

A Flexible 5 ps Bin-Width Timing Core for Next Generation High-Energy-Physics Time-to-Digital Converter Applications

Lukas Perktold

Graz University of Technology
Rechbauerstr. 12, 8010 Graz, Austria
Email: lukas.perktold@student.tugraz.at

Jørgen Christiansen

CERN
CH-1211 Geneva23, Switzerland

Abstract—A new flexible low-power timing core for high time resolution time-to-digital converter (TDC) applications is presented. The chosen architecture allows a high number of channels using only one instance of the timing core to be efficiently implemented. It uses a multi-stage conversion scheme employing a delay-locked-loop (DLL) in its first stage and a power efficient passive interpolation scheme in its second stage, to achieve bin-widths as low as 5 ps. A delay-cell, using an additional zero in its signal path, to achieve delays shorter than 20 ps in a 130 nm technology is described. The architecture, important design trade-offs and simulation results are presented.

I. INTRODUCTION

In upcoming high-energy physics experiments, there is an emerging need for high-resolution time measurements. Novel detectors such as 3D silicon detectors or multigap resistive plate chambers have reached far into the sub 100 ps time-resolution domain. For high-time resolution detectors, a high number of channels combined with low power consumption is critical for system level integration.

In contrast to other applications such as all-digital-pll [1], laser ranging applications [2] or test and measurement applications [3], high-energy physics (HEP) experiments rely on high precision, single shot, time measurements. They can not use repetitive measurements or noise shaping mechanisms to improve timing resolution.

Recently, many publications report TDCs which reach sub 10 ps-rms single-shot time resolutions. Such TDCs need to resolve very small time differences to achieve high precision time resolutions. For an ideal system, the rms quantization error for a single edge measurement is given by

$$\sigma_{rms} = \frac{\Delta t}{\sqrt{12}}$$

where Δt represents the minimum bin-width of the system. Most often such TDCs are not limited anymore by the quantization error but by noise and mismatch of the system. A good rule of thumb is to design the system with a bin-width in the same order of magnitude of the required rms-time resolution. This keeps the contribution of the quantization error reasonably small. Sub-10 ps-rms time resolution TDCs require bin-widths which are substantially smaller than the

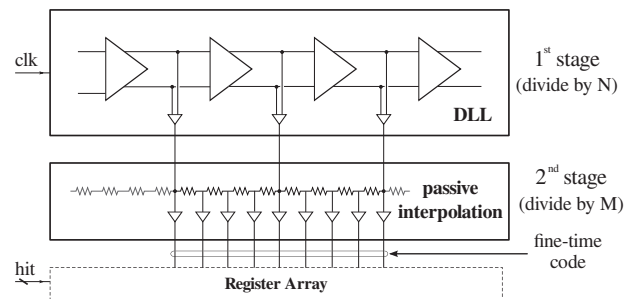


Fig. 1: Proposed architecture of the timing core employing a 2-stage interpolation scheme.

gate delay of modern CMOS processes. To overcome this intrinsic technology limit, architectures such as the vernier delay line [4], passive or active interpolation [5], [6], capacitive scaling [7], time difference amplifiers [8] or pulse shrinking [9] are in use for such kind of TDCs.

In this paper we present a timing core to achieve 5 ps bin-widths. The block has been designed to meet the demanding requirements set by the high energy physics community. Only one instance of the timing core needs to be implemented per TDC to serve all the channels. This allows for TDC designs with only a little power overhead per channel. Best practices have been employed to reduce noise and mismatch effects to achieve a rms-time resolution in the order of the bin-width of the system.

II. TIMING CORE ARCHITECTURE

The proposed architecture of the TDC is shown in Fig. 1. It consists of two stages which subsequently divide the input clock period T into smaller pieces. In the first stage an N element delay-line is used to generate N equally spaced signals with a bin-width equal to T/N . In a second stage, resistive division is employed to overcome the gate-delay limitation of the technology and to further reduce the bin-width of the system by a factor of M . With such a scheme an interpolation factor of NM and bin-widths as small as T/NM can be achieved. The scheme can easily be extended by a counter which counts completed clock cycles to increase the dynamic range of the TDC.

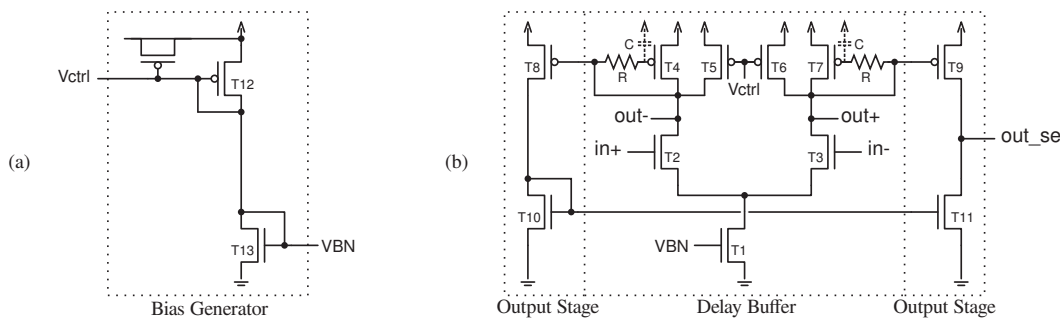


Fig. 2: Schematic representation of the bias generator and one delay cell.

In total a set of NM signals are generated which we refer to as the fine-time code. This fine-time code is sent to an array of registers to precisely determine the arrival time of a hit signal. The low power time capture registers and their configuration are beyond the scope of this paper and will be presented in future publications.

To reduce the number of elements within the delay line it is desirable to keep the interpolation factor NM as small as possible. To achieve small bin-widths at the same time, the delay line itself is required to run at a high input clock frequency. For a given interpolation factor we can choose N and M independently. For performance reasons it is advantageous to choose a high N and run the delay buffers as fast as possible. Running the delay buffers at a faster speed decreases their internal fall-/rise-times and makes the circuit more robust against coupled noise and mismatch. To keep the clock frequency within reasonable limits an interpolation factor of $N = 32$ in the first stage and $M = 4$ in the second stage has been chosen to achieve bin-widths as small as 5 ps when running the delay-line at 1.5625 GHz.

High-energy physics applications require all the channels of the TDC to be perfectly synchronized to a precise system reference clock and that the timing be held stable over process, voltage and temperature (PVT) variations. To keep the bin-width constant, the delay of the delay buffers is made adjustable and is controlled by a DLL. Making the delay buffers controllable assures a fixed and known bin-width of the fine-time code. This avoids the implementation of spare delay cells and their time capture registers, which would otherwise be necessary to be able to always cover a complete clock cycle in the presence of PVT variations. For a high channel count, TDC the implementation of spare registers for each channel is not acceptable as it would drastically increase power and area consumption of the TDC. Additionally, a control loop allows the delay cell to self adjust to different input clock frequencies and gives the possibility to trade off time resolution and power consumption.

While the delay-line is built up fully differential the distribution of the fine-time code is done in a single-ended manner. A fully-differential delay-line helps to reduce the circuit's sensitivity to coupled noise and improves matching. As most of the noise is introduced by the long delay of the delay line and only a little by the short propagation delay of the distribution circuit, it is feasible to use a fully-differential delay

line and distribute the fine-time code in a single-ended manner. Similar arguments are true regarding matching. Additionally, a fully-differential implementation of the delay-line allows delay buffers with very short propagation delay to be designed. Normally, for non-inverted output codes of the DLL, a pseudo-differential delay cell which increases power consumption or a non-inverting delay buffer which doubles the gate delay has to be used.

III. ZERO PEAKING BUFFER STAGE

Fully-differential delay buffers have been designed to be able to achieve bin-widths in the first stage smaller than 20 ps in a 130 nm technology. For comparison, the gate delay of a symmetrically designed inverter in the same technology, not employing any control circuitry, is of the order of 25 ps. The implemented circuit is shown in Fig. 2b and its corresponding dimensions are given in Table I.

The delay buffer is based on the circuit as reported in [10]. To reduce the delay of the buffer, a resistance in series with the gate capacitance of the diode connected PMOS loads is included. For the basic operation of the buffer the resistance can be neglected. The buffer works as follows. It uses the switches $T2$ and $T3$ to switch the current from one branch to the other. Thereby, the current flowing through either branch is defined by the NMOS bias voltage VBN . The low level of the differential output signal is defined by the top PMOS diode connected load. To be able to reach the top supply voltage an additional PMOS, controlled by the voltage $Vctrl$, is used. The total delay can be approximated by the relation

$$\delta \approx \frac{C_{eff} \cdot V_{Osc}}{I_{Bias}}$$

where C_{eff} represent the parasitic capacitance at the output node, V_{Osc} the oscillation voltage of the output signal and I_{Bias} the current defined by the tail NMOS current source. This relation just gives an indication of the speed as the effective speed will depend on both the charging and the discharging current, which both differ.

The speed of the buffer can be increased by increasing the width of the top PMOSs $T5$ and $T6$. This reduces the portion of current flowing through the PMOS diode connected load and effectively reduces the voltage V_{Osc} for a given tail current, I_{Bias} . Increasing the width of the voltage controlled PMOSs by a factor of 1.5 yields a 9% increase in speed. To further speed up the circuit a resistance R in series with

TABLE I: Device dimensions of the circuit shown in Fig. 2

Device	Width	Length
T1 & T13	18 μm	0.6 μm
T2 & T3	4 μm	0.12 μm
T5 & T6	4.5 μm	0.12 μm
T4 & T7	3 μm	0.12 μm
T8 & T9	3 μm	0.12 μm
T10 & T11	1.36 μm	0.12 μm
T12	7.5 μm	0.12 μm
R	11 k Ω	-

the PMOS's gate capacitance, C , is introduced [11]. This generates a zero at $-1/RC$ which hides the gate capacitance during the switching process and reduces C_{eff} at high frequencies. Another 11% gain in speed can be achieved with this technique.

To generate a single-ended version of the signal propagating in the delay line an output stage of a symmetrical operational-transconductance-amplifier (OTA) is used. The single-ended output is used to connect to the subsequent interpolating stage to generate sub-gate delay bin-widths. The connection to successive cells within the delay-line is made fully differential. To limit an increase in delay of the cell, additional loading at the output should be kept small. Since most of the delay mismatch between cells is introduced when converting from differential-to-single-ended signals, a trade-off between speed and matching needs to be made. For good delay matching across cells fast signals in the output buffer stage are required.

The biases of the cell are generated by the circuit shown on the left of Fig. 2a. The PMOS bias V_{ctrl} is directly connected to the big DLL loop filter capacitance, referenced to the upper supply voltage level. The NMOS bias V_{BN} is generated by a V-to-V converter. If operated at very high speeds (low V_{ctrl} voltages) the control gain is roughly equally shared among V_{ctrl} and V_{BN} and the loop filter PMOS varactor capacitance is maximized. This helps to reduce the supply noise sensitivity of the cell. For lower speed operation, noise is less critical.

IV. SUB-GATE DELAY BIN-WIDTHS

The fine-time interpolation is accomplished by resistive division between two edges with a small time difference as proposed by [6]. The concept of passive interpolation using resistive division is illustrated in Fig. 1. If the edges of two successive signals are overlapping, the resistive division will ideally generate signals as shown in Fig. 3a. In reality the resistive division will not work perfectly. Each node will see some parasitic capacitance which will cause an additional RC delay. This will slow down the propagation of the signal leading to larger bins at the beginning and to smaller ones at the end as shown in Fig. 3b. For a known network the RC delay of the resistive division can be compensated by designing a non-linear interpolation scheme successively increasing the resistances connected to the next node. However, such a non-linear interpolation can only compensate for constant RC delays and does not scale with different input clock frequencies.

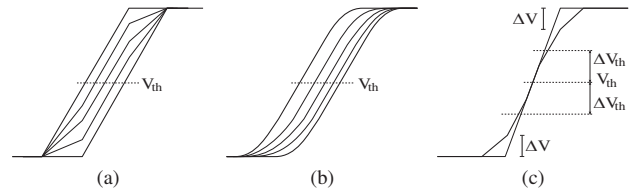


Fig. 3: Passive interpolation: (a) ideal resistive division (b) resistive division in presence of RC delay (c) resistive division threshold sensitivity

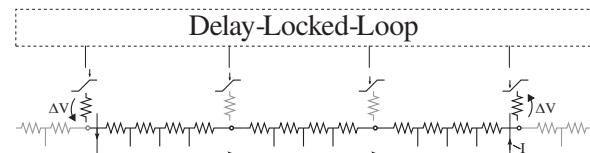


Fig. 4: Passive interpolation current flow illustration across bins.

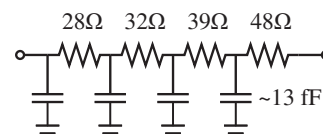


Fig. 5: Non-linear resistive division network.

As we want the interpolation to work reasonably well also for different input clock frequencies, the resistances are kept small.

If resistive interpolation is to be used to interpolate between more overlapping edges, as it would be the case in a DLL, there is a direct connection between successive nodes. This causes a current to flow across the resistances connecting the different nodes, as illustrated in Fig. 4. Ideally only one driver is driving the current flowing through the resistive ladder and only one driver is sinking its current. The number of bins involved in this effect is directly related to the rise-/fall-times of the signal. The bigger the rise-/fall-time the more bins are involved. Due to the finite output/input resistance of the drivers some of the current is flowing to adjacent nodes causing an additional voltage drop at that node. This leads to a degradation in rise-/fall-times of the signal and can cause additional time mismatches between bins in presence of voltage threshold mismatches as depicted in Fig. 3c. In order to avoid a big influence of this effect the internal driving resistance needs to be small compared to the sum of the resistances connecting successive nodes. A reduction of this effect also helps to reduce the number of delay elements needed at the beginning and the end of the delay line to reach uniform bin-widths. A trade-off between power consumption and rise-/fall-times of the signals has to be made. In any case, the power consumption of passive interpolation will be substantially lower than that for an active interpolation or vernier delay-line architecture, where a total $M \cdot N$ fast stages would need to be implemented to achieve the same interpolation ratio.

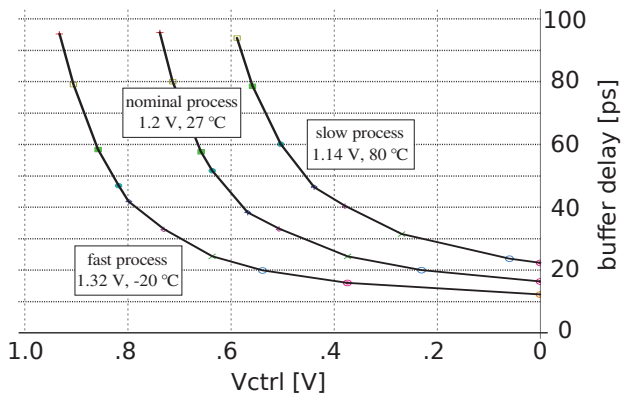


Fig. 6: Simulated buffer delay for different control voltages.

5 ps uniform bin-widths could be achieved using resistor sizes as shown in Fig. 5. On average, 13.4 fF parasitic capacitance is present at each node. 37 % due to parasitic capacitance of the resistors and 63 % due to the loading at the output.

V. SIMULATION RESULTS

Fig. 6 shows the post layout extracted simulated delay of the delay buffers. In a real system the delay cell can be operated with a bias V_{Ctrl} as low as 0.1 V, without putting the discharging current source of the loop filter capacitance into the linear-region. With this implementation a delay as small as 18 ps can be achieved. The bin-width across all generated 128 bins for different input frequencies is shown in Fig. 7. Simulation results show good uniformity for bin-widths reaching from 5 ps to 20 ps. The graph shows a clear degradation of uniformity for larger bin-widths. This is due to the constant RC delay of the passive interpolation network which is optimized to yield uniformity for 5 ps bin-widths.

The $1-\sigma$ variation of the bin-width across all 128 bins, more often referred to as the differential-non-linearity (DNL), was found through monte-carlo simulations to be 0.48 ps or approximately 10 %. This value does not include the variations someone would expect from the circuit necessary to distribute the fine-time code to the time capture registers. Although the variation of the distribution buffers has been kept low ($\sigma = 1.5$ ps), calibrating the delay of those buffers is necessary.

The current consumption of a delay cell running at 1.5625 GHz and 1.2 V supply voltage was simulated to be 987 μ A. The buffers to drive the passive interpolation network, to achieve rise-times of 125 ps, consume 286 μ A when running at 1.5625 GHz input clock frequency and 1.2 V supply voltage. The whole timing core consumes as much as 51 mA when running with a 1.5625 GHz clock and a 1.2 V supply voltage. In a final implementation the power consumption of the timing core is to be shared among 64 - 128 channels.

VI. CONCLUSION

A fine-code interpolation stage to generate bin-widths of 5 ps has been presented. The interpolating stage can be operated with different input frequencies allowing bin-widths reaching from 20 ps to 5 ps. Running at lower input frequencies allows a reduction in current consumption by up to a factor

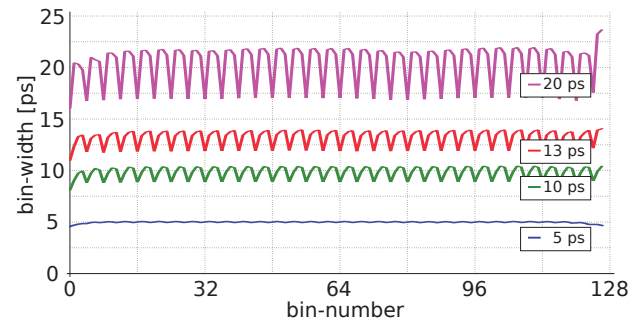


Fig. 7: Simulated bin-widths for different buffer delays.

of 4. All cells have been carefully dimensioned to reduce time mismatch between different bins and improve the noise sensitivity of the circuit. The presented design can be used to build up a flexible, low power, high time resolution TDC with many channels. A test chip for proof of concept and to evaluate the rms-time resolution has been submitted for prototype production in a commercial 130 nm technology.

ACKNOWLEDGMENT

I want to thank Kostas Kloukinas as my supervisor at CERN for his continuing support of my work and for the ever encouraging discussions.

REFERENCES

- [1] C.-M. Hsu, M. Straayer, and M. Perrott, "A low-noise, wide-bw 3.6ghz digital fractional-n frequency synthesizer with a noise-shaping time-to-digital converter and quantization noise cancellation," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, feb. 2008, pp. 340–617.
- [2] I. Niissinen, A. Mantyniemi, and J. Kostamovaara, "A cmos time-to-digital converter based on a ring oscillator for a laser radar," in *Solid-State Circuits Conference, 2003. ESSCIRC '03. Proceedings of the 29th European*, sept. 2003, pp. 469–472.
- [3] T. Hashimoto, H. Yamazaki, A. Muramatsu, T. Sato, and A. Inoue, "Time-to-digital converter with vernier delay mismatch compensation for high resolution on-die clock jitter measurement," in *VLSI Circuits, 2008 IEEE Symposium on*, june 2008, pp. 166–167.
- [4] L. Vercesi, A. Liscidini, and R. Castello, "Two-dimensions vernier time-to-digital converter," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 8, pp. 1504–1512, aug. 2010.
- [5] L. min Lee, D. Weinlader, and C.-K. Yang, "A sub-10-ps multiphase sampling system using redundancy," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 1, pp. 265–273, jan. 2006.
- [6] S. Henzler, S. Koeppe, D. Lorenz, W. Kamp, R. Kuenemund, and D. Schmitt-Landsiedel, "A local passive time interpolation concept for variation-tolerant high-resolution time-to-digital conversion," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 7, pp. 1666–1676, july 2008.
- [7] J.-P. Jansson, A. Mantyniemi, and J. Kostamovaara, "A cmos time-to-digital converter with better than 10 ps single-shot precision," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 6, pp. 1286–1296, june 2006.
- [8] S. Mandai, T. Iizuka, T. Nakura, M. Ikeda, and K. Asada, "Time-to-digital converter based on time difference amplifier with non-linearity calibration," in *ESSCIRC, 2010 Proceedings of the*, sept. 2010, pp. 266–269.
- [9] R. Rashidzadeh, R. Muscedere, M. Ahmadi, and W. Miller, "A delay generation technique for narrow time interval measurement," *Instrumentation and Measurement, IEEE Transactions on*, vol. 58, no. 7, pp. 2245–2252, july 2009.
- [10] J. Maneatis, "Low-jitter and process independent dll and pll based on self biased techniques," in *Solid-State Circuits Conference, 1996. Digest of Technical Papers. 42nd ISSCC., 1996 IEEE International*, feb 1996, pp. 130–131, 430.
- [11] S. Anand and B. Razavi, "A cmos clock recovery circuit for 2.5-gb/s nrz data," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 3, pp. 432–439, mar 2001.