in the spectrum of the output signal. Thus, applications such as data converters which are sensitive to tones, require even more time-to-digital converters to scramble the mismatched TDCs (mismatch shaping).

The multi-event time-to-digital converter [15] is an alternative approach that has the capability to measure multiple timing events in parallel as well as to start new measurements instantaneously. Therefore, only a single delay-line configured in a loop is required. The basic principle is illustrated in Fig. 4.17. The actual TDC is described by the thick loop. The bullets indicate the delay elements. For the sake of simplicity the loop counter(s) and the sampling elements are not drawn. Unlike the conventional looped TDC there are two inputs to inject timing events into the loop. The positions in the loop that are connected to the inputs are called injection points. The timing events are pulses generated by pulse generators which are inserted into the two input paths. The loop itself comprises check points before and after each injection point. When a timing event comes near to a certain injection point the checker before this injection point sets a 'block' signal. This 'block' signal is reset as soon as the timing event leaves the region around the respective injection point

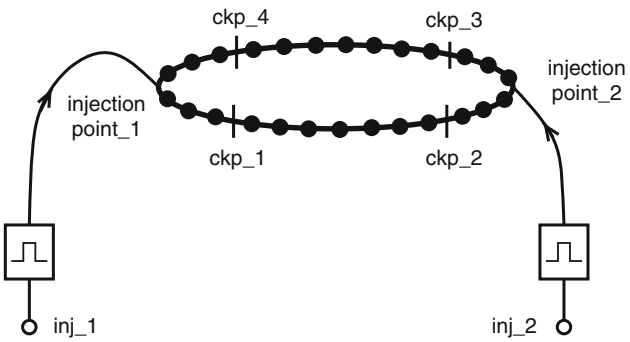

Fig. 4.17 Principle of multi-event TDC with two injection points

[4] Alternatively a measurement can be started at the time t_1 and stopped both at time t_2 and time t_3. This means that the state of the delay-line and the loop counter has to be sampled twice before the timing event is actually removed from the loop. This causes considerable non-linearity due to layout challenges as two comparators have to be placed and connected closely and symmetrically to the delay elements. Moreover, the requirement to start a new measurement instantaneously for the next period of the PWM signal persists.

again. This is accomplished by the check point after the injection point. The operation of the multi-event TDC is as follows: If a start signal occurs the control logic chooses an appropriate injection point and inserts a timing event into the loop. On the arrival of a new start event another timing event, namely another pulse, is injected into the same loop. The two (or more) pulses circulate independently within this single loop. 'Appropriate injection point' means that there is no other timing event in the vicinity of the respective injection point. The 'block' signals discussed above provide this information, i.e. disable a certain injection point while another timing event is passing by. This injection strategy is crucial as the two timing events could not be distinguished anymore if they interfered during the injection or at any time during the measurement.

The control challenge of the TDC is to select the injection points and to keep track of which pulse belongs to which measurement. The latter means that the sequence of various pulses inside the loop must be processed by the control logic. As long as there is a single pulse inside the loop a loop counter sees one trigger event within the loop cycle time. As soon as there is another event in the loop two trigger events occur in the same time interval. The control logic is then responsible to calculate how many loop cycles each pulse has completed. This can be done by multiple counters that are enabled just in the right moment by the control logic or by a single counter that is sampled before a new timing event is inserted into the loop. The sequence of the pulses is also important to assign the transitions in the pseudo thermometer code to the respective measurement. Of course the concepts can be extended for more than two injection points.

A timing diagram is given in Fig. 4.18 to review the operating principle again. At the beginning there is no pulse in the loop. On the arrival of the first start signal a timing event is injected. The control logic chooses injection point number one and blocks it immediately. The pulse starts cycling in the loop. On the arrival of the second start signal the first injection point is locked, so the control logic chooses the second one for the injection of the new event (thick pulse). Both pulses circulate without any interference. On the arrival of the stop signal the control logic computes the two measurement values by evaluating the pseudo thermometer code and the loop counter(s).

The open question is now what happens to the pulses after the measurement has been completed. They cannot persist inside the loop as the latter would overflow with new pulses. On the other hand the complete loop cannot be reset as for

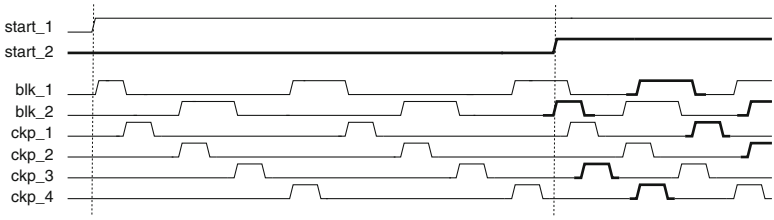

Fig. 4.18 Timing diagram describing the principle of the multi-event TDC

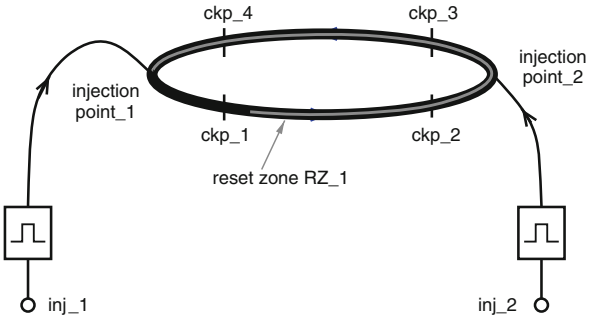

Fig. 4.19 Multi-event TDC with one partial-reset-zone highlighted

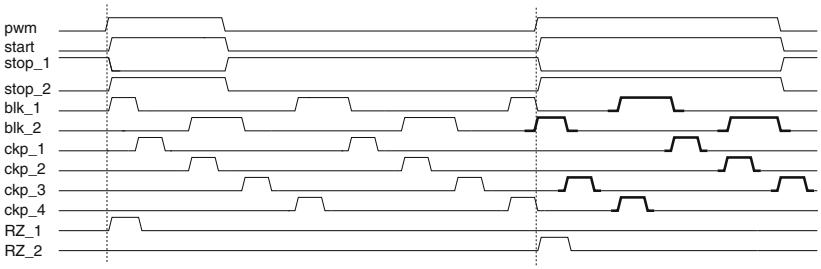

Fig. 4.20 Application scenario of the multi-event TDC with partial reset feature

consecutive measurements a new timing event is injected right when the previous measurement is stopped. A strategy to remove old pulses from the loop is the so called partial reset strategy that is illustrated in Fig. 4.19. Assume that a new measurement shall be started and that all old pulses shall be removed. For instance, if the control logic selects the first injection point, for the injection of the new timing event, it is assured that there is no timing event in the region between checkpoint ckp_4 and ckp_1 (counterclockwise). By forcing the outputs of all delay elements in the zone between ckp_1 and the first injection point to zero (again counterclockwise) all pulses except the one just injected are eliminated. This is a partial reset as only a fraction of all delay elements is forced to zero. Depending on the injection point, the control logic determines a suited reset zone. A timing diagram for the reset mechanism is shown in Fig. 4.20: A pulse width encoded signal is measured by starting a measurement on each rising edge of the PWM signal. The measurement is stopped twice, namely on the falling and on the next rising edge. Additionally, the next measurement is started on the next rising edge by injecting a new timing event. In the example injection point two is chosen and the corresponding partial reset zone is activated in order to remove all old timing events.

The advantage of the multi-event TDC is a considerably reduced area, but the control logic becomes more complicated. The main benefit, however, is the fact that only a single delay-line is used, so there is no mismatch. A certain asymmetry exists because of the different injection points. For long measurements this asymmetry becomes relatively small. Moreover, the signal statistic itself randomizes this

asymmetry, as it depends on the input signal which injection point is chosen. If this natural randomization is not sufficient a pseudo-random generator inside the control logic can be used to select one among multiple (non-blocked) injection points.

4.7 On-Chip Test and Characterization Engine

In Section 3.7 it has been described, that the gain and the offset error can be determined by two independent TDC measurements. To acquire the full converter characteristic, however, all possible input time intervals have to be applied as test time intervals to the TDC input. These time intervals must be reproducibly provided with a picosecond resolution. Two independent signal generators with a fixed skew are not feasible to produce the start and the stop signals because independent signal sources have uncorrelated jitter. Hence, the test time interval that should be actually constant varies according to the statistical sum of the two jitter processes. A single signal source can be used together with a tunable delay element to derive correlated start and stop signals. External measurement equipment that allows for delay tuning with a picosecond resolution is relatively expensive. Digital standard test equipment can definitively not be used. Hence, measuring a TDC characteristic under laboratory conditions is possible, but for production test, or even self-test during the system life time, an alternative solution is required. An on-chip measurement and characterization engine build of standard CMOS logic cells is depicted in Fig. 4.21. The time-to-digital converter to be tested is enclosed in the gray box. The circuitry in the dashed gray rectangle shall be neglected first. The remaining circuit serves as a stimulation engine for the TDC. A common characterization pulse is injected into two delay-lines. The outputs of these delay-lines are connected to the start and the stop inputs of the TDC, respectively. The first delay-line which is connected to the start terminal has a fixed length. It consists of N multiplexers followed by a chain

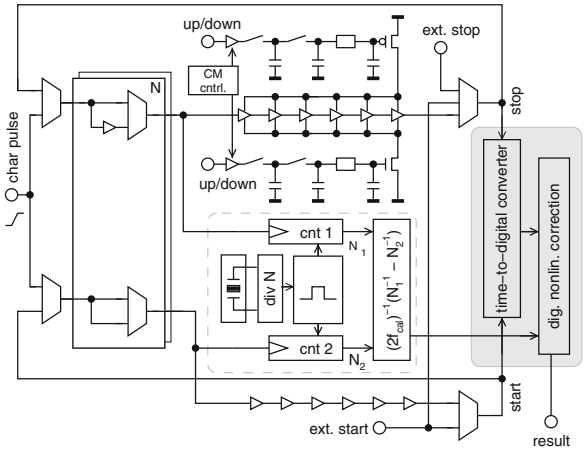

Fig. 4.21 On-Chip measurement and characterization engine for TDCs

of C^{2}MOS buffers. The footer and header devices of these buffers are connected to V_{DD} and V_{SS}, respectively. The second delay-line is connected to the TDC's stop terminal. Additional buffers can be inserted into the delay line for coarsely skewing the stop signal against the start signal. The C^{2}MOS buffers in the second part of the delay-line have now analog control voltages at the gates of the header and footer devices. Hence, the delay in this buffer chain and so the skew between the start and the stop signal can be tuned in a fine grained way. A control logic can gradually sweep the stop signal against the start signal and acquire the corresponding TDC measurement values. A coarse sweep is done by inserting additional delay elements and a fine grained adjustment is done by control voltages in the final delay-line. This measurement allows for a quasi-continuous acquisition of the complete converter characteristic. The problem is, that this characteristic only describes the relative TDC behavior, as the absolute skew, i.e. the time interval leading to a certain TDC output, is not known. To map the TDC characteristic to an absolute time axis the stimulation circuit has to be characterized. Therefore, the sub-circuit in the dashed gray box comes into play. It consists of a frequency divider with large divider factor N, two counters, and some arithmetic logic. For each setting of the two delay-lines, the lines are configured as ring-oscillators. Pulsed ring-oscillators are particularly advantageous in order to get insensitive against any difference between rising and falling propagation delays. The oscillation frequencies correspond to the inverse propagation delays of the two lines. To minimize quantization effects and to average noise induced delay variations, these two frequencies are measured over a relatively long time. Therefore, a stable reference clock is divided by N. One period $T_{cal} = N \cdot T_{ref}$ of this subdivided reference clock enables the two counters which count the number of ring-oscillator cycles. With these counter values N_1 and N_2 the delay difference of the two lines, i.e. the length of the test time interval ΔT applied to the TDC can be determined:

$$T_{cal} = N_1 T_1 = N_2 T_2 \tag{4.12}$$

$$\Delta T = T_1 - T_2 = T_{cal} \left(\frac{1}{N_1} - \frac{1}{N_2} \right) \tag{4.13}$$

$$= N \cdot T_{ref} \left(\frac{1}{N_1} - \frac{1}{N_2} \right) \tag{4.14}$$

Figure 4.22 shows a timing diagram of the TDC calibration. During calibration mode the timing in the two delay-lines circulates with slightly different free-running

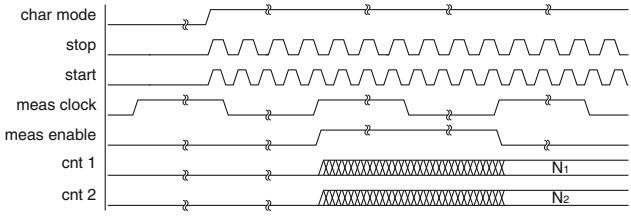

Fig. 4.22 Principle of absolute time measurement algorithm

frequencies, and the counters are enabled during a single period T_{cal} of the measurement clock. The two measurements, namely the stimulation of the TDC and the characterization of the stimulator enable the acquisition of the converter characteristic with an absolute time axis.

A common problem for both external and on-chip test signal generation is the fact that the TDC measures not only its own non-linearity but also the non-linearity of the stimulator. For the on-chip measurement and characterization engine, for instance, the non-linearity in the two delay-lines becomes indistinguishable from the non-linearity of the TDC itself. The same hold for noise in the stimulator and the TDC. Both noise components lead to a finite single shot precision but cannot be separated anymore.

4.8 Time Domain Quantizer

The time-to-digital converters discussed so far measure the time interval between a start and a stop event and quantize the measurement result. The start and the stop signals are only used as trigger signals and are not quantized themselves. On the other hand, the time-domain quantizer forces a time continuous signal with discrete values on a discrete time grid. Usually the discrete time instances are equally spaced. In the simplest case a single bit binary but time continuous signal, e.g. a pulse width modulated signal, is discretised along the time axes. As a side effect of most implementations the time intervals defined by the signal transitions that are quantized are measured as well. An exemplary signal that is quantized in the time domain is depicted in Fig. 4.23. The very first signal in the figure is the continuous time input signal $x(t)$. The very last one, namely $x_q(t)$, is the same signal with all signal

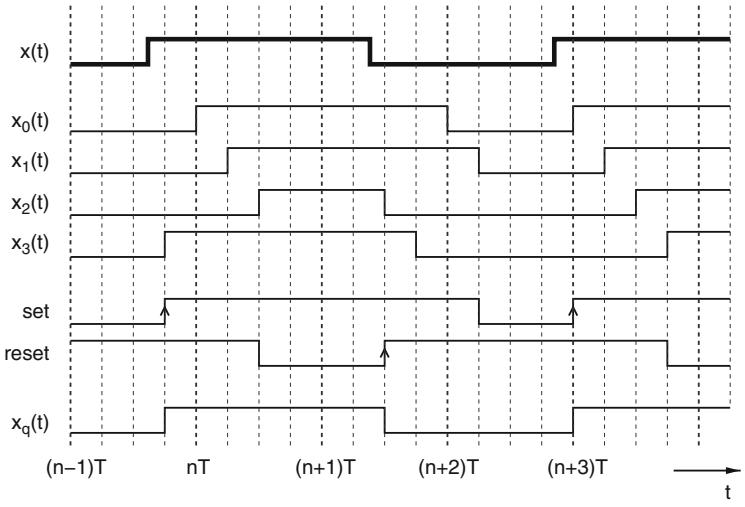

Fig. 4.23 Exemplary signal diagram of time domain quantizer

transitions forced to the discrete time raster. The vertical dashed lines represent the time raster. The thick lines span a coarse quantization raster with a periodicity T and the thin lines provide a fine resolution by interpolating within the coarse raster. The conventional approach of time domain quantization is to sample the signal in a flip-flop with a fixed clock. People often say that the signal is (re)-synchronized to the clock signal. For a high time domain resolution the same argumentation holds as for the time interval measurement in TDCs: A higher clock enables a higher resolution but increases the effort for the clock generation and distribution and may disturb sensitive circuitry in the vicinity. Moreover, the maximum resolution that can be achieved by feasible clock frequencies is still relatively low. A higher resolution can be achieved by a multi-phase clock generated by asynchronous delay elements. In a time-to-digital converter these multiple clock phases are used as input signals for latches/flip-flops that are triggered all at the same time. For time domain quantization, however, the multiple phases are used to clock latches/flip-flops that sample a single signal, namely the continuous time input signal. As the result each sampling element provides the value of the continuous input signal at the trigger instance (e.g. the rising edge) of the respective clock phase. This information can be directly used for the time interval measurement spanned by multiple features (e.g. signal transitions) in the continuous time signal. A quantization of the input signal itself has not been performed yet. In Fig. 4.23 the sampling process is described: The thick lines describe full clock periods and the thin lines the raster of the multiple phases. If the continuous signal goes high (low) the next clock edge sets the output of its respective sampling element (flip-flop) to high (low). As soon as any of the sampling elements goes to high, the quantized output signal should be set to high, too. This can be triggered by the rising edge of the logic OR conjunction of all flip-flop outputs. In the other direction, the quantized output signal should be set to low as soon as any of the sampling elements goes to low. This may be detected by a rising edge of the NAND conjunction of all sampled signals. As the switching commands are given by the signal transitions, the OR and NAND signals cannot be used directly to set or reset an output latch that provides the quantized signal. Hence, edge detectors, e.g. pulse generators, have to be added to the OR and the NAND gates. An exemplary schematic is shown in Fig. 4.24. A buffer delay-line is used to derive multiple clock phases from a single input clock CP. In a first step the multiple phases are used to sample the input signal and in a second step the sampling information is used to generate the quantized output signal. An alternative solution using two delay-lines is proposed in [9]. One is used to sample the original continuous time signal and the other one to sample the inverse signal. This yields some advantages with respect to symmetry of rising and falling signal transitions, but the basic principle is the same.

Similar as in the time-to-digital converter it is the fundamental nature of the time domain quantizer that the input signal is asynchronous to the sampling signals. Consequently, it is unavoidable that the setup condition of at least one sampling element is violated so the latter enters its meta-stability region. The clock-to-output delay of this sampling element becomes strongly dependent on the signal arrival time. As the sampled signal is used in the second step to set or reset the quantized signal, meta-stability causes non-linear effects during the time domain quantization.

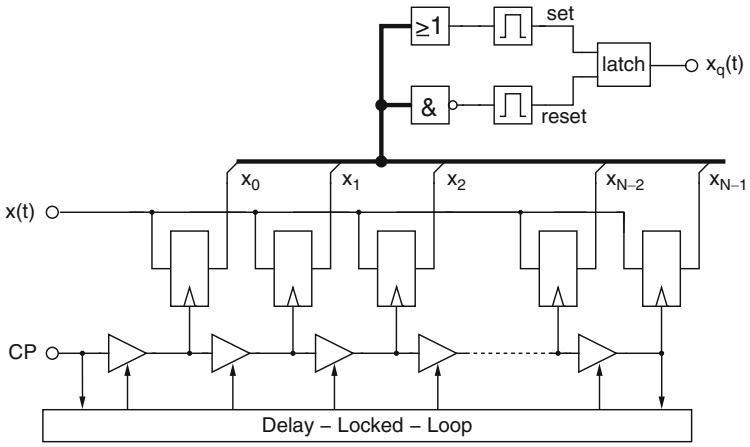

Fig. 4.24 Basic implementation of time domain quantizer

Re-sampling of the flip-flops or logic combination with later clock phases can reduce meta-stability effects considerably.

Variations (global and local process variations but also voltage and temperature variations) strongly affect the delay of digital circuits and so of the delay and sampling elements. Under fast conditions the multiple phases of the input clock are pushed together. This means that the signal is often sampled at the beginning of the clock phase but not at all at the end. Under slow conditions the multiple phases are drawn apart and may even not fit into a single clock period. Of course both effects cause considerable non-linearity. To equally subdivide the input clock interval in multiple phases a delay locked-loop is needed to assure that the overall delay along the delay line exactly corresponds to one clock period. If local process variations can be neglected each of the N delay elements in the delay-line has the same delay. The overall delay is the clock period so each delay element is tuned for a delay given by one over N times the clock period. The tuning of the delay elements, however, makes the time domain quantizer not a purely digital circuit. The main field of application so far is high-speed signal capturing, e.g. in serial links and the time domain quantizer in sigma-delta modulators based on time domain signal quantization [9]. In this application a single bit quantizer generates a pulse width encoded signal that is measured and quantized by a time domain quantizer. The measurement value is processed as conversion result in the digital domain and the quantized signal is used as feedback signal within the sigma-delta loop.

4.9 Summary TDC Architectures

In this chapter several TDC architectures have been presented. Each of them is optimized for a particular performance figure. The bipolar TDC, for instance, minimizes the offset error. The looped TDC enables long measurement times, i.e. a

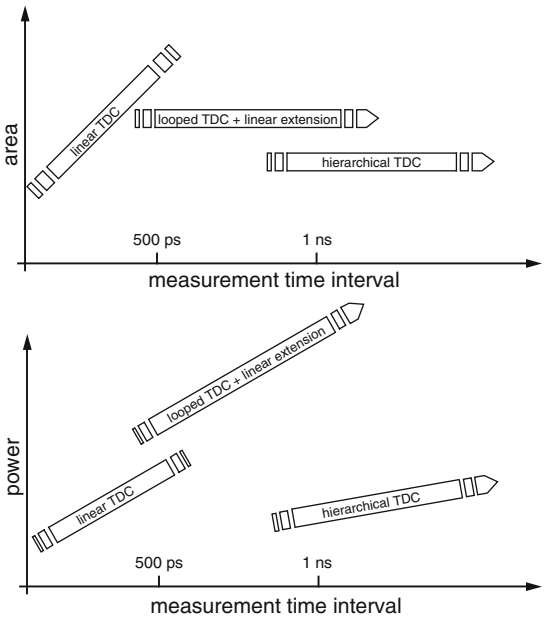

Fig. 4.25 Area and power consumption of TDC architectures depending on the application

high dynamic range. The linearly extended TDC improves the linearity for high dynamic range measurements, etc. To some extend the various concepts may be even combined. The variety of architectures, however, shows that there is not a single time-to-digital converter architecture that is suited and optimized for all applications. In fact the TDC architecture must be optimized for the specific application requirements. This is exemplary illustrated in Fig. 4.25: The most linear TDC architecture is the basic linear TDC. It has a very simple control logic and is suited best for short measurement times. For signed measurements a bipolar TDC consisting of two linear TDCs can be used. The linearity advantage, however, vanishes with increasing dynamic range. Furthermore, the area and power consumption increases with the measurement time so more advanced TDC architectures come into the play. For medium dynamic range a looped TDC can be used. This means that the area does not increase with the measurement time. The power consumption still increases with the measurement time because the full time interval is measured with the high resolution. Linearity challenges may be overcome by a linear extension. For very long measurement times a hierarchical approach is suited best. Its area is independent on the dynamic range and the power consumption scales much less than for the conventional looped TDC. Consequently, the application and its requirements should be carefully analyzed first. Then, in a second step, the most appropriate TDC architecture can be chosen.

Chapter 5
Time-to-Digital Converters with Sub-Gatedelay Resolution – The Third Generation

Abstract So far, the resolution of time-to-digital converters is limited by technology to one inverter delay. For higher resolution circuit techniques are required that do not depend on a single absolute gate delay. This chapter focuses on such techniques, i.e. on TDCs for sub-gate delay resolution. The main concepts are the TDC based on parallel scaled delay elements, the Vernier TDC, the pulse-shrinking TDC, and the local passive interpolation TDC. All concepts are explained in detail and analyzed with respect to resolution, dynamic range, area, power, and variability. Special emphasis is given to the last point in order to estimate the maximum feasible and achievable resolution under manufacturability considerations: In principle all concepts mentioned above can achieve an arbitrarily high resolution. In practice, however, the resolution is limited by process variations. The gated ring oscillator based TDC is discussed as a TDC architecture that has the ability to achieve a high resolution by means of oversampling and noise shaping. Finally, the TDC based on time interval amplification is presented as a concept that does not scale the quantizer references but the measurement residue.

Key words: Sub-Gatedelay Resolution, Vernier TDC, Pulse-Shrinking TDC, Local Passive Interpolation TDC, LPI-TDC, Time Amplification, Gated Ring Oscillator TDC

5.1 Sub-Gate Delay Resolution

A basic time-to-digital converter quantizes a time interval in multiples of a gate delay. Thus, the maximum resolution that can be achieved in a certain technology corresponds to one inverter delay. As this resolution depends only on the technology and not on any circuit design measures it is called the technology resolution T_{tech}. A higher resolution can be achieved only by a faster technology or by one of the circuit techniques discussed in this chapter. Any TDC that achieves a higher than the technology resolution is said to have a sub-gate delay resolution. The ratio

S. Henzler, *Time-to-Digital Converters*, Springer Series in Advanced Microelectronics 29,
DOI 10.1007/978-90-481-8628-0_5,

between the technology resolution and the actual resolution is the so called sub-gate delay interpolation factor, or simply interpolation factor IF:

$$IF := \frac{T_{tech}}{T_{LSB}} \tag{5.1}$$

TDCs with sub-gate delay resolution use sophisticated circuit techniques (resolution enhancement techniques) to circumvent the limitation of the technology delay. The circuit elements in such structures have another loading and other dimensions than in a basic TDC. Hence, the definition of the technology delay/resolution is not so obvious. A reasonable definition is simply the fan-out-2 inverter delay. This is slightly below the maximum resolution of an inverter based time-to-digital converter but clearly defined in all technologies.

5.2 Parallel Scaled Delay Elements

The delay t_d of a digital CMOS gate scales quite linearly with the transistor widths, i.e. the drive current, and the load capacitance. According to the logical effort formalism this can be expressed by

$$t_d = p + gh \tag{5.2}$$

where p is the parasitic delay, g the logical effort, and h the fan-out of the gate ($h = \frac{C_{load}}{C_{in}}$). In a delay-line TDC each gate is loaded by an identical gate so the delay which corresponds to the resolution becomes[1]

$$t_d = p + g \tag{5.3}$$

Configuring the gates not in a chain but in parallel yields the TDC depicted in Fig. 5.1. The start signal is applied to all delay elements in parallel. On the rising edge of the stop signal the outputs of all delay elements are sampled in parallel. If the delay elements DE_n are sized for the delays

$$t_{d,n} = t_d^0 + \Delta t_d \cdot n \tag{5.4}$$

the time difference between the start and the stop signal is quantized with a resolution $T_{LSB} = \Delta t_d$. According to eq. 5.2 the delays can be engineered either by modifying the width of the transistors responsible for the relevant output signal transition or by scaling the load capacitances. Both can be done with a very fine quantization, i.e. the delays can be tuned quasi-continuously. The parasitic delay and the effort delay for driving the sampling registers causes an offset but does not affect the measurement accuracy. This means that now a time interval measurement with a resolution below one gate (inverter) delay is possible. This feature of a TDC is named sub-gate-delay resolution.

[1] Due to the loading of the comparators the delay and so the resolution is even worse.

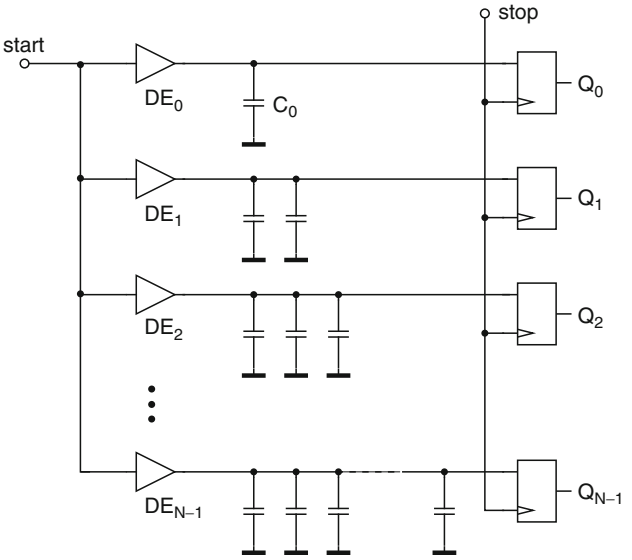

Fig. 5.1 Time-to-digital converter based on parallel scaled delay elements

As in the basic delay-line TDC the stop signal drives all sampling elements, i.e. the corresponding net has a high capacitance so a buffer tree is required. For linearity reasons the skew in this tree has to be minimized. An additional challenge in the TDC based on parallel scaled delay elements is the start signals which has to be connected to all parallel delay elements with a skew smaller than the resolution. Especially for a high dynamic range the balancing of the start and stop nets is challenging which also means that the layout is very critical.

For a maximum dynamic range T_{max}, $N = \frac{T_{max}}{T_{LSB}}$ parallel branches are required. This means N delay elements, N flip-flops, and $\frac{1}{2}N(N+1)$ tuning capacitors. The overall area[2] can be estimated according to

$$A_{core}^{parallel} = N \cdot \left(A^{inv} + A^{FF} + \frac{1}{2}(N+1)A^{cap} \right) \tag{5.5}$$

A^{inv}, A^{FF}, and A^{cap} describe the area of a single delay element, flip-flop, and tuning capacitor, respectively. If the comparators are reset at the beginning of each conversion the average power of the parallel TDC is given by

$$\left\langle P_{core}^{parallel} \right\rangle = f_{meas} \frac{T_{max}}{T_{LSB}} \left(E_{rise}^{inv} + E_{fall}^{inv} + \frac{1}{2}\left(E_{rise}^{FF} + E_{fall}^{FF} \right) + C_{tune}^{tot} V_{DD}^2 \right) \tag{5.6}$$

$$C_{tune}^{tot} = \frac{1}{2} C_0 N (N+1), \qquad N = \frac{T_{max}}{T_{LSB}} \tag{5.7}$$

[2] The core area considers the delay elements, the flip-flops, and the tuning capacitors, but not the thermometer-to-binary decoder and the control logic.

where f_{meas} is the measurement frequency, E_{rise}^{FF} and E_{fall}^{FF} are the energies to sample a logic one or a logic zero, respectively, and E_{rise}^{inv} and E_{fall}^{inv} are the switching energies of the inverters. Resetting of the sampling elements is advantageous to eliminate any history effect from prior conversions, and so unwanted correlation between subsequent samples. Without resetting the switching activity in the comparators is reduced (depending on the signal statistics) so the power consumption can be reduced.

The conversion results are immediately available after the rising edge of the stop signals. The conversion time and latency for the measurement of a time interval T are thus given by

$$T_{conv}^{parallel} = T \tag{5.8}$$

$$T_{latency}^{parallel} = 0 \tag{5.9}$$

These times are not dependent on the resolution which makes the embedding of the TDC into a system quite easy. The delays of the buffer trees and the comparator delay cause a constant offset. For many TDCs these times are even considerably smaller than the other delay components in the conversion time. Hence, buffer trees and comparators are neglected throughout this chapter.

5.2.1 Variability in TDC Based on Parallel Scaled Delay Elements

Local variations of the gate delays alter the switching instances of the parallel delay elements and may even disorder the switching sequence. Especially in deep submicron technologies this is a critical issue. Figure 5.2 shows the switching thresholds of the delay elements without any variations and for an example with local variations. The order of the delay elements may be mixed up if the variations be-

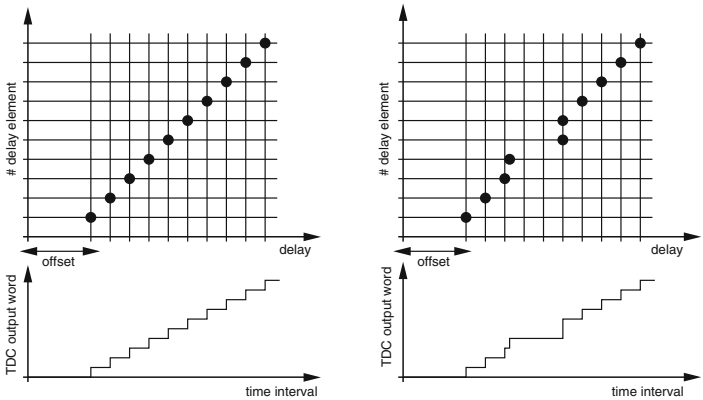

Fig. 5.2 Operating principle of a TDC based on parallel scaled delay elements

come larger than half the resolution. This leads to a high probability of bubble errors. The correction of multi-bit bubbles would require complex bubble correction logic. An alternative solution is to sum up all output bits of the TDC without considering the position of the respective bits [35,39]. This technique provides some robustness against local process variations but is not directly applicable if more than one signal transition occur in the block of parallel delay elements, e.g. for RF signals.

Let's assume that each delay varies with a standard deviation $std(\varepsilon)$. The maximum differential non-linearity is given by the sum of the delay variations of two consecutive delay elements, i.e.

$$max(DNL) = 2\frac{max(\varepsilon)}{T_{LSB}} \tag{5.10}$$

where ε is the delay variation. Statistically T_{LSB} is limited by

$$T_{LSB} > 3\sqrt{2}std(\varepsilon) \tag{5.11}$$

(ref. Section 3.7.4). Consequently, the theoretically infinite resolution is limited by local variations which is typical for TDCs. Figure 5.3 shows simulation (Monte Carlo) results for a 128 bit TDC with Gaussian delay variations (hypothetical variability). The black line with the square markers indicates the average DNL in dependence of the standard deviation of the delay variation. The grey error bars are the respective standard deviations of the differential non-linearity (DNL). Interestingly, the mean DNL scales only moderately with the delay variation. However, the maximum DNL is much larger than pretended by the mean value and its standard deviation. The black triangles show the minimum and maximum values that occurred during the simulation. Table 5.1 summarizes the properties of the TDC based on parallel scaled delay-elements.

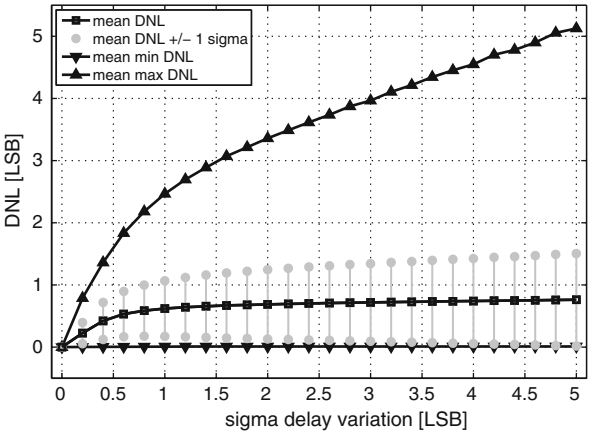

Fig. 5.3 Differential nonlinearity caused by local process variations in a TDC based on parallel scaled delay elements

Table 5.1 Performance summary of parallel scaled delay element TDC

Principle	Parallel delay elements with gradually increasing propagation delays are simultaneously sampled on the arrival of the stop signal
Resolution T_{LSB}	Quasi continuously tuned by transistor or load sizing of parallel delay elements
Number of Stages N	$\frac{T_{max}}{T_{LSB}}$
Core Area $A_{core}^{parallel}$	$N \cdot \left(A^{inv} + A^{FF} + \frac{1}{2}(N+1)A^{cap}\right)$
Average Power $\left\langle P_{core}^{parallel} \right\rangle$	$f_{meas} \frac{T_{max}}{T_{LSB}} \left(E_{rise}^{inv} + E_{fall}^{inv} + \frac{1}{2}\left(E_{rise}^{FF} + E_{fall}^{FF}\right) + C_{tune}^{tot} V_{DD}^2\right)$ $C_{tune}^{tot} = \frac{1}{2} C_0 N (N+1)$
Conversion Time $T_{conv}^{parallel}$	T
Latency $T_{latency}^{parallel}$	0
Loop Structure	No loop structure feasible
PROS	• Sub gate-delay resolution • Conversion time/latency independent from resolution
CONS	• Susceptible to variations • Not feasible for high dynamic range • Careful hand design of each stage

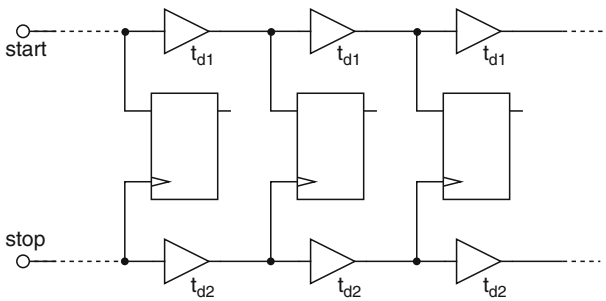

Fig. 5.4 Cut-out of a Vernier delay-line based TDC based on a first delay-line for the start signal and a second one for the stop signal. The arrival time at two corresponding nodes is assessed by early–late detectors such as flip-flops

5.3 Vernier TDC

A delay-line based TDC that is capable of measuring time intervals with a sub gate-delay resolution is the Vernier TDC [4, 6, 10, 24]. As shown in Fig. 5.4, it consist of two delay-lines: One for the start signal and one for the stop signal. The delay

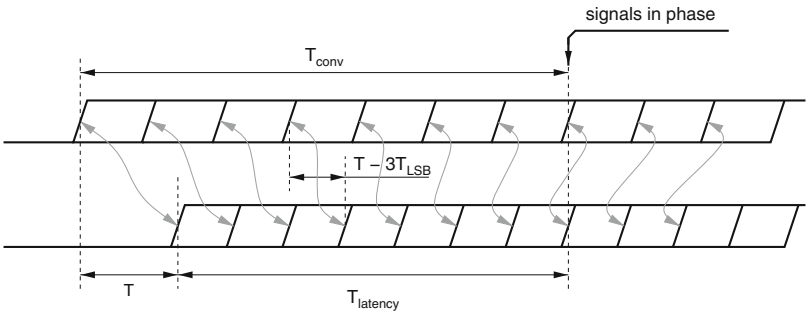

Fig. 5.5 Operating principle of Vernier delay-line based TDC. In each stage the skew between the start and the stop signals is reduced by one $T_{LSB} = t_{d1} - t_{d2}$

elements in the first delay-line have a delay t_{d1} which is slightly larger than the delay t_{d2} of the elements in the second chain. During a measurement the start signal propagates along the first delay-line. The stop signal occurs later, but the delays seen by this signal are smaller. Thus, the stop signal chases the start signal. In each stage it catches up by

$$T_{LSB} = t_{d1} - t_{d2} \tag{5.12}$$

The point (stage) where both signals are in phase is detected by early late detectors (ELDs), usually implemented as flip-flops. Figure 5.5 illustrates the operating principle graphically: The gray arrows indicate two corresponding signals, i.e. the delayed start and the delayed stop signals of a certain stage. At the beginning these signals are skewed by the measurement time T, but each stage reduces the time difference slightly. At the highlighted position the signals are in phase. For all later stages, the original order is reversed and the stop signal is leading the start signal. The granularity in the time domain, i.e. the TDC resolution, is given by the delay difference $T_{LSB} = t_{d1} - t_{d2}$ between the elements in the first and the second delay-line. In principle this difference can be made arbitrarily small, so the resolution does not depend on a gate delay, but on the difference of two gate delays. Consequently, the Vernier TDC provides a circuit technique to overcome the resolution limitations given by a certain technology. To estimate the price for this increased resolution the area, the power consumption and the latency are examined next: As the skew between the signals in the first and the second delay line is reduced by T_{LSB} in each stage, a maximum number of N stages with

$$N = \frac{T_{max}}{T_{LSB}} = \frac{T_{max}}{t_{d1} - t_{d2}} \tag{5.13}$$

is required to measure a maximum time interval T_{max}. Each stage consists of two buffers and a flip-flop. Hence, each LSB of the (core) dynamic range costs four inverters and one flip-flop. The TDC core area[3] is then

[3] The core area considers the delay-lines and the comparators but not the thermometer-to-binary decoder and the control logic.

$$A_{core}^{Vernier} = \frac{T_{max}}{T_{LSB}} \left(4A^{inv} + A^{FF}\right) \tag{5.14}$$

During a measurement, all inverters are switching, independent on the length of the measurement interval. Assuming equally distributed time intervals means that in average half of the flip-flops sample a logic one and the other half a logic zero. The power consumption can thus be estimated according to

$$\left\langle P_{core}^{vernier} \right\rangle = f_{meas} \frac{T_{max}}{T_{LSB}} \left[4\left(E_{rise}^{inv} + E_{fall}^{inv}\right) + \frac{1}{2}\left(E_{rise}^{FF} + E_{fall}^{FF}\right)\right] \tag{5.15}$$

where f_{meas} is the measurement frequency, E_{rise}^{FF} and E_{fall}^{FF} are the energies to sample a logic one or a logic zero, respectively, and E_{rise}^{inv} and E_{fall}^{inv} are the switching energies of the inverters. The latter terms contribute four times to the overall power consumption because each inverter switches once during the measurement, and once in the other switching direction when the circuit is prepared for the next measurement. It can be seen that both the area and the power consumption grow linearly with the maximum measurement interval T_{max} and linearly with T_{LSB}^{-1}. Hence, increasing the resolution by a factor of k increases the area and the power consumption by the same factor. The same holds for the conversion time and the latency that are calculated according to

$$T_{conv}^{Vernier} = \frac{T}{T_{LSB}} t_{d1} = \frac{t_{d1}}{t_{d1} - t_{d2}} T \tag{5.16}$$

$$T_{latency}^{Vernier} = T_{conv} - T = \frac{T}{T_{LSB}} t_{d2} = \frac{t_{d2}}{t_{d1} - t_{d2}} T \tag{5.17}$$

As these times become quite large for high resolution and long measurement intervals one might consider to inject new measurement signals into the delay-lines while the previous ones still propagate in the TDC. In principle this is possible but requires additional overhead for synchronization of the measurement results. The particular bits of the thermometer code become valid one after another and have to be re-sampled by the new measurement signal in order to avoid any overwriting.

Another issue is the dependence of the latency on the time interval to be measured. This has to be considered at the interfacing to the surrounding circuitry. Due to the consecutive arrival of the output bits the thermometer-to-binary decoder of a Vernier TDC is prone to glitches and so high power consumption.

5.3.1 Vernier TDC in Loop Configuration

As discussed in Section 4.2, long measurement times can be covered with reasonable area consumption by applying looped structures. In Vernier TDCs this is even more advisable as the length of the delay-lines grows with the resolution. The basic principle is shown in Fig. 5.6. The delay-line for the start signal (slow delay-line)

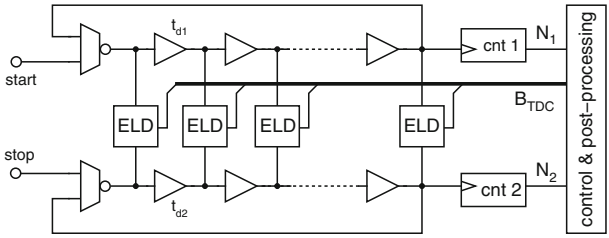

Fig. 5.6 Basic block diagram of a looped Vernier TDC

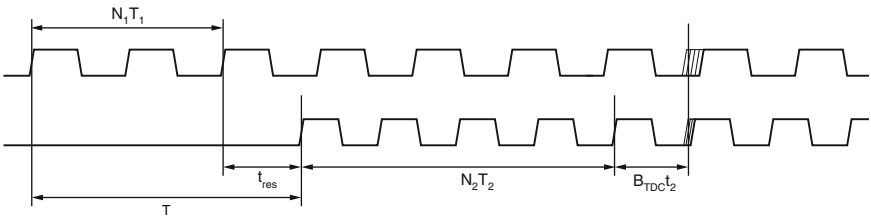

Fig. 5.7 Operating principle of looped Vernier TDC

and the delay-line for the stop signal (fast delay-line) are both configured in a loop configuration. Early–late detectors connect the outputs of each corresponding stage. Although the schematic topology seems trivial, this interconnected loop structure is the main disadvantage of the looped Vernier TDC: For a conventional looped TDC it is already difficult to keep the layout asymmetries and so the non-linearity reasonably small. For two interconnected loops with respectively matched delay elements this is even harder. Hence, the linearity of a looped Vernier TDC is usually degraded compared to its linear implementation. This is disappointing as a loop structure theoretically improves TDC linearity. The operating principle is shown in Fig. 5.7 by means of a signal diagram. On the arrival of the start signal a timing event is injected into the first loop. The second delay-line is quiet. A loop counter cnt_1 determines the number N_1 of full oscillation cycles before the arrival of the stop signal. Hence, the measurement interval T can be partitioned according to

$$T = N_1 \cdot T_1 + t_{res} \tag{5.18}$$

into the coarsely quantized interval $N_1 \cdot T_1$, where T_1 is the oscillation frequency of the slow loop, and a residue interval t_{res}. This residue interval is now quantized with the high resolution of the Vernier concept. Therefore, the stop signal is inserted into the fast delay loop. Early–late detectors indicate when and where inside the loop the two oscillating timing events are in phase. The position is described by B_{TDC}. In general the length of the Vernier delay-line is not sufficient to bring the signals in-phase. Hence, multiple oscillations are required before the two circulating timing events are in phase. A second loop counter cnt_2 determines the number N_2 of full oscillations in the second loop before in the subsequent cycle the in-phase condition

is detected. With M delay elements in the loop, the effective number $B_{vernier}$ of Vernier steps is then

$$B_{Vernier} = M \cdot N_2 + B_{TDC} \tag{5.19}$$

Finally, the time interval can be represented by the output word

$$B = \frac{N_1 \cdot T_1}{T_{LSB}} + B_{Vernier} = \frac{N_1 \cdot T_1}{T_{LSB}} + M \cdot N_2 + B_{TDC} \tag{5.20}$$

$B_{Vernier}$ directly results from the counter cnt_2 and the delay-line information B_{TDC}. The digital representation of the pre-quantization interval $N_1 \cdot T_1$, however, has to be computed from the slow gate delay T_1 and the resolution time T_{LSB}. As the loop delay T_1 is usually not a multiple of the quantization time T_{LSB} a second relative calibration, namely in between T_1 and T_{LSB} is necessary (internal calibration).

Beside the area, the looped Vernier TDC also reduces the conversion time and the latency, as only the residue time interval t_{res} is quantized according to the Vernier principle:

$$T_{conv}^{Vernier} = T + \frac{t_{res}}{T_{LSB}} t_{d2} \tag{5.21}$$

$$T_{latency}^{Vernier} = \frac{t_{res}}{T_{LSB}} t_{d2} \tag{5.22}$$

These times still depend on the time interval T, not linearly but rather periodically. A similar implementation [6] that greatly relaxes layout constraints does not use ELDs between all delay-elements but a single ELD at one position of the two Vernier delay loops. Now the complete loops have a delay difference corresponding to the resolution time. The conversion time goes up, but the linearity is usually better due to the more regular layout and the missing matching requirements inside the loops.

5.3.2 Variability in Vernier TDC

In this section the susceptibility of the Vernier TDC to local process variations is investigated according to the methodology presented in Section 3.7.4. Again two stages within the delay-line are considered. The start signal is injected into the first delay line at $t = 0$. It becomes visible in the nth and $(n+1)$th stage, respectively, at the times

$$t_{1,n} = n t_{d1} + \sum_{i=1}^{n} \varepsilon_{1,i} \tag{5.23}$$

$$t_{1,n+1} = (n+1) t_{d1} + \sum_{i=1}^{n+1} \varepsilon_{1,i} \tag{5.24}$$

where t_{d1} describes the nominal delay in the first delay-line and $\varepsilon_{1,i}$ the respective delay variations. The stop signal is injected into the second delay-line at the instance $t = T$ and becomes visible at the nth and $(n+1)$th stage at the times

$$t_{2,n} = T + nt_{d2} + \sum_{i=1}^{n} \varepsilon_{2,i} \tag{5.25}$$

$$t_{2,n+1} = T + (n+1)t_{d2} + \sum_{i=1}^{n+1} \varepsilon_{2,i} \tag{5.26}$$

The delay and the variation in the second delay-line are modeled by t_{d2} and $\varepsilon_{2,i}$, respectively. For the sake of simplicity the nominal blackout times of the registers are set to zero. The blackout time variations are given by

$$t_n^{BO} = \beta_n \tag{5.27}$$

$$t_{n+1}^{BO} = \beta_{n+1} \tag{5.28}$$

The nth stage samples a logic high signal under the condition

$$t_{2,n} - t_{1,n} > t_n^{BO} \tag{5.29}$$

$$\Leftrightarrow T > t_{s,n} := n\left(t_{d1} - t_{d2}\right) + \sum_{i=1}^{n} \varepsilon_{1,i} - \sum_{i=1}^{n} \varepsilon_{2,i} + \beta_n \tag{5.30}$$

and the consecutive stage $(n+1)$ under the condition

$$t_{2,n+1} - t_{1,n+1} > t_{n+1}^{BO} \tag{5.31}$$

$$\Leftrightarrow T > t_{s,n+1} := (n+1)\left(t_{d1} - t_{d2}\right) + \sum_{i=1}^{n+1} \varepsilon_{1,i} - \sum_{i=1}^{n+1} \varepsilon_{2,i} + \beta_{n+1} \tag{5.32}$$

The resulting differential non-linearity is thus

$$DNL_n = \frac{1}{T_{LSB}}\left(t_{s,n+1} - t_{s,n} - T_{LSB}\right) \tag{5.33}$$

$$= \frac{1}{T_{LSB}}\left(\varepsilon_{1,n+1} - \varepsilon_{2,n+1} + \beta_{n+1} - \beta_n\right) \tag{5.34}$$

This non-linearity varies with a standard variation of

$$\mathrm{std}(DNL_n) = \frac{1}{T_{LSB}}\sqrt{\mathrm{std}^2\left(\varepsilon_1\right) + \mathrm{std}^2\left(\varepsilon_2\right) + 2\mathrm{std}^2\left(\beta\right)} \tag{5.35}$$

Compared to the basic TDC of Section 3.7.4 the variability of the second delay-line additionally contributes to the DNL. With the same performance criterion, namely

$$3 \cdot std(DNL) < 1 \tag{5.36}$$

$$\Leftrightarrow 3\sqrt{\mathrm{std}^2\left(\varepsilon_1\right) + \mathrm{std}^2\left(\varepsilon_2\right) + 2\mathrm{std}^2\left(\beta\right)} < T_{LSB} \tag{5.37}$$

Table 5.2 Performance summary of Vernier TDC

Principle	Start and stop signals propagate in two delay-lines with slightly different delays.
Resolution T_{LSB}	$t_{d1} - t_{d2}$
Number of Stages N	$\frac{T_{max}}{T_{LSB}}$
Core Area $A_{core}^{Vernier}$	$\frac{T_{max}}{T_{LSB}} \left(4A^{inv} + A^{FF}\right)$
Average Power $\left\langle P_{core}^{vernier} \right\rangle$	$f_{meas} \frac{T_{max}}{T_{LSB}} \left[4 \left(E_{rise}^{inv} + E_{fall}^{inv}\right) + \frac{1}{2} \left(E_{rise}^{FF} + E_{fall}^{FF}\right)\right]$
Conversion Time $T_{conv}^{Vernier}$	$\frac{T}{T_{LSB}} t_{d1} = \frac{t_{d1}}{t_{d1} - t_{d2}} T$
Latency $T_{latency}^{Vernier}$	$\frac{T}{T_{LSB}} t_{d2} = \frac{t_{d2}}{t_{d1} - t_{d2}} T$
Loop Structure	Loop structure possible
PROS	• Sub gate-delay resolution • Modular structure • High DR feasible with loop structure
CONS	• Two long delay-lines • Conversion time and latency depend on measurement interval and resolution • Consecutive arrival of output bits • Glitches in thermometer-to-binary decoder

it becomes obvious that local variations become more critical in the Vernier TDC for two reasons: First the resolution is increased, i.e. T_{LSB} is reduced by the interpolation factor, and second there is the additional variability of the second delay-line. The properties and performance figures of the Vernier TDC are summarized in Table 5.2.

5.4 Pulse-Shrinking TDC

The pulse-shrinking TDC [5, 41, 56] is another approach to achieve sub-gate delay resolution by circuit techniques. The basic idea is to form a pulse with a rising edge defined by the start signal and a falling edge defined by the stop signal. Hence, the pulse-width T is equivalent to the time interval to be measured. The pulse generator can be omitted if the signal to be measured is already pulse-width encoded. For the purpose of time-to-digital conversion the measurement pulse propagates along a delay-line that has an intentional asymmetry in each stage. This asymmetry causes a modification (usually a reduction) of the pulse width. While the pulse propagates along the delay-line it becomes smaller and smaller and vanishes completely at some

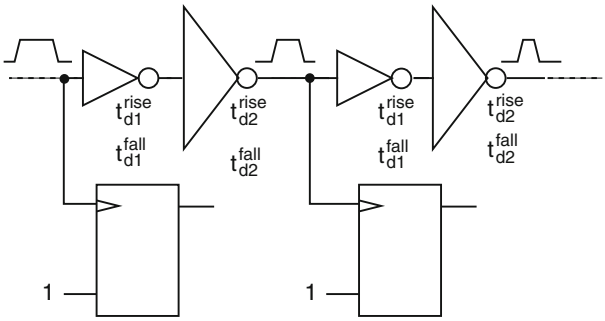

Fig. 5.8 Cut-out of pulse-shrinking TDC. The start and the stop signals form a measurement pulse with a pulse width that is reduced in each stage by one T_{LSB}

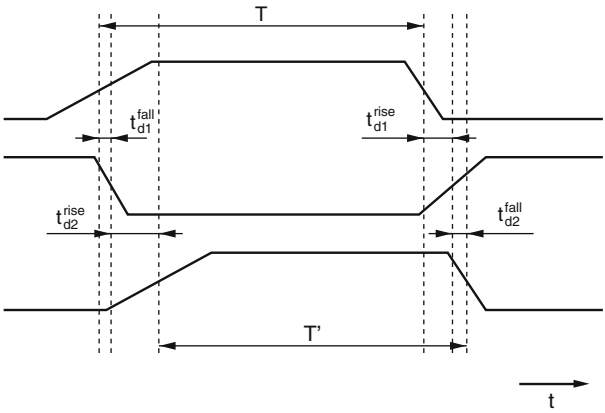

Fig. 5.9 Operating principle of the pulse-shrinking TDC. Input signal (top), intermediate signal (middle), and output signal (bottom) of pulse-shrinking element

position. Flip-flops that are reset before the measurement have a constant logic one at their data input and a clock input which is connected to the delay-line. After the measurement, the one-zero transition point in the thermometer code seen at the outputs of the flip-flops indicates the position where the pulse has vanished. As each stage reduces the pulse width by one T_{LSB}, this position is a measure for the time interval T. A cut-out of a basic pulse-shrinking TDC is depicted in Fig. 5.8. For a detailed understanding the signals in a single delay stage are illustrated in Fig. 5.9. The upper signal is the input and the lower the output signal of a single pulse-shrinking stage. The signal in the middle stands for the intermediate node. The leading edge of the pulse is inverted by the first inverter and experiences a delay of t_{d1}^{fall}. In the second inverter the signal is inverted again and is delayed by t_{d2}^{rise}. When the trailing edge passes the two inverters it experiences different delays (as the switching direction is opposite to the leading edge), namely t_{d1}^{rise} and t_{d2}^{fall}. The pulse width T' after the pulse shrinking stage can be calculated according to

$$\begin{aligned} T' &= T + t_{d1}^{rise} + t_{d2}^{fall} - t_{d1}^{fall} - t_{d2}^{rise} \\ &= T + \left[\left(t_{d1}^{rise} - t_{d1}^{fall} \right) - \left(t_{d2}^{rise} - t_{d2}^{fall} \right) \right] \end{aligned} \tag{5.38}$$

Thus, the original pulse width T is reduced by one T_{LSB} defined by

$$T_{LSB} = - \left(t_{d1}^{rise} - t_{d1}^{fall} \right) + \left(t_{d2}^{rise} - t_{d2}^{fall} \right) \tag{5.39}$$

This term shows that no pulse modification occurs if the rising and falling delays are equal in a certain stage or if the absolute delay difference is equal in each stage. Otherwise the pulse width is modified.[4] In principle, both a pulse-shrinking and a pulse-growing is possible. If the TDC is planned for shrinking pulses, it has to be assured, that the actual pulse shrinks under all process and environmental variations.

The pulse width after N stages is reduced to $T' = T - N \cdot T_{LSB}$, i.e. for a maximum measurement interval T_{max}

$$N = \frac{T_{max}}{T_{LSB}} = \frac{T_{max}}{- \left(t_{d1}^{rise} - t_{d1}^{fall} \right) + \left(t_{d2}^{rise} - t_{d2}^{fall} \right)} \tag{5.40}$$

stages are required. Each stage consists of two inverters and a flip-flop, so the core area of the pulse-shrinking TDC can be described by

$$A_{core}^{pulse-shrinking} = \frac{T_{max}}{T_{LSB}} \left(2A^{inv} + A^{FF} \right) \tag{5.41}$$

For the estimation of the average core power consumption equally distributed input time intervals are assumed. Only the inverters of pulse-shrinking stages where the pulse still exists do toggle, i.e. half of all inverters in average. The core power consumption is then

$$\left\langle P_{core}^{pulse-shrinking} \right\rangle = f_{meas} \frac{T_{max}}{T_{LSB}} \left(E_{rise}^{inv} + E_{fall}^{inv} + \frac{1}{2} E_{rise}^{FF} + \frac{1}{2} E_{reset}^{FF} \right) \tag{5.42}$$

As in the Vernier TDC the area and the power consumption linearly depend on maximum measurement time T_{max} and on the reciprocal resolution T_{LSB}^{-1}. When the pulse vanishes its leading and trailing edges coincide. This results in the conversion time and latency according to

[4] In fact each asymmetry in a delay-line causes a pulse shrinking or growing. In the pulse-shrinking TDC this is used for the conversion, however, in all other TDCs this is a parasitic effect. If a pulse propagates in a conventional TDC (for instance when a duty-cycle, a phase, or a period shall be measured) its pulse width is modified. This limits the maximum measurement interval, even if the TDC architecture in principle allows for arbitrary time intervals (e.g. looped TDC). Parasitic pulse shrinking gives also rise to a resolution versus measurement time trade-off, as high resolution usually means many delay stages that might modify the pulse width.

The only TDC known so far, that is totally robust against pulse shrinking is the linearly extended TDC with pulse generator that is discussed in Section 4.3.

$$T_{conv}^{pulse-shrinking} = \frac{T}{T_{LSB}} \left(t_{d1}^{fall} + t_{d2}^{rise} \right) \tag{5.43}$$

$$T_{latency}^{pulse-shrinking} = \frac{T}{T_{LSB}} \left(t_{d1}^{rise} + t_{d2}^{fall} \right) \tag{5.44}$$

Both the conversion time and the latency linearly depend on the measurement time and the reciprocal resolution. The dependence on the measurement time is an issue for the interfacing.[5] The strong dependence on the resolution causes a trade-off between the maximum measurement time and the resolution. A pipelining as for the Vernier TDC is not easily possible as it depends on the input signal whether a transition occurs at a certain stage during the next conversion or not. A critical issue for the pulse-shrinking TDC is the minimum pulse width that can be detected by the flip-flops. Before the pulse vanishes, it becomes very small and does not reach the full logic level anymore. This condition makes the sampling elements prone to any PVT variations.

5.4.1 Pulse-Shrinking TDC in Loop Configuration

For reduced area consumption and better linearity the pulse-shrinking TDC can be implemented in a loop structure, too. The principle is the same as for the conventional looped TDC (ref. Section 4.2) or the looped Vernier TDC (ref. Section 5.3.1). In a straightforward approach the measurement pulse is propagated in a ring/loop of pulse-shrinking elements. The drawback is that the complete pulse must fit into the loop, i.e. the minimum number of pulse-shrinking elements is given by

$$N_{min} = \frac{T_{max}}{t_{d1}^{fall} + t_{d2}^{rise}} \tag{5.45}$$

The delay of the two inverters in each pulse-shrinking element is approximately two times the resolution the technology can accomplish (T_{native}). Hence the loop structure reduces the number of delay elements by a factor of

$$\frac{N}{N_{min}} = \frac{2T_{native}}{T_{LSB}} = 2 \cdot IF \tag{5.46}$$

This is only a moderate reduction. A loop architecture that saves even more pulse-shrinking elements, i.e. area, works similar to the looped Vernier TDC: No explicit pulse generator is used in front of the loop. Moreover, the leading edge enters a short loop with an inverting characteristic. This edge oscillates in the loop until the trailing edge arrives. A loop counter determines the number of full loop cycles. During the oscillation the duty cycle is 50%, i.e. the pulse width corresponds to the

[5] If the pulse has vanished all subsequent flip-flops do not switch anymore. Not switching, however, is not an event that is usable to indicate the end of a measurement. Thus, in practice each measurement lasts as long as the measurement of the longest time interval.

delay of the loop. On the arrival of the trailing edge an additional signal transition opposite to the one currently propagating in the loop is injected. This forms a pulse that represents the residue time interval. The length of this residue interval is now determined via the pulse shrinking technique. The length of the loop is no longer determined by the measurement interval but by the latency in the control circuitry.

5.4.2 Variability in Pulse-Shrinking TDC

Local process variations also degrade the linearity of the pulse-shrinking TDC. To investigate the variation impact each gate delay is expressed by a constant term describing the nominal value and a variation induced perturbation ε:

$$t_{d1,i}^{rise} = t_{d1}^{rise} + \varepsilon_{d1,i}^{rise} \tag{5.47}$$

$$t_{d1,i}^{fall} = t_{d1}^{fall} + \varepsilon_{d1,i}^{fall} \tag{5.48}$$

$$t_{d2,i}^{rise} = t_{d2}^{rise} + \varepsilon_{d2,i}^{rise} \tag{5.49}$$

$$t_{d2,i}^{fall} = t_{d2}^{fall} + \varepsilon_{d2,i}^{fall} \tag{5.50}$$

The pulse width after n pulse-shrinking stages can be expressed as

$$T_n = T + \sum_{i=1}^{n} \left[t_{d1,i}^{rise} + t_{d2,i}^{fall} - t_{d1,i}^{fall} - t_{d2,i}^{rise} \right] \tag{5.51}$$

where T is the original pulse width defining the measurement interval. In the nth stage a logic high value is sampled if the pulse width T_n is larger than the minimum pulse width

$$T_{min,n} = T_{min}^{0} + \gamma_n \tag{5.52}$$

Local process variations in the sampling elements cause a variation of this minimum pulse width. For each sampling element this pulse width variation is modeled by γ_n. Using eq. 5.51, the condition for sampling a logic high value in stage n is

$$T > t_{s,n} := T_{min,n} + \sum_{i=1}^{n} \left[t_{d1,i}^{fall} + t_{d2,i}^{rise} - t_{d1,i}^{rise} - t_{d2,i}^{fall} \right] \tag{5.53}$$

Therewith, the differential non-linearity can be calculated according to

$$DNL_n = \frac{1}{T_{LSB}} \left(t_{s,n+1} - t_{s,n} - T_{LSB} \right) \tag{5.54}$$

$$= \frac{1}{T_{LSB}} \left(\gamma_{n+1} - \gamma_n + t_{d1,n+1}^{fall} + t_{d2,n+1}^{rise} - t_{d1,n+1}^{rise} - t_{d2,n+1}^{fall} - T_{LSB} \right) \tag{5.55}$$

$$= \frac{1}{T_{LSB}} \left(\gamma_{n+1} - \gamma_n + \varepsilon_{d1,n+1}^{fall} + \varepsilon_{d2,n+1}^{rise} - \varepsilon_{d1,n+1}^{rise} - \varepsilon_{d2,n+1}^{fall} \right) \tag{5.56}$$

This non-linearity varies with a standard deviation of

$$\mathrm{std}\left(DNL_n\right) = \frac{1}{T_{LSB}}\sqrt{2\mathrm{std}\left(\gamma\right)^2 + \mathrm{std}\left(\varepsilon_{d1}^{rise}\right)^2 + \mathrm{std}\left(\varepsilon_{d1}^{fall}\right)^2 + \mathrm{std}\left(\varepsilon_{d2}^{rise}\right)^2 + \mathrm{std}\left(\varepsilon_{d2}^{fall}\right)^2} \tag{5.57}$$

Using again a three sigma criterion results in a resolution that is limited according to

$$1 > 3 \cdot \mathrm{std}\left(DNL\right) \tag{5.58}$$

$$T_{LSB} > 3\sqrt{2\mathrm{std}\left(\gamma\right)^2 + \mathrm{std}\left(\varepsilon_{d1}^{rise}\right)^2 + \mathrm{std}\left(\varepsilon_{d1}^{fall}\right)^2 + \mathrm{std}\left(\varepsilon_{d2}^{rise}\right)^2 + \mathrm{std}\left(\varepsilon_{d2}^{fall}\right)^2}$$

In principle the variation of the pulse-shrinking TDC is similar to the Vernier TDC. However, the variability of the minimum pulse width is often larger than the variation of the blackout time. The properties and performance figures of the pulse-shrinking TDC are summarized in Table 5.3.

Table 5.3 Performance summary of pulse-shrinking TDC

Principle	Start and stop signal form a pulse that propagates in a delay-line. Its width is reduced in each stage of the delay line. The position where the pulse vanishes is detected and used as measure for the time interval.
Resolution T_{LSB}	$-\left(t_{d1}^{rise} - t_{d1}^{fall}\right) + \left(t_{d2}^{rise} - t_{d2}^{fall}\right)$
Number of Stages N	$\frac{T_{max}}{T_{LSB}}$
Core Area $A_{core}^{pulse-shrinking}$	$\frac{T_{max}}{T_{LSB}}\left(2A^{inv} + A^{FF}\right)$
Average Power $\left\langle P_{pulse-shrinking}^{core}\right\rangle$	$f_{meas}\frac{T_{max}}{T_{LSB}}\left(E_{rise}^{inv} + E_{fall}^{inv} + \frac{1}{2}E_{rise}^{FF} + \frac{1}{2}E_{reset}^{FF}\right)$
Conversion Time $T_{conv}^{pulse-shrinking}$	$\frac{T}{T_{LSB}}\left(t_{d1}^{fall} + t_{d2}^{rise}\right)$
Latency $T_{latency}^{pulse-shrinking}$	$\frac{T}{T_{LSB}}\left(t_{d1}^{rise} + t_{d2}^{fall}\right)$
Loop Structure	Loop structure possible
PROS	• Sub gate-delay resolution • Modular structure • High DR feasible with loop structure
CONS	• Conversion time and latency depend on measurement interval and resolution • Consecutive arrival of output bits • Minimum pulse width variation • Glitches in thermometer-to-binary decoder • No stop event

5.5 Local Passive Interpolation TDC

Another approach [20] for sub gate-delay resolution that does not depend on the ratio of two delays is the local passive interpolation TDC (LPI-TDC). The basic principle is illustrated in Fig. 5.10. Assume that the two thick lines are two logically equivalent signals within a delay-line with a skew corresponding to one inverter delay t_d^{rise}. Higher temporal resolution means that there are further signals in between these two base signals (thin lines). One important fact for the understanding of the interpolation principle is the circumstance that the rise time of a signal produced by a single stage CMOS gate is in the order of the gate-delay. This means that the output signal is switching while the corresponding input signal has not settled yet. Linear interpolation according to

$$V_{int,i} = V_B + a_i (V_A - V_B) \tag{5.59}$$

leads to a new signal $V_{int,i}$ that lies right in between the base signals. The interpolation coefficient a_i weights the two base signals and determines whether the new signal is nearer to V_A or nearer to V_B. In the simplest case one interpolated signal is generated, and the interpolation coefficient is $a_i = \frac{1}{2}$. The resolution is doubled, i.e. the interpolation factor IF is equal to two. For $IF = 3$ two interpolated signals have to be generated and the interpolation factors are $a_i = \{\frac{1}{3}, \frac{2}{3}\}$. For $IF = 4$, $a_i = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ and in general $a_i = \frac{i}{IF} \quad i = 1 \ldots (IF - 1)$. In the region where both base signals are switching the interpolated signal(s) is (are) parallel to the original signals, i.e. appear(s) as a shifted version(s) of one of the base signal. For an interpolation factor IF, all interpolated signals are parallel to the base signals around the midlevel if the rise time and the delay follow the relation

$$t_d^{rise} < \frac{1}{2} t^{rise} \frac{IF}{IF - 1} \tag{5.60}$$

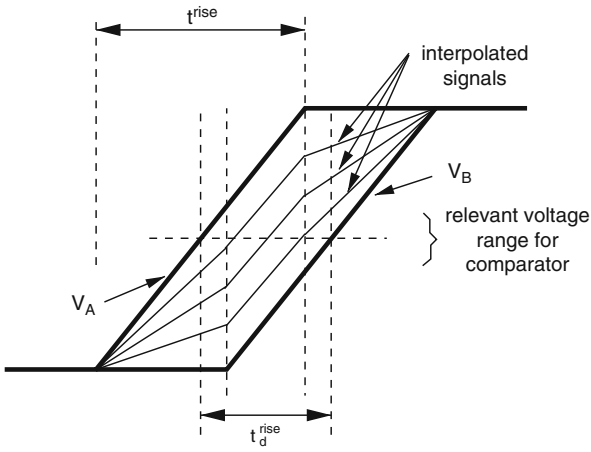

Fig. 5.10 Principle of local passive interpolation TDC

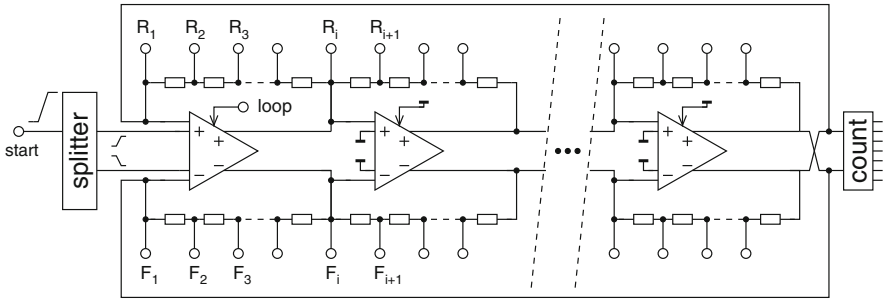

Fig. 5.11 Basic circuit diagram of local passive interpolation TDC

The midlevel is particularly important, as it usually indicates the threshold of the comparators that sample the delay-line.

Conventional CMOS gates require at least two inverter delays to generate two logically equivalent signals. As discussed above, the interpolation principle requires that the delay is small enough compared to the rise time. Hence, a circuit is required that provides two logically equivalent signals with a skew of just one inverter delay. This can be achieved by a differential delay-line that propagates both the start as well as the inverted start signal. Figure 5.11 shows a basic circuit diagram of a local passive interpolation TDC with $IF = 4$. The delay-line consists of differential delay elements. In the simplest case each of these delay elements comprises just two parallel inverters. Coupling elements between these inverters, however, are strongly recommended in order to avoid a drift of the complementary signals. At the input of the delay-line a differential start signal is injected for propagation along the line. This means that after each stage an inverted and a non-inverted copy of the start signal is available. The signals of two subsequent stages have each a skew of only one inverter delay. Hence, the prerequisites for the interpolation according to Figure 5.10 are fulfilled.

The interpolation is done by resistive voltage dividers connected in between the inputs of the differential delay elements and their respective logically equivalent outputs. Integrated resistors are known for high variability (typically $\pm 20\%$). For the interpolation, however, only the ratio of two resistances is relevant, so global variations cancel out. This means, that for the interpolation itself, the absolute values of the resistors are not relevant. This is only true, as long as the time constants caused by the parasitic capacitances in between two interpolation resistors are negligible with respect to the rise time of the delay elements. A detailed discussion of that issue is provided in Section 5.5.2. Ohmic resistors, in particular poly-silicon resistors, are suited best for the interpolators. The reasons are linearity, symmetry of the device, and the absence of any threshold voltage. Alternatively, transistors or diodes can be used. However, as high switching speed usually requires very low resistance values, the area advantage of active devices is small.

On the arrival of the stop signal the state of the delay-line has to be frozen. Therefore, comparators/flip-flops can be connected to the set of rising signals $\{R_i\}$ or

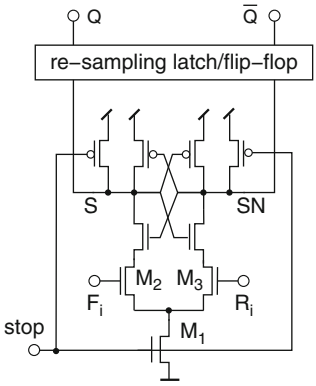

Fig. 5.12 Fully differential sense amplifier based comparator

to the set of falling signals $\{F_i\}$. To take full advantage of the differential delay-line, fully differential comparators such as the sense-amplifier based comparator in Figure 5.12, can be connected to both $\{R_i\}$ and $\{F_i\}$. While the stop signal is low, the internal nodes S and SN are pre-charged to V_{DD}. All other nodes reach a level of approximately $V_{DD} - V_{tn}$, where V_{tn} is the NMOS threshold voltage. By this pre-charge any memory effect from prior comparisons is removed. On the rising edge of the stop signal, the pre-charge devices are turned off and the footer device is activated. Although the footer is not operated as a current source, the differential pair acts somehow like a differential amplifier during the comparison. Depending on which input node carries a higher potential either S or SN is discharged faster than the other node. The regenerative structure (cross coupled inverter pair loading the differential pair) amplifies this trend until one node is fully discharged to V_{SS} and the other one is re-charged to V_{DD}. Due to symmetry the switching threshold of this comparator is exactly in between V_{DD} and V_{SS} and does not depend on the switching threshold of any gate. Hence, as long as the matching among corresponding transistors in the comparator is good the circuit is robust against process, voltage, and temperature variations.

In the LPI-TDC, IF steps are resolved within one stage. Compared to the Vernier or pulse-shrinking TDC, the number of delay stages

$$N = \frac{T_{max}}{t_d} = \frac{1}{IF}\frac{T_{max}}{T_{LSB}} \tag{5.61}$$

is thus reduced by the interpolation factor IF. As N depends on the stage delay t_d (not on T_{LSB}) the resolution can be increased without making the delay-line longer. This is a great advantage as the delay-line length versus resolution trade-off does not exist anymore. Another result is the conversion time

$$T_{conv}^{LPI} = T \tag{5.62}$$

which is only dependent on the measurement interval T but not on the resolution T_{LSB}. Immediately after the arrival of the stop event the conversion result becomes visible at the TDC output, i.e.

$$T_{latency}^{LPI} = 0 \tag{5.63}$$

A zero latency makes the embedding of a TDC in its surrounding system environment easy. Thus, the LPI-TDC enables sub gate-delay resolution but can be used as a conventional delay-line TDC without resolution enhancement techniques. Its area can be estimated according to

$$A_{core}^{LPI} = \frac{T_{max}}{T_{LSB}} \left(\frac{2.2}{IF} A^{inv} + A^{FF} + 2A^{R} \right) \tag{5.64}$$

The factor 2.2 takes the effort for the differential structure, e.g. the coupling elements, into account. The estimation of the average power consumption can be qualitatively done by

$$\begin{aligned} \left\langle P_{core}^{LPI} \right\rangle = f_{meas} \frac{1}{IF} \frac{T_{max}}{T_{LSB}} \Big(& 2.2(E_{rise}^{inv} + E_{fall}^{inv}) + IF \cdot E^{FF} \\ & + IF \cdot E_{precharge}^{FF} + 4E_{x-curr} \Big) \end{aligned} \tag{5.65}$$

An analytic separation of the inverter energies (E_{rise}^{inv} and E_{fall}^{inv}) and the cross-current energy (E_{x-curr}) is not that easy, as smaller interpolation resistors mean higher cross-currents but also stronger inverters. Hence, eq. 5.65 should be used to identify the particular power components. For precise power analysis a simulation of the delay cell should be done.

5.5.1 LPI-TDC in Loop Configuration

The input–output behavior of a LPI-TDC corresponds exactly to a conventional delay-line TDC except that the resolution is improved. The conversion time and the latency are not degraded by the resolution enhancement. Consequently, the loop architecture of an LPI-TDC is straight forward and can be done as described in Section 4.2. Non-inverting loops and pulsed operation are preferable to avoid cross currents in the interpolators during idle mode or signal injection.

5.5.2 Resistor Sizing and Interpolation Accuracy

The interpolators of the LPI-TDC are resistive voltage dividers. The divider ratio is given by a resistance ratio and does not depend on the absolute resistance. Yet, this is only true if the voltage divider is not loaded, i.e. if no current is drawn from its intermediate nodes. In a TDC, however, comparators are connected to each node

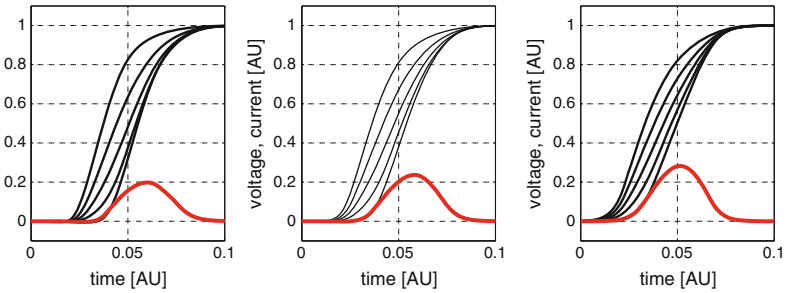

Fig. 5.13 Trade-off between linearity and power consumption in LPI-TDC

in the delay-line and each interpolated node. The input capacitance of these comparators causes a load at each interpolated node. The interpolators become reactive sub-circuits and the voltage divider formula is not exact anymore. In fact the step response of each interpolated node is not a step, but a decaying exponential function. This exponential functions have a time constant that depends on the interpolation resistance. The deviation from the ideal interpolation factors reflects in a systematic non-linearity, precisely a DNL inside each interpolation interval. Hence, the theoretically irrelevant value of the resistors have to be low enough to keep the DNL small with respect to the quantization time T_{LSB}. On the other hand the cross-currents and so the power consumption in the interpolators grow with decreasing interpolation resistance: A design trade-off that is illustrated in Fig. 5.13. The left graph shows a scenario where the resistors have large values. The current profile is low, but the time interval between the two base signals is by far not equally partitioned by the interpolated signals. A nearly perfect interpolation is shown in the right graph, but the power consumption is considerably increased. A compromise that enables acceptable DNL with reasonable power consumption is shown in the middle. It depends on the application and the available power budget how this trade-off can be resolved. An analytical strategy to investigate the linearity of the passive interpolation is shown in Fig. 5.14. According to the superposition theorem the stimuli are separated into the sources V_1 and V_2. Both cases are analyzed separately and combined later on. Each signal source representing a signal transition between V_{SS} and V_{DD} (with rise time t_{rise}) is further decomposed into two ramp signals. The overall result is thus the superposition of four ramp responses. To calculate each of these ramp responses, a node voltage analysis is performed:

$$\begin{bmatrix} 0 & G & 0 & 0 & 0 \\ G & sC & -G & 0 & 0 \\ 0 & -G & sC+2G & -G & 0 \\ 0 & 0 & -G & sC+2G & -G \\ 0 & 0 & 0 & -G & sC+2G \end{bmatrix} \underline{v_k} = \begin{bmatrix} GV \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5.66}$$

Solving for the particular node voltages $\underline{v_k}$ and dividing by the stimulus V results in the transfer functions

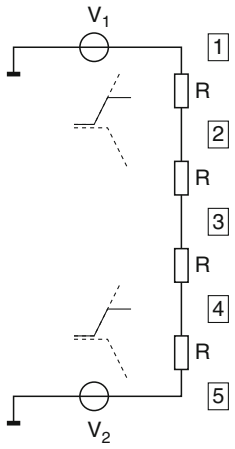

Fig. 5.14 Theoretical investigation of interpolator network

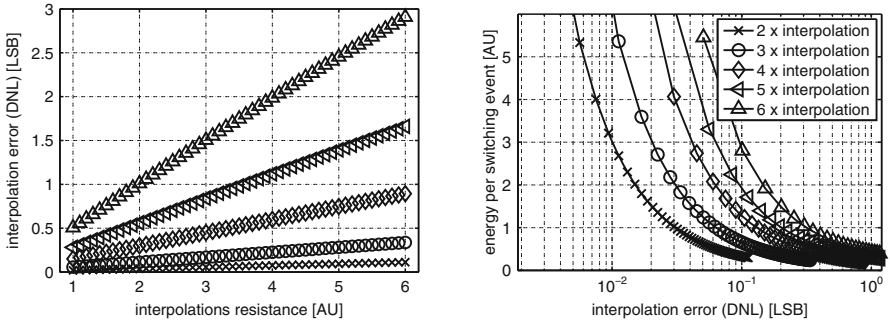

Fig. 5.15 Power versus linearity trade-off for the choice of the interpolation resistance of an LPI-TDC: Differential non-linearity over interpolation resistance for $IF = 2$ to $IF = 6$ (left) and energy per switching event over differential non-linearity (right)

$$H_k(s) = \frac{v_k(s)}{V} \tag{5.67}$$

Multiplication of these transfer functions with $\frac{V_{DD}}{t_{rise}} \frac{1}{s^2}$ yields the frequency domain representation of the ramp responses, i.e.

$$v_k^{ramp}(t) = \mathscr{L}^{-1}\left(\frac{V_{DD}}{t_{rise}} \frac{H_k(s)}{s^2}\right) \tag{5.68}$$

The analytic formulation of the interpolated signals is advantageous to investigate the power versus linearity trade-off. This trade-off is shown in the two graphs of Fig. 5.15. The left graph shows the maximum differential non-linearity that occurs in the interpolator depending on the interpolation resistance for interpolation factors from $IF = 2$ to $IF = 6$. It can be seen that the DNL grows quite linearly with the

interpolation resistance. For the minimum resistance a reasonable value from a layout perspective is chosen. It is a reasonable constraint for most applications to limit the maximum DNL to half an LSB. The right graph shows the energy per switching event over the maximum differential non-linearity in the interpolator. With decreasing resistance, i.e. with decreasing DNL, the energy grows faster and faster. It is also worth mentioning that for a given power consumption the DNL gets worse with increasing interpolation factor IF. The axis are normalized, as the systematic DNL depends on the gate delay t_d that is subdivided. This is due to the fact that the time constant is given by RC and does neither depend on the gate delay nor on the rise-time of the driving inverters.

5.5.3 Variability in LPI-TDC

For the local passive interpolation TDC, the investigation of the effects of local variations is slightly different from the other TDCs discussed so far. This is due to the fact that some of the delayed signals come from the interpolators and not directly from the delay elements. Hence, the number n of a delayed signal is decomposed according to

$$n = n_c(n) + n_i(n) \tag{5.69}$$

$$n_c(n) = n \operatorname{div} IF \tag{5.70}$$

$$n_i(n) = n \bmod IF \qquad n_i \in \{0, \ldots, IF-1\} \tag{5.71}$$

where n_c indicates the number of the delay element and n_i the number of the interpolated signal associated with this delay element. If the start signal is injected into the delay-line at $t = 0$ it arrives at the n_cth delay element at the time

$$t_{n_c} = n_c \cdot IF \cdot T_{LSB} + \sum_{i=1}^{n_c} \varepsilon_i \tag{5.72}$$

where ε_i describes the delay variation of the ith element caused by LPV. Therewith, the nth signal crosses the trigger threshold of its comparator at

$$t_n = t_{n_c(n)} + \frac{n_i(n)}{IF} t_{d,n_c+1} + \delta_n \tag{5.73}$$

$$= [n_c(n) \cdot IF + n_i(n)]\, T_{LSB} + \sum_{i=1}^{n_c(n)} \varepsilon_i + \frac{n_i(n)}{IF} \varepsilon_{n_c+1} + \delta_n \tag{5.74}$$

$$= n \cdot T_{LSB} + \sum_{i=1}^{n_c(n)} \varepsilon_i + \frac{n_i(n)}{IF} \varepsilon_{n_c+1} + \delta_n \tag{5.75}$$

The propagation delay of any delay element i is modeled by $t_{d,i}$. The timing error caused by local variations of the passive interpolators is described by δ_n. The start

signal has been injected into the delay-line at $t = 0$, thus the stop signal arrives at the time instance

$$t_s = T + \varepsilon_s \tag{5.76}$$

For the sake of simplicity no latency along the stop tree is modeled. However, ε_s describes a hypothetical variation of this latency and/or of the stop arrival time. Assuming a zero blackout time with a local variation, i.e. $t_n^{BO} = \beta_n$, results in the following condition for the sampling of a logic high value in the nth stage:

$$t_s - t_n > t_n^{BO} \tag{5.77}$$

$$T > t_{s,n} := nT_{LSB} + \sum_{i=1}^{n_c(n)} \varepsilon_i + \frac{n_i(n)}{IF}\varepsilon_{n_c+1} + \delta_n + \beta_n - \varepsilon_s \tag{5.78}$$

From this follows the differential non-linearity according to

$$DNL_n = \frac{1}{T_{LSB}}\left(t_{s,n+1} - t_{s,n} - T_{LSB}\right) \tag{5.79}$$

$$= \frac{1}{T_{LSB}}\left(\frac{1}{IF}\varepsilon + \delta_{n+1} - \delta_n + \beta_{n+1} - \beta_n\right) \tag{5.80}$$

As the resistance of the interpolators is usually quite small, the area of each resistor is relatively large. This makes the interpolators relatively robust against local process variations,[6] so δ_i can be neglected usually. The differential non-linearity becomes then

$$DNL = \frac{1}{T_{LSB}}\left(\frac{1}{IF}\varepsilon + \beta_{n+1} - \beta_n\right) \tag{5.81}$$

If this result is compared to the DNL of the conventional delay-line TDC (eq. 3.69) it becomes obvious that the local variations of the delay elements are subdivided by the interpolation factor IF. Variations of the sampling elements contribute in the same way to the non-linearity. The reduced impact of delay element variations is another advantage of the LPI-TDC. Another advantageous effect that cannot be seen from the equations above arises from the series connection of the interpolators. The current flowing through the interpolation resistors contributes to the charging currents of the respective nodes in the delay-line. This reduces the effective delay variation of the delay elements themselves. A similar effect results from the differential structure of the LPI delay-line.

Finally, the same three sigma criterion is applied to estimate the maximum resolution of the LPI-TDC:

$$1 > 3 \cdot \mathrm{std}\left(DNL\right) \tag{5.82}$$

$$\Leftrightarrow T_{LSB} > 3\sqrt{\frac{1}{IF^2}\mathrm{std}^2\left(\varepsilon\right) + 2\mathrm{std}^2\left(\beta\right)} \tag{5.83}$$

[6] Remember, global variations cancel out (first order) as the voltage divider formula is the ratio of two resistances.

5.5.4 Implementation Example

An exemplary LPI-TDC with an interpolation factor $IF = 4$ has been implemented in a 90 nm low power CMOS technology and is operated with a supply voltage of 1.2 V. The 7-bit delay-line is characterized by an on-chip measurement and calibration engine such as the one described in Section 4.7. The resulting converter characteristic is depicted in Fig. 5.16 (left). A zoom into the characteristic is given in the right graph. The converter achieves a resolution of 4.7 ps at a supply voltage of 1.2 V and 3.9 ps at 1.4 V. The measured non-linearity is illustrated by the INL and the DNL plot given in Fig. 5.17. Throughout the complete characteristic the maximum integral non-linearity is always smaller than $INL = 1.2\,LSB$. The maximum differential non-linearity is $0.6\,LSB$. For repeated measurement of constant time intervals a single-shot precision of $0.7\,LSB$ is achieved. This is shown in Fig. 5.18 for a cut-out of the converter characteristic with a fine granular delay sweep. The circles indicate the average measurement value and the error bars the single-shot precision. For repeated measurements of slowly varying signals, the finite single-shot precision

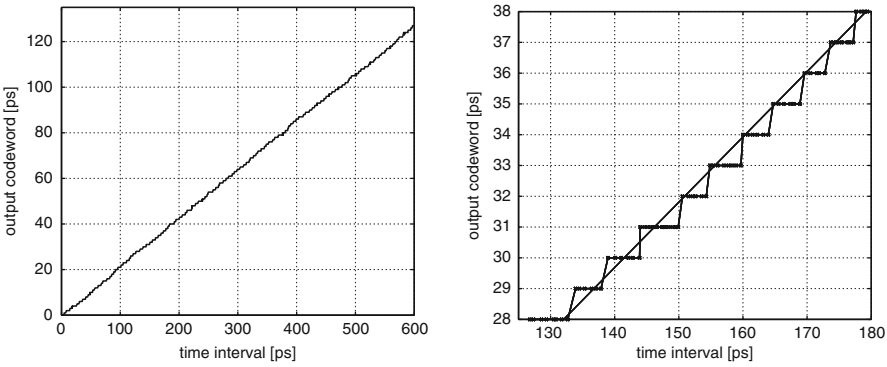

Fig. 5.16 Input-output characteristic of LPI-TDC implemented in 90 nm CMOS technology

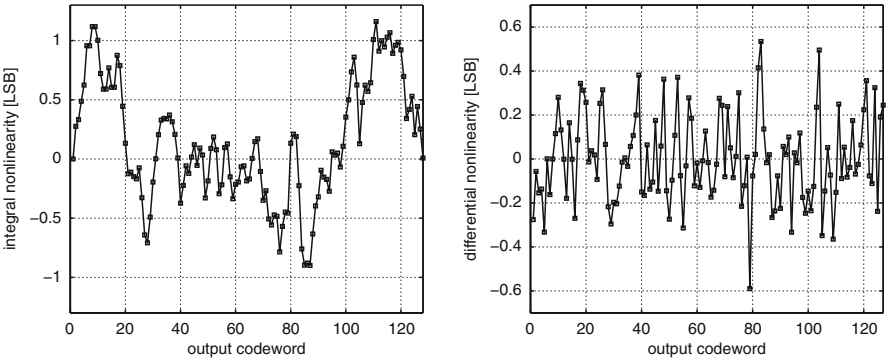

Fig. 5.17 Integral (left) and differential (right) non-linearity of measured 90 nm LPI-TDC

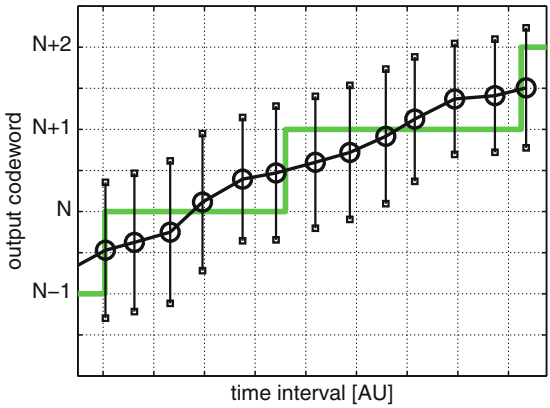

Fig. 5.18 Measured single shot precision of 90 nm LPI-TDC

Table 5.4 Performance summary of LPI-TDC

Principle	Complementary start signal propagates in differential delay-line, intermediate signals are generated by passive voltage dividers.
Resolution T_{LSB}	$\frac{1}{IF} t_d^{inv}$
Number of Stages N	$\frac{T_{max}}{t_d} = \frac{1}{IF} \frac{T_{max}}{T_{LSB}}$
Core Area A_{core}^{LPI}	$\frac{T_{max}}{T_{LSB}} \left(\frac{2.2}{IF} A^{inv} + A^{FF} + 2A^{R} \right)$
Average Power $\langle P_{LPI}^{core} \rangle$	$f_{meas} \frac{1}{IF} \frac{T_{max}}{T_{LSB}} \left(2.2(E_{rise}^{inv} + E_{fall}^{inv}) + IF \cdot E^{FF} + IF \cdot E_{precharge}^{FF} + 4E_{x-curr} \right)$
Conversion Time T_{conv}	T
Latency $T_{latency}$	0
Loop Structure	Loop structure possible
PROS	• Sub gate-delay resolution • Modular structure • Scalable resolution ($IF = 2 \ldots 6$) • No latency • Short delay-line • Reduced variability in delay-line
CONS	• Cross current in interpolator

dithers the output values, so averaging of multiple measurement results reduces the quantization error, i.e. increases the effective resolution. For the TDC core the measured energy consumption is 19 pJ/conversion. Table 5.4 summarizes the properties and performance figures of the LPI-TDC.

5.6 Gated Ring Oscillator TDC

In the field of analog-to-digital and digital-to-analog converters oversampling is a well known technique to achieve high resolution even with coarse quantizers. The basic idea is to sample (convert) a slowly varying signal $x(t)$ with a bandwidth f_0 more often than required by the Nyquist criterion and to remove high frequency components later on by filtering. Under the assumption of white quantization noise the noise power is equally spread over all frequencies in between $-\frac{f_s}{2}$ and $+\frac{f_s}{2}$, where f_s is the sampling frequency. For a signal band of interest given by

$$-\frac{f_s}{2 \cdot OSR} < f < \frac{f_s}{2 \cdot OSR} \tag{5.84}$$

where OSR is the oversampling ratio defined by

$$OSR = \frac{f_s}{2 f_0} \tag{5.85}$$

all noise components outside of this band can be filtered out. This gives an oversampling term in the formula 3.12 of the ideal SNR:

$$SNR = 6.02\ \text{dB} \cdot M + 1.76\ \text{dB} + 10\ \text{dB} \cdot lg(OSR) \tag{5.86}$$

In addition to oversampling noise shaping is a technique that increases the SNR (the resolution) further. The quantization power at low frequencies is pushed out of the signal band towards higher frequencies. As high frequency components are filtered out according to the oversampling concept even less noise power is present in the signal band. The simplest form of noise shaping, the so called first order noise shaping, arises if the measurement device is built such that the output contains the difference of two samples of the quantization error. It is important to understand that this improved signal-to-noise ratio, i.e. the increased resolution is not achieved for a single measurement value but for multiple averaged samples. Thus, a TDC based on oversampling cannot achieve a high single shot resolution but a high effective resolution after post-processing (filtering) the results.

Quantization in time-to-digital converters arises from the fact that the comparators can distinguish only whether a certain delay stage has been passed by the timing event or not. The timing event itself propagates in the delay-line even if the measurement has been stopped. For a new measurement all state variables from the previous measurement are reset. Hence, there is no information about the quantization error that can be used to improve the resolution. Of course oversampling can be used but no noise shaping. According to the theoretical considerations of Section 3.6 it would be desirable to start the new measurement from exactly that instance where the previous one has been stopped. The gated delay-line technique proposed in [11, 12] achieves this by sampling and freezing instead of only sampling a TDC delay-line. The principle is illustrated in the signal diagram of Fig. 5.19. Each signal is one phase of a five stage ring oscillator, i.e. each phase is shifted by one inverter

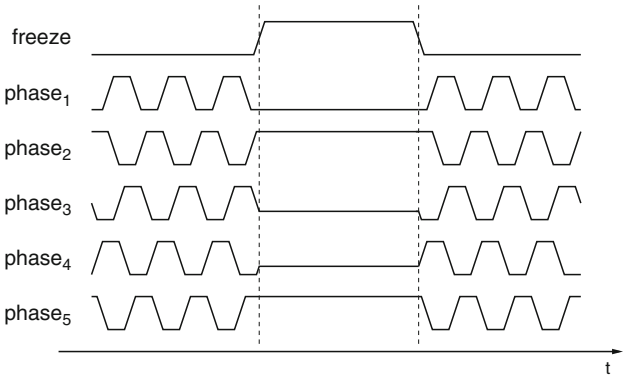

Fig. 5.19 Operating principle of gated ring oscillator

delay with respect to its neighbors. The period of the oscillation is ten inverter delays. Next, it shall be assumed that there is a control signal *freeze* that can stop and restart the signal propagation in any state. When the *freeze* signals goes high in Fig. 5.19 the first, the second, and the fifth phase have a valid logic value. It is obvious that these digital states can be kept constant simply by gating the signal propagation in the delay-loop. The third and fourth phase, however, switch while the *freeze* signal becomes active. They do not have a valid logic level but an analog state, namely a voltage in between the levels representing a logic low and a logic high level. It shall be further assumed that this analog state is preserved while the *freeze* signal is active. When the *freeze* signal goes low again, all delay elements are activated in exactly that state where they have been stalled. The oscillation (the signal propagation) is thus continued in the same phase without additional transients. In addition to the freezing of the delay-loop comparators sample the state in response to the *freeze* signal, i.e. the freeze signal is the stop signal of the TDC. The quantization, i.e. the addition of the quantization error, happens in the comparators. The stalled delay-line, however, still contains the continuous time information without quantization error. If the next measurement starts from this continuous state, the quantization error of the sampling is considered in the next measurement. This corresponds to the conditions theoretically discussed in Section 3.6, i.e. the quantization error is subject to a first order noise shaping.

A technical implementation of a time-to-digital converter based on gated delay-lines is depicted in Fig. 5.20. It is often called gated ring oscillator (GRO) time-to-digital converter. Freezing and releasing the state of the delay-loop means that the individual measurement results accumulate in the TDC. Differentiation of the measurement values results in the digital representation of the measured time intervals. The delay elements require means for freezing the analog state. The challenge is to implement these switching means thus that there is no injection of error signals, such as clock feed-through, into the analog state. A basic delay element enabling analog gating is the C^2MOS inverter. The inner devices form an inverter and the outer devices the gating devices. This device order preserves not only the output voltage but

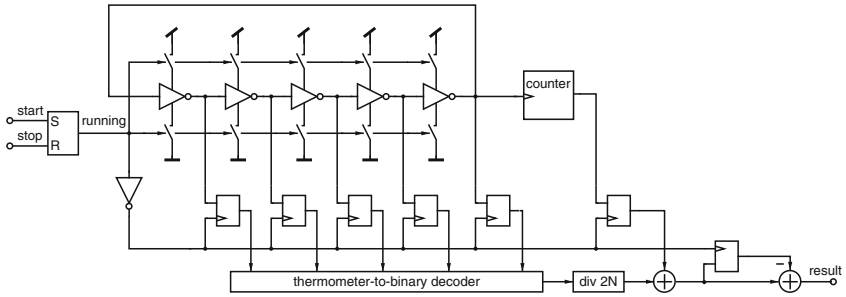

Fig. 5.20 Principle of gated ring-oscillator based TDC

also the intermediate node voltages of the transistor stacks (approximately). If the transistors were conversely connected the internal nodes would be discharged. This would result in large phase error during the activation.

Leakage currents are another serious impairment of the gated ring oscillator TDC: The continuous state information is stored as charge on parasitic capacitors in the delay-line. For long idle time this dynamically stored information is degraded or even lost [13, 14]. Especially in deep sub-micron technologies this requires a relatively high measurement rate.

5.7 Time-to-Digital Converter with Time Amplification

In this section, the last time-to-digital converter technique enabling sub-gate delay resolution is presented, namely time amplification [1]. All TDCs discussed so far achieve a higher single shot resolution by creating more narrowly spaced time events. This means that the resolution is increased by reworking the quantizer references. (The gated ring oscillator based TDC is not discussed in this context as it achieves the high resolution by oversampling with noise shaping.) Time amplifier based TDCs use a completely different technique: They take a time interval, usually the residue of a coarse quantization, and stretch it along the time axis. The enlarged interval can then be quantized with a conventional TDC approach. Consequently, the time interval itself is manipulated, not the references.

The basic principle of a time interval amplifier (TA)[7] that has been proposed in [1,29,30] is depicted in Fig. 5.21. It consist of a conventional NAND latch followed by an arbiter. Time propagates continuously, time intervals, however, are defined by discrete time instances. Hence, the circuit processes discrete, not continuous events. This is a fundamental difference between a TA and an analog amplifier such as a

[7] In literature this technique is sometimes called time amplification. This is a little bit sloppy as the time is the independent variable of all dynamic processes in our world and cannot be stretched or amplified (TDCs are considered as non-relativistic devices ☺). What is amplified is the time difference between two time instances, so the better name is time interval amplifier.

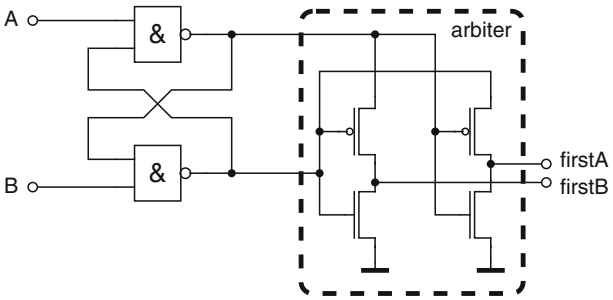

Fig. 5.21 Basic principle of time interval amplifier (TA)

voltage amplifier. Before the measurement, more precisely before an amplification event, both input signals *A* and *B* are low. Thus the outputs of both NAND gates are high. On the arrival of a rising edge on one of the inputs the associated NAND gate starts to discharge its output. As soon as this output is below the switching threshold of the NAND gates the output of the other gate is fixed to a high potential, whatever switching activity occurs at the other input. The arbiter evaluates which latch output goes low (first) and sets the appropriate output to high. This switching sequence has a certain fixed delay. However, if the second input receives a rising edge before the switching sequence is completed, the propagation delay between the arrival of the input and the output signals becomes dependent on the skew between the two input edges. When the first input goes high the associated NAND output voltage decreases. When the second input goes high the output of the second NAND gate decreases as well. With decreasing voltage of any of the two NAND output nodes the respective other node is pulled up by the p-mos device and its discharge is impeded by the closing n-mos device in the pull-down stack. This contention is dominated by the gate whose input has switched first. The closer the switching events the harder the contention. Thus, with decreasing skew the settling time of the circuit increases exponentially. If both inputs switch at the same time the circuit behavior is indefinite: Both nodes simultaneously discharge and reach an instable mid-level until noise and external coupling events unbalance the circuit slightly so that the contention can start. The period while both gates compete for the logic low level is called the meta-stable state of the latch. Usually meta-stability is an unwanted effect that occurs during sampling of asynchronous signals or in synchronous systems if the setup time constraint is violated. In the time interval amplifier meta-stability is used to prolong the time difference between the input signals. It is obvious that this technique only works if the time interval is smaller than the settling time of the same latch with a single switching event at its inputs. Figure 5.22 shows switching trajectories for several input time intervals. For small intervals both node voltages come near to the mid-level. After a while one gate dominates, i.e. one node goes low and the other goes back high. It is the purpose of the arbiter to shield this meta-stable state to the environment. This is important to avoid cross currents and undefined logic states.

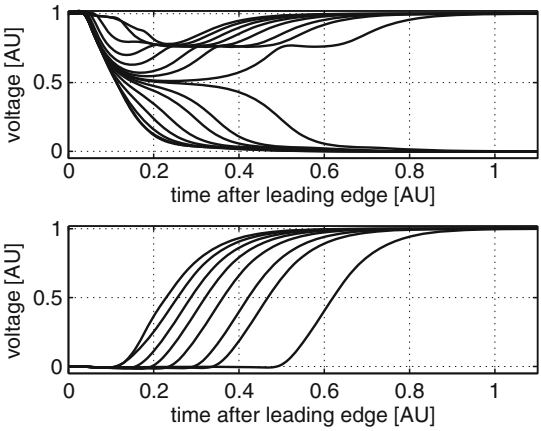

Fig. 5.22 Internal node voltages and output signal of basic time amplifier

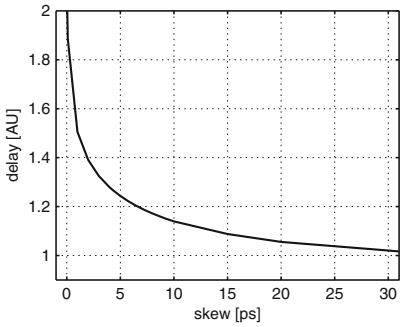

Fig. 5.23 Input to output delay of basic time amplifier in dependence on the skew of the two input signals, i.e. on the time interval to be amplified

The skew dependent delay of the basic time interval amplifier is depicted in Fig. 5.23. An ideal amplifier with constant gain would have a linear characteristic. Yet, the actual characteristic is quite non-linear. A basic linear analysis for the first phase of the contention suggest a logarithmic delay curve [1]. As real gates behave pretty non-linear a simulation approach has been chosen here in order to obtain more realistic results. A linearization concept has been proposed in [1] and improved in [30]. The basic idea is to use two time interval amplifiers with contrarily shifted characteristics. The results of the two conversions are subtracted later on to get a more linear gain curve. Figure 5.24 demonstrates how this linearization is done. Each delay curve has an even symmetry due to the topological symmetry of the circuit with respect to its inputs. Delaying one input causes an asymmetry that shifts the curve either to the left or to the right. This can be either done by inserting additional delay elements in one signal path or by an asymmetrical sizing of the two NAND gates. The curve in the lower sub-graph results from subtracting the two non-linear amplifier characteristics. The good news is that the resulting

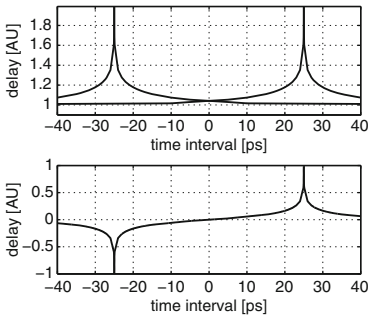

Fig. 5.24 Linearized delay of time amplifier

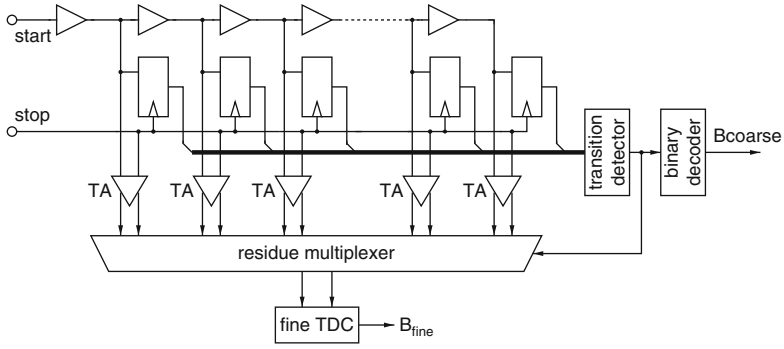

Fig. 5.25 Time-to-digital converter based on time-amplification

characteristic is quite linear. The bad news is that this linear region is very small (some tens of pico-seconds in this exemplary implementation). Obviously, the challenge is the realization of well matched amplifiers and the subtraction of the two time delays.

In a time to digital converter TAs could be used to amplify the input time interval before the quantization in a conventional delay-line based TDC. Due to the very small input dynamic range of the time amplifiers such TDCs are useful only for very small time intervals. The approach presented in [30] uses a coarse quantization based on a buffer delay-line and amplifies the residue. A cut-out of the architecture is shown in Fig. 5.25. A conventional delay-line TDC coarsely quantizes the time interval spanned by the start and the stop signal. Time amplifiers are connected to each tap and stretch the time interval between the respective delayed version of the start signal and the stop signal. The implementation of one TA per tap is necessary as time intervals cannot be stored within the time domain directly. This makes a sequential two step approach impossible where the residue is processed not until the coarse quantization. Hence, a speculative amplification of all possible residue intervals is required. A transition detector identifies the stage of the one-zero transition and selects the appropriate time amplifier later on. The amplified residue is passed to a second TDC stage (fine TDC) for further quantization. Due to the small

dynamic range of the TAs the latency in the transition detector is a critical point. This is particularly true as the residue multiplexer is controlled by the comparator results (in the coarse TDC) which may experience meta-stability during sampling. The coarse TDC is essentially linear, i.e. the overall linearity may be acceptable. The residue quantization, however, is subject to non-linear distortion in the time interval amplifiers. This locally degrades the linearity of the overall TDC characteristic.

Beside the non-linearity noise is a critical issue for time interval amplifiers. The meta-stable state is a very sensitive state of a latch. Any disturbance by noise or (capacitive) coupling favors one of the inputs, so reflects in a delay uncertainty of the output signal. Flicker noise in the transistors of the NAND gates, for example, temporarily changes the threshold of the TA and shifts the characteristic. Supply noise is another critical issue, especially in SOC with high digital switching activity. (Local) process variations seriously influence the amplifier characteristic which makes it very difficult to assure a certain (linear) operating region. Reflecting the motivation behind time domain signal processing discussed in Chapter 2, the idea is to exploit the advanced robustness and scaling properties of digital circuits instead of classical analog circuits. From a digital perspective, however, meta-stability is a parasitic effect that is sensitive to everything, i.e. the time stretching is a non-linear analog effect itself. Hence, the story line to go into the time domain with many applications seems not convincing for the time interval amplifier. Altogether, the susceptibility to noise and variations causes serious challenges and it is doubtful whether time interval amplification is suitable for robust products and mass production.

Chapter 6
Applications for Time-to-Digital Converters

Abstract Amongst other things, time-to-digital converters are used in high energy and particle physics, for measurement and instrumentation applications, for time of flight measurement, digital PLLs, and data converters. This chapter focuses on emerging applications in main stream micro-electronics. In the first part time-to-digital converters are discussed in the context of digital phase-locked loops. Both the classical PLL architecture based on a feedback divider as well as the recently proposed all-digital PLL architecture are discussed. The second part addresses time-to-digital converter based analog-to-digital converters.[1] The classical dual slope converter is revisited for motivation purposes. Next an ADC based on pulse position modulation is discussed. Finally, sigma delta modulator based ADCs are investigated. A continuous time sigma delta modulator with time domain quantizer is presented as a classic approach and an ADC based on an asynchronous sigma delta modulator with feed-forward TDC as quantizer is presented as promising new architecture for ultimately scaled CMOS. It is the intention of this chapter to provide an overview on concepts and architectures, not to focus on implementation details.

Key words: TDC Applications, Digital Phase-Locked-Loop, TDC Based Analog-to-Digital Converter

6.1 Digital Phase Locked Loop

The first and currently most important TDC application in integrated electronics is definitively the digital phase-locked-loop (PLL). The purpose of phase-locked-loops is to generate a (quasi-)periodic signal, i.e. an output frequency, that has a fixed frequency ratio with respect to a reference signal received at its input. Moreover, the phase relation between the reference signal and the locally generated output signal is fixed. Amongst others, the field of applications covers clock generation, cleaning of

[1] This chapter discusses only a choice of the most promising concepts for high linearity, high resolution, high bandwidth, and robust implementation.

S. Henzler, *Time-to-Digital Converters*, Springer Series in Advanced Microelectronics 29,
DOI 10.1007/978-90-481-8628-0_6,

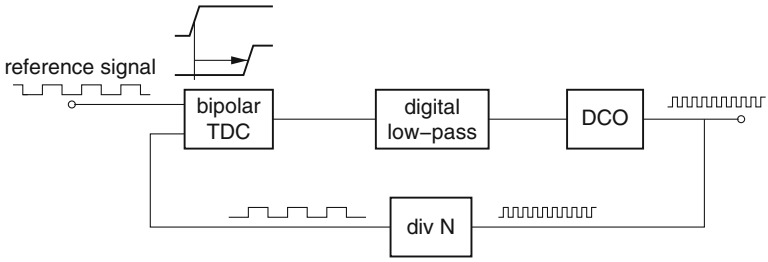

Fig. 6.1 Basic digital phase-locked-loop derived from classical analog structure

noisy clocks, de-skewing of signals, frequency multiplication, clock-data-recovery, and RF carrier generation. A PLL forms a closed-loop control circuit that works on the phase of quasi-periodic signals. The output signal is locally generated by a controlled oscillator. This controlled oscillator, classically a voltage controlled oscillator (VCO), provides a quasi-periodic signal with an instantaneous frequency that is configured by the control signal at its input. A phase-(frequency) detector (PD/PFD) compares the locally generated signal with the reference signal in terms of phase. The low frequency components of the phase-detector output describe the phase relation of the two signals. A low-pass filter extracts this information and steers the controlled oscillator accordingly. If the reference signal leads, the frequency of the controlled oscillator is increased. If the reference signal lags, the oscillator frequency is decreased. For general details on phase-locked-loops the reader is referred to the literature, e.g. [14, 46].

This section focuses on digital PLLs, especially on TDCs used as phase detectors. A classical analog PLL contains the voltage controlled oscillator and the loop filter as clearly analog blocks. In the case of rectangular signals, the phase detector is already quite digital. It may consist of a simple XOR gate (PD) or of a flip-flop like state machine (PFD). The output of the phase detector is pulse-width encoded. Thus, the phase detector consists of digital building blocks, but does not provide a digital output signal in the sense of a time discrete binary value. For frequency multiplication a divider is inserted into the feedback path. Usually this is a digital circuit as well. However, analog design methodologies apply when high frequencies have to be processed.

With this knowledge one can consider which building blocks can be digitized. The oscillator itself is a pure time-continuous block. When the frequency is somehow changed, the instantaneous output frequency evolves from the previous frequency towards the new frequency by means of time continuous transients. This fundamental nature of an oscillator cannot be changed. Even if we measure the instantaneous frequency at discrete time instances, e.g. at the zero crossings of the output signal, the oscillation itself is continuous. In the context of a digital PLL digitizing the oscillator thus means the addition of a digital interface to the oscillator. Therewith, the frequency control information can be provided at discrete time instances and by binary words. The digital interface can be a DAC that translates the digital control information into the control voltage of a classical VCO. More advantageous is the integration of this DAC right inside the oscillator, i.e. the use of a parametric

oscillator. In the case of an LC oscillator, for instance, tank capacitors may be added or removed in response to digital control signals. A controlled oscillator with digital control interface is called digital controlled oscillator (DCO). The oscillator itself remains analog but this analog behavior is now hidden behind a digital interface. Therewith the phase detector and the loop filter are the remaining candidates for further digitization. In general the purpose of the phase detector is to measure the phase difference between the feedback and the reference signal. This corresponds to the time interval measurement between rising edges of these two signals together with an appropriate normalization to the instantaneous period. A device that can measure this time interval and that provides a purely digital result is however a TDC. If the phase detector provides digital result words and the DCO expects digital control words a digital implementation of the desired loop filter is straight forward.

Figure 6.1 shows a digital phase-locked-loop derived from the classical analog architecture by replacing each element by its digital equivalent. All building blocks of a classical analog PLL can be identified clearly. The frequency/phase steering information is still contained in the reference signal, namely in its phase and frequency. A digital implementation of PLLs is manifold advantageous. Apart from the DCO there are no more passive devices. In particular the large capacitors of an analog charge pump PLL can be omitted. Although digital PLLs currently tend to be larger than analog implementations[2] their superior scaling behavior will enable future PLLs that are more robust and smaller. Even more attractive is the addition of novel functionality that would not be possible with an analog implementation: The filter transfer function can be easily adjusted during the operation of the circuit by loading new coefficients or by changing the order of the filter. In some sense this would be also possible with analog trimming, but not with the digital flexibility, accuracy and reproducibility. Only possible with the digital realization are the storing and loading of internal states [17] (not coefficients).[3]

In the PLL of Fig. 6.1 either the reference or the feedback signal may lead. Hence, a bipolar TDC is required. Usually, the reference frequency is quite low, e.g. some MHz. The counter reduces the frequency of the feedback signal to the same low frequency. Consequently, the dynamic range of the TDC has to be very high. This makes the TDC not only large but also prone to noise. A completely new architecture that has been proposed in [34, 51, 53] is shown in Fig. 6.2. This circuit

[2] This is due to one time effects of transferring analog to digital building blocks. A TDC, for instance, consumes considerable more area than a conventional phase-frequency detector consisting of some flip-flops and a push–pull stage.

[3] This may be useful for fast (periodic) jumps between different frequencies. For example consider a frequency hopping between two frequencies. First the PLL is configured for the first frequency. When the PLL has locked, all internal state variables are stored in a digital memory. Then the second frequency is configured, the PLL settles, and finally the state variables for the second frequency are stored. For going back to the first frequency the PLL is not only configured with the new coefficients but also the state variables are loaded again. Hence, the loop is initialized in its steady state of the previous phase where the first frequency was generated. Hopping back to the second or even to a third or fourth frequency can be done accordingly. If the time intervals are long the operating conditions may have changed. The loaded state is thus only an estimation of the new steady state. However, the error should be small and a fast (linear) settling can be expected [17].

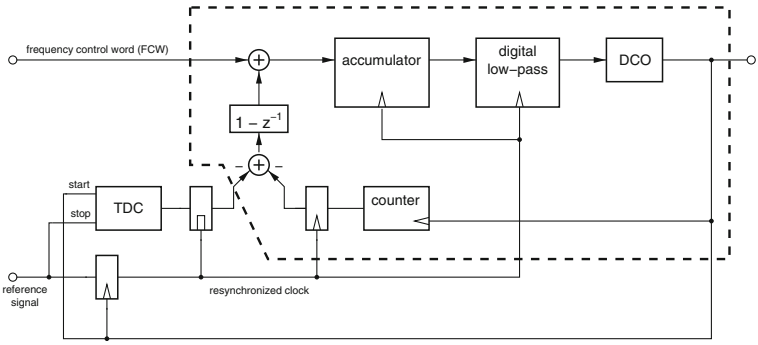

Fig. 6.2 All digital phase-locked-loop architecture proposed in [34, 51, 53]

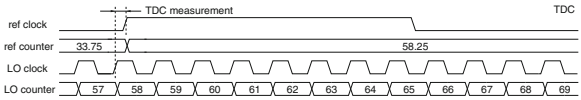

Fig. 6.3 Signal diagram illustrating operation of TDC based phase detector in the all digital PLL

is known as all-digital phase-locked-loop (ADPLL). For the sake of simplicity only the sub-circuit inside the dashed box is considered at the moment. The remaining blocks will be added later. At first glance only the low-pass filter and the DCO is common to the previous implementation. The divider in the feedback path has been removed. Instead a digital counter counts the cycles of the DCO continuously. For now, let's assume that all digital circuit blocks are clocked by the reference signal. The counter increment from one reference cycle to the next one is compared with a digital frequency control word (FCW), that may be updated in each reference cycle. The comparison is done by a digital subtraction of the two words, namely the FCW and the counter value. The difference represents the phase error that is accumulated and low-pass filtered. The accumulation reveals that the loop dynamics is similar to an analog type-II PLL. Figure 6.3 illustrates the operation with the aid of a signal diagram. The counter determines how many full cycles of the local oscillator (LO) occur within one reference cycle and the frequency control word provides the information how many cycles should be within this cycle. The quantization error can be decreased with a TDC that measures the phase relation between the locally generated high-speed signal and the low frequency reference signal. If the TDC result is normalized to one LO period and added to the counter value the resulting digital word describes the fractional number of output cycles inside one reference period. The big hit is the fact that also the FCW can be fractional. Hence, a fine grained frequency control is possible in each cycle, not only in average as for classical fractional-N PLLs. Another great advantage of this counter plus TDC approach is the fact that no measurement information is contained in the pulse width and phase of a divided signal but directly available as digital word. In the implementation of Fig. 6.2 not the reference clock but a derived clock synchronized to the fast LO clock is used for synchronization of all digital blocks. Beside meta-stability

issued this allows a TDC measurement while there is no digital switching activity. First the TDC is stopped and then the digital blocks are triggered.

The TDC in the all digital phase-locked-loop measures the time interval between the last rising LO edge that occurs before the reference clock edge and this reference edge. This is a directed measurement, so no bipolar TDC is required. The maximum time interval corresponds to the period of the digital controlled oscillator. Compared to the TDC in the implementation of Fig. 6.1 the dynamic range is considerably reduced. This reduces the TDC noise and allows for more simple converter architectures. In RF frequency synthesizer, even linear time-to-digital converters may cover the full measurement range. Apart from the quantization error the phase detector in the ADPLL provides the phase error without any high frequency components. In a conventional phase detector the phase error is contained in the average output signal. However, there is also a high frequency component, that arises from the periodic input signals. The lack of this high frequency components in the ADPLL has a positive impact on the loop filter that must only filter the quantization noise and create the loop dynamics.

A general advantage of the TDC in a PLL is the combination with the loop filter. Due to the single shot precision of each TDC a single shot measurement always contains a noise induced error. The loop filter provides some averaging, so filters out high-frequency components of this TDC noise. The main challenge for the TDC in the ADPLL is probably the calibration, i.e. the normalization of the measurement result to one reference period. This is particular true if the frequency is changed. In this case the calibration has to be continuously done in the background.

6.2 TDC Based Analog-to-Digital Converter

6.2.1 Dual-Slope Analog-to-Digital Converter Revisited

One of the very first analog-to-digital converters (ADC) based on a counting technique is the so called dual-slope ADC. It consist of a single operational amplifier (OP), a comparator, a counter, and some control logic. Due to its simplicity it is suited for both, discrete and integrated implementations. A basic block diagram is depicted in Fig. 6.4. At the beginning of each conversion, the integrator, comprising the operational amplifier, is reset, i.e. the capacitance C_1 is discharged by the switch S_2. Then the negative[4] input voltage to be measured ($-V_{in}$) is connected to the integrator via the switch S_1 and integrated for a constant time interval T_1. At the end of the integration the output voltage V_x of the OP is given by

$$V_x = \int_0^{T_1} \frac{1}{R_1 C_1} V_{in} dt = \frac{T_1}{R_1 C_1} V_{in} \tag{6.1}$$

[4] The input signal is inverted, as the RC-integrator is inverting itself.

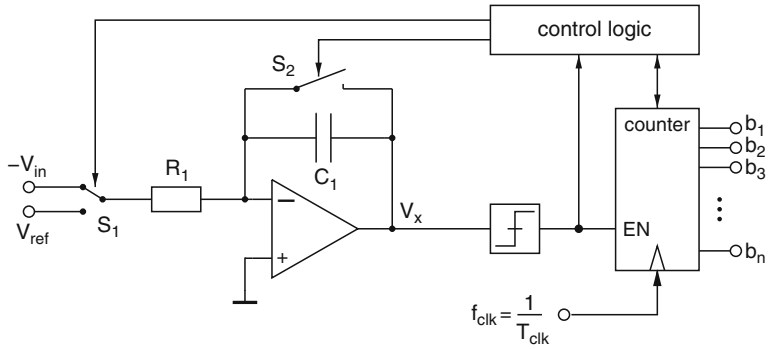

Fig. 6.4 Circuit block diagram of classical dual-slope ADC

This means that V_x is proportional to the input voltage. The integration constant is given by the product of the passives, namely the input resistance R_1 and the integration capacitance C_1. The last equation is only valid for a constant input signal, i.e. for a sampled or at least a slowly varying input signal. When the integration is completed the control logic disconnects the input and connects it to a reference voltage V_{ref}. Therewith, the circuit continues its integration but now with a fixed and negative slope. The output signal can be described by the following equation:

$$V_x(t) = \frac{T_1}{R_1 C_1} V_{in} - \int_{T_1}^{t} \frac{V_{ref}}{R_1 C_1} dt \qquad t \geq T_1 \tag{6.2}$$

The length T_2 of this second integration phase is measured by a counter. The comparator connected to the output of the integrator detects when V_x is equal to zero and stops the counter. The length T_2 of the second interval is then given by

$$\begin{aligned} V_x(T_1 + T_2) &= 0 \\ T_2 &= \frac{V_{in}}{V_{ref}} T_1 \end{aligned} \tag{6.3}$$

It is worth mentioning that this time is independent on the passives, which makes the dual-slope technique very robust against variations of device parameters. If the first time interval T_1 is chosen according to $T_1 = 2^N T_{clk}$, where T_{clk} is the reference clock period, the length of the second interval can be expressed by

$$T_2 = \frac{V_{in}}{V_{LSB}} T_{clk} \tag{6.4}$$

where the fundamental relation $V_{ref} = 2^N \cdot V_{LSB}$ has been used. In terms of the clock period T_{clk} this is exactly the counter value B,[5] i.e.

[5] Neglecting the quantization error.

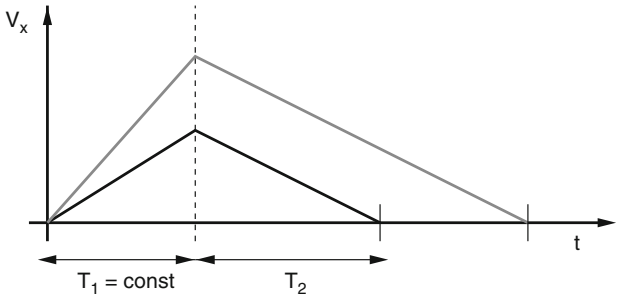

Fig. 6.5 Principle of classical dual-slope ADC

$$B = \frac{T_2}{T_{clk}} = \frac{V_{in}}{V_{LSB}} \tag{6.5}$$

Consequently the counter value describes the input voltage in terms of V_{LSB}, thus is a digital representation of the input voltage V_{in}. An exemplary signal diagram for two different input voltages is shown in Fig. 6.5. The first interval has a fixed length and the slope of the signal depends on the input level. The second interval has a variable length but the signals decrease with a constant slope.

For a high resolution, a long measurement time is required, as both T_1 and T_2 exponentially grow with the number of bits. If the measurement time cannot be prolonged, the clock frequency must be increased. Maximum feasible clock frequencies are (currently) limited to a small single digit number of GHz. Hence, there is a fundamental trade-off between the sampling rate and the resolution of the dual-slope ADC. This trade-off is the reason why the dual-slope converter has been regarded as slow and old-fashioned for the last couple of years. With the emergence of time-to-digital converters in integrated micro-electronics, however, it might experience a renaissance, as high resolution time interval measurement does not mean ultra high frequency clocks anymore. Hierarchical time-interval measurement based on TDCs for instance enables long measurement times and high resolution even with moderate clock frequencies.

6.2.2 Pulse Position Modulation Analog-to-Digital Converter

A pulse position modulation (PPM) analog-to-digital converter has recently been proposed by Naraghi et al. [35]. The principle is explained on the basis of the schematic circuit diagram depicted in Fig. 6.6. An exemplary signal diagram is provided in Fig. 6.7. The integration capacitance C is periodically discharged by a reference clock signal CP. The discharged capacitance C is then charged by a constant current I leading to a continuous time voltage ramp

$$v(t) = \frac{I}{C} \cdot t \tag{6.6}$$

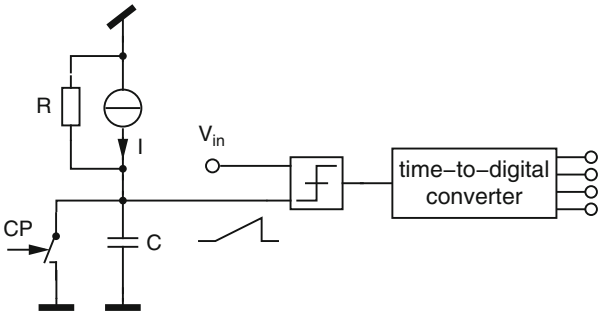

Fig. 6.6 Schematic block diagram of pulse-position modulation ADC

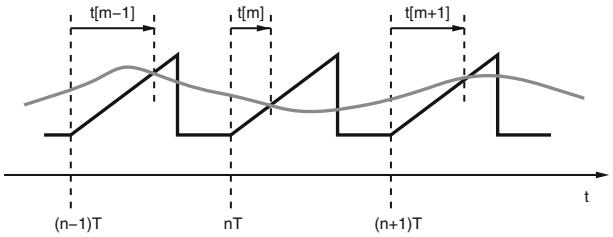

Fig. 6.7 Principle signal diagram of pulse-position modulation ADC

This ramp[6] signal is compared to the continuous time input signal in an asynchronous comparator. A time-to-digital converter is synchronously started on the beginning of the charging process and stopped when the input voltage $v_{in}(t)$ is equal to the ramp voltage $v(t)$. The measured time difference $t[n]$ can be inserted into the linear equation 6.6 to compute the corresponding input voltage $v_{in}(t_s)$. Obviously, the sampling instance

$$t_s[m] = nT + t[m] \tag{6.7}$$

becomes signal dependent, i.e. the input signal is non-uniformly sampled by the PPM-ADC. As a conversion is only performed during rising reference ramps, the matching requirements for the analog front end in the PPM ADC are pretty much relaxed. Linear decimation, i.e. conventional low-pass filtering, of the ADC output values requires considerable oversampling to keep the error caused by the non-uniform sampling reasonably low. More advanced non-linear recovery algorithms that take the special nature of non-uniformly sampled signals into account allow for better results with lower OSR and so lower power consumption and higher signal bandwidth. However, the digital post processing effort for these algorithms increases. As in heavily scaled technologies analog functionality becomes increasingly problematic and digital functionality comes nearly for free, this is definitively a trend-setting approach. In this context it should be noted that there is only a very

[6] Other reference signals are also possible. A ramp signal, however, can be generated easily and yields the linear relation eq. 6.6.

small analog part in the PPM-ADC, namely a current source and a continuous time comparator. No opamps or samplers are required. This makes the PPM-ADC robust against variations, scalable, and power efficient.

In spite of everything, the PPM-ADC has also some challenges that have to be carefully considered during design: First of all, an absolute time measurement is required, i.e. the TDC has to be calibrated in regular calibration intervals. An ideal current source yields the linear relation eq. 6.6 between the measured time intervals $t[m]$ and the analog input voltage $v_{in}(t)$. In reality each current source has a finite output resistance R, i.e. the ramp signal is in fact an exponential signal with a time constant RC:

$$\begin{aligned} v(t) &= RI\left(1 - e^{-\frac{t}{RC}}\right) \\ &= \frac{I}{C} \cdot t - RI \sum_{n=2}^{\infty} (-1)^n \frac{t^n}{(RC)^n n!} \end{aligned} \tag{6.8}$$

The second equation separates the linear term, i.e. the ramp signal, from non-linear distortion terms. These non-linear terms directly contribute to the non-linearity of the ADC. A large integration capacitance reduces non-linearity but slows down the converter – a conventional trade-off. Asynchronous comparators suffer from a rise time dependent propagation delay. While the ramp signal has a constant slope, the rate of change at the other comparator input is signal dependent. This is another source of non-linearity that has to be considered during the design of the comparator. Finally the clock signal is an impairment for the overall system performance, as any clock jitter adds to the time measurement, thus translates into a noisy signal according to eq. 6.6.

6.2.3 Sigma Delta Modulator with Time Domain Quantizer

The very first approach of trading voltage resolution against time domain resolution is the sigma delta modulator. Even with a single bit quantizer the analog input signal can be quantized with a very high resolution by means of oversampling and noise shaping. While the voltage resolution is quite low (only one bit in the extreme case) the oversampling means that the time domain resolution is very high. The peak SNR of a first order sigma delta modulator, for instance, is given by

$$SNR_{max} = 6.02\,\text{dB} \cdot N - 3.41\text{dB} + 30\,\text{dB}\lg(OSR) \tag{6.9}$$

N denotes the number of bits in the quantizer and OSR is the oversampling ratio. Obviously a high resolution can be achieved by a large oversampling ratio. If the bandwidth of the input signal is increased the oversampling ratio is decreased. Hence, high resolution and high bandwidth are somehow contradictory. The way out of this conflict is an increased resolution in the quantizer, i.e. a larger N. This, however, is a conceptual step backwards because the system depends on an increased voltage

domain resolution again. Sigma delta converters with higher order quantizers, i.e. $N > 1$, are so called multi-bit sigma delta modulators. The challenges are the power consumption and the linearity of the quantizer, comprising a multi-bit FLASH converter, and the multi-bit feedback DAC. While a one bit quantizer as well as a one bit DAC are inherently linear, multi-bit quantizers/DACs for high resolution sigma delta modulators often require dynamic element matching and/or calibration techniques. This is particularly true for implementations in nanometer scale technologies with low supply voltages.

In order to enable the scalability of power, resolution, and technology another consequent step towards the time domain was required. The replacement of the classical voltage quantizer by a time domain quantizer is such a step. Sigma delta modulators based on time domain quantizers have been recently demonstrated [9]. The basic idea is to quantize the output signal of the loop filter not in the voltage but in the time domain. Therefore, the voltage provided by the loop filter is converted into time intervals that are measured and quantized in a second step. A very basic voltage-to-time converter is a voltage controlled oscillator (VCO). The drawback of most VCOs is the weak linearity which limits the linearity of the overall converter. The approach discussed here circumvents the linearity issue by using a linear pulse width modulator (PWM) as voltage-to-time converter. Figure 6.8 shows a schematic block diagram of the modulator. The quantizer is replaced by a pulse width modulator and a time domain quantizer (ref. Section 4.8). The feedback multi-bit DAC is missing completely. In each cycle of the sampling clock CP, the pulse width modulator converts the slowly varying output voltage of the loop filter into a pulse with a pulse width ΔT. The synchronous clock assures a constant sampling rate at the modulator output. If the filter output is sampled by the PWM input stage the sampling instance is perfectly defined. For a continuous PWM the exact sampling instance varies within the clock interval which results in non-linear distortion. After the PWM the signal is represented in the time domain but is still continuous. A time domain quantizer transfers this continuous time signal into a discrete time signal

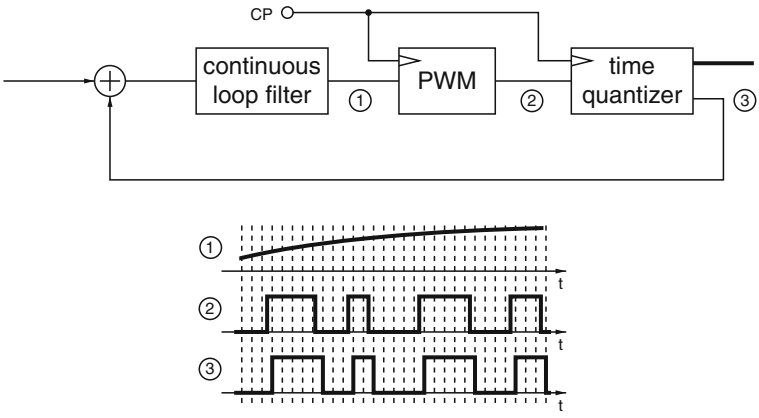

Fig. 6.8 Sigma delta modulator with time domain quantizer

as well as into a digital representation. The digital word is passed to the output of the modulator whereas the discrete time domain signal serves as feedback signal of the loop. As the feedback signal only comprises a single bit, the DAC is trivial and perfectly linear. Still, a high feedback resolution is achieved by the fine quantization of the pulse width. The signal diagram in Fig. 6.8 explains the feedback mechanism schematically: The continuous time PWM signal is mapped onto an equally spaced high resolution time grid and then passed to the feedback. Consequently the time domain quantizer fulfills two functions: First it accurately quantizes the feedback signal which corresponds to the function of the multi-bit DAC. Second it converts the continuous signal into a digital representation that can be further processed by the system.

The advantage of the sigma delta modulator with time domain quantizer is a high resolution that can be achieved with components that are suited for low supply voltages and heavily scaled technologies. No power hungry multi-bit FLASH ADC, multi-bit DAC, and multi-bit voltage references are required. The price is a high frequency clock for the time domain quantization or (what is recommended) a time-domain quantizer. Other sigma delta modulators based on time domain quantizers have been proposed in [7, 23].

References

1. Abas, A., Bystrov, A., Kinniment, D., Maevsky, O., Russell, G., Yakovlev, A.: Time difference amplifier. Electronics Letters **38**(23), 1437–1438 (2002). DOI10.1049/el:20020961
2. Abaskharoun, N., Roberts, G.: Circuits for on-chip sub-nanosecond signal capture and characterization. Custom Integrated Circuits, 2001, IEEE Conference on. pp. 251–254 (2001). DOI10.1109/CICC.2001.929766
3. Andreani, P., Bigongiari, F., Roncella, R., Saletti, R., Terreni, P., Bigongiari, A., Lippi, M.: Multihit multichannel time-to-digital converter with 1 Solid-State Circuits, IEEE Journal of **33**(4), 650–656 (1998). DOI10.1109/4.663573
4. Chan, A., Roberts, G.: A deep sub-micron timing measurement circuit using a single-stage vernier delay line. Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002 pp. 77–80 (2002). DOI10.1109/CICC.2002.1012770
5. Chen, P., Liu, S.L., Wu, J.: A cmos pulse-shrinking delay element for time interval measurement. Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on **47**(9), 954–958 (2000). DOI10.1109/82.868466
6. Chen, P., Zheng, J.C., Chen, C.C.: A monolithic vernier-based time-to-digital converter with dual plls for self-calibration. Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005 pp. 321–324 (2005). DOI10.1109/CICC.2005.1568670
7. Corporales, L., Prefasi, E., Pun, E., Paton, S.: A 1.2-mhz 10-bit continuous-time sigmadelta adc using a time encoding quantizer. Circuits and Systems II: Express Briefs, IEEE Transactions on **56**(1), 16–20 (2009). DOI10.1109/TCSII.2008.2008524
8. Daniels, J., Dehaene, W., Steyaert, M., Wiesbauer, A.: A/d conversion using an asynchronous delta-sigma modulator and a time-to-digital converter. In: Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on, pp. 1648–1651 (2008). DOI10.1109/ISCAS.2008.4541751
9. Dhanasekaran, V., Gambhir, M., Elsayed, M., Sanchez-Sinencio, E., Silva-Martinez, J., Mishra, C., Lei, C., Pankratz, E.: A 20mhz bw 68db dr ct sigma delta adc based on a multi-bit time-domain quantizer and feedback element. In: Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International, pp. 174–175,175a (2009). DOI10.1109/ISSCC.2009.4977364
10. Dudek, P., Szczepanski, S., Hatfield, J.: A high-resolution cmos time-to-digital converter utilizing a vernier delay line. Solid-State Circuits, IEEE Journal of **35**(2), 240–247 (2000). DOI10.1109/4.823449
11. Helal, B., Straayer, M., Wei, G.Y., Perrott, M.: A low jitter 1.6 ghz multiplying dll utilizing a scrambling time-to-digital converter and digital correlation. VLSI Circuits, 2007 IEEE Symposium on pp. 166–167 (2007). DOI10.1109/VLSIC.2007.4342700

12. Helal, B., Straayer, M., Wei, G.Y., Perrott, M.: A highly digital mdll-based clock multiplier that leverages a self-scrambling time-to-digital converter to achieve subpicosecond jitter performance. Solid-State Circuits, IEEE Journal of **43**(4), 855–863 (2008). DOI 10.1109/JSSC.2008.917372
13. Henzler, S.: Power Management of Digital Circuits in Deep Sub-Micron CMOS Technologies. Springer (2005)
14. Henzler, S.: High-speed digital cmos circuits. Lecture Notes online Chapter on Phase-Locked-Loops, Institute for Technical Electronics (2007-2010). Www.lte.ei.tum.de/homes/henzler
15. Henzler, S.: Method and device for measuring time intervals. Infineon AG, var. Patent Applications (4/2008)
16. Henzler, S.: Determining a time interval based on a first signal, a second signal, and a jitter of the first signal. Infineon AG, var. Patent Applications (9/2007)
17. Henzler, S., Koeppe, S.: Frequency synthesizer and method. Infineon AG, var. Patent Applications (2006)
18. Henzler, S., Koeppe, S.: Apparatus and system with a time delay path and method for propagating a timing event. Infineon AG, var. Patent Applications (2/2008)
19. Henzler, S., Koeppe, S., Kamp, W., Mulatz, H., Schmitt-Landsiedel, D.: 90nm 4.7ps-resolution 0.7-lsb single-shot precision and 19pj-per-shot local passive interpolation time-to-digital converter with on-chip characterization. Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International pp. 548–635 (2008). DOI 10.1109/ISSCC.2008.4523300
20. Henzler, S., Koeppe, S., Lorenz, D.: Time delay circuit and time to digital converter. Infineon AG, US Patent (3/2007)
21. Henzler, S., Koeppe, S., Lorenz, D., Kamp, W., Kuenemund, R., Schmitt-Landsiedel, D.: Variation tolerant high resolution and low latency time-to-digital converter. 33rd European Solid State Circuits Conference, 2007. ESSCIRC pp. 194–197 (2007). DOI10.1109/ESSCIRC.2007.4430278
22. Henzler, S., Koeppe, S., Lorenz, D., Kamp, W., Kuenemund, R., Schmitt-Landsiedel, D.: A local passive time interpolation concept for variation-tolerant high-resolution time-to-digital conversion. Solid-State Circuits, IEEE Journal of **43**(7), 1666–1676 (2008). DOI 10.1109/JSSC.2008.922712
23. Hernandez, L., Prefasi, E.: Analog-to-digital conversion using noise shaping and time encoding. Circuits and Systems I: Regular Papers, IEEE Transactions on **55**(7), 2026–2037 (2008). DOI10.1109/TCSI.2008.918003
24. Hwang, C.S., Chen, P., Tsao, H.W.: A high-precision time-to-digital converter using a two-level conversion scheme. Nuclear Science Symposium Conference Record, 2003 IEEE **1**, 174–176 Vol.1 (2003). DOI10.1109/NSSMIC.2003.1352024
25. Hwang, C.S., Chen, P., Tsao, H.W.: A high-precision time-to-digital converter using a two-level conversion scheme. Nuclear Science, IEEE Transactions on **51**(4), 1349–1352 (2004). DOI10.1109/TNS.2004.832902
26. Jansson, J., Mantyniemi, A., Kostamovaara, J.: A delay line based cmos time digitizer ic with 13 ps single-shot precision. Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on pp. 4269–4272 Vol. 5 (2005). DOI10.1109/ISCAS.2005.1465574
27. Jansson, J.P., Mantyniemi, A., Kostamovaara, J.: A cmos time-to-digital converter with better than 10 ps single-shot precision. Solid-State Circuits, IEEE Journal of **41**(6), 1286–1296 (2006). DOI10.1109/JSSC.2006.874281
28. Karadamoglou, K., Paschalidis, N., Stamatopoulos, N., Kottaras, G., Paschalidis, V., Sarris, E.: A cmos time to digital converter for space science instruments. Solid-State Circuits Conference, 2002. ESSCIRC 2002. Proceedings of the 28th European pp. 707–710 (2002)
29. Lee, M., Abidi, A.: A 9b, 1.25ps resolution coarse-fine time-to-digital converter in 90nm cmos that amplifies a time residue. VLSI Circuits, 2007 IEEE Symposium on pp. 168–169 (2007). DOI10.1109/VLSIC.2007.4342701
30. Lee, M., Abidi, A.: A 9 b, 1.25 ps resolution coarsefine time-to-digital converter in 90 nm cmos that amplifies a time residue. Solid-State Circuits, IEEE Journal of **43**(4), 769–777 (2008). DOI10.1109/JSSC.2008.917405

31. Lin, H., Taylor, K., Chong, A., Chan, E., Soma, M., Haggag, H., Huard, J., Braat, J.: Cmos built-in test architecture for high-speed jitter measurement. Test Conference, 2003. Proceedings. ITC 2003. International **1**, 67–76 (2003)
32. Maatta, K., Kostamovaara, J.: A high-precision time-to-digital converter for pulsed time-of-flight laser radar applications. Instrumentation and Measurement, IEEE Transactions on **47**(2), 521–536 (1998). DOI10.1109/19.744201
33. Mantyniemi, A., Rahkonen, T., Kostamovaara, J.: A nonlinearity-corrected cmos time digitizer ic with 20 ps single-shot precision. In: Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on, vol. 1, pp. I–513–I–516 vol.1 (2002). DOI10.1109/ISCAS.2002.1009890
34. Muhammad, K., Ho, Y.C., Mayhugh, T., Hung, C.M., Jung, T., Elahi, I., Lin, C., Deng, I., Fernando, C., Wallberg, J., Vemulapalli, S., Larson, S., Murphy, T., Leipold, D., Cruise, P., Jaehnig, J., Lee, M.C., Staszewski, R., Staszewski, R., Maggio, K.: The first fully integrated quad-band gsm/gprs receiver in a 90-nm digital cmos process. Solid-State Circuits, IEEE Journal of **41**(8), 1772–1783 (2006). DOI10.1109/JSSC.2006.877271
35. Naraghi, S., Courcy, M., Flynn, M.P.: A 9b 14w 0.06mm2 ppm adc in 90nm digital cmos. Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International pp. 168–169,169a (2009). DOI10.1109/ISSCC.2009.4977361
36. Nissinen, I., Kostamovaara, J.: Time-to-digital converter based on an on-chip voltage reference locked ring oscillator. Instrumentation and Measurement Technology Conference, 2006. IMTC 2006. Proceedings of the IEEE pp. 250–254 (2006). DOI10.1109/IMTC.2006.328409
37. Nutt, R.: Digital time intervalometer. Review of Scientific Instruments **39**(9), 1342–1345 (1968). DOI10.1063/1.1683667
38. Ouzounov, S., Roza, E., Hegt, J., van der Weide, G., van Roermund, A.: Analysis and design of high-performance asynchronous sigma-delta modulators with a binary quantizer. Solid-State Circuits, IEEE Journal of **41**(3), 588–596 (2006). DOI10.1109/JSSC.2005.864147
39. Paulus, C., Bluthgen, H.M., Low, M., Sicheneder, E., Bruls, N., Courtois, A., Tiebout, M., Thewes, R.: A 4gs/s 6b flash adc in 0.13 m cmos. VLSI Circuits, 2004. Digest of Technical Papers. 2004 Symposium on pp. 420–423 (2004)
40. Rahkonen, T., Kostamovaara, J.: The use of stabilized cmos delay lines for the digitization of short time intervals. Solid-State Circuits, IEEE Journal of **28**(8), 887–894 (1993). DOI 10.1109/4.231325
41. Raisanen-Ruotsalainen, E., Rahkonen, T., Kostamovaara, J.: A low-power cmos time-to-digital converter. Solid-State Circuits, IEEE Journal of **30**(9), 984–990 (1995). DOI10.1109/4.406397
42. Raisanen-Ruotsalainen, E., Rahkonen, T., Kostamovaara, J.: A high resolution time-to-digital converter based on time-to-voltage interpolation. Solid-State Circuits Conference, 1997. ESSCIRC '97. Proceedings of the 23rd European pp. 332–335 (1997)
43. Raisanen-Ruotsalainen, E., Rahkonen, T., Kostamovaara, J.: An integrated time-to-digital converter with 30-ps single-shot precision. Solid-State Circuits, IEEE Journal of **35**(10), 1507–1510 (2000). DOI10.1109/4.871330
44. Ramakrishnan, V., Balsara, P.: A wide-range, high-resolution, compact, cmos time to digital converter. VLSI Design, 2006. Held jointly with 5th International Conference on Embedded Systems and Design., 19th International Conference on pp. 6 pp.– (2006). DOI 10.1109/VLSID.2006.28
45. Razavi, B.: Principles of Data Conversion Systems. IEEE Press (1994)
46. Razavi, B.: RF Microelectronics. Prentice Hall (1997)
47. Robertson, D.: Specifications and Figures of Merit for Mixed-Signal-Circuits. ISSCC Tutorial (2002)
48. Roza, E.: Analog-to-digital conversion via duty-cycle modulation. Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on **44**(11), 907–914 (1997). DOI 10.1109/82.644044
49. Safi-Harb, M., Roberts, G.: Embedded narrow pulse measurement in digital cmos. Instrumentation and Measurement Technology Conference, 2006. IMTC 2006. Proceedings of the IEEE pp. 1195–1200 (2006). DOI10.1109/IMTC.2006.328449

50. Sansen, W.M.: Analog Design Essentials. Springer (2006)
51. Staszewski, R., Muhammad, K., Leipold, D., Hung, C.M., Ho, Y.C., Wallberg, J., Fernando, C., Maggio, K., Staszewski, R., Jung, T., Koh, J., John, S., Deng, I.Y., Sarda, V., Moreira-Tamayo, O., Mayega, V., Katz, R., Friedman, O., Eliezer, O., de Obaldia, E., Balsara, P.: All-digital tx frequency synthesizer and discrete-time receiver for bluetooth radio in 130-nm cmos. Solid-State Circuits, IEEE Journal of **39**(12), 2278–2291 (2004). DOI10.1109/JSSC.2004.836345
52. Staszewski, R., Vemulapalli, S., Vallur, P., Wallberg, J., Balsara, P.: 1.3 v 20 ps time-to-digital converter for frequency synthesis in 90-nm cmos. Circuits and Systems II: Express Briefs, IEEE Transactions on **53**(3), 220–224 (2006). DOI10.1109/TCSII.2005.858754
53. Staszewski, R., Wallberg, J., Rezeq, S., Hung, C.M., Eliezer, O., Vemulapalli, S., Fernando, C., Maggio, K., Staszewski, R., Barton, N., Lee, M.C., Cruise, P., Entezari, M., Muhammad, K., Leipold, D.: All-digital pll and transmitter for mobile phones. Solid-State Circuits, IEEE Journal of **40**(12), 2469–2482 (2005). DOI10.1109/JSSC.2005.857417
54. Stevens, A., Van Berg, R., Van der Spiegel, J., Williams, H.: A time-to-voltage converter and analog memory for colliding beam detectors. Solid-State Circuits, IEEE Journal of **24**(6), 1748–1752 (1989). DOI10.1109/4.45016
55. Swann, B., Blalock, B., Clonts, L., Binkley, D., Rochelle, J., Breeding, E., Baldwin, K.: A 100-ps time-resolution cmos time-to-digital converter for positron emission tomography imaging applications. Solid-State Circuits, IEEE Journal of **39**(11), 1839–1852 (2004). DOI10.1109/JSSC.2004.835832
56. Tisa, S., Lotito, A., Giudice, A., Zappa, F.: Monolithic time-to-digital converter with 20ps resolution. Solid-State Circuits Conference, 2003. ESSCIRC '03. Proceedings of the 29th European pp. 465–468 (2003). DOI10.1109/ESSCIRC.2003.1257173

Index