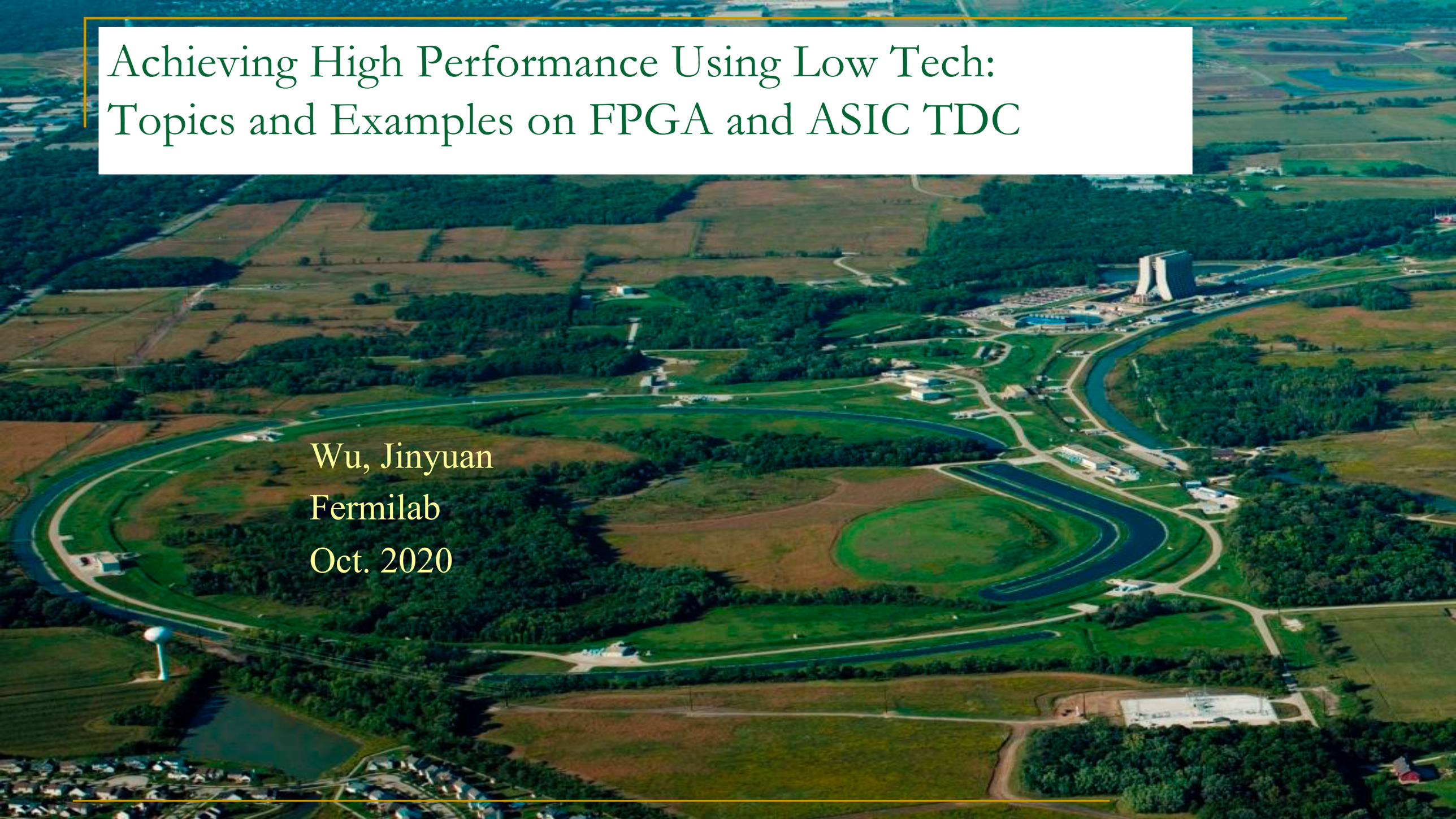
Achieving High Performance Using Low Tech: Topics and Examples on FPGA and ASIC TDC



Wu, Jinyuan
Fermilab
Oct. 2020

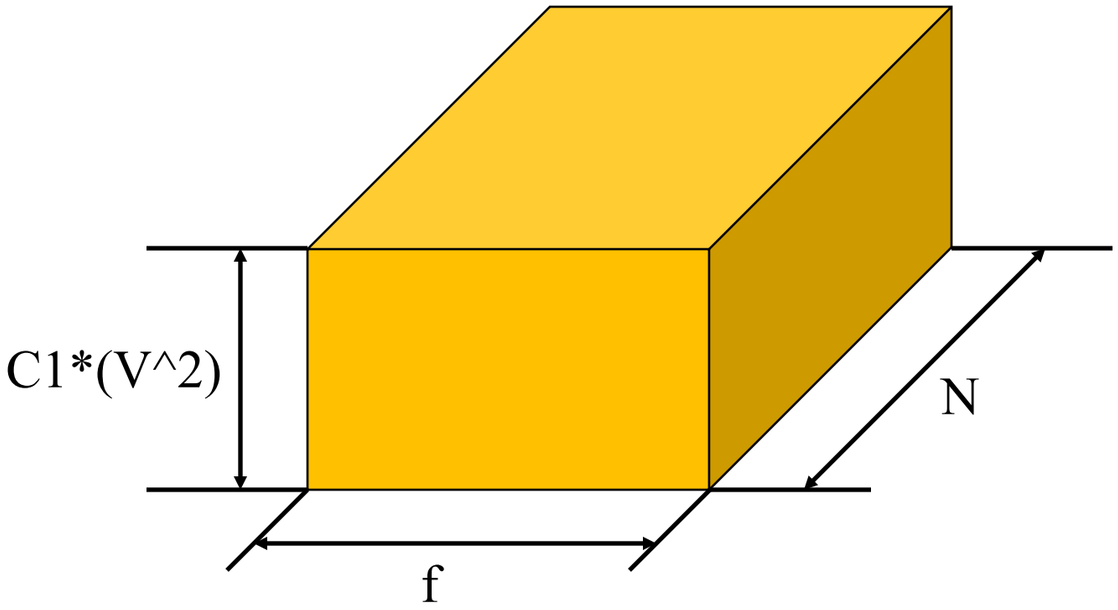
Introduction

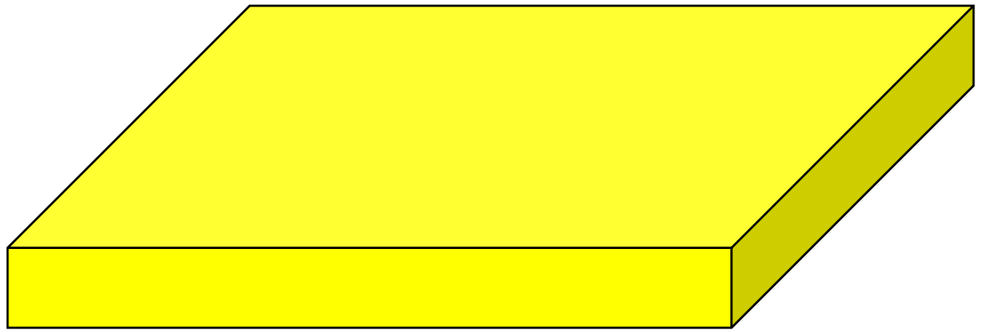
- Silicon technologies have been constantly progressing but will never satisfy demands of performance in HEP projects, so currently available technologies can always be considered low tech.
- FPGA can be considered low tech comparing with ASIC for TDC implementation.
  - Delay line based TDC was developed in ASIC in 1990's.
  - The TDC was transplanted into FPGA in 2000 to 2010 with simplifications.
  - FPGA TDC tricks are transplant back to ASIC.
- Timing uncertainty confinement is a necessary practice for good TDC implementation.
- Power saving can be achieved with operation reduction practices including event triggering using gated ring oscillator.

Resource Usage Awareness

New Tech and Resources

$$P = N*C1*(V^2)*f \qquad \text{Number of Operations/second} = N*f$$

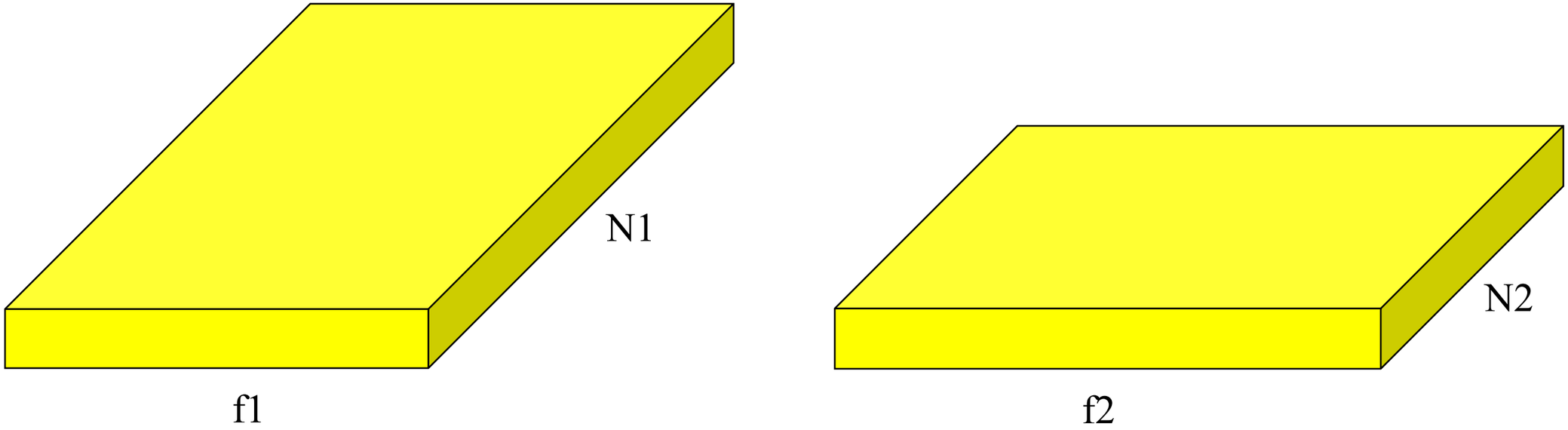




- N: number of gates; C1: average capacitance of each gate; V: voltage; f: frequency.
- New technology reduces C1 and V and increases maximum f.
- As technologies progress, higher rate of operations become affordable with same P.

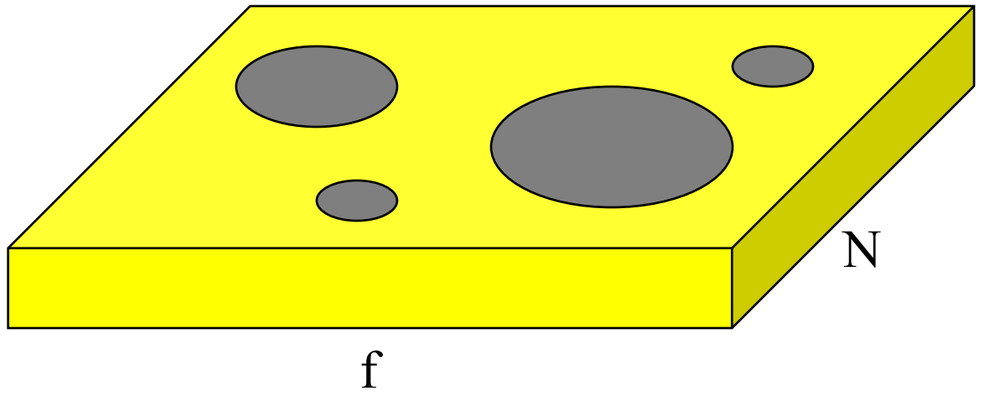
Silicon Area and Frequency

$$P = N*C1*(V^2)*f \qquad \text{Number of Operations/second} = N*f$$



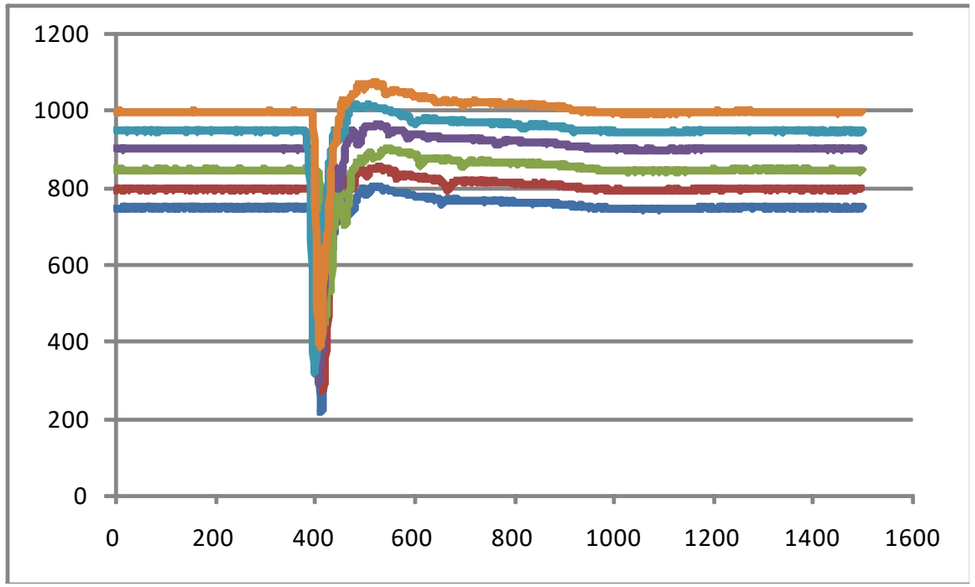
- One may duplicate functional blocks to handle high operation rate.
- One may also use higher operating frequency to save silicon area and to reduce cost.
- The power consumption will remain the same.

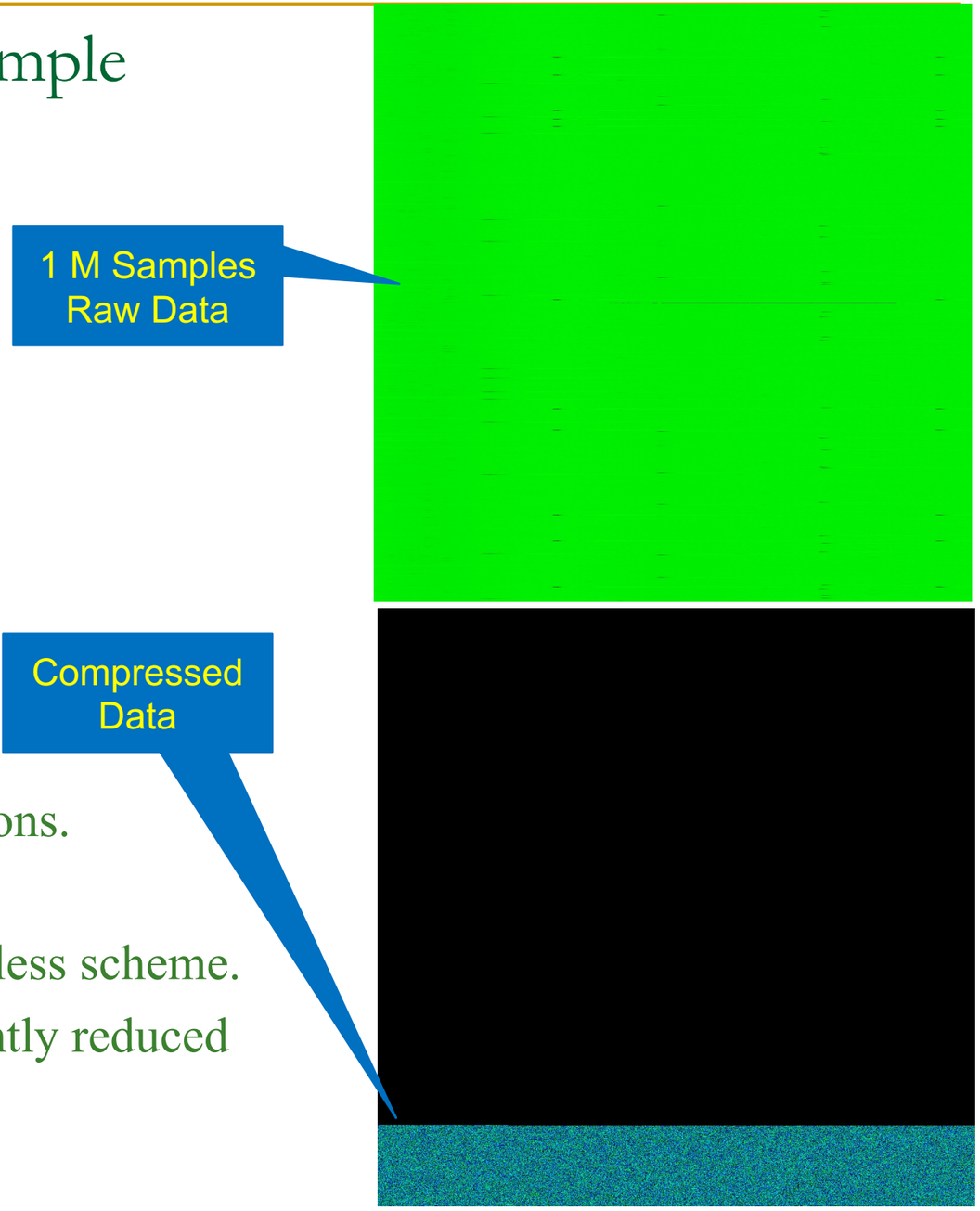
Resource Saving in Given Task



- Large number of operations may not be necessary.
- Eliminating unnecessary operations can reduce power consumption.
- Eliminating unnecessary functional blocks can reduce cost as well as develop time.

Resource Saving in a Broader View, an Example

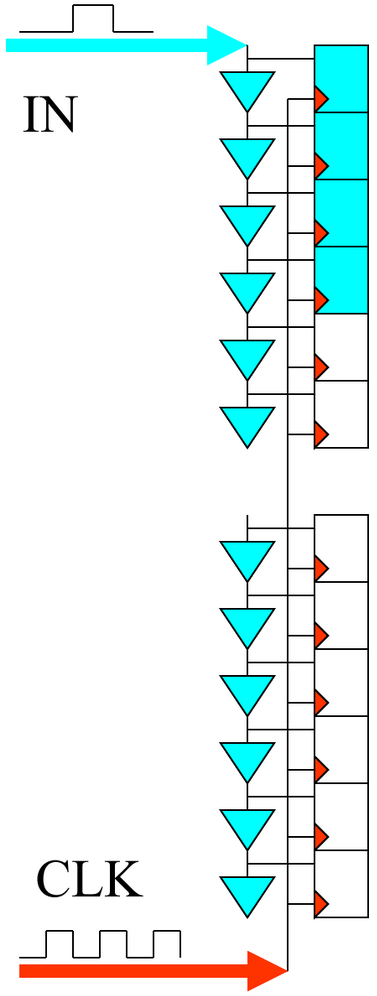


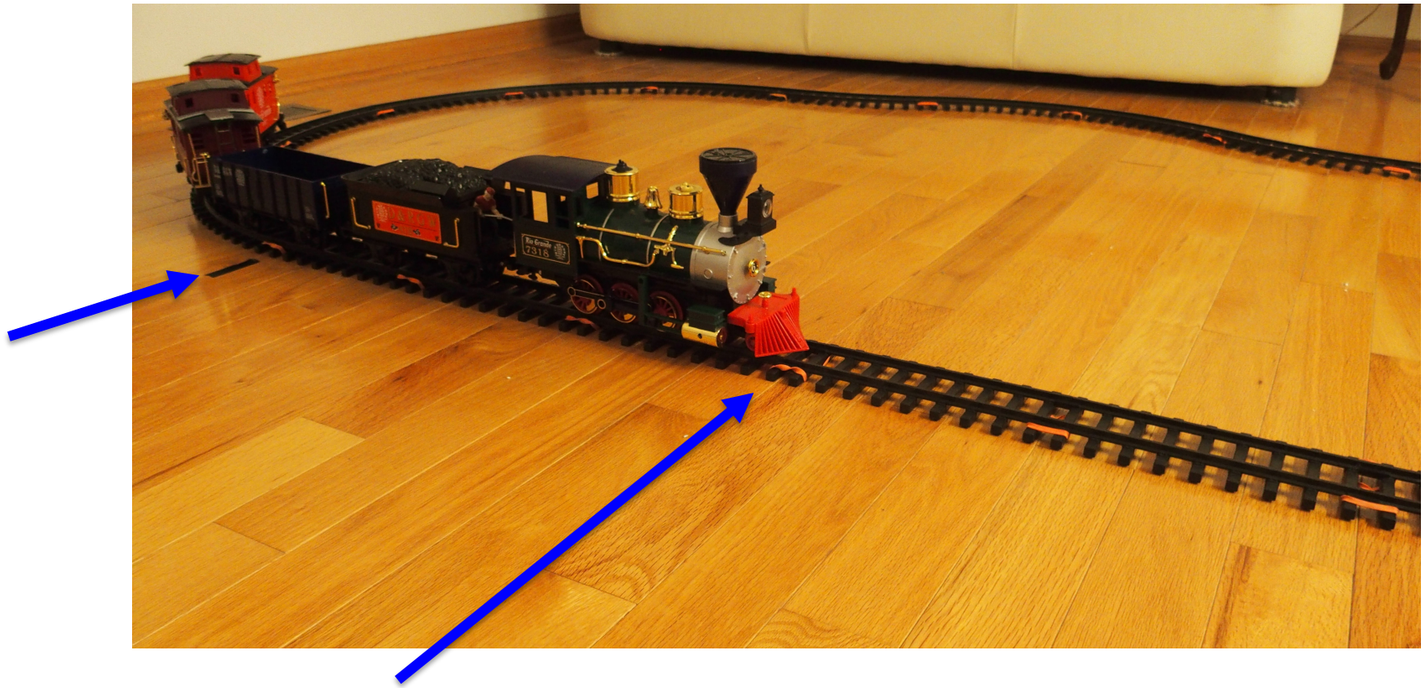


- Pulse shape digitization is needed for many applications.
- The raw ADC samples represents large data volume.
- The data can be compressed using a very simple lossless scheme.
- Buffer storage, transmission bandwidth are significantly reduced after compression.

TDC: From ASIC to FPGA and Back to ASIC

Delay Line Based TDC





- The input signal propagates in the delay chain.
- The “snapshot” is taken into the register array.
- The position of the signal transition is encoded into arrival time relative to the CLK.

Mainstream ASIC TDC in History (& Today)

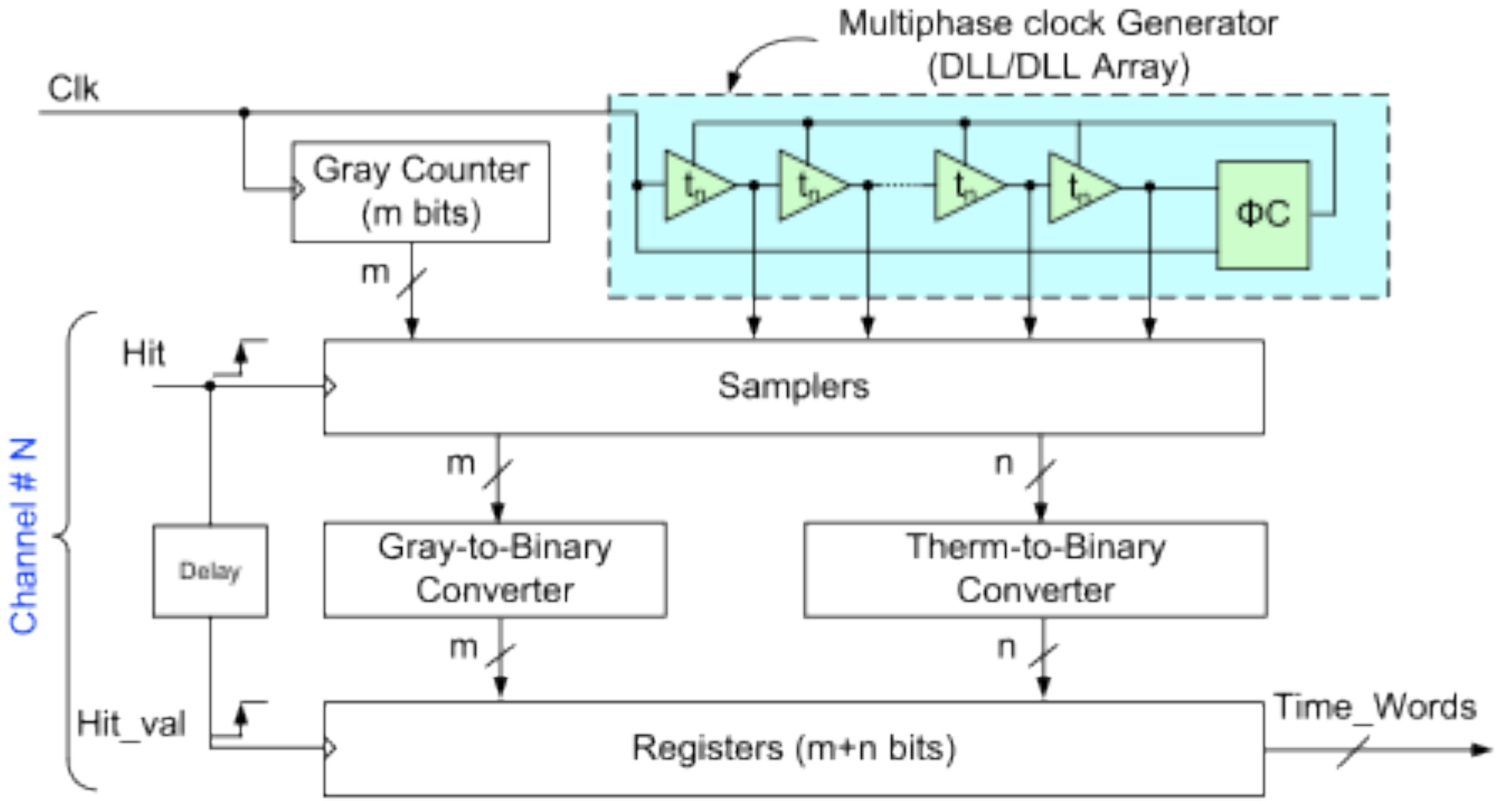
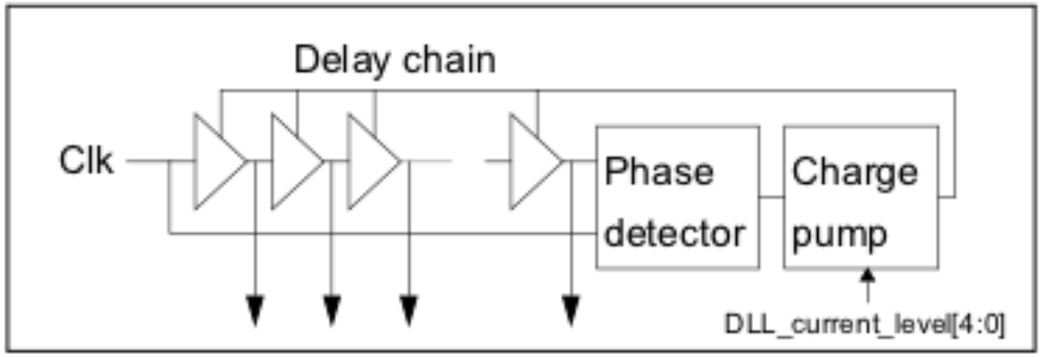
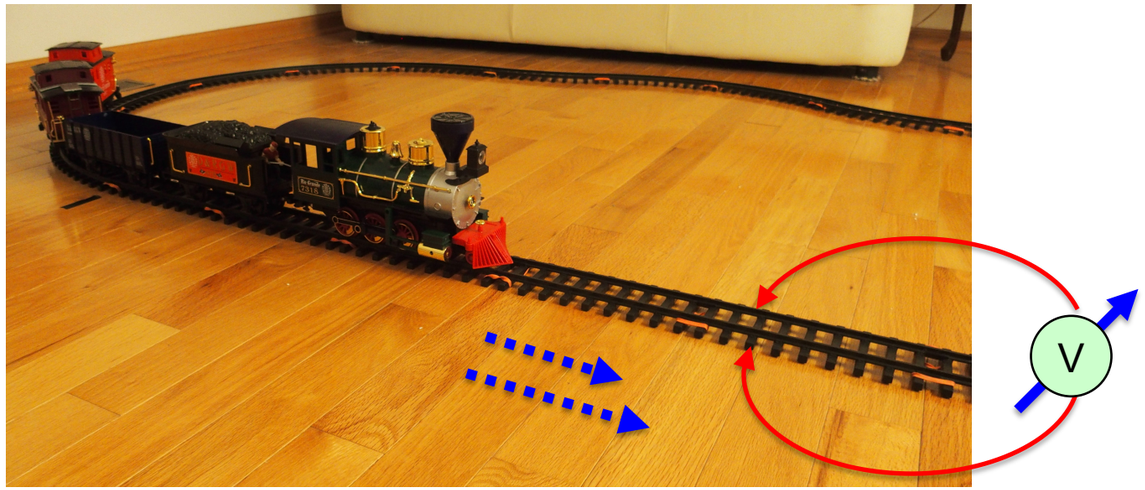


Figure 1: Tranditional architecture of a flash TDC.

- A fast clock (e.g. 320 MHz) is sent to the delay chain.
- The delay cells are adjusted to match the clock period.
- The outputs of the delay taps are routed to a set of FF registers.
- The leading edge of the HIT signal captures the delay pattern.

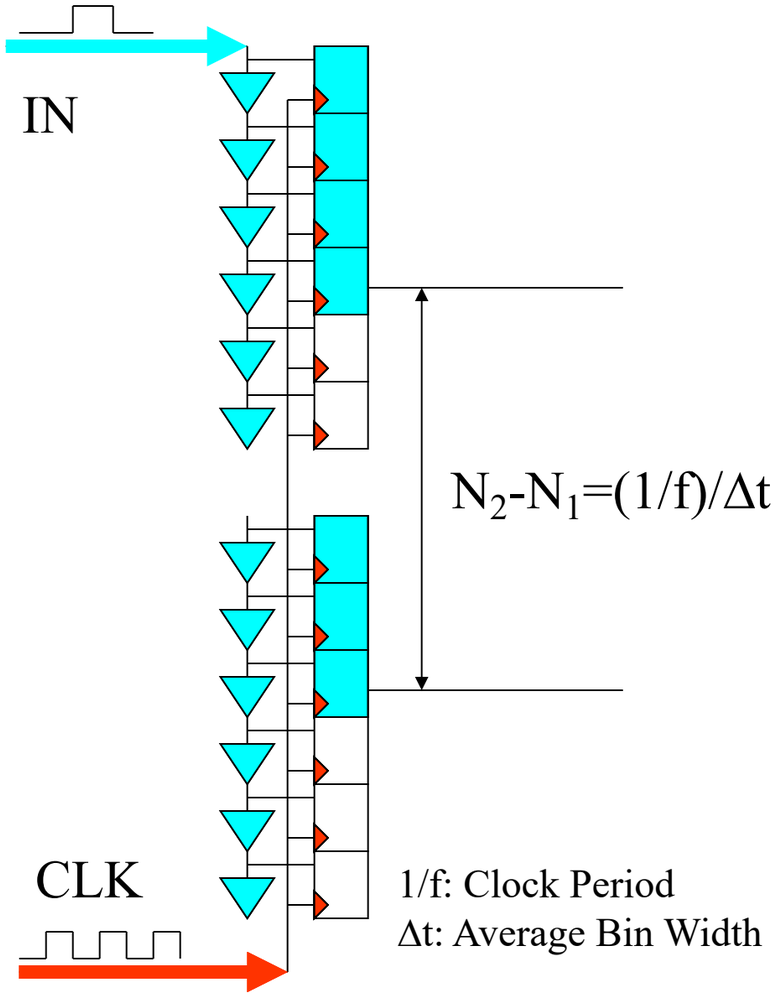
Good Things in ASIC But Not in FPGA

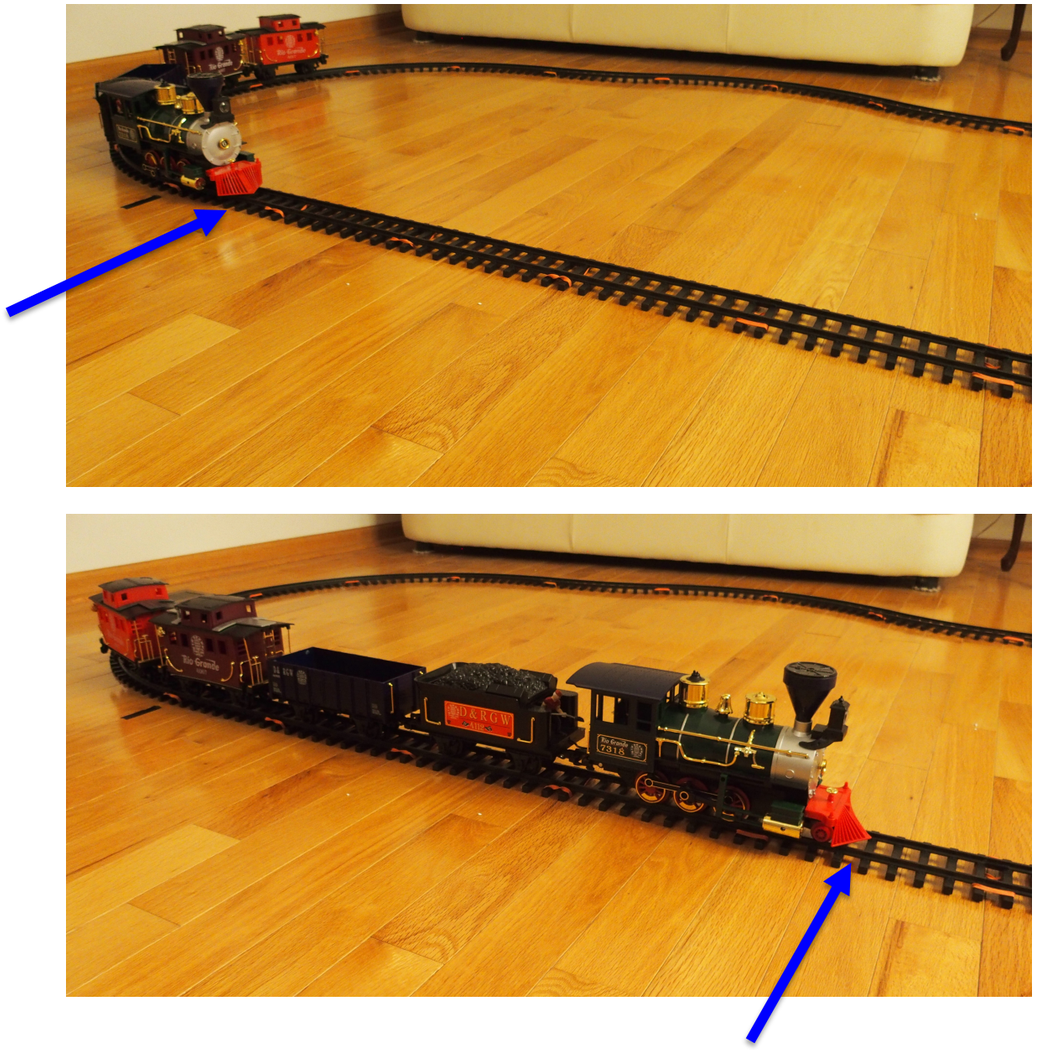




- The phase detector, charge pump and voltage-controlled delay cells are available in ASIC but difficult to implement in FPGA.
- Pro:
  - The delay cells are adjusted to a known speed. (e.g., 100 ps/tap)
- Con:
  - The delay cells are slowed down, rather than running at the full speed.
  - Charge pump consumes extra power.

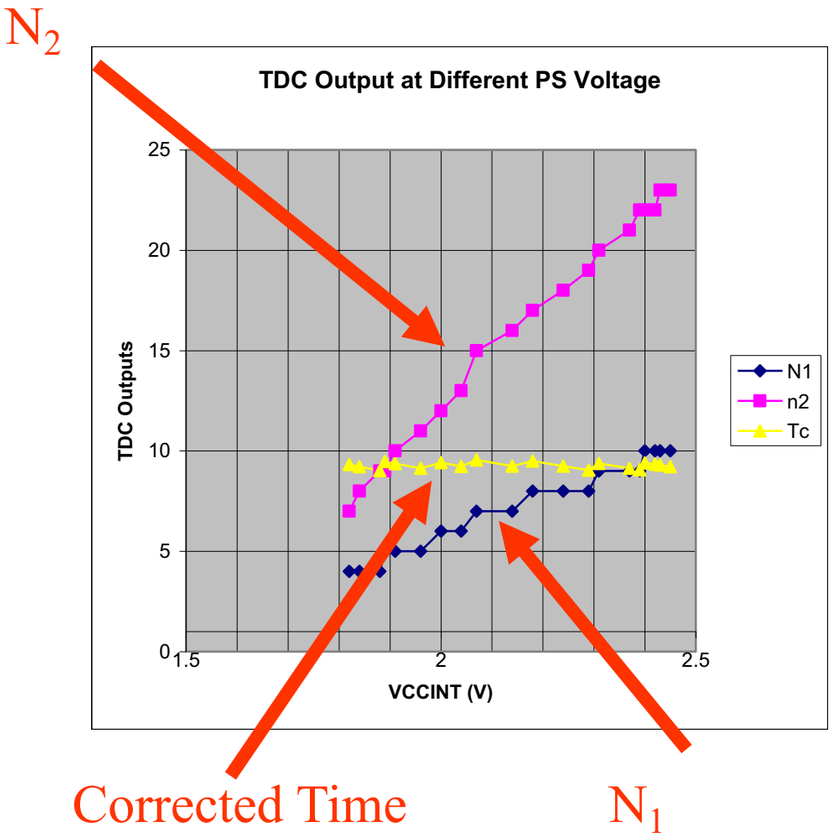
Un-adjustable Delay Lines FPGA





- For each input, make two measurements.
- The extra measurement can be used to calibrate temperature variation as well as increase measurement precession.

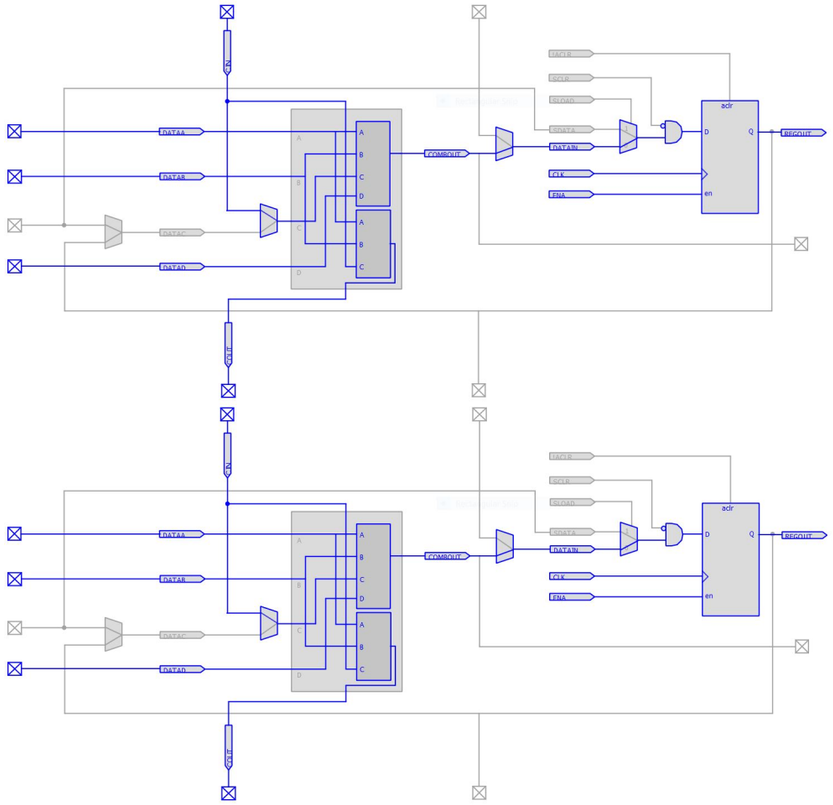
Temperature and Power Supply Voltage Calibration

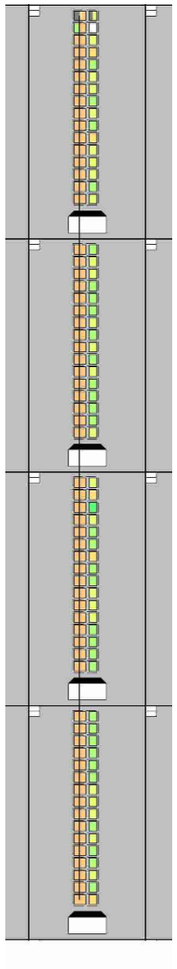


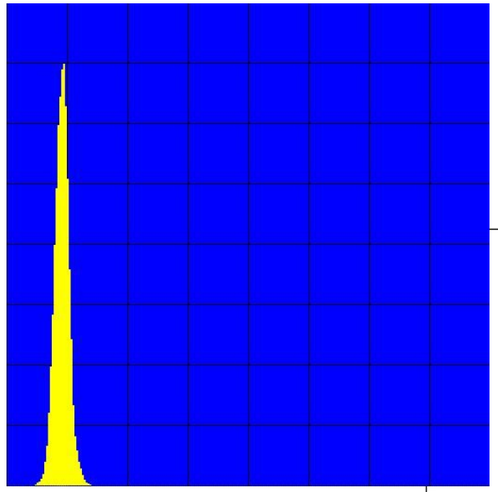
$$T_c = T\frac{(N_1 + N_0)}{\langle N_2 - N_1 \rangle} = \frac{T}{L}(N_1 + N_0)$$

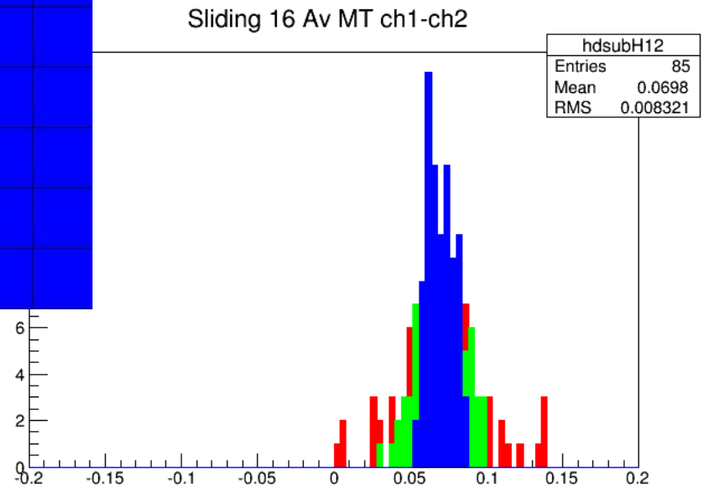
- Temperature and power supply voltage will cause the delay cell speed change.
- With multiple measurements, the variation can be corrected.
- The delay cell always run at full speed rather than being slowed down.

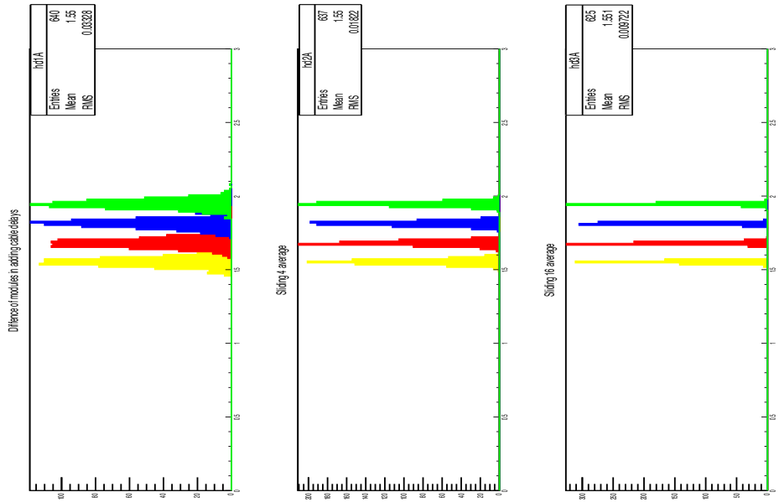
FPGA TDC





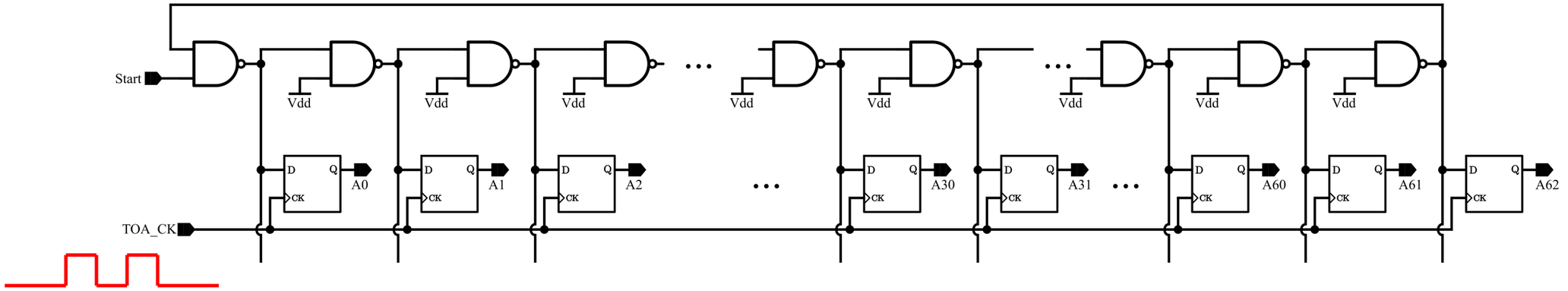






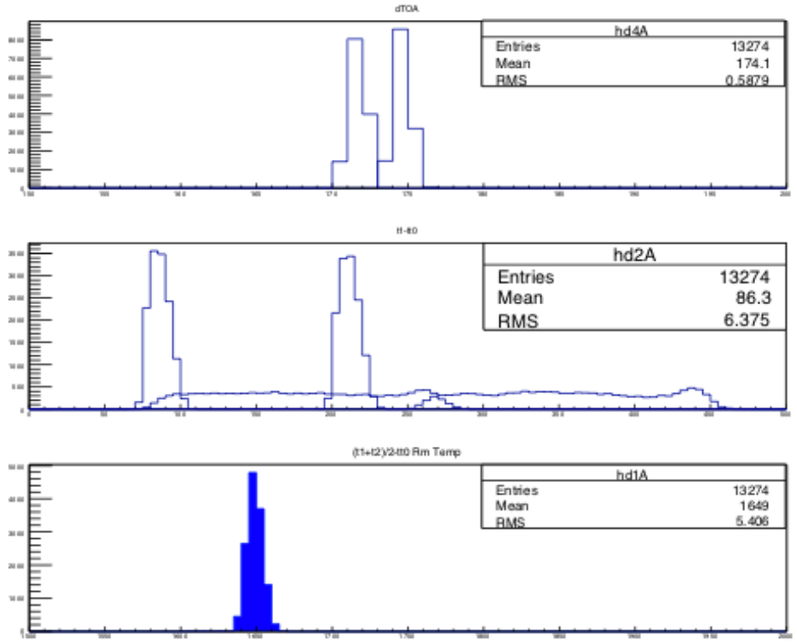
- The delay chain is implemented using carry chain inside FPGA.
- Today FPGA TDC precision is better than 20 ps, some high-end ones better than 6 ps.

Transplanting Back to ASIC TDC

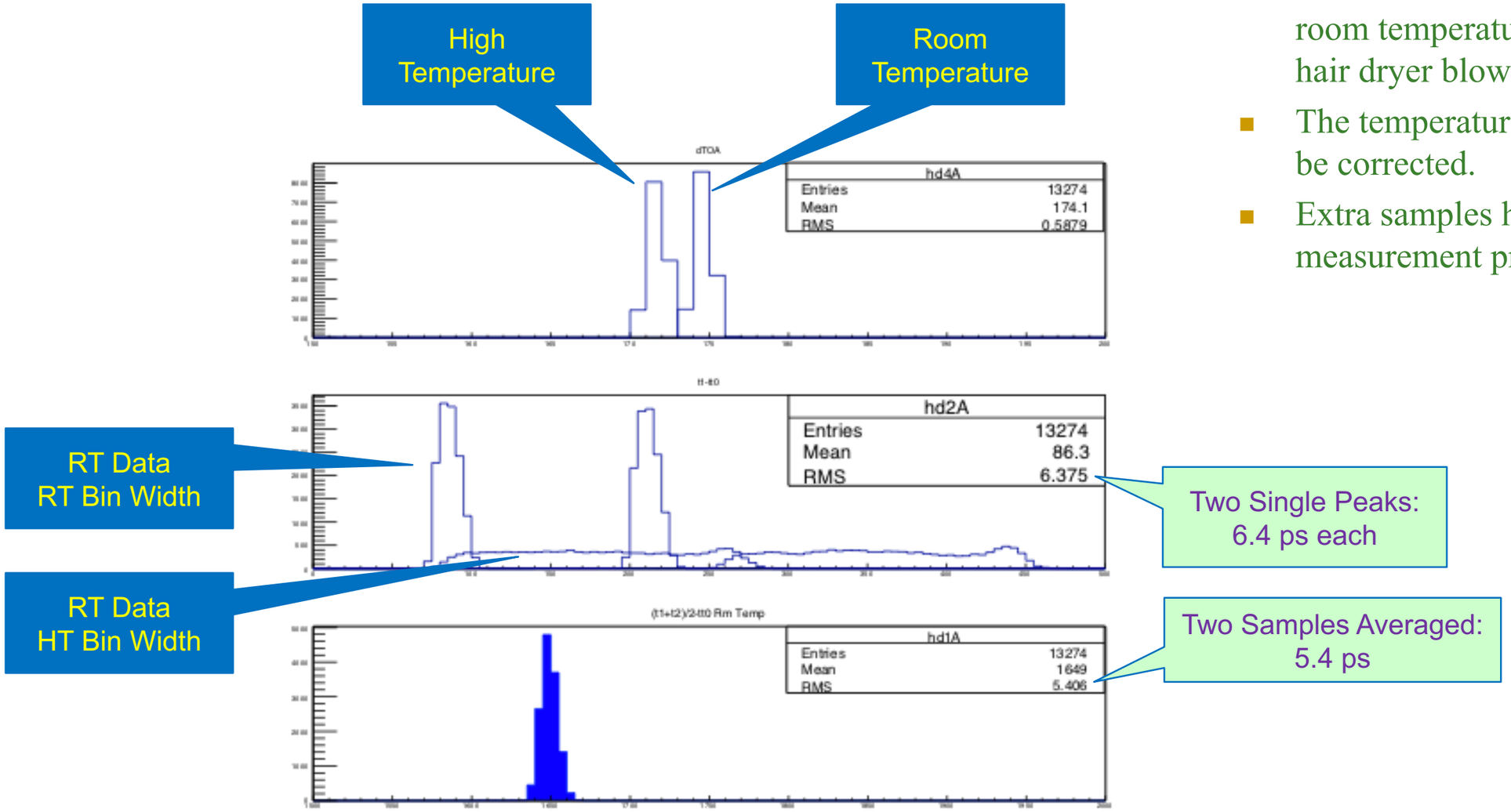


- The delay chain is not controlled so that it runs at full speed.
- Temperature variations and uneven width of the bins are corrected in digital domain.
- Two samples also help improving measurement precision.
- Note that the delay cells are inverting NAND gates with single gate delay (17 ps in 65 nm Technology).
- No hit no flip: very low power consumption.

- It works.

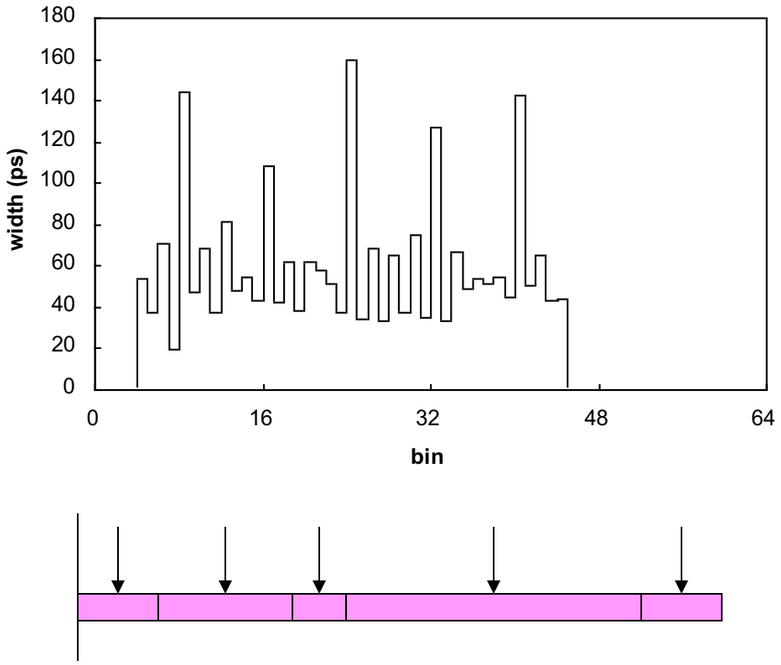


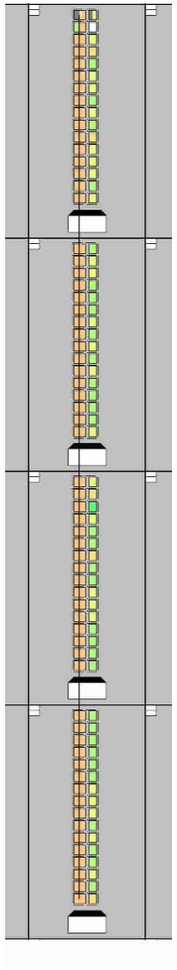
Histogram



- The measurements are taken at room temperature and under hair dryer blow.
- The temperature variation can be corrected.
- Extra samples helps improving measurement precision.

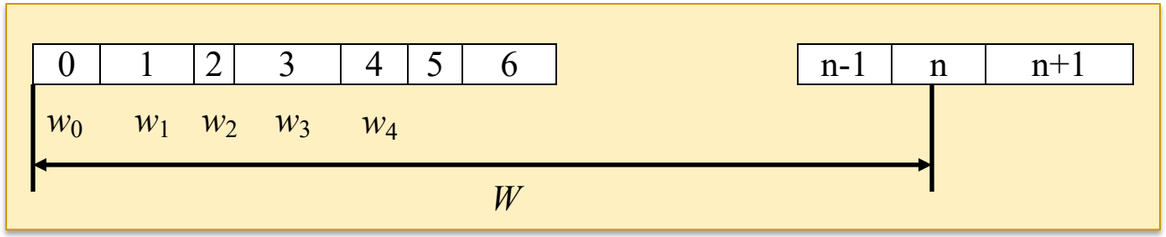
Uneven Bin Width

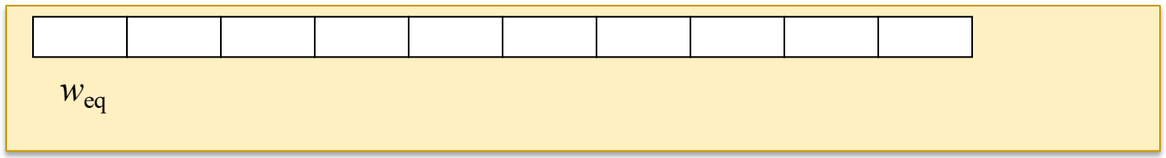




- Another disadvantage of FPGA delay line is that the bin widths are uneven due to the carry chain structure and clock region.
- Bin-by-bin calibration is usually necessary for high-end TDC implementation.
- With bin-by-bin calibration, all bin widths are known and the time value at the center of each bin is the best estimate of the measurement.

Equivalent Bin Width





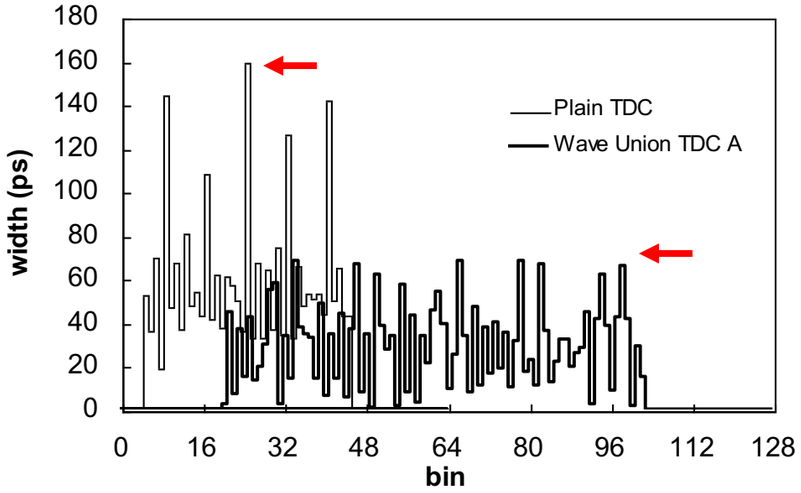
$$\sigma_{eq}^{\ 2} = \sum_i \left(\frac{w_i^{\ 2}}{12} \right) \left(\frac{w_i}{W} \right) \qquad w_{eq} = \sigma_{eq} \sqrt{12}$$

Error contributed by the i-th bin

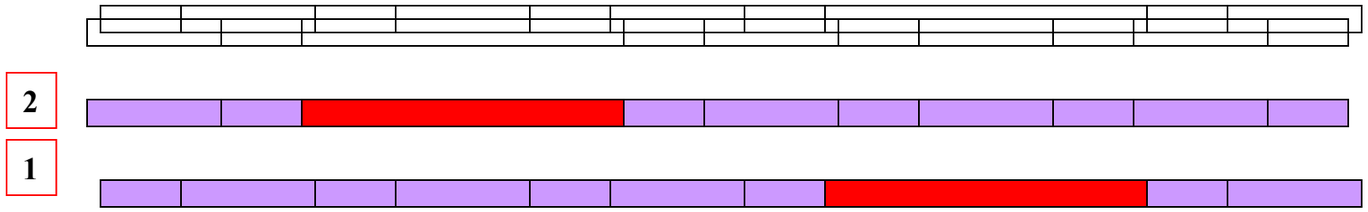
Probability of falling into the i-th bin

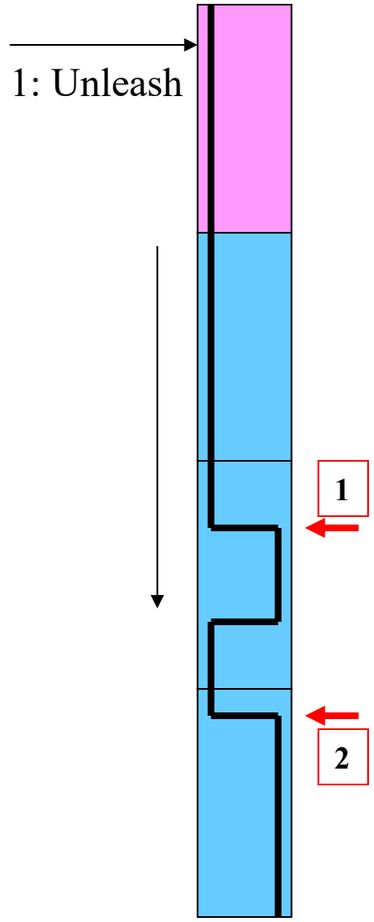
- To compare digitizers with uneven bin widths, a concept called “equivalent bin width” is defined.
- The equivalent bin width is NOT a regular average bin width.
- Both bin width and probability of falling into the bin are considered while calculating contribution of a bin to the measurement error.

Ultra-wide Bins and Multiple Measurements



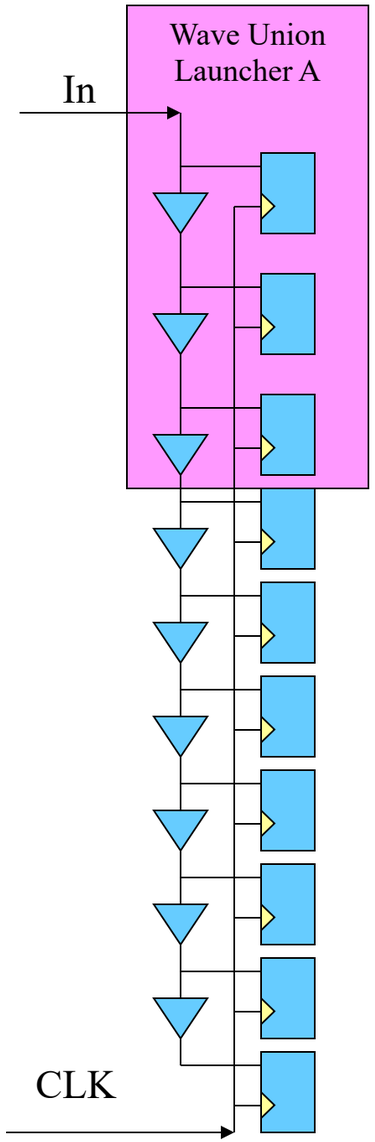
$$\sigma_{eq}^{\ 2} = \sum_i \left(\frac{w_i^{\ 2}}{12} \right) \left(\frac{w_i}{W} \right)$$

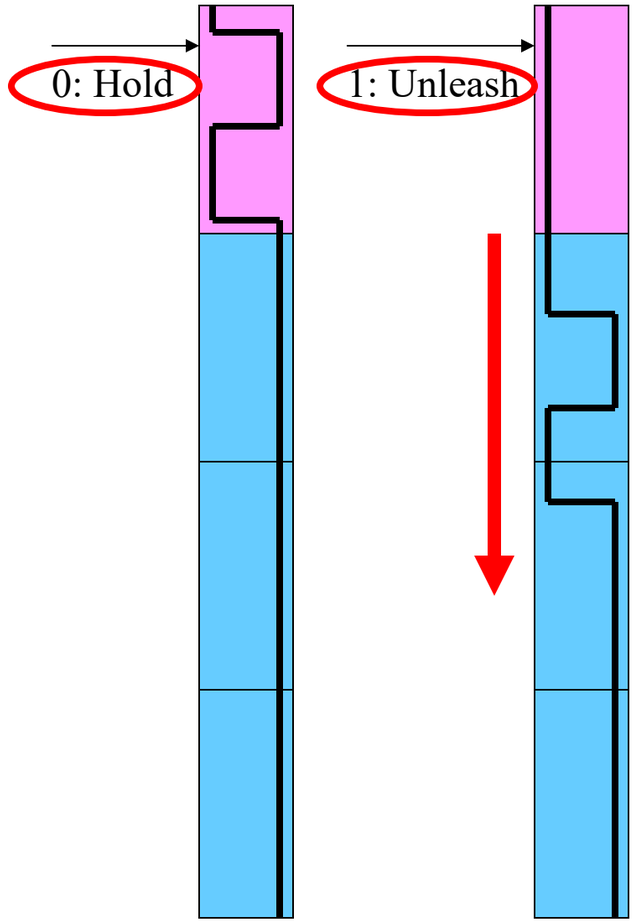


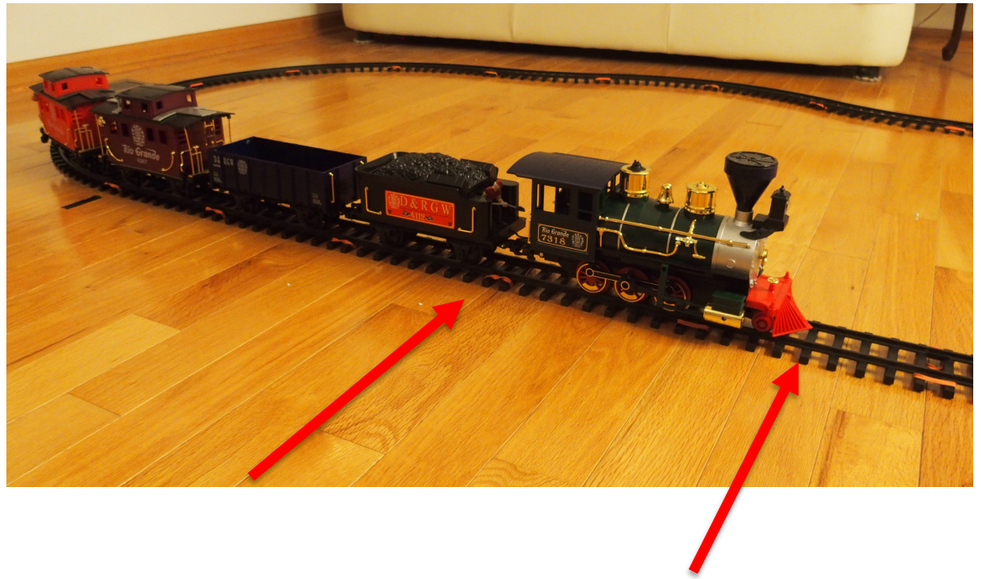


- The ultra-wide bins contribute to measurement errors significantly.
- Using multiple measurements may cut ultra-wide bins and improves measurement performance.

Wave Union TDC (A Type)

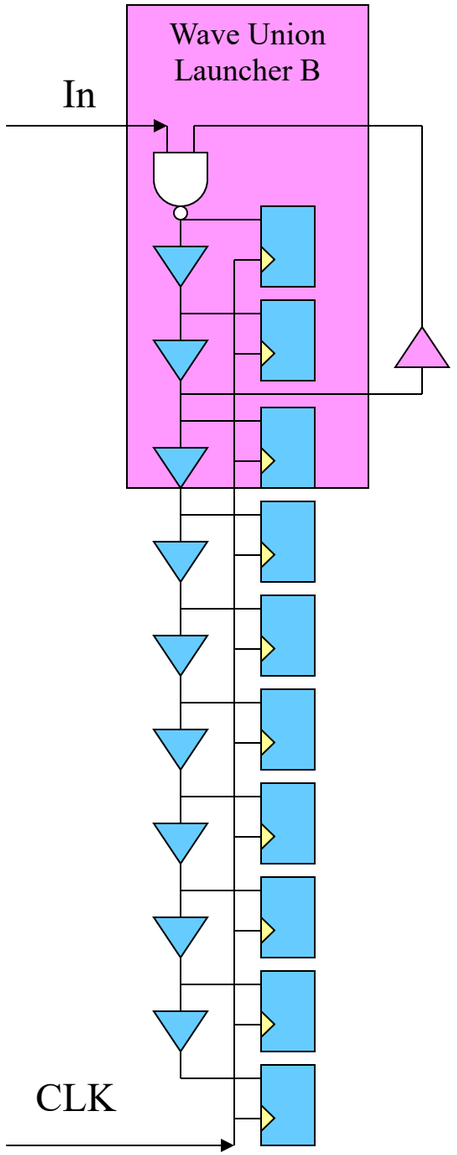


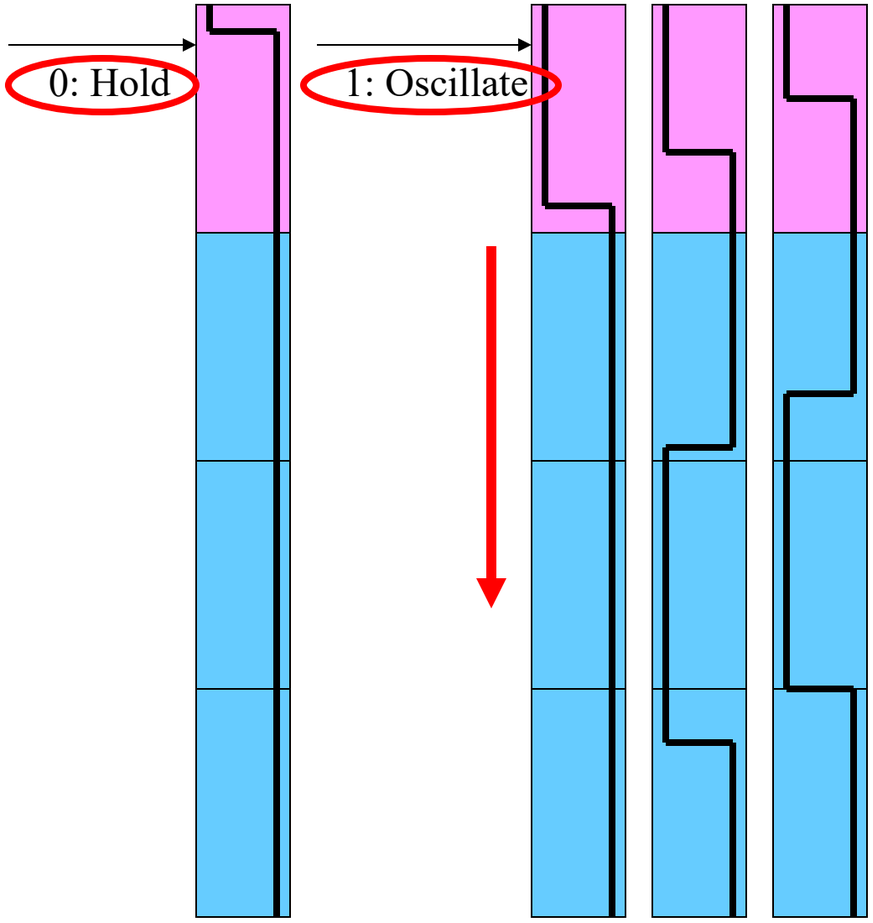




- Wave Union TDC is a method of make multiple measurements.
- In Wave Union TDC, a Wave Union, rather than just a single logic transition in regular TDC is launched in the delay line.

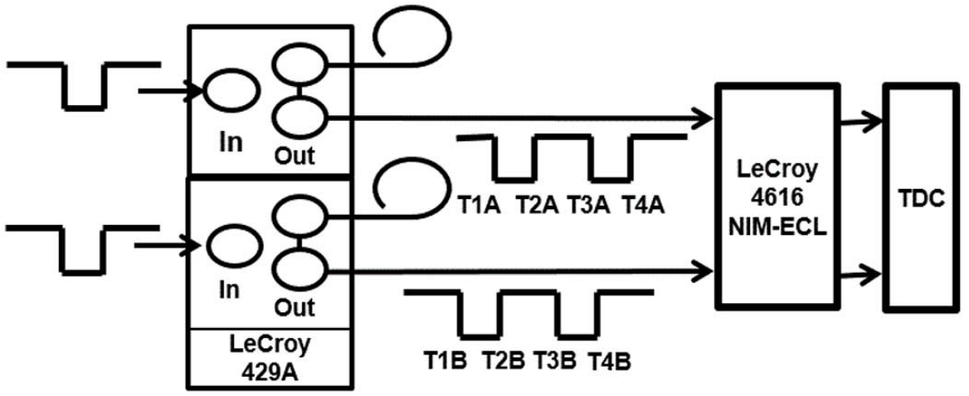
Wave Union TDC (B Type)



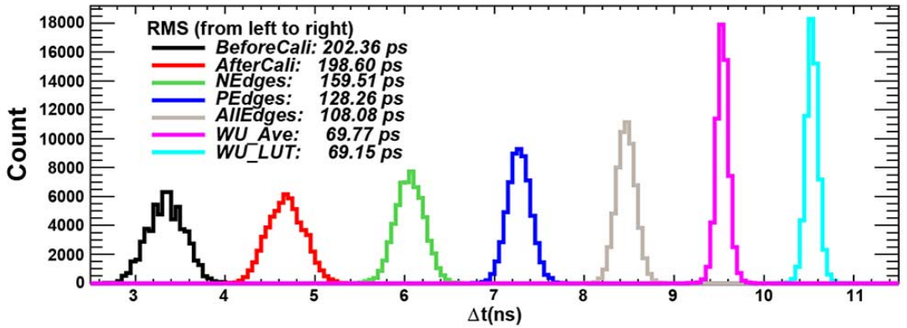


- The Wave Union Launcher can also be a gated ring oscillator.
- Many “snapshots” can be taken.

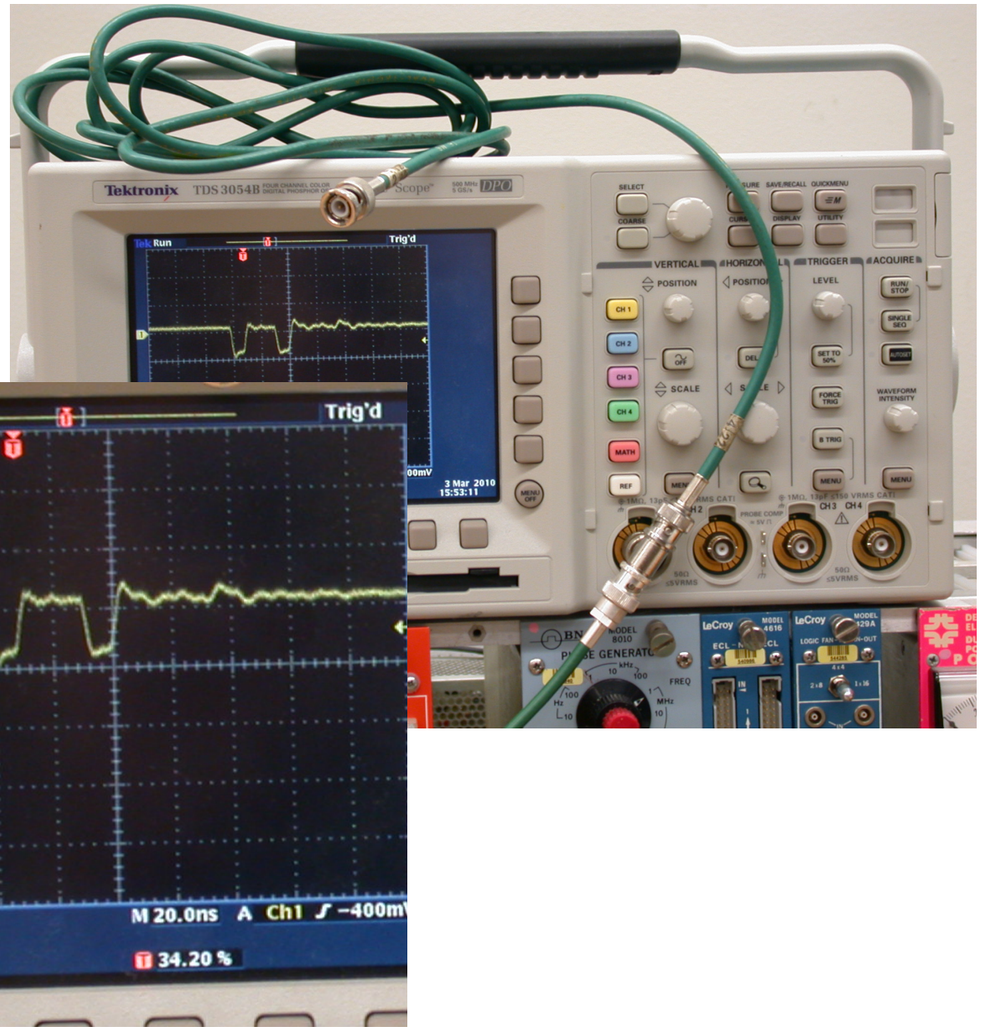
External Wave Union Launcher



(a)



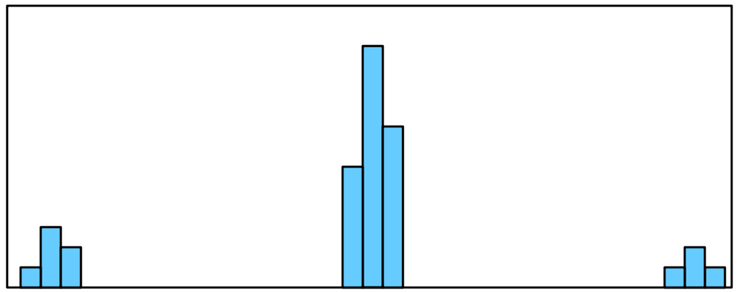
(b)

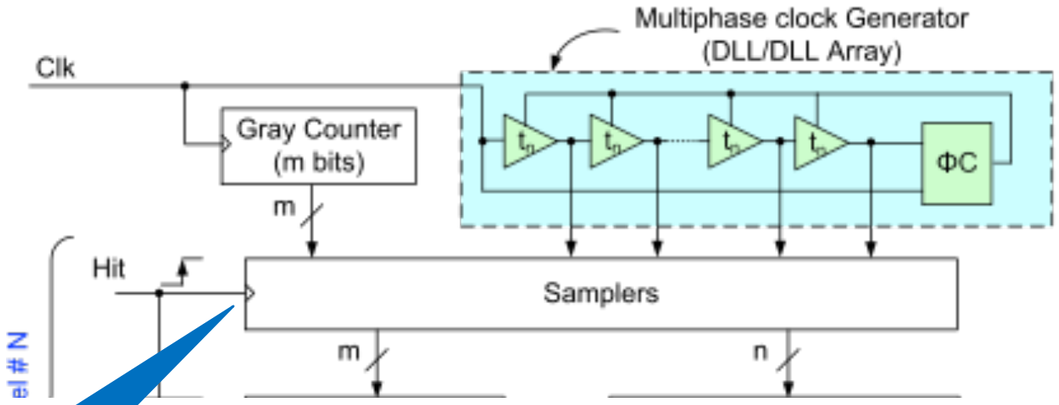


- An external Wave Union Launcher can be added to any multi-hit TDC.
- Measurement precisions can be improved with minimum efforts.

Timing Uncertainty Confinement

“Outlier”





Hit to the CK port
Timing uncertainty
while capturing
coarse time

- New TDC designers may see “outlier” issue, i.e., ghost peaks +- 32 (64) bins.
- The “outlier” issue is due to timing uncertainty while capturing coarse time when “Hit” is fed to the CK port of the D-FF.
- There are solutions, but the whole issue can be avoided from beginning.

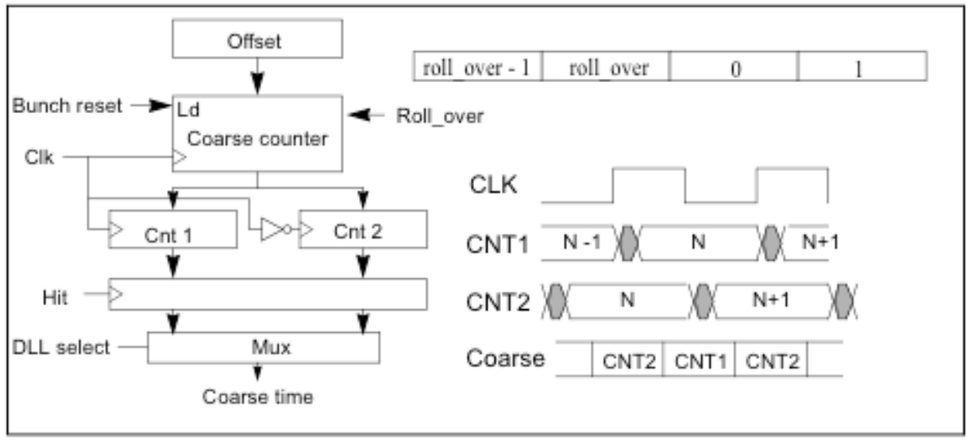
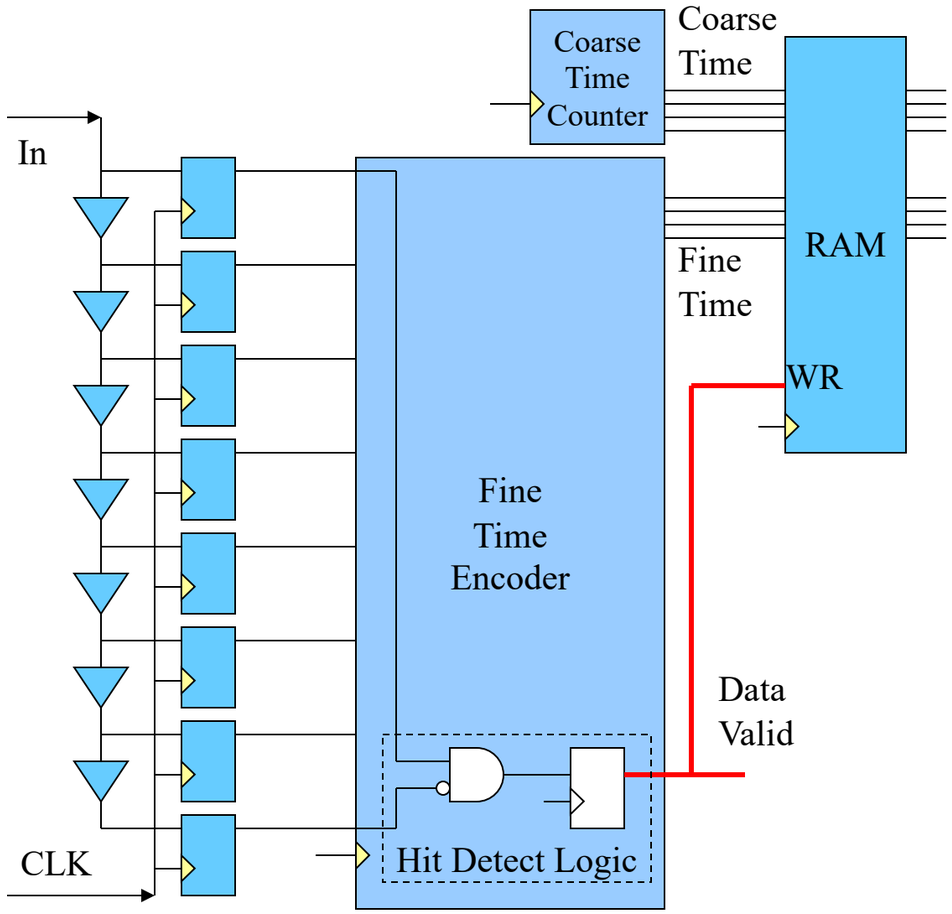


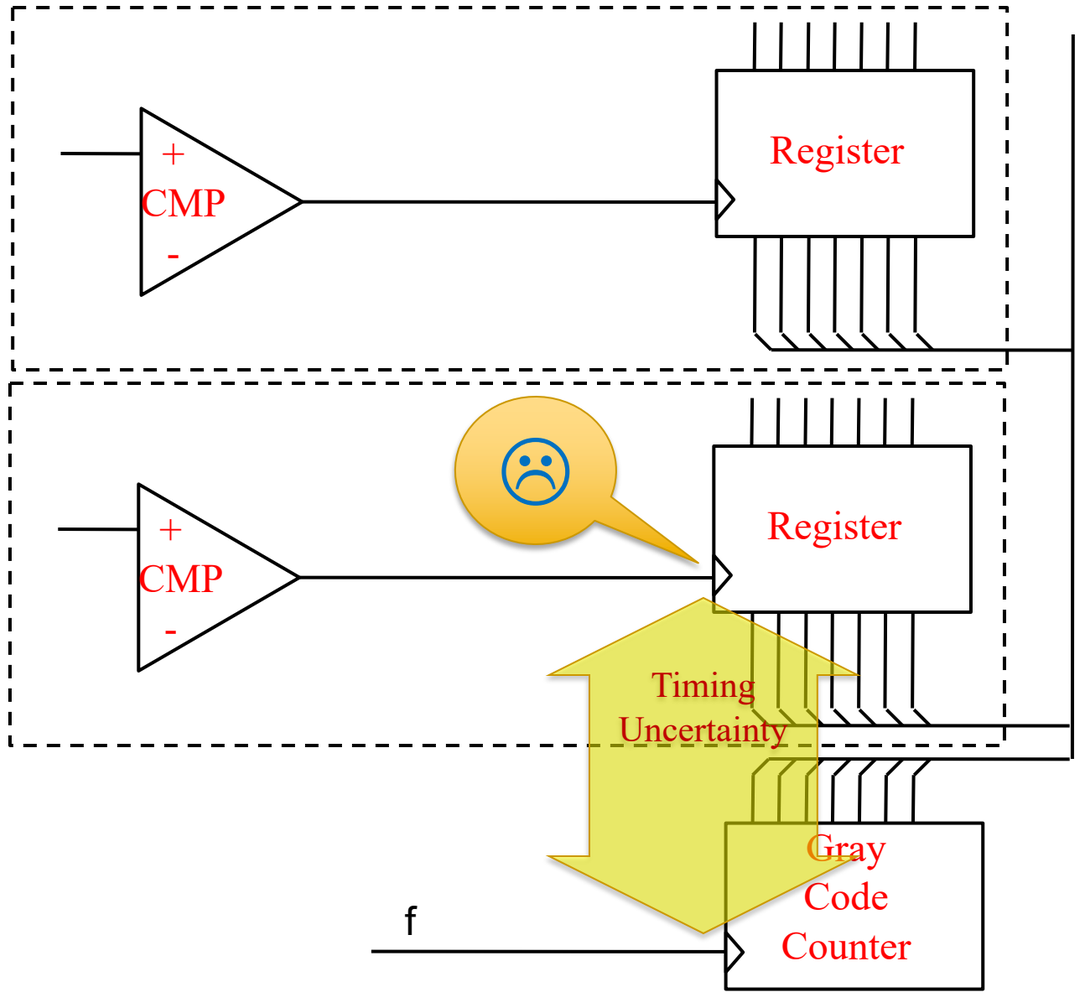
Fig. 6 Phase shifted coarse time counters loaded at hit.

Coarse Time Counter in FPGA TDC

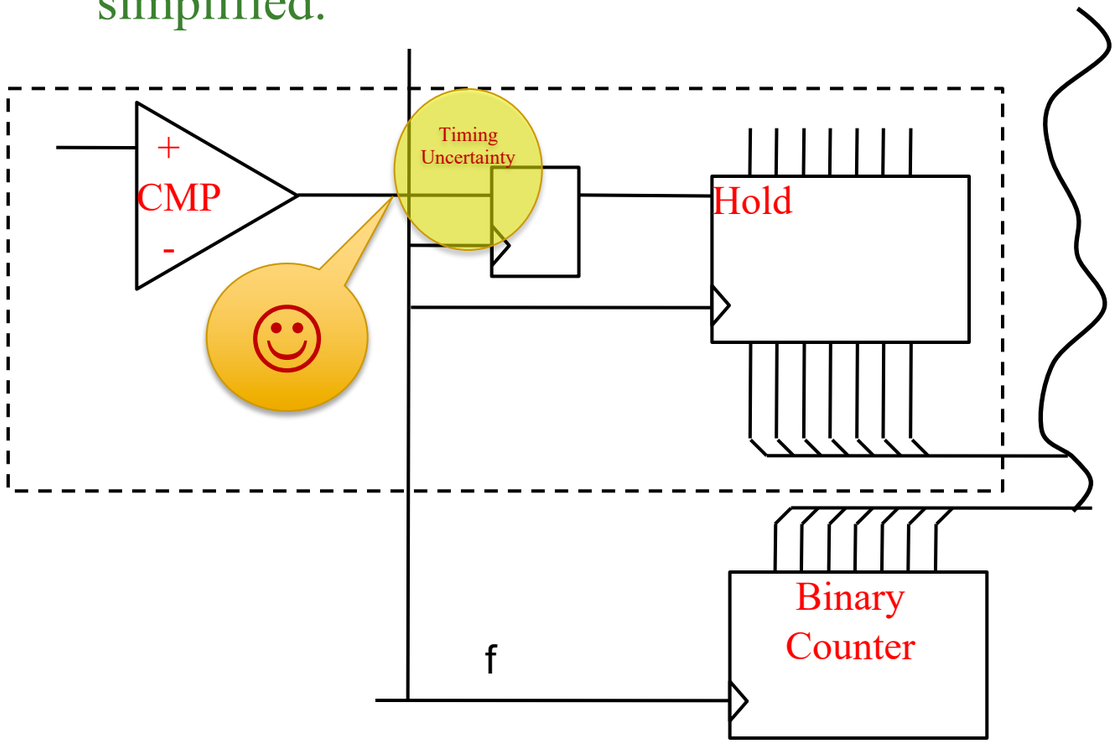


- The FPGA structure forced designers to feed CLK into the CK port of the D-FF.
- Timing uncertainty is confined at the input of the register array.
- The coarse time counter timing condition is satisfied automatically.

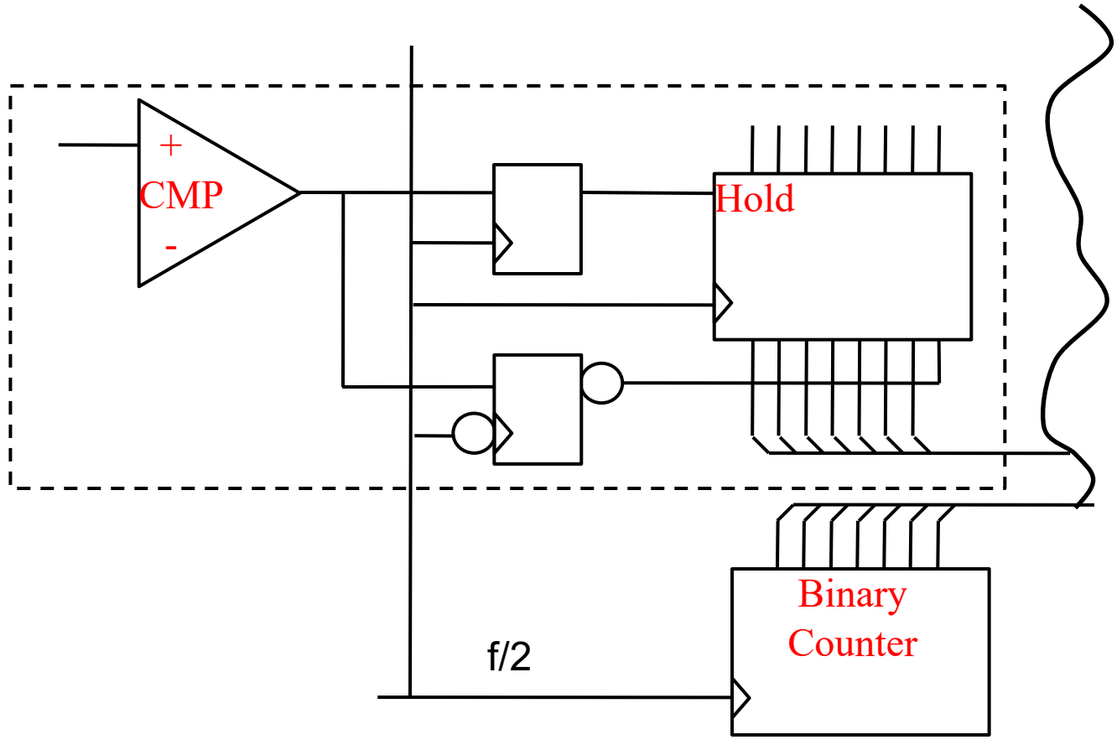
Gray Code Counter?



- Some low resolution TDC may use Gray Code Counter.
- All bits must be distributed with good delay match and the output results must be translated.
- With small circuit adjustment to confine the timing uncertainty, many things are simplified.



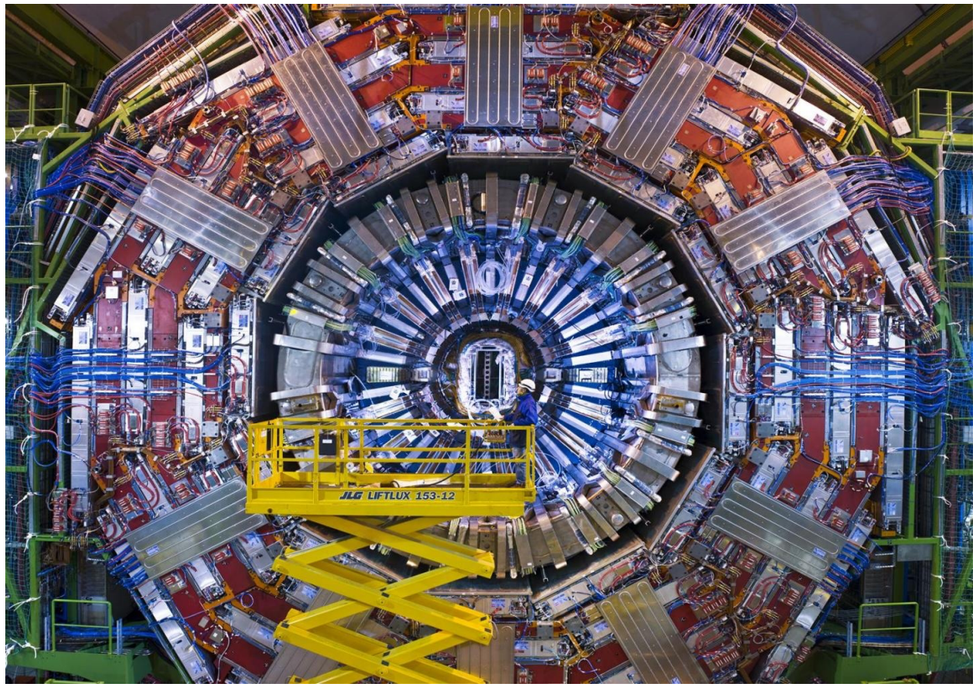
Counter-Based TDC

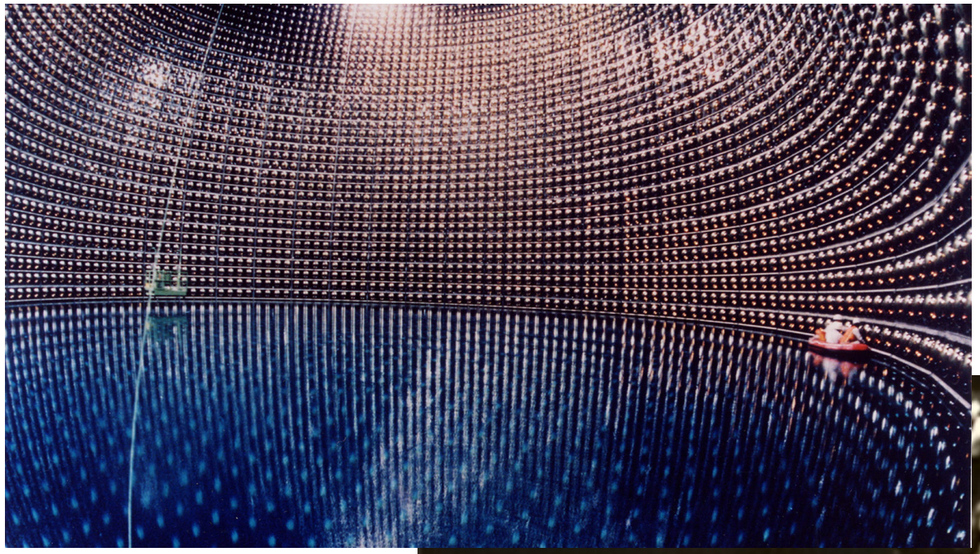


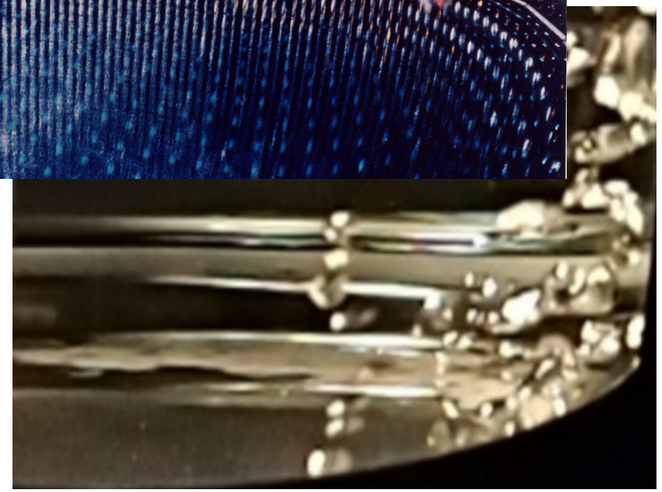
- The counter-based TDC may be further improved by sampling both clock edges.
- With the same resolution, the clock frequency can be reduced by 1/2 which reduces power consumption accordingly.

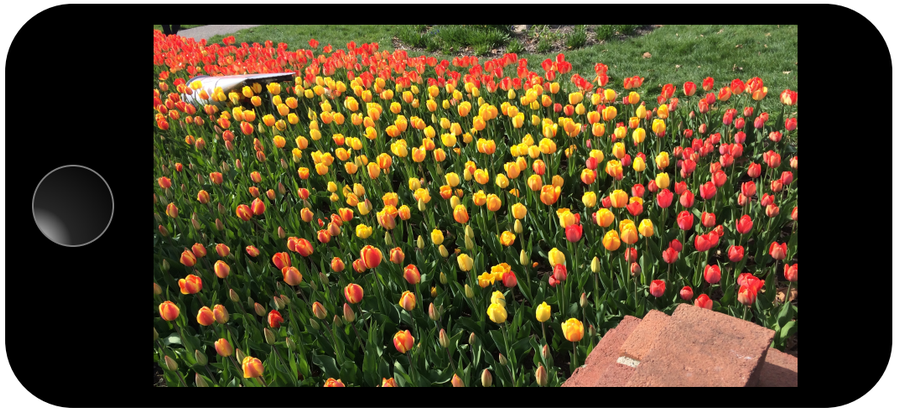
Power Saving Approaches

Power Consumption Matters



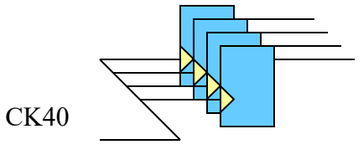






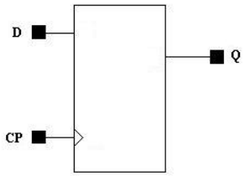
- Inside Highly Packed Detector
- Under Water
- In Liquid Argon
- Portable Devices

Power Consumption at CK Port



Clock: 40 MHz
32 bits
65 nm
P: 14.5 uW

DFQDx
D Flip-Flop, Single Output



Power Consumption (Input Pin Power)(unit:pj)

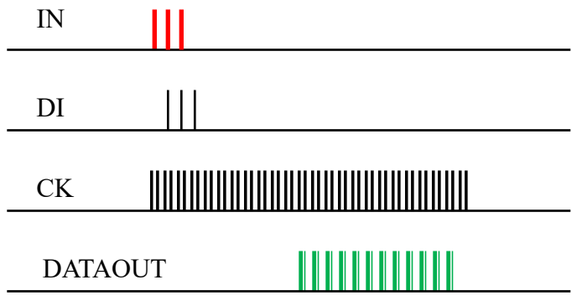
PG Pin=VDD

Cell Name	Parameter	Input Pin	
		CP	D
DFQD1	Rise	0.0036	0.0008
	Fall	0.0077	0.0023

Cell Name	Parameter	Input Pin	
		CP	D
DFQD2	Rise	0.0036	0.0008
	Fall	0.0077	0.0023

Cell Name	Parameter	Input Pin	
		CP	D
DFQD4	Rise	0.0041	0.0008
	Fall	0.0088	0.0023

P = 32*(40 MHz)*(0.0036 pJ + 0.0077 pJ)
= 14.5 uW

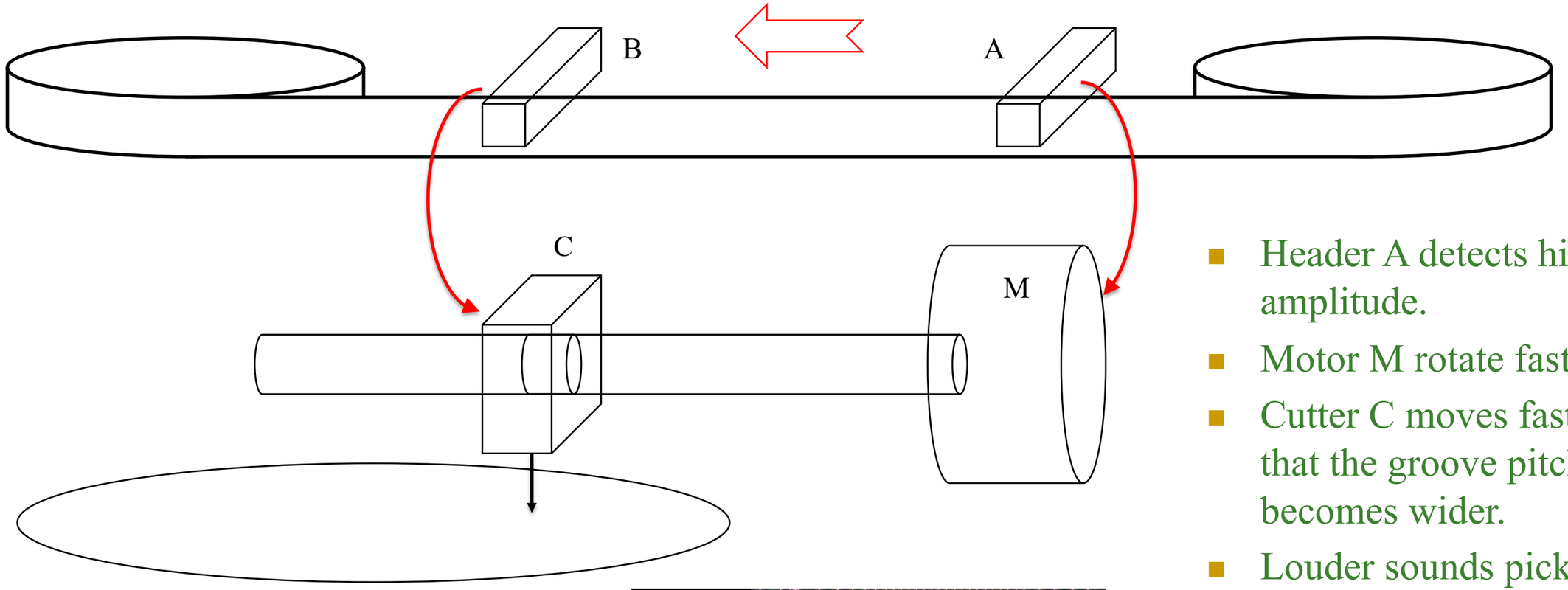


- The CK ports of D-flip-flop consume large amount of power if they are connected to a continuous clock signal.
- It would be ideal if the CK ports of DFF for wide data bus can be kept static most of time while only be driven when the data are to be processed.

Inspiration From Old Tech

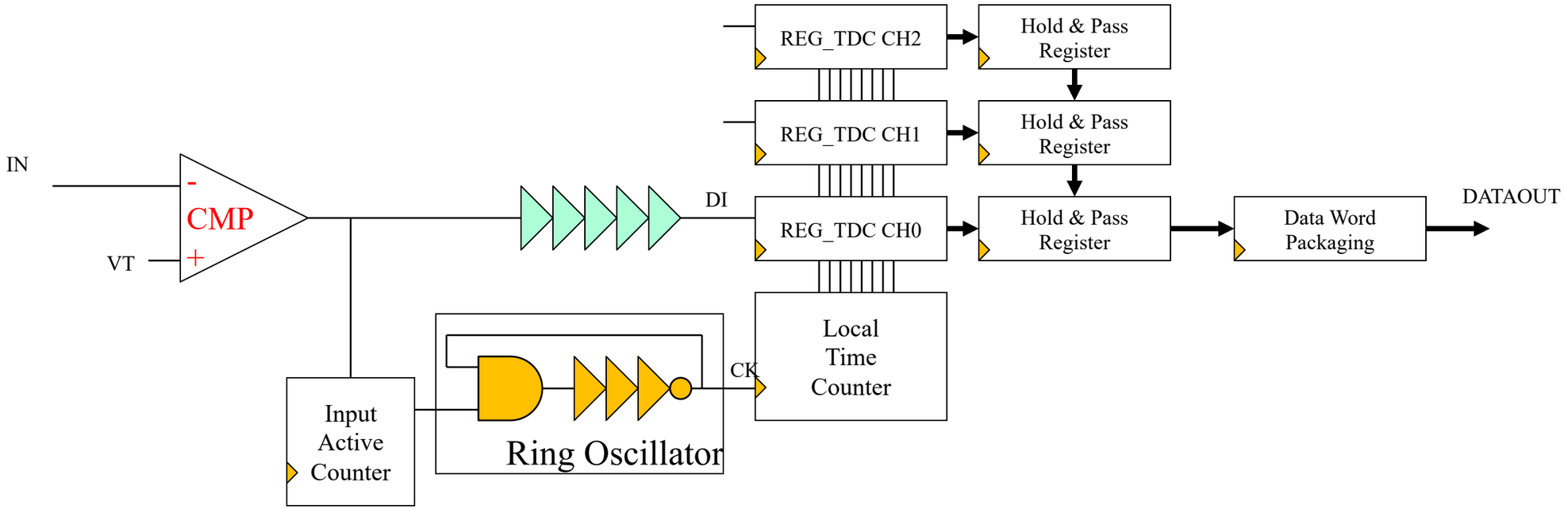
- In LP records, sounds are stored in grooves with different pitches.
- Quieter sounds use finer pitch to save disc area resource.

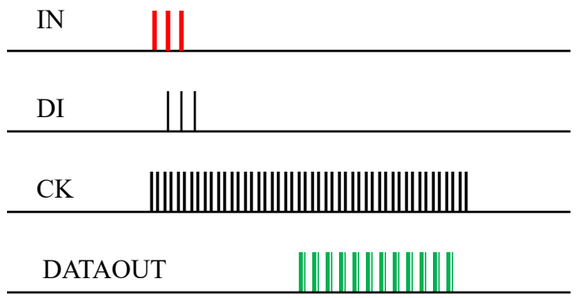
The LP Record Cutting Scheme



- Header A detects higher amplitude.
- Motor M rotate faster.
- Cutter C moves faster so that the groove pitch becomes wider.
- Louder sounds picked up by header B are sent to cutter C to be stored in the wider pitch grooves.

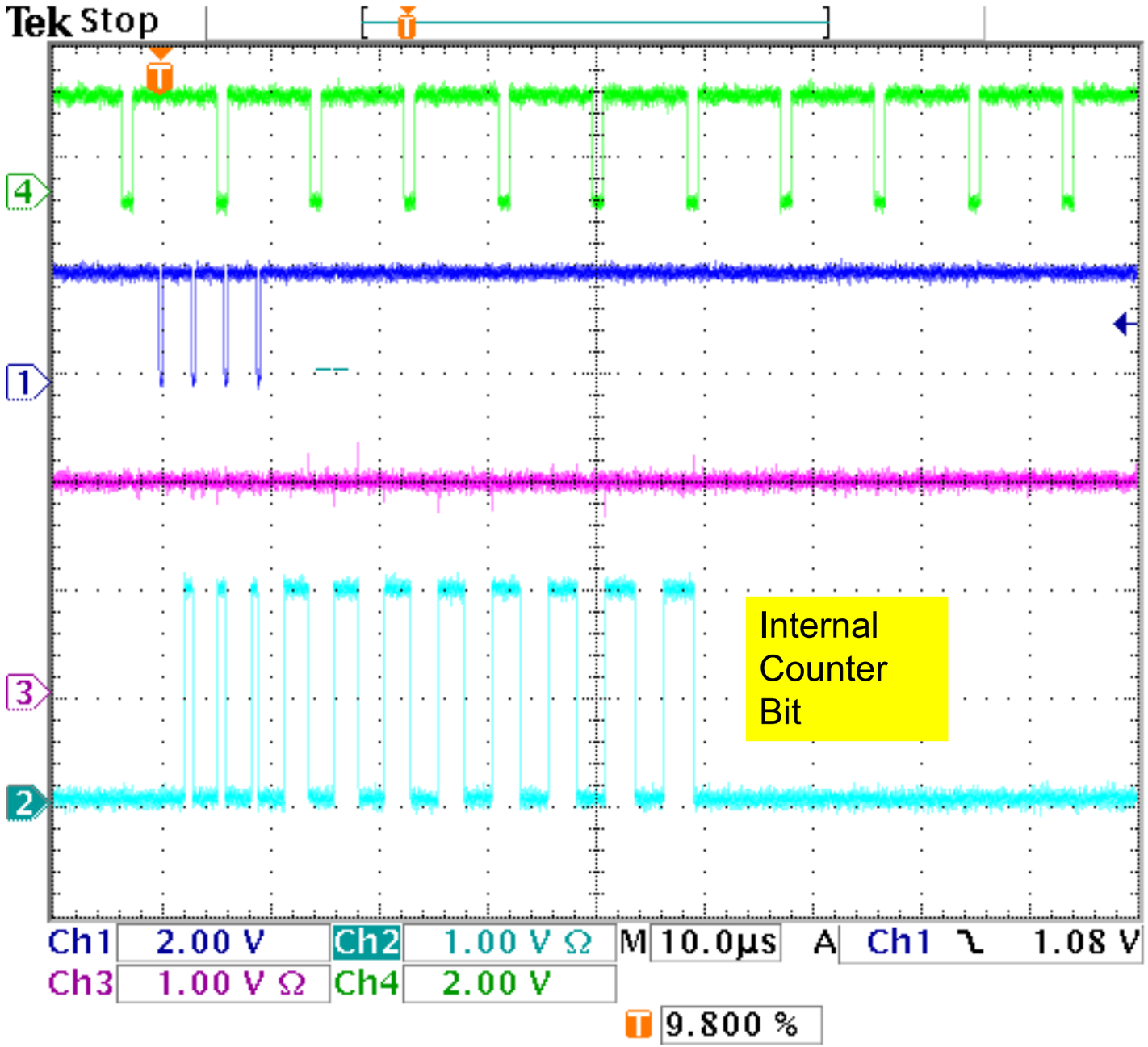
Self Triggering TDC





- When there is no hit, nothing oscillates so that dynamic power consumption is zero.
- When an input burst starts, the gated ring oscillator runs for certain period and the input pulse times are digitized.
- The data are concentrated and output.

Test Waveform



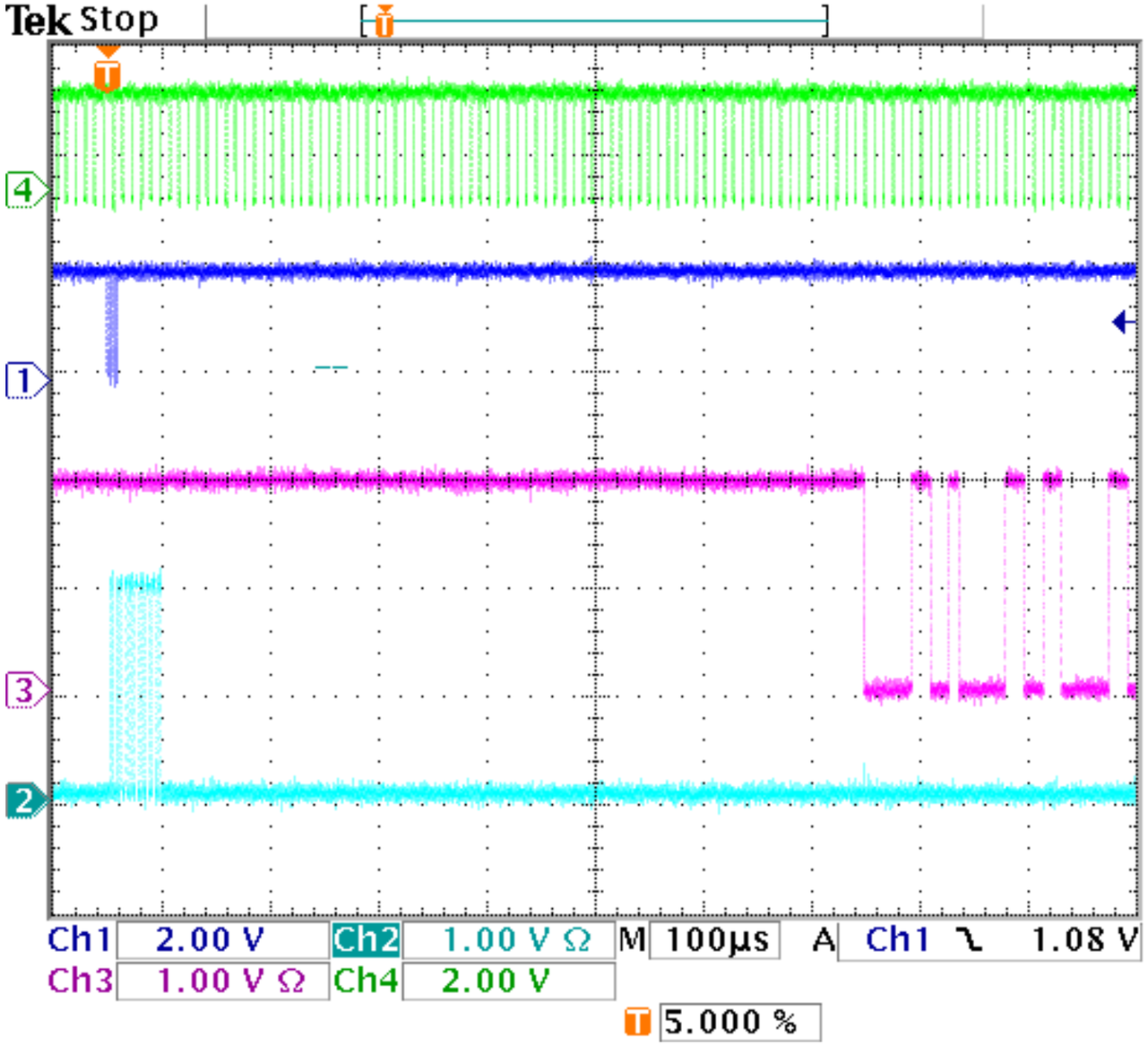
Timing Markers

Input Burst

Internal Counter Bit

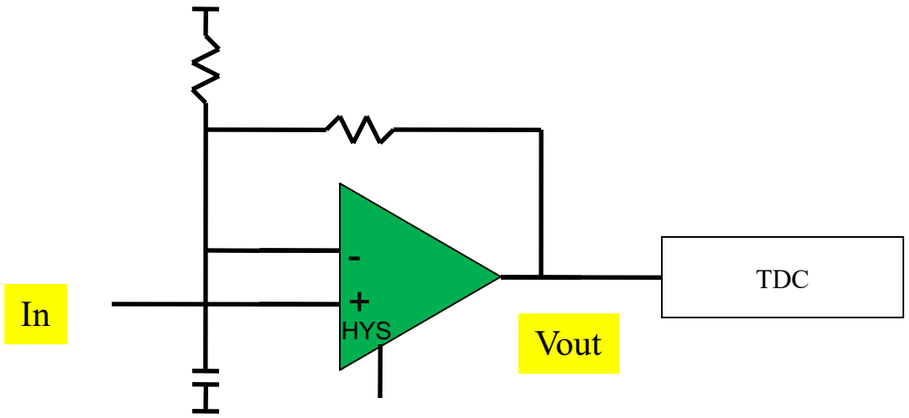
Test Waveform (2)

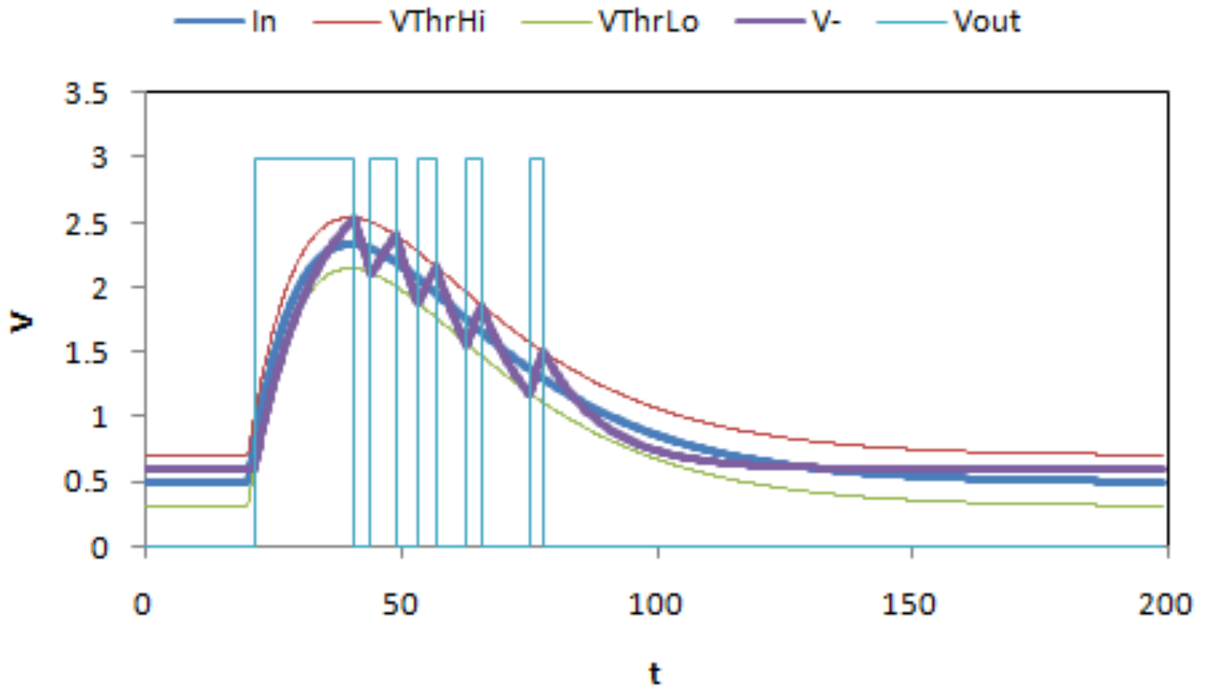
Input Burst

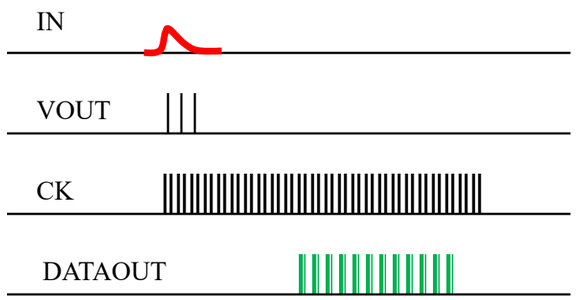


Data output

An Example: The Time Over Oscillating Threshold (TOOT)

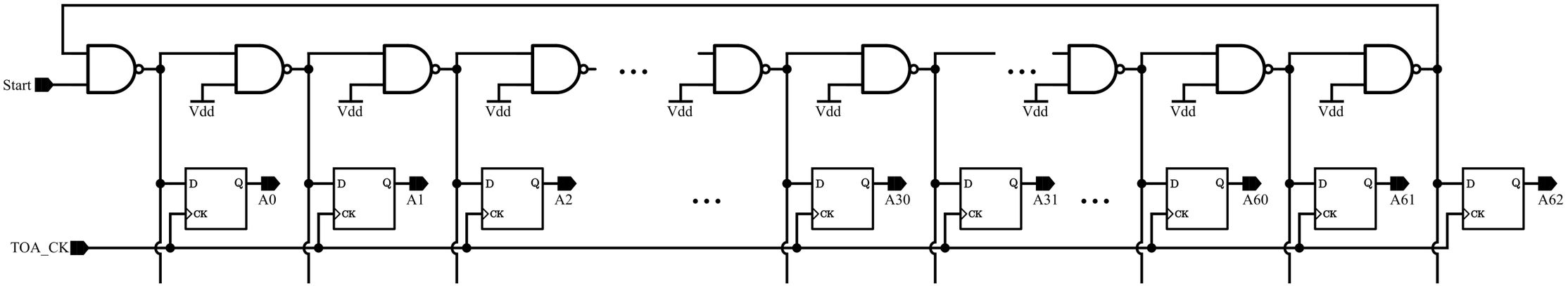






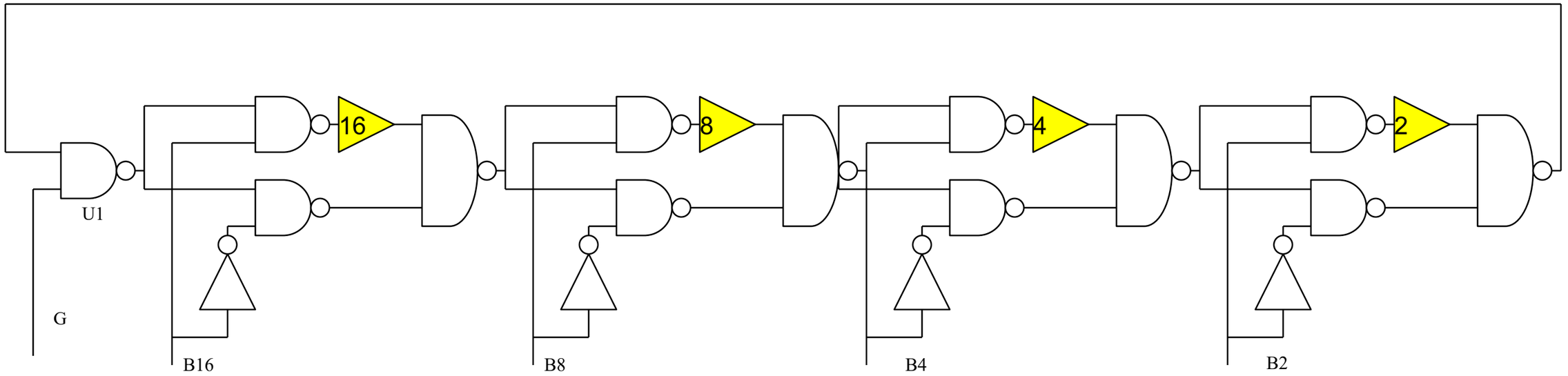
- When the input is below threshold, entire circuit is quiet.
- When an input pulse passes through the threshold, the comparator flips.
- The threshold ramps causing the comparator to oscillate.
- From times of the transitions, the pulse shape can be reconstructed.
- Digitized results includes not only pulse arrival time, but also pulse shape and total charge etc.

Another Example: TDC with Built-In GRO



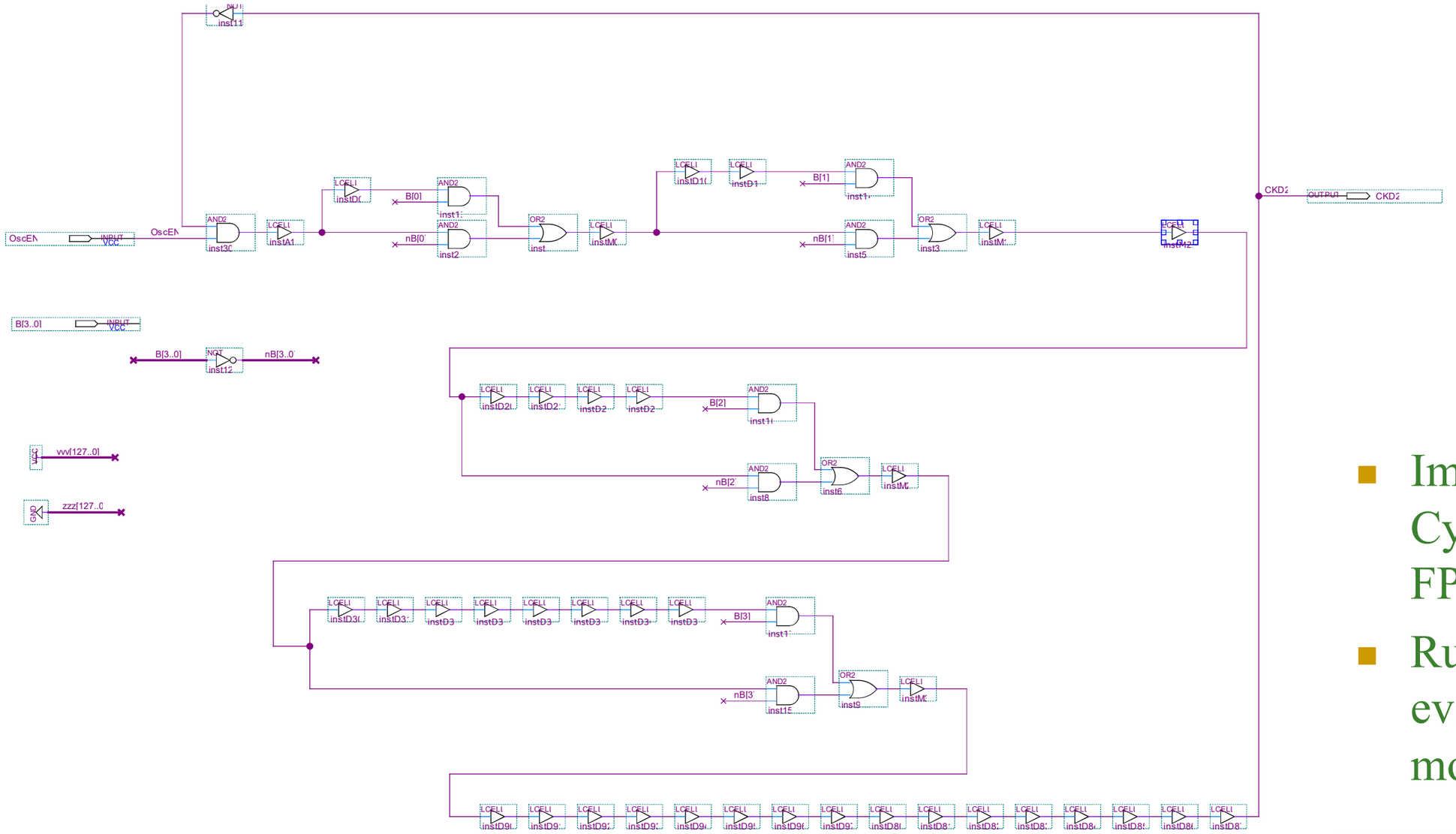
- The delay chain is looped back to create a gated ring oscillator.
- This scheme shortens the delay chain/register array structure.
- No-hit-no-flipping, the power consumption is small when the hit rate is low.

Building Element: Purely Digital Adjustable Ring Oscillator



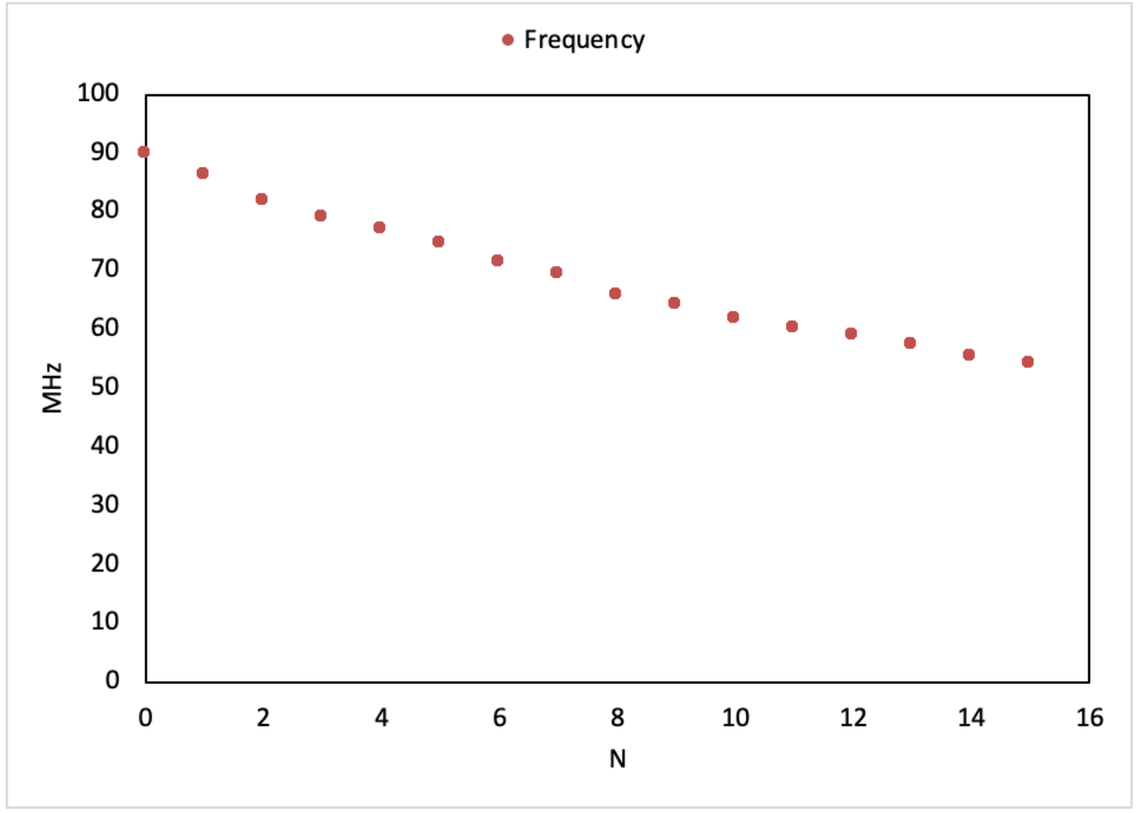
- Ring oscillator frequencies may need to be adjusted.
- There are analog adjustment approaches but there are also purely digital ones.
- Oscillation frequency can be adjusted with different setting of B2 to B16 so that the delay elements can be included or excluded.

Test in FPGA



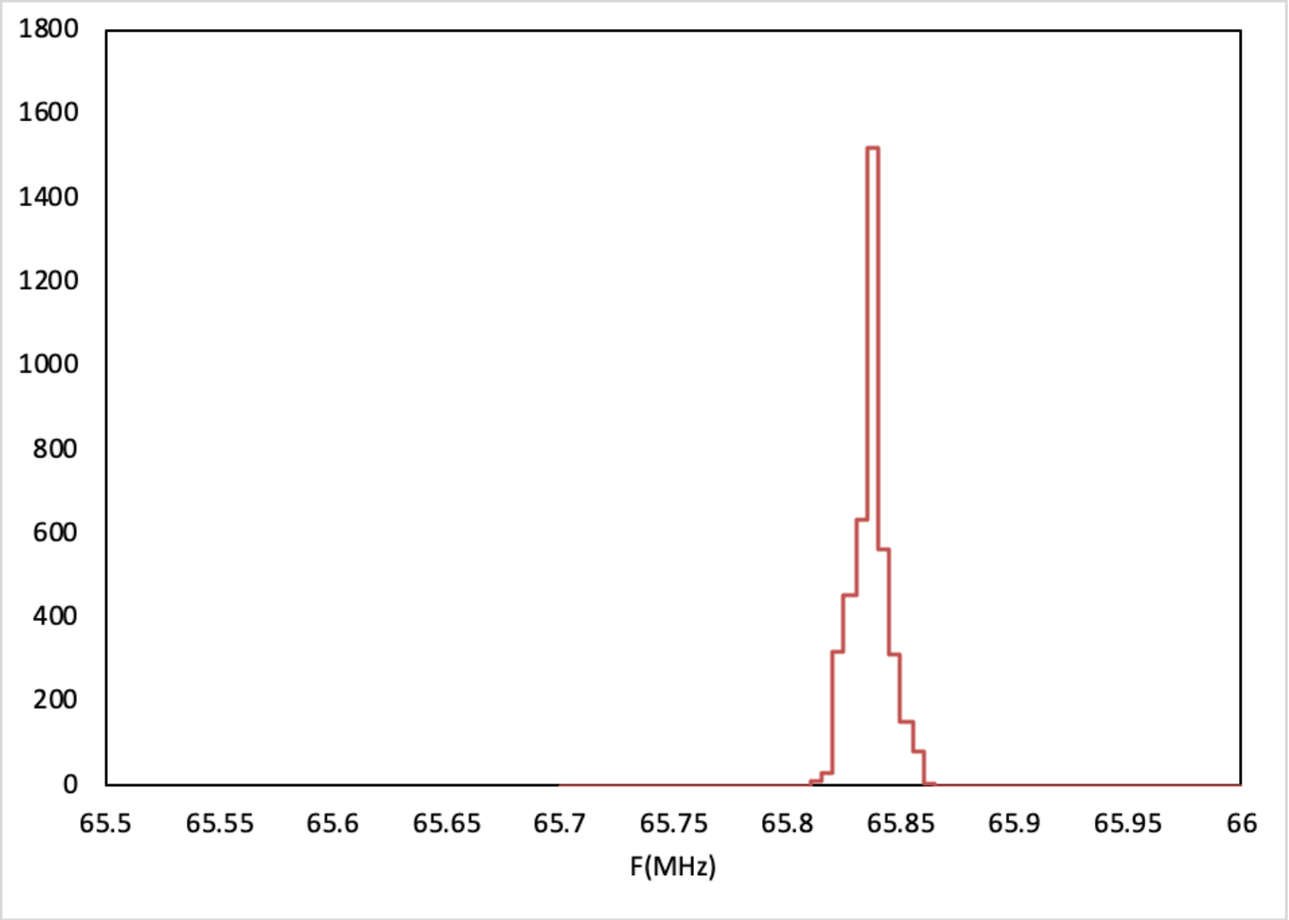
- Implemented in Cyclone V FPGA.
- Run with an evaluation module.

Test Results



- Frequency can be adjusted with delay line length setting bits.

Frequency Stability



- Frequencies are measured over > 300 sec.
- The stability is better than 124 ppm (sigma) even in noisy FPGA environment.

Summary

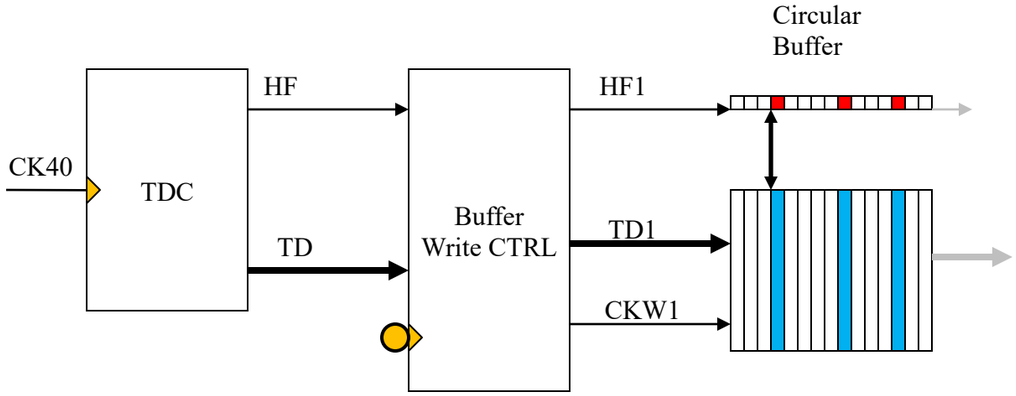
- It is very common that a device with top technology is not available or even the top tech device will not satisfactory for our demand.
- In our daily design jobs, there are rooms for the designers to put in various good thinking to eliminate unnecessary elements or operations to save silicon resource.
- Saving silicon resource will dramatically reduce chip size, power consumption and costs for given requirements.
- Such improvements may enable functionalities that otherwise would be impossible.

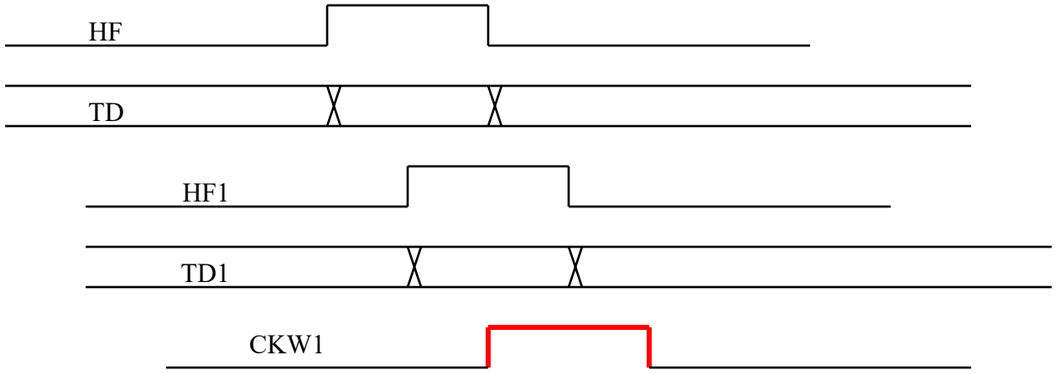
HAAI: (Human Assisted Artificial Intelligence)

- A: Algorithm
  - Parameters/inputs sensitivity pre-selection
  - FFT-like computation strategies (Inner-products of large matrices)
  - Multiplier-less approaches (Least-squares track fitting can use ML method.)
  - Shift-add for small contribution terms
- B: Building Blocks
  - Stop/start processing engines (Gated ring oscillator driven processing units)
  - Data accessing with less read/write operations (Register-like block RAM)


The End
Thanks

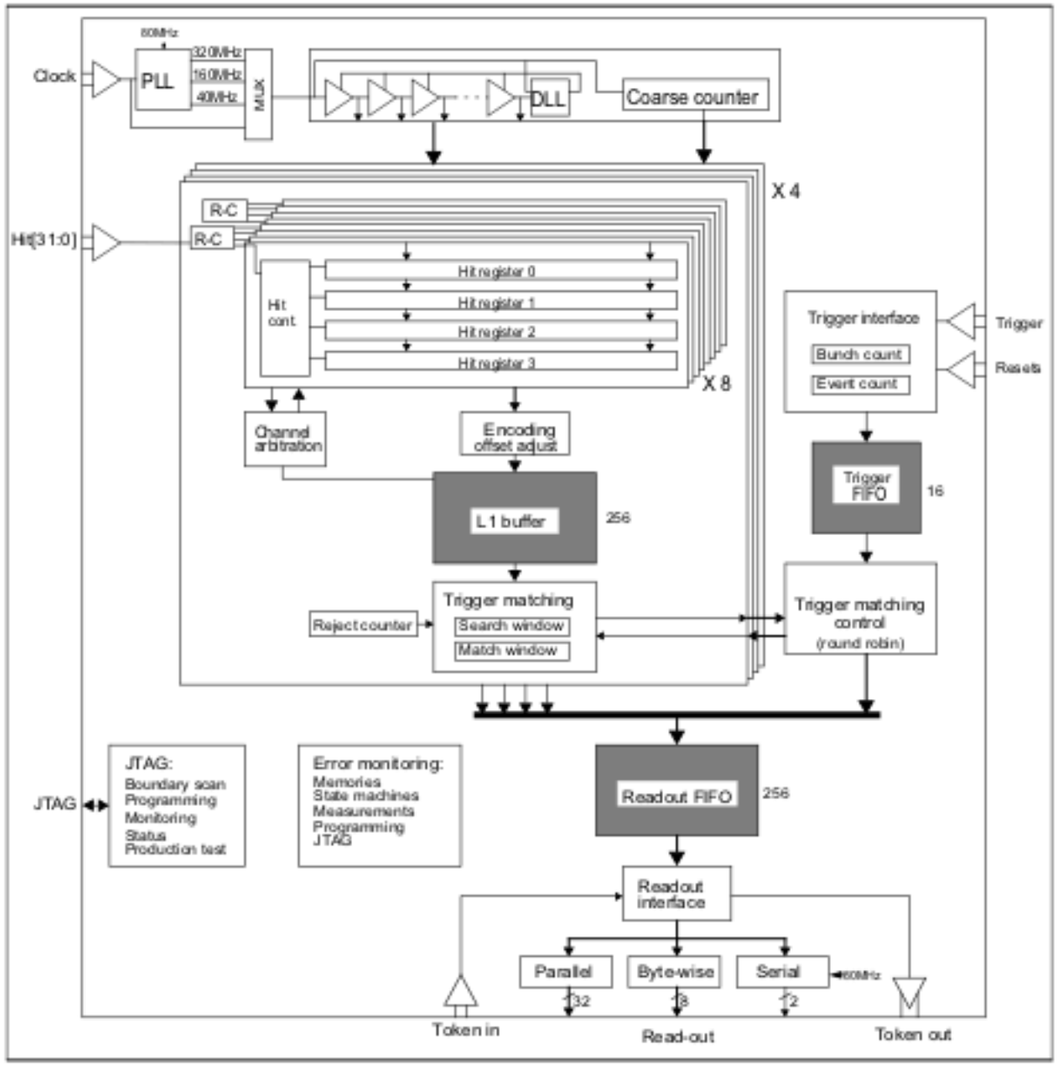
Circular Buffer Writing Operation

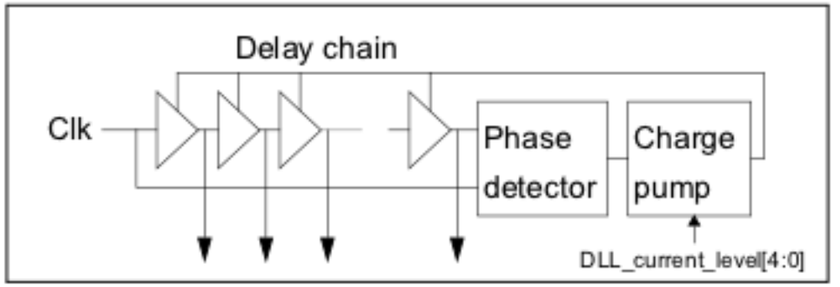


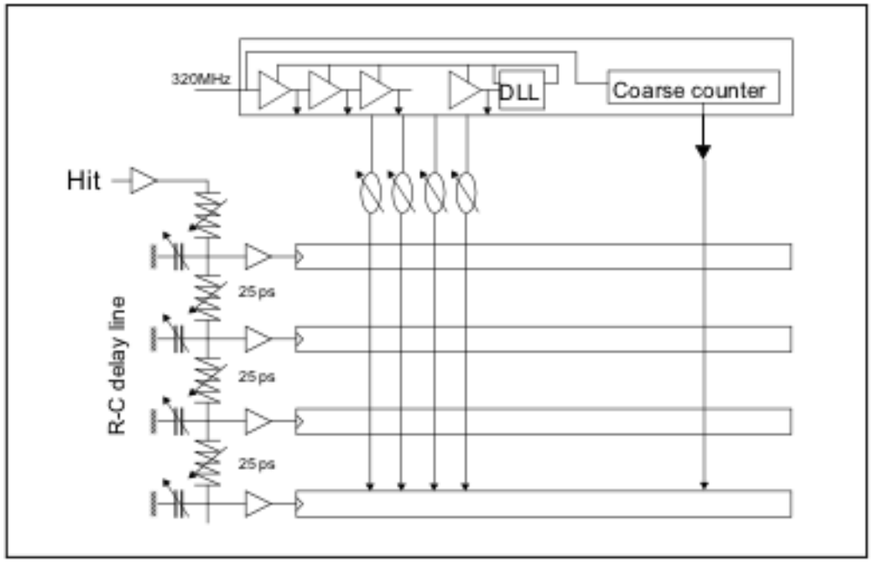


- The TDC generates 1-bit hit flag and about 30 bits of time data every clock cycle.
- A circular buffer is used to store hit flag and data to wait for level 1 trigger.
- The HF1 must be written into circular buffer every clock cycle, but the 30-bit data is only valid when HF1 = 1, which happens <1% of the cycles.
- Toggling the CK port of the RAM at 40 MHz consumes large amount of power.
- It is ideal to generate the data writing clock CKW1 after seeing HF = 1.

Mainstream ASIC TDC in History (& Today)







- A fast clock (e.g. 320 MHz) is sent to the delay chain.
- The delay cells are slowed down (to about 100 ps) so that the delay time of 32 cells matches a clock period.
- The outputs of the delay taps are routed to a set of FF registers.
- The leading edge of the HIT signal captures the delay pattern.