An improved technology for eliminating nondeterministic latency in the L1 trigger system (postprint)

Authors: DU Zhongwei, SU Hong, KONG Jie, ZHAO Xingwen, QIAN Yi

Date: 2023-06-18T00:00:00+00:00

Abstract

Gamma Ray Array Detector (GRAD) is one of external target facility subsystems in the Cooling Storage Ring on the Heavy Ion Research Facility at Lanzhou (HIRFL-CSR). The trigger subsystem of GRAD is required to make a fast L1 trigger decision with a fixed latency for the data acquisition. Because the hit signals from the detector are asynchronous with the local clock of the trigger system, a nondeterministic latency (the value changes between zero and one clock period) is generated when the synchronous receivers of the conventional trigger system process the hit signals. In this paper, an improved trigger logic based on a field-programmable gate array is developed, and comprised of zero-delay broadening circuits as receivers and an improved adding circuit designed for the new receivers. Software simulation and experimental measurement have been conducted. Comparison with the conventional trigger logic, the improved trigger logic has the advantage of eliminating the nondeterministic latency and reducing the total processing latency.

Full Text

Preamble

An Improved Technology for Eliminating Nondeterministic Latency in the L1 Trigger System

Zhongwei Du[12], Hong Su[2],*, Jie Kong[2], Xingwen Zhao[2], Yi Qian[2]
[1]State Key Laboratory of Particle Detection & Electronics, University of Science and Technology of China, Hefei 230026, China
[2]Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China

Abstract
The Gamma Ray Array Detector (GRAD) is one of the external target facility subsystems in the Cooling Storage Ring on the Heavy Ion Research Facility at Lanzhou (HIRFL-CSR). The trigger subsystem of GRAD is required to make a fast L1 trigger decision with fixed latency for data acquisition. Because hit signals from the detector are asynchronous with the local clock of the trigger system, conventional trigger systems generate nondeterministic latency (varying between zero and one clock period) when synchronous receivers process these signals. This paper presents an improved trigger logic based on a field-programmable gate array, comprising zero-delay broadening circuits as receivers and an improved adding circuit designed for the new receivers. Software simulation and experimental measurements have been conducted. Compared with conventional trigger logic, the improved design eliminates nondeterministic latency and reduces total processing latency.

Key words
Nondeterministic latency, Field-programmable gate array, L1 trigger, Zero-delay broadening

Supported by the Important Direction Project of the CAS Knowledge Innovation Program (No. KJCX2-YW-N27) and the National Natural Science Foundation of China (No. 11079045)
Corresponding author. E-mail address: suhong@impcas.ac.cn
Received date: 2012-08-10

Introduction

A reliable trigger system is critically important for large detector electronics in particle physics. A universal architecture for L1 trigger systems can be simplified from several particle experiment electronics designs [Figure 1: see original paper], such as BESIII[1], ATLAS[2], and ALICE[3]. The basic logic for making an L1 trigger decision in a field-programmable gate array (FPGA) consists of three components: a receiving module, an adding module, and a comparator that compares the hit-number with the trigger condition. When a collision event occurs, multichannel analog signals from the detector are converted to digital hit signals by threshold discriminators in the front-end electronics (FEE) and delivered to the L1 trigger system. The receiving module synchronizes and aligns these hit signals, the adding module counts the total number of responding channels (hit-number) in the event, and the comparator evaluates whether the acquired hit-number meets the trigger condition to generate an L1 trigger decision. Simultaneously, position information and other useful data are delivered to the L2 trigger system for further processing. These multilevel trigger decisions collectively select valid events for the data acquisition (DAQ) system.

The L1 trigger latency consists primarily of two components: receiving latency and adding latency. Due to the nondeterministic timing of random collision events, the arrival time of each hit signal is asynchronous with the local clock of

the L1 trigger system. Conventional trigger receiving modules employ registers to synchronize hit signals by latching them on clock edges (e.g., a FIFO with a 41.67-MHz local clock is used as the receiving module in the BESIII TOF trigger subsystem[4]). The hit signals are latched on the clock edge, generating receiving latency equal to the time difference between signal arrival and the clock edge. This receiving latency is nondeterministic because its value varies randomly between zero and one local clock period across different events. While the average nondeterministic latency (half a clock period) can be reduced slightly by increasing clock frequency, higher frequencies increase logic design difficulty in meeting timing slack, and the maximum clock frequency is limited by FPGA hardware constraints.

The GRAD electronics in HIRFL-CSR[5] comprises three components: a FEE based on ASIC chips named MATE[6], a DAQ subsystem, and a 64-channel L1 trigger subsystem. Due to the track-hold processing mechanism in the MATEs (which requires sampling to start at a specific time for accurate measurement), the DAQ subsystem demands a fast L1 trigger decision with fixed latency for accurate offline calibration. Conventional synchronous receivers cannot satisfy this requirement due to their nondeterministic receiving latency. To eliminate this nondeterministic latency, zero-delay broadening circuits have been developed as receivers, and an improved adding module has been designed to match these new receivers. The design and performance of this improved L1 trigger logic are described below.

Improved L1 Trigger Logic

[Figure 2: see original paper] shows the improved L1 trigger logic implemented in the FPGA. The kernel trigger logic consists of zero-delay broadening circuits as receivers and an improved adding circuit. The new receivers can capture and broaden multichannel hit signals without introducing nondeterministic latency. The improved adding circuit sums all hit signals together to acquire the hit-number (with each channel signal represented as 1-bit data). This hit-number is then compared with the trigger condition to generate a fast L1 trigger decision that initiates DAQ subsystem readout. The L1 trigger decision latency is low and fixed because no synchronization occurs in the receiving and adding processes. Simultaneously, conventional synchronous trigger logic is employed to acquire trigger information for the L2 trigger system, whose slower decision determines whether the event data acquired by the DAQ system should be stored.

2.1 Zero-Delay Broadening Circuits

When charged particles strike the corresponding scintillator for each signal channel, arrival times differ due to hit position variations[4]. Consequently, signals from hit channels arrive at the trigger system at different times. In any given event, a kernel channel determines the trigger latency (e.g., if the trigger threshold for hit-number is 5 and the actual hit-number is 10, the 5th received hit

signal is the kernel; only the latency contributed by receiving this 5th hit signal matters, while latencies of the 4th and remaining signals are irrelevant). Additionally, a valid arrival time window (tmax) represents the maximum time difference across all channel paths (e.g., tmax is 20 ns in GRAD, meaning all hit signals received from the first hit signal arrival until 20 ns later belong to the same event). Due to variations in arrival time and pulse width of hit signals, the receivers' first operation must align all hit signals. Instead of using registers as in conventional synchronous receivers, zero-delay broadening circuits perform this alignment. Each hit channel signal is broadened to a width exceeding tmax, creating a period during which all hit signals presented to the adding circuit maintain a high voltage level. The combined adding circuit can then count the hit-number during this period—this constitutes asynchronous alignment.

[Figure 3: see original paper] shows a zero-delay broadening circuit constructed from D-type flip-flops (DFF). To meet the requirement that broadening width exceed tmax, the minimum number of DFFs used per channel is:

$$n > \frac{t_{max}}{t_{clk}}$$

where tclk is one clock period. When a hit signal pulse is delivered to the clock pin of the first DFF, the Q pin output immediately transitions to high level due to the high-level connection at the D input pin. Simultaneously, subsequent DFFs receive and transmit the high-level signal from the first DFF in sequence. Synchronized with the local clock, each following DFF generates a latency of one clock cycle (the second DFF's latency is less than one clock period because a synchronizing operation must be conducted). After a total latency exceeding tmax, the clear pin of the first DFF receives a high-level feedback signal from the Q pin of the last DFF. The first D flip-flop resets, and its output signal returns to low level until the next hit signal arrives. Consequently, the hit signal is broadened to a width exceeding tmax with zero-delay, regardless of input signal changes during this process.

[Figure 4: see original paper] shows timing simulations comparing synchronous and zero-delay receivers. A synchronous receiver with several registers can synchronize and store hit signals for a few clock periods, but two issues arise: first, the latency for channel 1 does not equal that of channel 3; second, the signal from channel 2 (with pulse width less than one clock cycle) is missed. In contrast, the zero-delay receiver exhibits fixed and lower receiving latency compared to synchronous receivers. This latency, composed only of pin-to-pin delay and gate delay, depends on the FPGA's speed grade and contains no nondeterministic latency in the receiving process. Furthermore, the zero-delay receiver can capture short-pulse-width signals, though false triggering may occur if such short pulses are noise. In practice, this does not happen due to reliable FEE design. Additionally, the broadening width time represents a dead time for the improved trigger logic (any new hit signal is ignored during this period). Therefore, the zero-delay receivers' broadening width should just satisfy

physical requirements—when tmax is 20 ns, as shown in [Figure 4: see original paper].

2.2 Improved Adding Circuit

When the adding circuit receives 64 broadened signals as 64 1-bit data (each 1-bit data becomes "1" when a hit occurs on the corresponding scintillator), it must perform 63 adding operations to calculate the hit-number, requiring 63 adders. The adding circuit is a combinational logic circuit. Since the zero-delay receivers are asynchronous circuits without registers, the adding circuit must also avoid nondeterministic synchronization latency. Therefore, both nondeterministic latency and race hazards must be addressed.

[Figure 5: see original paper] shows two types of adding circuits: serial and parallel adders. The serial configuration is the FPGA default if logic is compiled without special design, while the parallel configuration serves as the improved adding logic and is superior to the serial approach.

First, the serial adding circuit introduces new nondeterministic latency. As shown in [FIGURE:5(a), input channels have different path lengths (the channel 1 signal passes through 63 adders, while channel 64 passes through only 1 adder). Because the kernel channel changes randomly across events, the actual processing delay (which depends only on the kernel channel) varies with nondeterministic latency. In [FIGURE:5(b), the parallel circuit eliminates this problem because every input signal passes through exactly 7 adders.

Figure 6: see original paper shows the latency for different channels (with input signals for channels 1, 12, 24, 36, 48, and 56 from left to right). The serial circuit's latency changes with different input channels, while the parallel circuit's latency remains fixed.

Second, the serial circuit generates more glitches. Due to asymmetric paths across 64 channels, more junctions exist when the serial circuit is compiled into look-up tables (LUTs) in the FPGA. If a signal passing a junction traverses two paths of different lengths and reaches two LUTs or two inputs of one LUT, a race hazard occurs, causing a glitch. Because all channel paths in the parallel circuit are identical, fewer glitches are generated. Figure 6: see original paper shows glitches for both adding circuits in an extreme state. The parallel circuit produces fewer and shorter glitches than the serial circuit. In practice, since the hit-number in one event is smaller than 10 in gamma-ray energy measurement experiments, actual glitches are fewer than in simulation.

A grounded capacitor connected to the trigger decision output pin eliminates glitches. The capacitor charging voltage is:

$$V = E \times (1 - e^{-\frac{t}{RC}})$$

where E is the FPGA pin output voltage, R is its output impedance (approximately 10 Ω), C is capacitance, and t is charging time. If RC > t, the capacitor voltage remains below 0.63×E, eliminating glitches with pulse widths less than t in LVTTL signals (where the high-level threshold exceeds 0.63×E). For the improved trigger logic, a 10 pF grounded capacitor eliminates glitch widths below 1 ns, which is sufficient for the parallel adding circuit. The capacitor also delays the trigger decision output signal by a value equal to RC. Consequently, circuits with more glitches require larger capacitors, generating greater trigger latency. The improved parallel adder circuit thus offers lower and fixed latency compared to the serial adding circuit.

3 Experiment Measurement

To quantify the latency difference between conventional synchronous and improved asynchronous trigger logics, a test system was assembled. A 60Co source and CsI scintillator from GRAD generate hit signals, which FEE modules convert from analog to digital. The 64-channel L1 trigger module receives and processes these signals using test logic implemented in an FPGA. The test logic contains both conventional trigger logic with synchronous receivers and improved trigger logic with zero-delay receivers. Both logics use parallel adders to minimize other differences. The two logics receive hit signals simultaneously and generate L1 trigger decisions, which are delivered to three counters measuring latency difference, trigger frequency, and false triggering rate. A latency timer operating with a 250-MHz clock from a phase-locked loop (PLL) starts counting upon receiving the improved trigger decision and stops upon receiving the conventional trigger decision, thus measuring the latency difference between the two logics. Every 0.25 seconds, data from the three counters are packed and stored in a FIFO, after which the counters reset. When the FIFO approaches full capacity, data are transmitted to the host computer via a peripheral component interconnect (PCI) bus. Software calculates the average latency difference per event, as shown in [Figure 7: see original paper].

shows the latency difference between improved and conventional trigger logics at various clock frequencies. Theoretically, the average latency difference equals half a clock period, but several factors introduce error. First, the timer clock frequency is limited to 250 MHz (constrained by the FPGA), yielding 4 ns measurement precision. Second, because timer and trigger clocks originate from the same PLL and clock source, a timer clock period delay occurs when the timer captures the synchronous trigger decision (e.g., a consistent 4-ns delay exists at 10-MHz trigger clock when the timer stops counting). Despite limited measurement precision, results demonstrate that the improved logic reduces latency by approximately half a clock cycle. The reduced latency matches the expected average nondeterministic value and varies randomly between zero and one clock cycle. Since the test only differs in receiver design, the results confirm that the improved trigger logic eliminates nondeterministic latency.

References

[1] Feng C Q, Liu S B, An Q. IEEE Trans Nucl Sci, 2010, 57:
[2] Garvey J, Hillier S, Mahout G, et al. Nucl Instrum Meth A, 2003, 512: 506–516.
[3] Bourrion O, Guernane R, Boyer B, et al. J Instrum, 2010, 5: 12048–12048.
[4] Liu S B, Shen Q, Zheng W, et al. IEEE Trans Nucl Sci, 2010, 57: 625–629.
[5] Bocharov V, Bubley A, Boimelstein Y, et al. Nucl Instrum Meth A, 2004, 532: 144–149.
[6] Baron P, Atkin E, Blumenfeld Y, et al. IEEE Nuclear science symposium-conference record, Portland, 2004:

4 Conclusions

The improved L1 trigger logic with zero-delay broadening circuits as receivers eliminates nondeterministic latency and reduces total latency by half a clock period. This improved L1 trigger logic has been implemented in the GRAD trigger subsystem with reliable performance and is suitable for other L1 trigger systems requiring fixed, low-latency trigger decisions.

Acknowledgements
This work was supported by the Important Direction Project of the CAS Knowledge Innovation Program (No. KJCX2-YW-N27) and the National Natural Science Foundation of China (No. 11079045). The authors thank all HIRFL collaborators for their support and useful suggestions.

Note: Figure translations are in progress. See original paper for figures.