The Phase-1 Upgrade of the ATLAS Level-1 Endcap Muon Trigger

Shunichi Akatsuka*, on behalf of ATLAS Collaboration
*Kyoto University Graduate School of Science, Kyoto, Japan

Abstract—For Run 3 (from 2021), the LHC will undergo a significant increase in instantaneous luminosity to 1.5 times its current value which will lead to larger collected statistics and an enhanced sensitivity to new physics. The Phase-1 upgrade of the ATLAS Level-1 endcap muon trigger is essential to keep the physics acceptance at Run 3. A new trigger logic to take coincidence with detectors inside and outside the magnetic field is described and is shown to reduce the trigger rate to lower than the required level in Run 3. A new trigger board, NSL, to integrate all the information from various detectors, has been developed, and the implementation of the new logic on the FPGA has been successfully demonstrated.

I. INTRODUCTION

The Standard Model of particle physics offers the current best explanation of our universe, however there still remain some unsolved problems, such as the hierarchy problem and a lack of a dark matter candidate. These mysteries imply the presence of a Beyond Standard Model (BSM) physics. The collider experiment is one of the most promising approaches to probe BSM. The Large Hadron Collider (LHC) [1], located in Switzerland, is a proton-proton collider with center-of-mass energy at $\sqrt{s} = 13$ TeV. LHC provides the proton bunch collision at 40 MHz frequency, with its peak instantaneous luminosity at $2.1 \times 10^{34}$ cm$^{-2}$s$^{-1}$. LHC is planned to undergo a further upgrade of its luminosity and the center-of-mass energy, as shown in Fig. 1. Run 3 will start in 2021, with an instantaneous luminosity at $3.0 \times 10^{34}$ cm$^{-2}$s$^{-1}$, which is about 1.5 times larger than the instantaneous luminosity in the current run (Run 2).

ATLAS [3] is a multi-purpose detector to measure the particles produced at the interaction point (IP), located in one of the collision sites at LHC. The final data recording rate of the ATLAS experiment is limited to 1 kHz, therefore an online event selection (trigger) is performed in order to decrease the event rate below the limitation, while keeping the acceptance to new physics. ATLAS trigger system consists of 2 parts; the Level-1 (L1) trigger and the High-Level trigger (HLT). L1 trigger is a hardware-based trigger system that receives coarse detector information and reduces the event rate to smaller than 100 kHz. For those events that passed the L1 trigger, the full-detector data is transferred to the Readout System, and more precise information is passed to the HLT where final trigger decision is made. The limitation of the L1 rate comes from the data transfer bandwidth, from the front-end electronics to the readout system. This L1 rate limitation cannot be raised until a major update of the front-end electronics itself (which in fact will take place in Phase-2 Upgrade, from 2024).

Under this limitation, the requirements on the L1 trigger system will be severe at Run 3. Despite the higher event rate, the L1 rate need to be kept at the same level. The ATLAS detector needs an upgrade before LHC Run 3, to enhance its performance to cope with these high luminosity conditions. Collectively this effort is known as the Phase-1 Upgrade.

II. PHASE-1 UPGRADE OF THE ATLAS LEVEL-1 MUON TRIGGER

The L1 trigger rate assigned to the un-prescaled lowest-threshold trigger for the single muon (primary muon trigger) in Run 3 has been defined to be 15 kHz, considering other triggers and physics requirements [4]. Current primary muon trigger has a transverse momentum ($p_T$) threshold of 20 GeV, with its trigger rate of 20 kHz at instantaneous luminosity of $2.0 \times 10^{34}$ cm$^{-2}$s$^{-1}$. With an extrapolation from the trigger rate at the current luminosity, the rate at Run 3 luminosity at $3.0 \times 10^{34}$ cm$^{-2}$s$^{-1}$ will be 30 kHz, which is twice as much as the requirement. To keep the trigger rate below the requirements with current the trigger system, the $p_T$ threshold would have to be raised to 40 GeV [4], which will immediately lead to decrease the physics acceptance significantly. One of the examples that could be influenced strongly by the trigger threshold is the process where Higgs boson is produced in association with a W or Z boson. In this production channel, for the Higgs decay modes $H \rightarrow b\bar{b}$ [5] and $H \rightarrow c\bar{c}$ [6], the most important trigger will be the lepton trigger, which triggers the events by identifying leptons from the W or Z decay. If the muon trigger threshold is raised to 40 GeV, the trigger efficiency for $WH \rightarrow lvbb$ will decrease by more than 30%, as shown in Fig. 2.

![Fig. 1. LHC upgrade plans](From ATL-DAQ-PROC-2018-007. Published with permission by CERN.)
In order to keep the primary muon trigger threshold to 20 GeV, upgrade of the L1 muon trigger is essential. The main strategy of the upgrade is to reduce the muon trigger rate at regions with $|\eta| > 1.0$, the endcap region.

Figure 3 shows the pseudo-rapidity ($\eta$) distribution of the Level-1 muon primary trigger candidate (L1 MU20). This plot reveals that more than 90% of the muon trigger candidates in the endcap region are due to background events, i.e. events without muon with $p_T$ above the threshold. About 50% of these background triggers are due to events with no associated reconstructed muon. These background triggers are known as “fake” triggers, caused by charged particles emerging from the beam pipe. Other background triggers are due to muons with $p_T$ below the threshold. The main strategy of the upgrade is to eliminate the fake triggers and the low-$p_T$ muons, by implementing some new trigger algorithms which make use of several detectors that will be introduced in Run 3.

The detectors for the endcap muon trigger in Run 3 are shown in Fig. 4. The toroidal magnetic field bends the muon tracks, so that the $p_T$ can be calculated from the track angles in the Thin-Gap Chambers (TGC) [9] installed outside the magnetic field. These chambers are called TGC Big Wheel (TGC BW), as it forms a wheel-like structure. As shown in Fig. 4 while the fake particles create muon-like tracks in the TGC BW, they do not make any hits in the detectors inside the field. Thus, the fake triggers can be rejected by requiring a coincidence between hits in the TGC BW and the detectors inside the field. In Run 3, there are 4 detectors inside the magnetic field that can be used in L1 muon trigger: New Small Wheel (NSW) [7], TGC EI [9], RPC BIS 7/8 [10], and Tile Calorimeter [11]. NSW is a combined detector of micromegas [12] and sTGC (small-strip TGC), and will be newly installed in Run 3. The position and angle resolutions in the information that NSW can provide to L1 system is 0.005 in the $\eta$ plane, 10 mrad in the $\phi$ plane, and 1 mrad in the $\theta$ plane. These are dramatically high compared to the current inner chamber at the same position, which has resolutions of 0.15 in the $\eta$ plane and 65 mrad in the $\phi$ plane. TGC EI and Tile Calorimeter have already been installed since the beginning of the experiment. RPC BIS 7/8 will be newly installed in Run 3. The region with $1.3 < |\eta| < 1.9$ is covered by the NSW, and the region with $1.0 < |\eta| < 1.3$ is covered by TGC EI, RPC BIS 7/8 and Tile Calorimeter each covering different $\phi$ regions. Run 3 L1 endcap muon trigger will integrate information from all the detectors described above to make the final trigger decision.

### III. Trigger Algorithm

In this section, the algorithms of the Level-1 Endcap muon trigger in Run 3 is summarized.
**A. TGC BW local coincidence**

TGC BW consists of three stations, M1, M2 and M3, from the inner to the outer layers. Hit segment in the outermost M3 station is used as the trigger seed. If a line is drawn from the IP to a hit in M3, this straight line corresponds to infinite-momentum track. The deviations from this path in the M2 and M1 planes are used to calculate $p_T$. For the lower $p_T$ muons, the tracks will be bent more by the magnetic field, so the deviations from the straight track become larger. This trigger logic is implemented on a FPGA chip by a pre-defined Look-Up-Table. The deviation in the $R$ direction, $dR$, and the deviation in the $\phi$ direction, $d\phi$, are handed to the LUT, and then the LUT immediately returns the $p_T$ value. The correlation between the deviation angle and the $p_T$ is quite different depending on the trigger seed position, because the toroidal magnetic field is not uniform. Therefore the LUT must be defined depending on the position, to maximize the performance.

**B. Position Matching**

The main concept of the position matching with the inner detector is simple: that is, when the TGC BW finds a trigger candidate by its local coincidence, to confirm the decision by requiring hit in detectors inside the magnetic field. The position matching algorithm requires hits at appropriate position in the inner detector. As shown in Fig. 5 position matching will not only reject the fake triggers, but also reject low-$p_T$ candidates if the resolution of the inner detector is high enough. NSW, with the largest coverage in endcap region and high granularity, is expected to have an impact on the trigger rate by rejecting low $p_T$ muons.

![Fig. 5. Schematic diagram for the position matching algorithm. With a good position resolution at inner station, the segment position at inner station will have better sensitivity to the muon $p_T$.](image)

**C. Angle Matching**

In addition to the position matching algorithm, it is also possible to make use of the angle information at the inner station (angle matching). In the NSW trigger processor, $d\theta$ is defined as the angle of the segment with respect to the direction pointing at the detector center. For a muon emitted at the detector center and arrived to the inner station with a straight track, $d\theta$ should be ideally zero. However, we need to consider the beam spot size in $z$-direction ($O(10cm)$) and the multiple scattering with the detector materials, especially in the calorimeters. These effects allow the low-$p_T$ muons to fake the high-$p_T$ muons, as explained in Fig. 6. Here, a low-$p_T$ muon is indicated by the blue line, and high-$p_T$ muon is indicated by the red line. In Fig. 6, we are considering the situation when there is a trigger seed in a specific position at M3 plane. The low-$p_T$ track can fake the hit position of the high-$p_T$ tracks if it was scattered in the calorimeter region. The hit position in NSW and BW are very similar, but in this case the $d\theta$ information should differ significantly. Note that $d\theta$ information on itself cannot distinguish the $p_T$ of the muon, whereas can be combined with hit position in NSW and BW to retrieve more accurate $p_T$ information. By combining $d\theta$ and the position information, especially the $z$ position at NSW, low-$p_T$ muons can be eliminated effectively.

![Fig. 6. Schematic diagram for the angle matching algorithm. By combining the position and angle information at inner station, it is possible to further distinguish low-$p_T$ muons with straight-like hit position.](image)

**D. Trigger performance**

The trigger efficiency as a function of $p_T$ of the offline-reconstructed muon is shown in Fig. 7. This plot shows the relative trigger efficiency compared to the Run 2 trigger efficiency, calculated using single muon MC samples. The NSW track segment reconstruction efficiency is assumed to be 97%, and is included in this study. A significant increase in trigger rejection power for low-$p_T$ muon candidates are seen, for example additional 50% of the 10 GeV muons are rejected by taking both position and angle matching. Estimation of the trigger rate is shown in Fig. 8. The original $p_T$ distribution of the muons that pass the L1 MU20, shown by the dashed black line, is retrieved from the 2016 data. The detailed conditions can be found in [8]. Note that the fake triggers are not included in this plot, as the fake triggers are defined as triggers that do not match with offline reconstructed muons, and so that the $p_T$ of the muons cannot be defined.
The distribution after the NSW coincidence is obtained by multiplying the relative trigger efficiency in Fig. 7 to the original distribution.

Figure 8 shows the $p_T$ distribution of the muons that passed Level-1 muon primary trigger [8]. The distribution in dashed line is obtained from 2016 data. The distributions after applying position and angle matching are calculated by multiplying relative efficiency estimated by MC.

Figure 9 shows the pseudo-rapidity ($\eta$) distribution of L1 MU20 candidate [8]. The fake reduction by NSW and BIS 7/9 is estimated using other muon detector segments, while low-$p_T$ rejection is estimated from MC.

IV. DEVELOPMENT OF THE NEW SECTOR LOGIC BOARD

Figure 10 shows the block diagram of the new Sector Logic board (NSL). In this section, we summarize the hardware specification of the NSL board.

a) FPGA: The main processor FPGA of NSL is required to have many user I/Os, mainly for the 14 G-LINK [13] connections. These G-LINK connections are used to receive data from TGC BW. G-LINK uses 21 input user I/Os per channel, including the control/flag bits, therefore 14 G-Links will use 294 input I/O ports. Adding other I/O ports, the total user I/O required for New Sector Logic FPGA will be greater than 400. Another requirement for the FPGA is to have some Multi-Gigabit transceivers in order to receive data from NSW and other detectors.

Kintex-7 FPGA [14] XC7K410T-1FFG900 made by Xilinx Inc. was chosen as a main processor. XC7K410T has 500 user I/O pins, which is the largest number in the Kintex-7 series. XC7K410T supports GTX [15], and contains 16 GTX transceivers. GTX is a multi-gigabit transceiver implemented on Xilinx Kintex-7 series FPGA. 16 transceiver port is enough to fulfill the requirements.

This FPGA has 795 Block RAMs (BRAMs) [16], which is approximately 20 times as many as the current Sector Logic. BRAM is a large memory block which can contain up to 38 Kb of data, and is used to implement trigger LUTs in Sector Logic. Large number of BRAMs means a larger capability of trigger logic implementation. The BRAM resource needed to implement the trigger logic will strongly depend on what kind of logic we will use, and also on how we implement them on the BRAMs.

b) CPLD: CPLD is used for the VME bus control. Xilinx CoolRunner-II XC2C256-7Q208C [17] has been chosen, which is the same chip as the one used for the current Sector Logic. Because the CPLD is a non-volatile memory, FPGA on the NSL board can be configured via VME as soon as
the power is turned on. FPGA configuration via BPI is also controlled by CPLD.

c) **BPI**: Micron JS28F256P30T has been chosen for memory to contain a firmware design. This memory can contain up to 256 Mb data, which is enough to contain the firmware design.

d) **G-LINK Receiver chip**: HDMP-1034A chip [13] has been selected for G-LINK receiver chip. This chip receives serial data in G-LINK protocol, deserialize then transmit it as parallel signals. Same chip as the one used in current Sector Logic has been chosen.

e) **Ethernet PHY chip**: Microchip Tech. LAN8810i-AKZE [18] is used for ethernet connection. Using SiTCP [19] technology to connect this PHY chip to FPGA appropriately will allow data transmission by TCP/IP.

f) **Clock Jitter Cleaner**: Silicon Labs Clock Jitter Cleaner Si5334 [20] is used to create a reference clock for the GTX. The chip is semi-customized (Si5334C-B05812-GM) to input frequency of 40.08 MHz, which is the exact LHC bunch crossing frequency. The output of this jitter cleaner is a low-jitter 160 MHz clock, which is used for the GTX reference clock.

V. **FIRMWARE IMPLEMENTATION AND PERFORMANCE**

The trigger logic explained in Section III need to be processed on the FPGA, to be used in the L1 trigger. The requirement on the logic is that the trigger decision must be processed in fixed and short latency. The latency from the collision to the L1 trigger output of the Sector Logic is limited to 53 LHC clocks, including all the latency such as TOF, digitization, signal serialization, data transfer, trigger logic etc. The latency that can be assigned to inner coincidence logic is, considering other latencies, limited to 2 LHC clocks.

On the FPGA, we need to implement 3 types of LUTs: TGC BW local coincidence, position coincidence between BW and inner stations, and angle matching between BW and inner stations. A problem is that there can be several inner coincidence candidates for one BW trigger seed. For example, for the NSW segments, 16 candidates at the maximum can come into the scope for one BW trigger seed. To implement this kind of logic in a fixed-latency scheme, a naive method would be to implement 16 same LUTs so that the trigger processing can be performed in parallel. In such an implementation, we would need 16 times the FPGA resource, which is not acceptable in the current chip. Another way is to process the 16 candidates in serial, one-by-one. In this way the resource usage can be minimized, however the latency will be 16 times longer than the original latency.

To overcome these problems, we have decided to pick the good points of the above two ideas. Fig. 11 shows the brief schematic diagram of the implementation. In our implementation, two identical LUTs (LUT pair) were placed in parallel. Each LUT is operated at 320 MHz clock, which is 8 times faster compared to the LHC clock. Therefore, the LUT can process 8 candidates in one LHC clock. The 8 outputs values are then compared, and the highest $p_T$ candidate is chosen as the final output. This selection is done in 2 steps: in Fig. 11 they are referred to as “re-calc.” and “selector”. The first step simply compares the output of the 8 candidates from one of the LUTs, and then choose a highest $p_T$ candidate. The second step, “selector”, compares the 2 candidates from each LUT “$p_T$ re-calc.” algorithm, and then send the higher one back on the 40 MHz clock domain, as the final $p_T$ value. This implementation enables us to process maximum 16 candidates within 2 LHC clocks, and also within reasonable amount of RAM resource usage.

![Schematic diagram of LUT implementation](image)

Fig. 11. LUT implementation schematic. The different colors of the lines in the figure shows different clock domains. Up to 16 NSW candidates can be provided for one BW trigger seed, and they are split into 2 groups and are handed to two LUTs which have exactly the same content. The LUT is operated at 320 MHz clock to process all the candidates in 2 LHC clocks. The 2 candidates from each LUT “$p_T$ re-calc.” algorithm, and then send the higher one back on the 40 MHz clock domain, as the final $p_T$ value. This implementation enables us to process maximum 16 candidates within 2 LHC clocks, and also within reasonable amount of RAM resource usage.

![Vivado simulation output](image)

Fig. 12 shows a Vivado simulation output of the trigger logic part. This test was performed under a very simple situation, where only one of the LUT pair is used. The LUT is initialized to simply return the input value. The different to colors of the lines in the figure shows different clock domains. In this simulation, the original $p_T$ calculated in the BW local coincidence logic is “7”, as shown in “BW $p_T$” column. Eight inner segments information are handed to the LUT, shown as “NSW1 $d$” to “NSW8 $d$”. The 8 candidates are handed to the 320 MHz clock domain in series, and are
processed one-by-one, from “NSW1_deta” to “NSW8_deta”, by the same LUT. The output of the LUT is referred to as “NSW_pos_LUT_out” in Fig. 12. These output values are then compared, and the maximum \( p_T \) value is given to the “High_pT” register. The “High_pT” value is chosen as “Final_pT”, and then are handed back on the 40 MHz clock domain. All these processes are completed within two LHC clocks.

We have performed similar tests for various input patterns and LUT initialization values. These tests were done on logic simulation with Vivado, and also on an actual FPGA chip. The output of the trigger logic was consistent with our expectations, thus it is confirmed that we succeeded in handling the coincidence logic for maximum 16 inner segments, in a fixed latency and reasonable FPGA resource consumption.

VI. CONCLUSION

Phase-1 upgrade of the ATLAS Level-1 endcap muon trigger is essential to keep the physics acceptance at LHC Run 3. A new trigger algorithm to take coincidence with detectors inside- and outside- the magnetic field was suggested. The new algorithm consists of position and angle matching, to reduce the trigger rate to 14.2 kHz at 3.0 \( \times 10^{34} \) cm\(^{-2}\)s\(^{-1}\), which meets the Run 3 requirements. A new trigger board, NSL, to integrate all the information from various detectors, was developed. NSL has 12 GTX input port and 14 G-LINK input port, to receive data from all the five detectors; TGC BW, NSW, TGC EI, RPC BIS7/8 and Tile Calorimeter. The firmware implementation of the position and angle matching is realized by combining two ideas. The first is to place identical LUTs in parallel, which allows to process two candidates at the same time. The second is to re-use the LUT 8 times each, on a fast 320 MHZ clock, so that 16 candidates can be processed within the limited latency of two LHC clocks, while keeping the resource usage to acceptable level. The test of the trigger logic was successfully done on simulation, as well as on the real NSL board.

REFERENCES

[1] L. Evans and P. Bryant, “LHC Machine,” JINST, vol. 3, p. S08001, 2008.
[2] “HL-LHC High Luminosity Large Hadron Collider: The HL-LHC project.” [Online]. Available: http://hilumilhc.web.cern.ch/about/hl-lhc-project
[3] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” JINST, vol. 3, p. S08003, 2008.
[4] ——, “Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System,” Geneva, Sep. 2013. [Online]. Available: https://cds.cern.ch/record/1602235
[5] ——, “Evidence for the \( b \rightarrow b \) decay with the ATLAS detector,” JHEP, vol. 12, p. 024, 2017.
[6] ——, “Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment,” Phys. Rev. Lett., vol. 120, p. 211802, 2018.
[7] T. Kawamoto et al., “New Small Wheel Technical Design Report,” Geneva, Jun. 2013. [Online]. Available: https://cds.cern.ch/record/1552862
[8] ATLAS Collaboration, “L1 Muon Trigger Public Results.” [Online]. Available: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/L1MuonTriggerPublicResults
[9] ——, “ATLAS muon spectrometer: Technical Design Report,” Geneva, May 1997. [Online]. Available: https://cds.cern.ch/record/331068
[10] ATLAS Muon Collaboration, “Proposal of upgrade of the ATLAS muon trigger in the Barrel - End Cap transition region with RPCs,” in PoS TIPP2014, 2014, p. 117.
[11] ATLAS Collaboration, “ATLAS tile calorimeter : Technical Design Report,” Geneva, Dec. 1996. [Online]. Available: https://cds.cern.ch/record/331062
[12] Y. Giomataris, P. Rebourgeard, J. Robert, and G. Charpak, “Micromegas: A high granularity position sensitive gaseous detector for high particle flux environments,” Nucl.Instrum.Meth., vol. A376, p. 2935, 1996.
[13] “Agilent HDMP 1032-1034 Transmitter-Receiver Chip-set Datasheet,” Agilent Technologies. [Online]. Available: http://www.physics.ohio-state.edu/~cmfe/bd/hdmp1032.pdf
[14] “7 Series FPGAs Data Sheet: Overview,” Xilinx Inc. [Online]. Available: https://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf
[15] “7 Series FPGAs GTX/GTH Transceivers User Guide,” Xilinx Inc. [Online]. Available: https://www.xilinx.com/support/documentation/user_guides ug476_7Series_transceivers.pdf
[16] “7 Series FPGAs Memory Resources,” Xilinx Inc. [Online]. Available: https://www.xilinx.com/support/documentation/user_guides/ug473_7Series_Memory_Resources.pdf
[17] “XC2C256 CoolRunner-II CPLD,” Xilinx Inc. [Online]. Available: https://www.xilinx.com/support/documentation/data_sheets/d094.pdf
[18] “GMI 10/100/1000 Ethernet Transceiver with HP Auto-MDIIX Support,” Microchip Technology. [Online]. Available: http://ww1.microchip.com/downloads/en/DeviceDoc/000001870C.pdf
[19] T. Uchida, “Hardware-Based TCP Processor for Gigabit Ethernet,” IEEE Trans. Nucl. Sci., vol. 55, pp. 1631–1637, 2008.
[20] “IN-CONTROLLED ANY-FREQUENCY, ANY-OUTPUT QUAD CLOCK GENERATOR,” Silicon Labs. [Online]. Available: https://www.silabs.com/documents/public/data-sheets/Si5334.pdf
[21] “Vivado Design Suite HLx Editions,” Xilinx Inc. [Online]. Available: https://www.xilinx.com/support/documentation/backgrounders/vivado-hlx.pdf