Testing New RPC Singlet Electronics in ATLAS

And

Calorimeter Based Muon Identification Algorithm in ATLAS – CaloMuonScore WP tuning

CERN Summer Student Program 2023

Godfrey Ho Ching Luk

The Chinese University of Hong Kong

Supervisor:

Luca Pizzimento

Luis Roberto FLORES-CASTILL

Project Report

2023/8/18

Testing New RPC Singlet Electronics in ATLAS

2023 CERN Summer Students: Godfrey Ho Ching Luk (The Chinese University of Hong Kong)
e-mail: hochingluk@gemail.com
Supervisor: Luca Pizzimento
Hin Kwong Leung, Kin Yip Fung, Minlin Wu, Tsz Chun Wong

***Abstract* – To match the harsh working condition in the high luminosity LHC experiment, the current RPC needs to be upgraded. The upgraded front-end electronics will include an analog to digital converter and with better amplifier and discriminator. The performance of a prototype of the upgraded electronic will be tested and find out any possible defect.**

***Keywords* – ATLAS, HL-LHC, RPC**

I. INTRODUCTION

A. Working Principle of RPC

Resistive Plate Chambers (RPC) is a detector used to detect muon. The RPCs in the ATLAS detector uses a parallel plate configuration. This means the gas gap is sandwiched between two high resistivity plates. The gas gap is made out of insulated polycarbonate. Then, they are sandwiched between two graphite layers which are connected to a high voltage supply. Finally, there will be a series of parallel strip along θ direction on one side and a series of parallel strip along η direction on the other side, which the strip in both sides is perpendicular to each other. Every strip act as one output channel. When a charged particle passes through, the gas molecules are ionized. Electrons and ions are created along the path. Affected by the electric field created by the parallel graphite layer, the ionization process triggers an avalanche of electrons. The avalanche of electrons will induce a signal on both θ and η strip. So, the position of the muon pass through can be known. A schematic diagram of the RPC is shown in Figure 1.

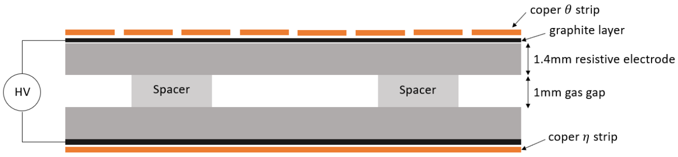

Fig. 1. A schematic diagram of the RPC detector.

The reason why there is a resistive plate is that the resistive plate help to protect the strip from the avalanche. Without the resistive plate, all electron created by the avalanche will reach the strip and create the signal directly. Although the signal will be larger than the induced signal, a larger signal will have a longer duration on the strip and making the strip not able to read another signal. So, the larger the signal, the lower rate capability the RPC has. That is why a resistive plate is needed to match the high collision rate in the LHC experiment.

There is another strip configuration other than the θ and η strip, which the strip at both side is parallel to each other. The disadvantage of having perpendicular strip is that electronic is need in both θ and η direction. This may cause trouble in installation or when multiple RPC is needed to configurate side by side to form a sector. Also, in the RPC that is in a long rectangle shape, the θ and η strip configuration will make the long side have a lot of strips and electronics. For the parallel strip configuration, the position is calculated using the time difference between the signal arrive the read out electronic. For example, if both ends receive the signal at the same time, the muon must be passed through the center of the strip. The RPC singlet that are being tested in this work have a long rectangular shape and its strip on two sides is parallel to each other. Currently only one electronic board is tested. That means only 8 strips on one side will be used. A photo of the RPC that will be tested is shown in Figure 2.

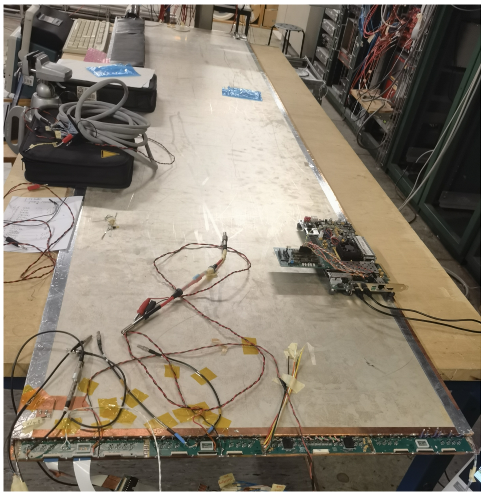

Fig. 2. A photo of the RPC that will be tested. Only the leftmost electronic board out of the four, which consist of 8 readout strips, is on.

B. RPC in ATLAS

The ATLAS muon detector system consists of 4 main types of detectors. They are Thin Gap Chambers (TGC), Resistive Plate Chambers (RPC), Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC). The position of the detectors in ATLAS is shown in Figure 3. The RPC is mainly for triggering and 2nd coordinate measurement in the central region. RPCs parallel plate structure makes it respond quicker than MDT. A parallel plate creates a constant electric field. This makes the electron created by the muon pass through accelerate and avalanche immediately. While MDTs tube shape creates an electric field that is inversely proportional to the radius. That means the electron will start an avalanche when it reaches close to the center. This makes RPC have a faster response than MDT and can been used as trigger.

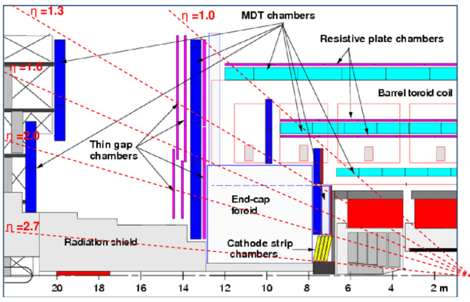

Fig. 3. Above shown the position of the muon detectors in ATLAS.

C. RPC Upgrade

The current RPC needs to be upgraded to meet the high luminosity LHC experiment (HL-LHC). In HL-LHC, the integrated luminosity is expected to reach $3000fb^{-1}$. Such harsh condition not only increase the required rate capability, but it also increases pileup and radiation background. So, the RPC will be aging faster, and the rate capability will drop quicker over time. The current installed RPCs are certified to work at $100Hz/cm^2$ for 10 years. And the current rate capability of the RPC is around $1kHz/cm^2$. However, the 10 years performance guarantee may not be held at the harsh condition of HL-LHC. So, an upgrade is needed to improve the rate capability.

The rate capability of the RPC is controlled by 3 quantities by the equation.

$$rate\ capability = \frac{1}{\rho d < Q >}$$

where ρ is the resistivity of the electrode, d is the thickness of the electrode and <Q> is the average charge per count. So, one of the three quantities has to be decreased in order to increase the rate capability. It is hard to reduce ρ since it is determined by the material that are currently using. Reducing ρ means a new material is needed. A new material will have the risk of increasing the material cost or increasing the aging problem due to more current driven. And it is unlikely to find a material that has resistivity 1 order of magnitude lower than the current one. It is also hard to reduce d. The thickness of the electrode is currently 1.4mm. It is technologically difficult to reduce the thickness by a large amount. So, the only possible quantity to reduce is the average charge per count. The problem of reducing <Q> is that the signal will also become much smaller. So, in order to detect such a small signal, a very sensitive front-end electronic with good signal to noise ratio is needed. The system also needs to suppress the noise induced inside the detector by electronics or externals source. An upgraded electronic prototype is produced, and it can in theory detect the signal even if the charge produced inside the gas gap is a factor of 10 smaller. Such electronic will allow an increase of the rate capability up to $10kHz/cm^2$.

Apart from increasing the rate capability, this electronic upgrade also aims at putting the analog to digital converter in the front-end electronic board. Currently, the front-end electronic board can only send out analog signal to an external analog to digital converter. However, transmitting analog signal may cause the signal to translate or distort. Introducing systematic error to the timing. However, the advantage of transmitting digital signal is that digital signal will not be affected by a distortion of shape. So, putting the analog to digital converter in the front-end board allows the RPC to send out digital signal directly.

D. Upraged Electronic

A simplified electronic configuration of the RPC is shown in Figure 4. The RPC original signal will first pass through a charge integration amplifier, such amplifier can integrate the current input to remove noise. Then the signal is passed through a discriminator. A threshold is set for the discriminator so that the signal will be reconstructed as a square wave by setting all voltage above the threshold to be the threshold value while all voltage below the threshold will be 0. Then the signal will pass through the analog to digital converter.

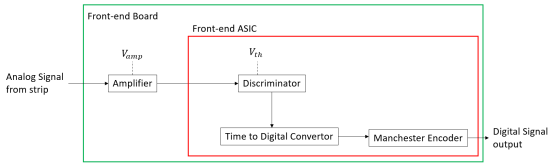

Fig. 4. A simplified electronic configuration of the RPC. The solid line represent the signal flow while the dotted line represent some of the important low voltage input. V_{amp} control the degree of amplifacation in the amplifier and V_{th} control the threshold of the discriminator.

The analog to digital converter consists of two main parts, Time to Digital Converter (TDC) and a Manchester encoder. The simplified structure of the TDC is shown in Figure 5. The Voltage Control Oscillator (VCO) will produce a periodic square wave with a period of 50ns. The square wave will then pass through the Divider to become 8 square waves with period each divided by 2. This is the 8-bit clock that will be used to record time as illustrated in Figure 6. The Scalar will record the binary number when the signal from discrimination arrives. An external clock is needed to reset the Scaler memories every 50ns. The Manchester encoder will then transfer the 8-bit binary number to Manchester word.

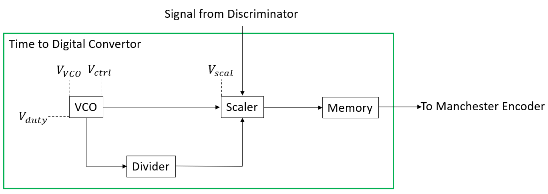

Fig. 5. A simplified electronic configuration of the TDC. The solid line represent the signal flow while the dotted line represent some of the important low voltage input. V_{VCO}, V_{ctrl}, V_{duty} control the degree of amplitude, frequency and the shape of the square wave created by the VCO respectively. And V_{Scal} control the threshold of the Scaler to read the square wave produced by VCO.

The advantage of using Manchester code is that charge will not accumulate during transmission. Transmitting using non-zero voltage bit as 1 and zero voltage bit as 0 will cause charge to accumulate since the reading electronics will set the reference voltage to be 0.5 to read both 1 and 0. When the number of 1 and 0 is not equal in a word, the average voltage will not be 0.5 and charge will accumulate. The Manchester code avoids the problem by using transition to represent 0 and 1. In 1 bit, from zero voltage to non-zero voltage is 1 and the reverse is zero as shown in Figure 7. This can ensure the average voltage is always 0.5 and no charge will be accumulated.

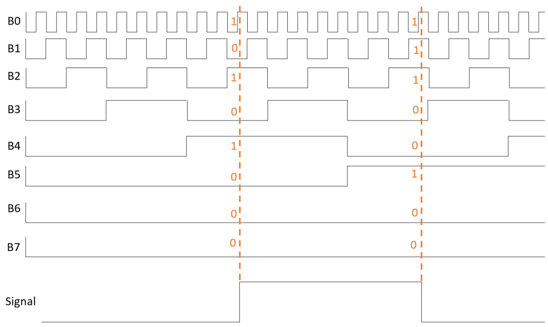

Fig. 6. Above illustrate the working principle of the 8-bit clock. B0 to B6 is the 7 square waves obtained by dividing the initial B7 square wave from VCO. Then the signal arrive, the leading time and the trailing time will be recorded. The example signal have the leading time 00010101 and the trailing time 00100111. One binary digit represent time that is half the period of B0 square wave.

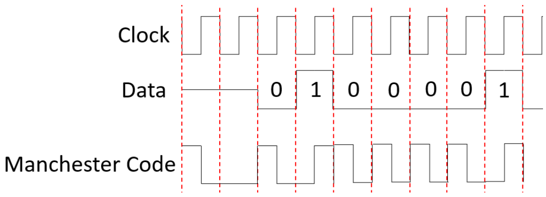

Fig. 7. Above shows an example of Manchester encoding.

Before the Manchester word started, there a Manchester error is printed. A Manchester error is a sequence that is not possible to appear in a word, which is a 1.5 bit of zero voltage, to indicate the start of a word. Then the Manchester word of the RPC electronics is printed. The word consists of 19 bits. The first bit is always zero. The next two bits indicate how many clock cycles does the signal last for. Since the RPC signal is around 4.5ns, these two bits will also be zero. The next 8 bits will indicate the leading time while the last 8 bits will indicate the trailing time.

An example of the whole electronic process is shown in Figure 8.

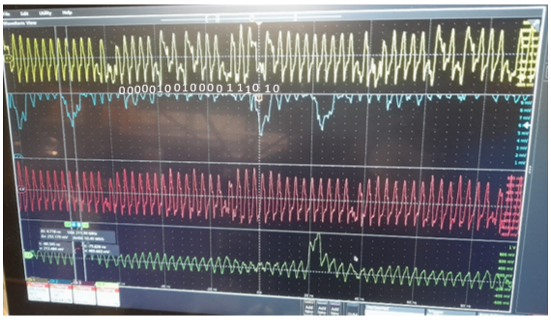

Fig. 8. An example of the whole electronic process. The white 0 and 1 is marked to shows how to do Manchester decoding manually. The signal width is measured by the cursor in the green channel.

In the Figure, the blue channel is the injected signal. The green channel is after passing through the amplifier and the discrimination. The red channel is the B3 clock since its period is 16 times smaller than the 50ns clock reset period. The yellow channel is the Manchester word. Decoding the word give the 0 00 00100100 00111010. Since the clock reset every 50ns, 1 binary digit represents $50/2^8 ns$, which is 0.195ns. So, the leading time is 7.4ns while the trailing time is 12.0ns. Because they are in the same clock cycle, the width of the signal gives 4.6ns, which is consistent with the signal in the green channel.

E. Purpose of the testing

With the electronic prototype, the aim of the testing is to check whether it works properly and the cause of any poor performance. The final goal of the study will be finding out what combination of low voltage input will the electronics perform the best, which is impossible to known by theory since the voltage work with each other and affect each other, like V_{th} and V_{amp} or V_{vco} and V_{scal}. So, although there is a standard value of the low voltage inputs, testing is needed to tun the voltage value such that the electronic can perform the best.

II. TESTING

A. Low Voltage Setup Test

To power the RPC electronic system, 16 individual low voltage is needed, 3 of them is ground. The voltage applied will be controlled by a multi-channel power supply. The power supply and the printed circuit board (PCB) are connected by a 16-pin flat wire using an IDE to ZIF adaptor shown in Figure 9. The purpose of this test is to ensure that the different channels from the power supply are connected to the PCB correctly, which is verifying the configuration in Table 10.

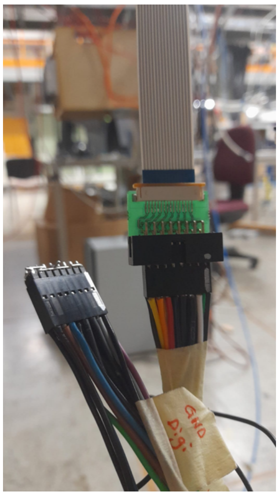

Fig. 9. The bottom wires are comes from the multi-channel power supply and the top is the 16-pin flat wire.

The test is performed by turning on the power supply channel one by one to 1V. When only one channel is on, the voltage of all 16-pin in the flat wire and the adaptor is measured. Only the correct pin in the flat wire and adaptor should get 1V while all other 15-pins should have no voltage. During the testing, the pin of V_{Dig} and V_{driver} only give 0.6V when connected to the PCB while they give 1V when disconnected to the PCB. All other pins give the correct voltage. This shows the adaptor, and the flat wire works properly but there is some problem with the PCB. After further investigation by Luca, turns out the circuit of V_{Dig} and V_{driver} are not properly connected to ground in the PCB. This happens because the PCB and the chip are still a prototype and there will be such kind of defect. This problem is solved by welding out a resistor in the PCB. The welding is done by Hin Kwong.

Low Voltage Name	Pin no. in flat wire (Left to Right)	Channel no. in power supply
GND_{analog}	1	Ground
V_{TH}	2	1A0
V_{amp}	3	1A1
$Memery$	4	1A2
V_{ctrl}	5	2A0
$Bias$	6	2A1
V_{Dis2}	7	2A2
GND	8	Ground
V_{Dig}	9	2A3
V_{Driver}	10	2B0
V_{SCAL}	11	2B1
V_{DDout}	12	2B2
V_{sync}	13	2B3
V_{Dis1}	14	1A3
V_{VCO}	15	1B0
GND_{Dig}	16	Ground

Fig. 10. The configuration of the 16 low voltage supply. Relating the order of the flat wire and the order of the multi-channel power supply.

So, through this test, the connection between the multichannel power supply and the PCB is checked, and it has been confirmed that the voltage set in the power supply will be the same as the voltage received by the electronics.

B. Gas Gap Quality Test

Before starting any testing in the RPC, the quality of the gas gap must be tested since after a long time of not using a gas gap, the gas gap may deform or accumulate dust.

The quality of the gas gap is characterized by the current between the graphite layer with only the high voltage supply between the graphite layer is on, that means all electronics are off. There are two effects that will contribute to a current in the RPC even if no muon is passing through. The first effect is sparks. When the voltage is high, any spike or impurities on the surface of the electrode will create a huge electric field. Gas molecules will then be ionized and create sparks. This effect creates current in the order of μA and the current growth exponentially with the high voltage supplied. The second effect is parasitic resistance. Inside the gas gap, there are spacers that create support to prevent compression of the chamber due to air pressure or gravity. Although the spacers are made out of insulating polycarbonate, layers of dust, oil or impurities may be attached on the surface of the spacers. These parasitic resistances will connect the two electrodes and create current when the high voltage is provided based on Ohms law. Since the current is governed by Ohms law, the current created by this effect increase linearly with the high voltage and it is in the order of μA. During the testing, the effect of cosmic muon is negligible is it will only create current in the order of pA.

When the current between the graphite layer is measured, the combined effect is shown. So, the expected current against high voltage supply graph will show a linear behavior at low voltage and exponential increase at high voltage like in Figure 11. The quality of the gas gap can be quantified by looking at the turning point of the current against high voltage supply curve. If the turning point is too high, that means the linear effect is strong and there is too much parasitic resistance. This indicated that many impurities accumulated on the surface of the spacer, or the gas gap is inflated.

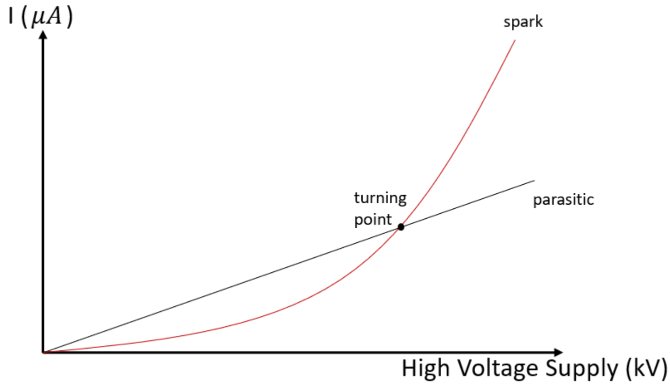

Fig. 11. The expected current against high voltage supply graph due to the effect of sparks and parasitic.

The current between the electrode is too small to be measured. Instead, the voltage across an 100k external resistor that is in series is measured as shown in Figure 12. So, the current between the electrode can be obtained using Ohms law. A capacitor is added to the circuit so that the current from sparks can be smooth instead of a sharp current peak.

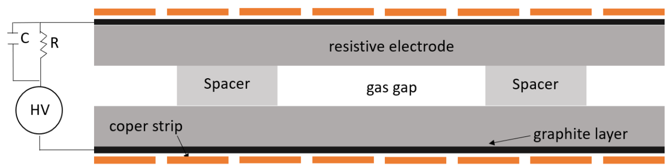

Fig. 12. The setup that are use to calculate the current by measuring the voltage across an 100k external resistor.

The result of the measurement is shown in Figure 13. A proper gas gap should have a turning point less than 5A. Since our gas gap have the turning point around 0.05A, the gas gap is in a pretty good condition.

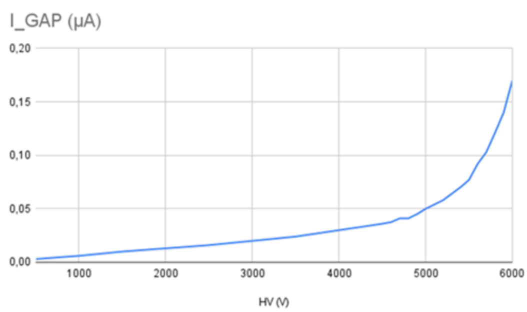

Fig. 13. The result of the current against high voltage supply curve, which clearly show the linear behaviour at first then exponential behaviour.

C. Noise Level Test

After confirming the gas gap is in good quality, the noise level test is performed when turning on the amplifier and the discriminator. The high voltage supply is set to be 5.8kV, which is the operation voltage of the old RPC. The source of the noise now only comes from cosmic muon and sparks since the amplifier will remove the effect of the parasitic resistance.

This noise level test has two purposes. First is to ensure the noise level is low when compared with the signal level. The primary method for RPC to remove noise is to use RPC doublet and perform time coincident. That is only when two RPC measure a signal as the same time will be count as a muon pass through. But this method only works when the noise frequency is much lower than the signal frequency. The second purpose is to study the effect of the low voltage on the noise level, especially V_{amp}. Increasing V_{amp} will increase the amplification of the signal. This makes the noise level increase with V_{amp} since more signal can pass through the threshold of the discriminator. Since the value of V_{amp} will change in later test, it is important to make sure the variation of V_{amp} will not greatly affect the noise level.

To measure the noise level, the output signal is directed into an oscilloscope. The oscilloscope is set to show a duration of 1s without any trigger. The number of signals within 1s is counted manually 100 times. An example is shown in Figure 14. The average count per second is obtained and it is converted to Hz/cm^2 using the area of the RPC.

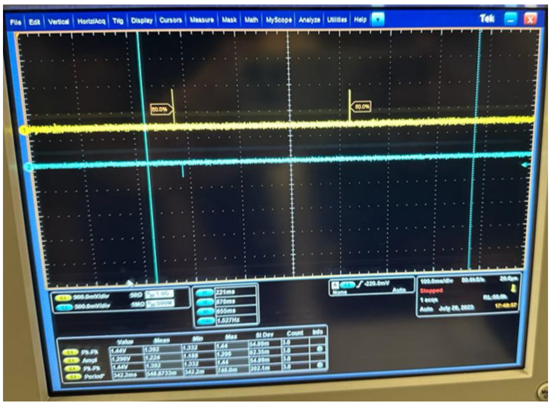

Fig. 14. A example of counting the noise level of the RPC. The noise is shown in the yellow channel. Two noise is observed and the first one is from cosmic muon since the scintillator from blue channel shows a signal. And the seond noise come from spark.

Then only the amplifier and the discriminator are on, the measured noise is around 0.00006 to 0.00013 Hz/cm^2 when V_{amp} increase from 1.1V to 1.5V. When all electronic is turned on, the noise level remain at around 0.0017 Hz/cm^2. Compared to the expected muon rate in ATLAS, which is around $100Hz/cm^2$, the noise level is extremely small. Variating the voltage input of electronics has a negligible effect on the noise level. So, the electronic voltage supply can be changed in further test without worrying about the increases of noise level.

D. Efficiency Test

After testing the quality of the gas gap and the noise level, the efficiency of the RPC with only amplifier and discriminator is on is tested. The efficiency of an RPC detector depends on the high voltage supply. A muon will be missed if it cannot produce a large enough signal. The higher the voltage supply, the larger signal will be produced when a muon passes through. So, the efficiency increases with the high voltage supply as a sigmoid function shape. However, increasing the high voltage will increase the amount of noise. The convention working voltage of an RPC is the voltage that can achieve 98% efficiency. The purpose of the test is to vary low voltage supplies to obtain the different efficiency curve. And find the low voltage supply setting which gives the lowest working voltage, that is finding a combination of the 16 low voltage supply that will give the RPC best performance.

To measure the efficiency of the RPC, the actual amount of muon that is passing through the RPC is needed. Since cosmic muons are used as the source, a scintillator pair is used as the trigger. The setup of the scintillator is shown in Figure 15. The signal from the RPC and scintillator is directed to the oscilloscope and the trigger of the oscilloscope is set on the scintillator. One efficiency curve is obtained by increasing the high voltage setting from 4.8kV to 6kV in step of 0.1kV. For every high voltage, 1000 triggered event is recorded and saved as csv file. A python script is written by Tsz Chun to read the csv files and calculate how many events will have a signal by RPC given that there is a signal by scintillator. The efficiency of 3 different sets of low voltage is measured by varying V_{th} and V_{amp}. The result of the efficiency curve is shown in Figure 16. Note that when V_{th} and V_{amp} increased, the efficiency curve shifted left. This is expected since increasing V_{th} and V_{amp} will give increase the signal amplification by the discriminator and amplifier. So, more signals can be captured. However, there is still some problem in the efficiency curve which needs further investigation before varying other low voltage supplies.

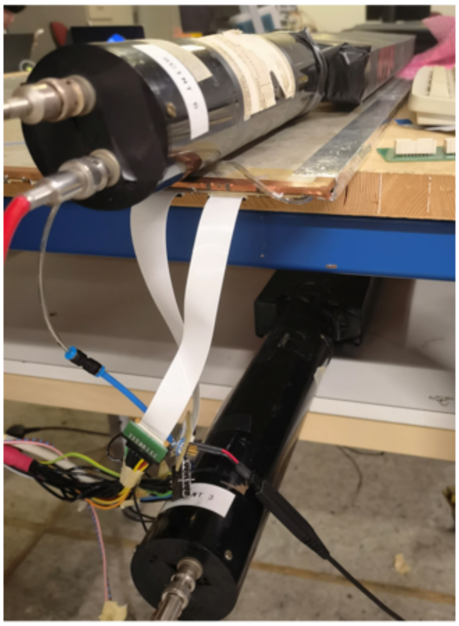

Fig. 15. Above shows the setup of the scintillator. The black detector at top and bottom of the RPC is the scintillator.

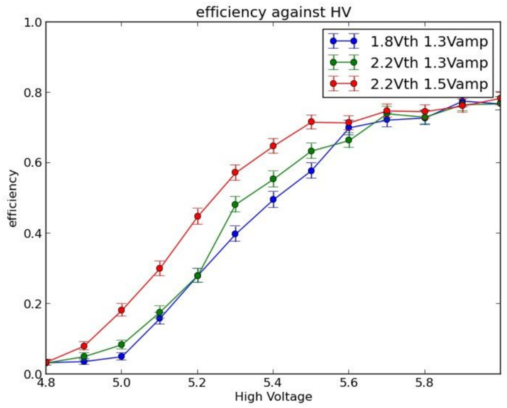

Fig. 16. The three efficiency curves obtained by changing V_{amp} and V_{th}.

First, the starting point of the curve is not 0%. The efficiency of all 3 curves is 3% at 4.8kV. The non-zero efficiency at the starting tail may means the background noise is too high that it makes sone sparks have time coincident with the scintillator. So, some things went wrong if the efficiency kept at 3% when the high voltage further decreased. To confirm the noise level is in an acceptable range, the efficiency of 4.6kV and 4.7kV is also measured and the efficiency dropped further dropped to 1%. This means the efficiency will keep dropping when lower voltage is supplied, which is normal.

The second problem is the saturation efficiency is not near 100%. At all 3 curves, the saturation efficiency is around 80%, which is unacceptably low. A typical working RPC should have an efficiency approach to 100% as very high voltage. There are two possible explanations of such low saturation efficiency. It could be the RPC is not completely covered by the scintillator pair geometrically. So, the acceptance of the setup is not 100%. Some cosmic muons that pass through the scintillator do not pass through the RPC. Another possible reason is that one of the channels in the RPC is broken so that it does not produce signal even if a muon passes through.

The third problem is the sudden jump near high voltage, like 5.6kV to 5.7kV jump for the green curve. The jump is not caused by statistical fluctuations since 1000 event is taken for each data point. This makes the statistical error of each data point very small as shown in the graph. The jump may be caused by a broken readout strip. At low voltage, a broken readout strip can not produce any signal. But when a higher voltage is applied, the avalanche size will be increased, and the signal can be readout by nearby strip. So, the broken strip is being covered by nearby strips and suddenly a lot more muon can be detected. This will cause a sudden jump in efficiency.

The fourth problem is the efficiency curve increases linearly at the middle, not in a sigmoid function shape. This again indicated the acceptance of the setup is not 100%. The muon that initially missed the RPC can be detected since the avalanche size increase when high voltage supply increase, which means the effective area of the RPC is increasing. This effect causes efficiency to increase linear with the high voltage supply.

So, before varying low voltage setting to obtain more efficiency curves and find out what is the best combination of low voltage setting, the above problem needs to be solved. The broken strip problem can be checked by measuring signal from strip individually. The acceptance problem can be checked by using a doublet RPC to do self-time coincident trigger instead of singlet RPC plus scintillator trigger. Since overlapping two RPC together will guarantee 100% acceptance. But both checking need to change to setup and rewelding wires to electronics which takes some time for the technician to complete. Currently, a doublet is in preparation. So, the efficiency test is terminated for now.

E. TDC Setup Test

The next test aimed to set up the time to digital converter (TDC) and the Manchester encoder. A pulse generator is used to provide a signal to simulate a muon pass through. The simulated signal will process through the whole set of electronics shows in Figure 4. The pulse generator also provides a periodic pulse as the clock for the TDC. Four signals are directed into the oscilloscope as shown in Figure 8. However, the standard value of the 16 low voltages does not produce a meaningful Manchester encoded signal and tuning is needed for the 16 low voltage values. The low voltage values are varied from the standard value one by one. After the Manchester signal is decoded manually, the performance of the TDC and the Manchester encoder can be known since the timing and the width of the injected signal is known.

After trying several low voltage combinations, some combination gives a non-decodable Manchester signal, such as the one shown Figure 17, while some give an unreasonable leading time and trailing time. Since the injected signal is known to be 4.5ns width, there is some systematic problem when the Manchester code gives a width of that is off by 2ns. Some Manchester code will even give a leading time larger than the trailing time, which is not possible. Not a single low voltage combination will give a correct Manchester coded signal.

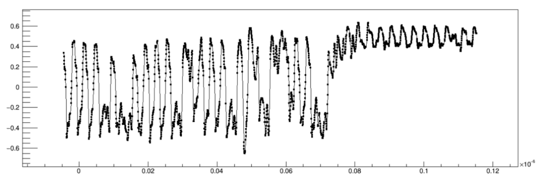

Fig. 17. An example of non-decodable Manchester signal.

To solve the problem, Luca improved the grounding system of the electronics to reduce electronic noise. He replaced the pulse generator with a new one that is more reliable and performs better at high frequency since operating as a clock needs a stable periodic signal with a very high frequency. The new pulse generator is shown in Figure 18. After these two changes, the system gives a reasonable Manchester signal output for the standard low voltage values and for the voltage value that has around 1 Volt deviation from standard value. So, after making sure that the whole electronic system is working, further test can be performed to investigate the performance of the TDC and Manchester encoder.

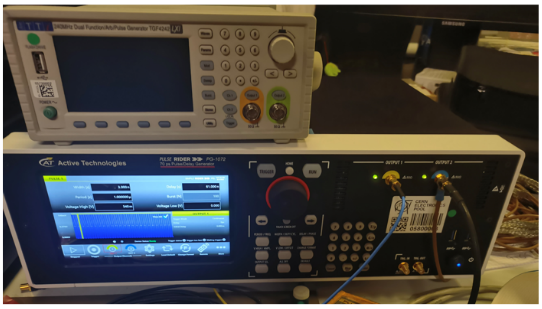

Fig. 18. A photo of the old pulse generator (top) and the new pulse generator (bottom).

F. Manchester Encoding Test

After making sure the output Manchester encoded signal is reasonable, a large amount of event can be recorded and decoded using a package called RPC_upgrade_manchester developed by Kin Yip Fung. After passing thousands of events to the decode package, 3 problems are found.

First, a few Manchester encoded signal is non-decodable with some reason, such as the example is given in Figure 17. In theory, every injected signal is the same so the digital signal output should be the same for all 767 events. So, the efficiency of the system is around 99.5%. It is unsure these Manchester words is causes by poorly injected signal, unstable clock, fail of Manchester encoder or other things. Second, when the leading time of all signals is plotted as a histogram, one peak is expected since all signal is injected at the same time. However, two peaks are observed in the resulting histogram, see Figure 19. This maybe cause by an unstable clock or an unstable signal injection. Thirdly, the mean value of each bit should be 0.5, that is the reason why the Manchester code is used. But looking at the mean value of the bit in the word, they have a huge mean value deviation as shown in Figure 20. This indicates that although the encoding is successful, the encoded digital signal is not of a good quality. All three problems need further investigation to find out the cause of the problem and the way to solve it. But the overall performance of the TDC and the Manchester encoder is not bad.

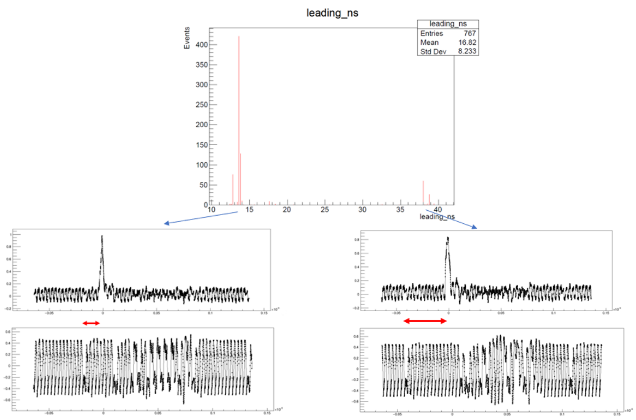

Fig. 19. Above shows the distribution of the signal leading time. The bottom graphs shows two different the distance between the leading time and the clock reset time. Since the trigger of the oscilloscope is based on signal, the unstable can happen on the signal or the clock.

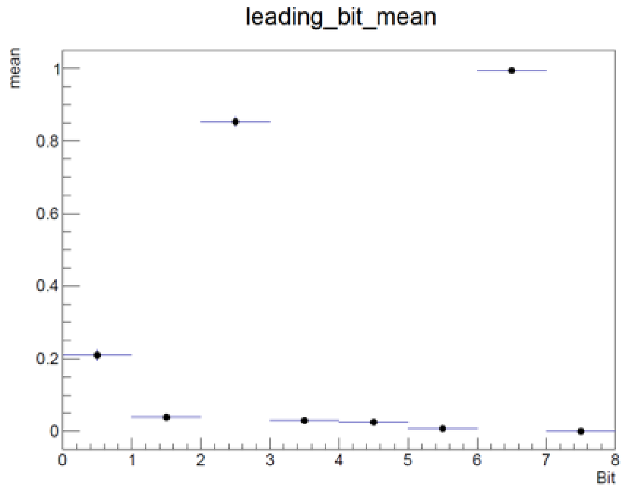

Fig. 20. Above shows the voltage average of the 8 leading bit from 767 Manchester word.

III. CONCLUSION

As discussed in all the testing mentioned above, the performance of the RPC electronic is acceptable as a prototype. Still there is some poor performance with unknow reason. Further investigation is needed to find out the cause and the way to improve the electronics.

ACKNOWLEDGMENT

This work will not be possible without Luca. He provided a detail instruction of task and well rounded physics background behind the task. Also, special thanks to Him Kwong for the electronic technical support and Tsz Chun and Minlin for writing the analysis code. Lots of the data taking or setup work is a collaboration work of me and Tsz Chun.

DECLARATION

The reference of this report comes from Luca presentation slides about Upgrade of the ATLAS muon spectrometer for the High Luminosity phase of LHC and also his explanation during day to day work. So, the content of this report maybe over simplified or not perfectly accurate. This report should only act as a experimental notes and should not be used as a reference of others work.

Calorimeter Based Muon Identification Algorithm in ATLAS (CaloMuonScore - Cut Value Tuning)

2023 CERN Summer Students: Godfrey Ho Ching Luk (The Chinese University of Hong Kong)
e-mail: hochingluk@gemail.com
Supervisor: Luis Roberto FLORES-CASTILLO, Tong Pan, Kin Yip Fung

***Abstract* – CaloMuonScore is an algorithm that aims at recovering efficiency in the ATLAS muon detector system, which is an inner detector system and muon spectrometer system. CaloMuonScore identifies muon from background by processing the energy deposits pattern of the particle in calorimeter using convolutional neural network. CaloMuonScore is proven to be outperforming the old identification algorithm, a energy threshold based CaloMuonIDTag. However, the performance of CaloMuonScore need to be investigate closer before using it in analysis. This study will investigate and propose solution to two potential defects of CaloMuonScore. First, the performance of CaloMuonScore is inconsistent in different transverse momentum. Second, the background efficiency shows huge fluctuation at high transverse momentum with unknown reason.**

***Keywords* – ATLAS, CALOMUONSCORE, CERN, CONVOLUTIONAL NEURAL NETWORK, MUON IDENTIFICATION**

I. INTRODUCTION

A. Muon Reconstruction in ATLAS

In ATLAS, two main groups of detectors are detecting muon. They are the Inner Detector (ID) and the Muon Spectrometer (MS). A schematic view of both detectors is shown in Figure 1 and 2 [1]. When a muon passes through the detectors, tracks will be reconstructed within ID and MS. Two tracks are combined in a global fit and the momentum of the muon can be known by measuring the curvature of the track given by the ATLAS magnetic system. However, the efficiency of the system is not 100% since the geometrical acceptance of the detector is low in some regions [2]. The acceptance MS is shown in Figure 3. To recover some of the efficiency, a Muon identification algorithm based on calorimeter is developed.

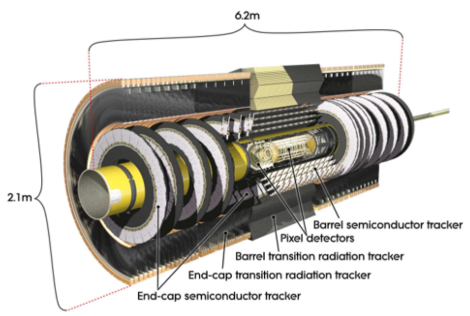

Fig. 1. A schematic view of the Inner Detector. Inner Detector is located close to the beam line.

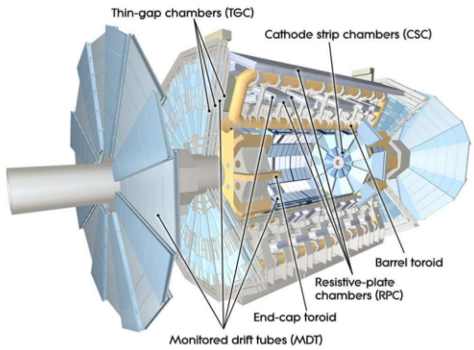

Fig. 2. The schematic view of Muon Spectrometer. Muon Spectrometer is located at the outermost layer of ATLAS.

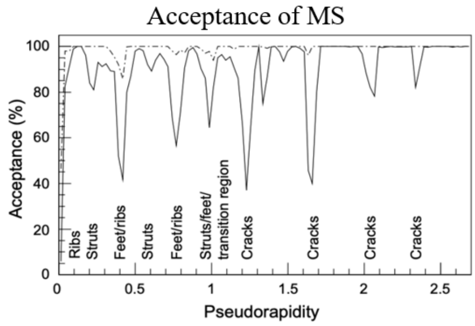

Fig. 3. The acceptance of MS in ATLAS. The MS has nearly zero acceptance for $|\eta| < 0.03$ due to a 30-cm gap required for cryogenic pipes and other service lines to the electromagnetic calorimeter. The reduced acceptance of other regions is due to the mechanical structures that support barrel magnets or cracks between the non-perfectly overlapped end-cap MDT chambers.

B. Muon Identification in Calorimeter

Since muon is not the only particle that will deposit energy in the hadronic calorimeter, an algorithm is needed to separate muon from background like pion and kaon.

Initially, the CaloMuonIDTag algorithm, which is a simple energy threshold cut algorithm, is used [3]. The inner detector tracks are extrapolated to the electromagnetic and hadronic calorimeters, and the energy deposits in the calorimeter cells traversed are recorded. Since muon generally deposits less energy in the calorimeters compared with hadron, muon can be identified by setting an energy threshold, E_{th}, with

$$E_{th} = \frac{50MeV}{sin^2\theta}$$

where θ is the track scattering angle, which is the polar angle measured between the beamline and the muon ID track. The threshold is θ depending since the energy deposit increases with the path length of the muon in calorimeter and the muon with low θ have a longer path length. So, particle deposit energy less than E_{th} will be classified as a muon.

Later, a Convolutional Neural Network (CNN) based algorithm called CaloMuonScore was developed by Ricardo Wolker [1]. CNN is initially an algorithm applied for image recognition. In CaloMuonScore, instead of inputting the 3 channels from RGB images, 7 channels are used as input representing the energy deposit profile of the muon when it passes through 7 layers of calorimeter. An example of an energy deposit profile is shown in Figure 4 [1]. The algorithm will give a score between 0 to 1 to every particle. Then, a cut value is applied and all particles with a score higher than the cut value will be identified as muon.

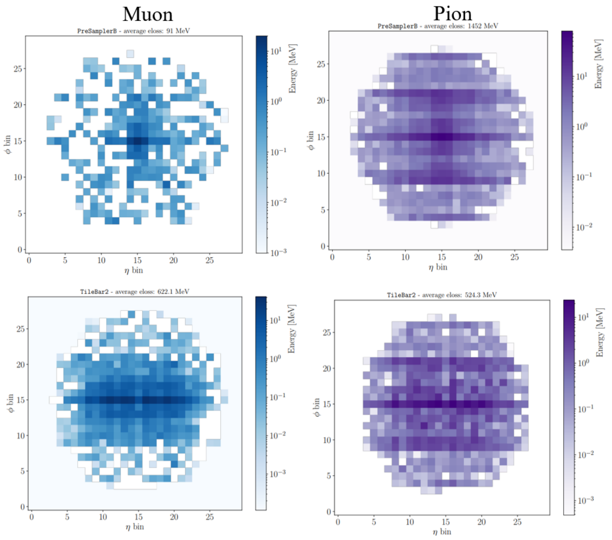

Fig. 4. Above shown the average energy deposit profile of 2000 muons or pions in calorimeter layer 1 (top) and layer 7 (bottom). The total energy deposit and the deposit pattern for muons and pions are different and this is used to distinguish muons from pions in the algorithm.

C. CaloMuonIDTag vs CaloMuonScore

To compare the performance of the algorithms, the quantities signal efficiency ε_{sig} and background efficiency ε_{bkg} are used.

$$\varepsilon_{sig} = P(ideatified\ as\ muon|real\ muon)$$

$$\varepsilon_{bkg} = P(ideatified\ as\ muon|not\ muon)$$

Signal efficiency is defined as out of all real muons, how many muons can the algorithm successfully identify. Background efficiency is defined as out of all fake muons, how many fake muons are misidentified as muons by the algorithm. So, the background efficiency is also called the fake rate. Ideally, the signal efficiency should be as high as possible, and the background efficiency should be as low as possible.

One of the advantages of using CaloMuonScore over CaloMuonIDTag is that CaloMuonIDTag gives a binary label to muon but CaloMuonScore gives a sliding scale. This mean CaloMuonIDTag can only have a fix ε_{sig} and ε_{bkg}. But for CaloMuonScore, one can change ε_{sig} and ε_{bkg} by changing the cut value based on the purpose of the analysis. A higher cut value represents a tighter selection while a lower cut value represents a looser selection. For example, if an analysis only needs a fraction of muons, a higher cut value can be applied such that both ε_{sig} and ε_{bkg} will decrease. This means less muon will be identified but the fake rate will be lower. So, the performance of the CaloMuonScore in a testing sample is shown as a receiver operating characteristic (ROC) curve in Figure 5 [1]. The test sample consists of Monte Carlo Simulation result of pion from 2×10^5 ttbar event and muon from 2×10^4 Zmumu event. Clearly some cut value in CaloMuonScore can give a higher ε_{sig} with a lower ε_{bkg} than CaloMuonTag.

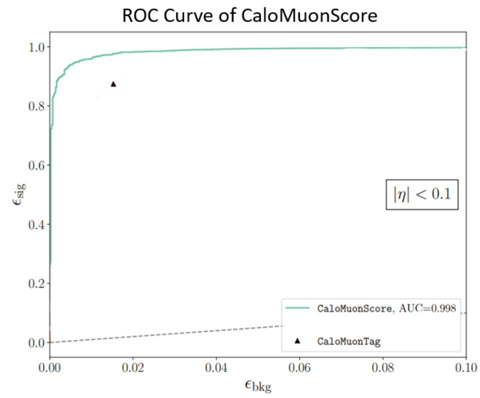

Fig. 5. The green line showns the ROC curve of CaloMuonScore. The higher the ε_{sig} the higher the ε_{bkg} since increasing the cut value will make the algorithm able to identify more muon. But at the same time misidentify more non-muon as muon. The dashed line represents the theoretical ROC curve one would obtain by making random guesses.

The current cut value used for analysis is aimed to achieve a flat ε_{bkg} of 6.8% for all transverse momentum (pT). To achieve a flat ε_{bkg}, the cut value has to be pT depending since most fake muon has a low pT. So, the cut value has to be higher at low pT to achieve the same fake rate. The current cut value curve of the algorithm is shown in Figure 6 [1].

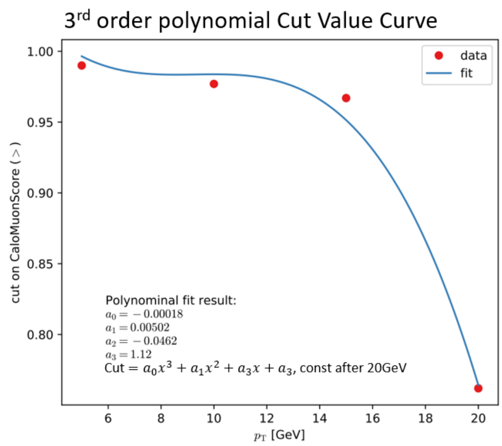

Fig. 6. Above shows the cut value curve that are currently using. It consist of a 3rd-order polynomial function at low pT and a constant value starting at 20GeV.

D. Performance of the Algorithms

To better compare the performance of the algorithms. Two groups of Monte Carlo Simulation results are prepared. They are ttbar sample and Zmumu sample which consists of 1.0×10^8 and 2.8×10^8 events respectively. The ε_{sig} and ε_{bkg} curve of algorithms being applied to both samples are shown in Figure 7 and 8.

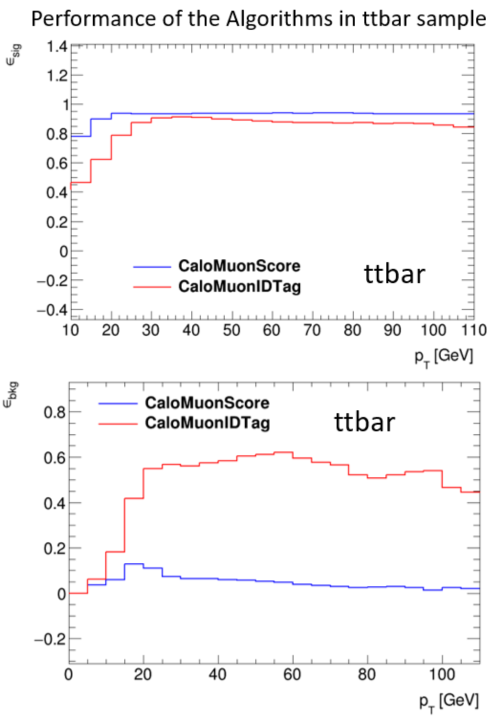

Fig. 7. The ε_{sig} and ε_{bkg} of CaloMuonScore and CaloMuonIDTag when applied to ttbar sample.

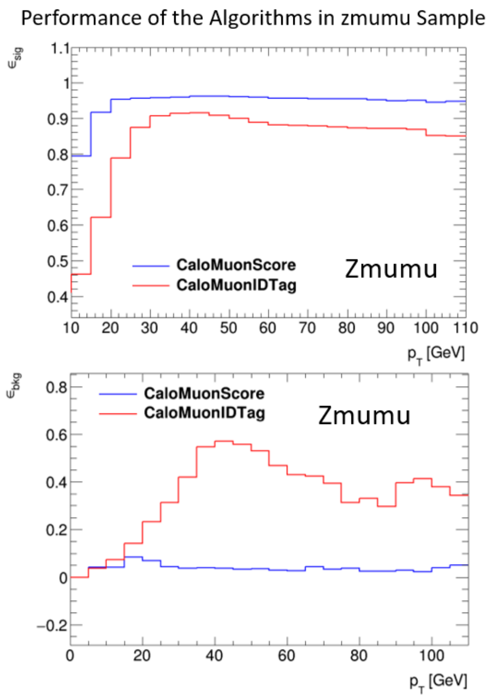

Fig. 8. The ε_{sig} and ε_{bkg} of CaloMuonScore and CaloMuonIDTag when applied to zmumu sample.

From the plot, clearly CaloMuonScore outperforms CaloMuonIDTag with a significantly lower ε_{bkg} and higher ε_{sig} in all pT region. This is expected since CaloMuonIDTag only takes the total energy deposit in 1 calorimeter layer as input while CaloMuonScore can access to the energy deposit profile of 7 layers.

E. Goal of the Study

While CaloMuonScore outperforms CaloMuonIDTag, a few problems in the performance of CaloMuonScore need to be further investigated before it can be used. Figure 9 show with a smaller pT bin plot of the performance of CaloMuonScore on the ttbar and zmumu events. There are two problems worth investigating.

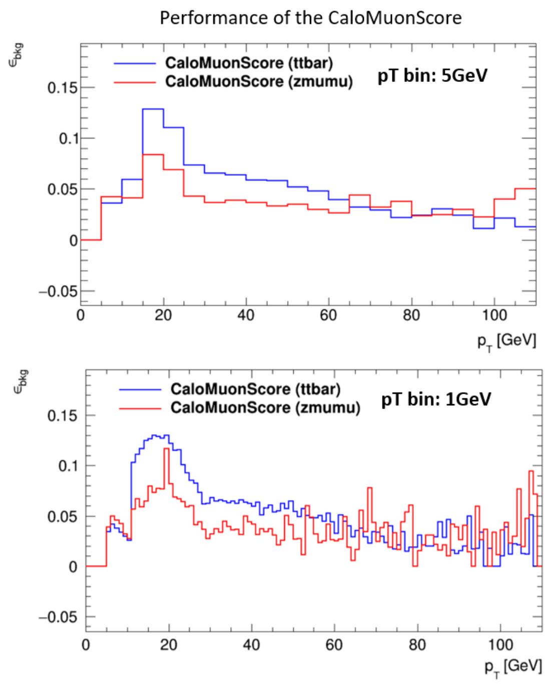

Fig. 9. The ε_{sig} and ε_{bkg} curve of CaloMuonScore when applied to both sample shown in 1GeV and 5GeV bin. Even at 5GeV pT bin size, the background efficiency showed a peak around 20GeV. In 1GeV pT bin size, the fluctuation at high pT is significant.

First, the background efficiency is not a constant value of 6.8% at low pT, there is a 10% difference between ε_{bkg} at 10GeV and 20GeV. This is not ideal since a pT dependent ε_{bkg} will create bias when those misidentified muons are used in analysis. Secondly, the ε_{bkg} shows huge fluctuation at high pT and the reason is unknown. So, the goal of this study is to find a cut value curve that achieves a flat fake rate of 6.8% and find out the cause of the fluctuation at high pT.

II. METHOD

A. Concept

To produce a pT-depending cut value curve that can achieve a flat fake rate. One needs to consider the intersection point between multiple constant cut value background efficiency curves and a horizontal line in the background efficiency plot representing the desired background efficiency.

In Figure 10, it shows the background efficiency of a ttbar sample achieved by CaloMuonScore when the cut values are constant value of 0.9, 0.95, and 0.99. The background efficiency curve is obtained by linear joining the point from the background efficiency histograms. So, the separation between points in the background efficiency curve is the same as the bin size of the histogram, which is 1GeV. Notice that when the cut value increases, the background efficiency curve drops, which is expected since increasing the cut value will make less muon able to pass through the cut value threshold, and less fake will be misidentified as muon. For example, by considering the seven red-marked intersections between the three background efficiency curves and the desired background efficiency horizontal line, the seven CutValue-pT pair can be obtained. At pT=10.8, 10.9, 11 GeV, the cut value should be 0.9, 0.95, and 0.99. Then the cut value at pT=17, 22, 25 GeV should be 0.99, 0.95, and 0.9 respectively. Following these seven CutValue-pT pair, a cut value against pT curve can be obtained by linear joining the six points. This cut value curve, shown in Figure 11, will theoretically give a constant background efficiency when it is applied to the ttbar sample.

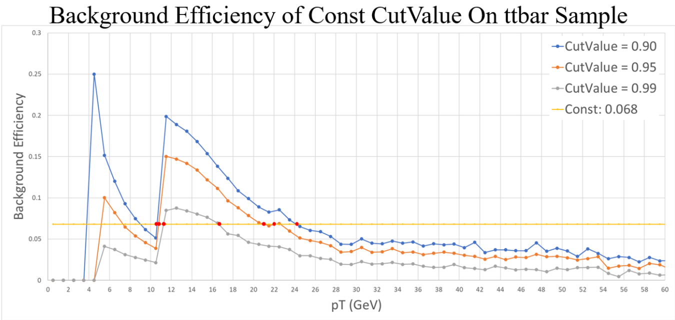

Fig. 10. The ε_{bkg} curve of CaloMuonScore when a constant cut value is applied. The peaks at low pT indicate that most fake muon have pT<20GeV.

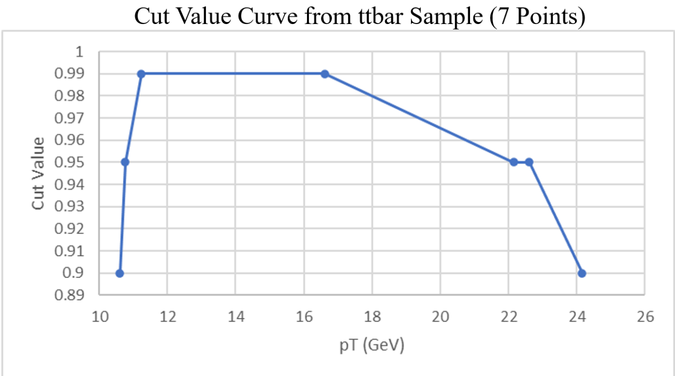

Fig. 11. The cut value curve obtained from ttbar sample by only considering seven interception point from three constant cut value ε_{bkg} curve.

To get the full cut value curve, the background efficiency curves of constant cut value are produced from 0.3 to 0.99 in the step of 0.01 and from 0.99 to 0.999 in the step of 0.001, instead of just the three curves in Figure 10. Less separation between curves will give more intersection and thus give a denser cut value curve.

B. Package and Coding

The package called rel22validation developed by Tong Pan is used as the bases of the code. Rel22validation will loop over the sample ntuples and select the muon that is passing the CaloMuonScore. The package is modified to produce background efficiency plots with a constant cut value. Then, a separate Python code is written to calculate all the intersection points and produce a cut value curve. The resulting cut value curve obtained from the ttbar sample is shown in Figure 12. Note that the cut value curve at high pT has a lot of fluctuation and this will be explained in Section III C.

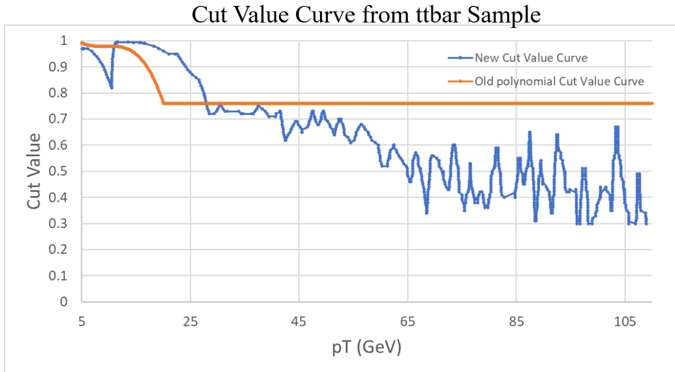

Fig. 12. The new cut value curve obtained from ttbar sample and the old 3rd-order polynomial curve. They shows very different shapes.

III. RESULT and ANALYSIS

A. Result on ttbar Sample

The result of applying the cut value curve obtained from ttbar sample to the ttbar sample is shown in Figure 13.

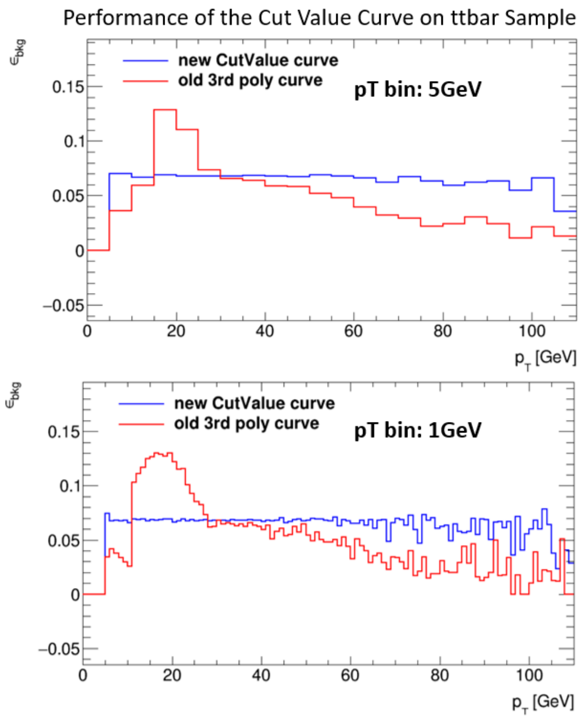

Fig. 13. The ε_{bkg} performance of the two cut value curve when applied on the ttbar sample. The same histogram are shown in two different bin size, 5GeV (top) and 1GeV (bottom).

Not surprisingly, the background efficiency is much more flatted than the old 3rd-order polynomial cut value curve. Even though the fluctuation in background efficiency at high pT remains, the average of the background efficiency stays static at 6.8%, like considering 5GeV pT bin. This shows that the code and the idea are working properly.

B. Result on zmumu Sample

The result of applying the cut value curve obtained from ttbar sample to the zmumu sample is shown in Figure 14. Surprisingly, the background efficiency of the new cut value curve shows an even larger fluctuation than the old 3rd-order polynomial cut value curve.

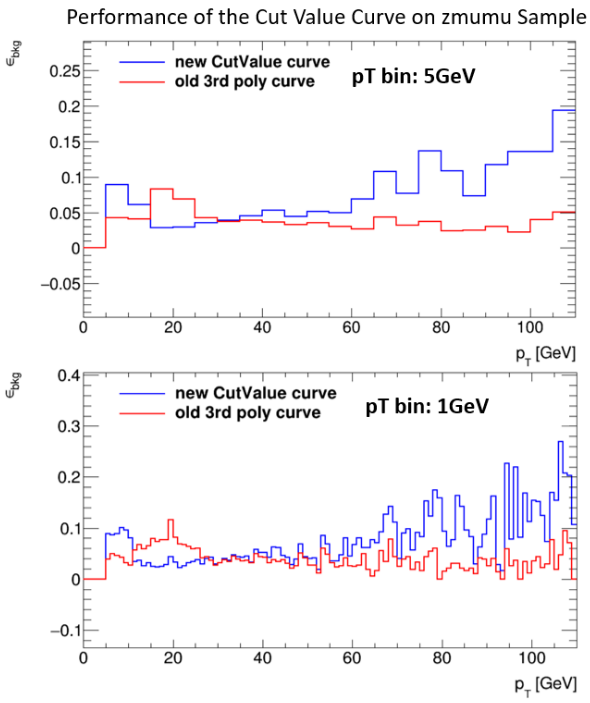

Fig. 14. The ε_{bkg} performance of the two cut value curve when applied on the zmumu sample. The same histogram are shown in two different bin size, 5GeV (top) and 1GeV (bottom).

To understand the reason behind, the pT distribution of fake muon in ttbar and zmumu sample needs to be considered, shown in Figure 15. Note that the pT distribution of fake muon in both events is very different. Fake muon in zmumu sample is more concentrated at low pT while fake muon in ttbar sample is more widespread. Since the basic idea of having a pT-depending cut value is that some pT has more fake muon so the cut value needs to be increased to achieve a flat background efficiency, when two samples have different pT distribution of fake muon, it naturally requires two different cut value curve to achieve flat background efficiency. So, the old 3rd-order polynomial curve should not be used as a cut value of CaloMuonScore for every event. For example, if the old 3rd-order polynomial curve is applied to a sample that the fake muon is concentrated in pT equal to 30GeV, then there will be a huge difference between background efficiency at 10GeV and 30GeV since the cut value curve create a tighter selection at 10GeV and a looser selection at 30GeV. Instead, the cut value curve needs to be produced according to the background event that the particle is coming from.

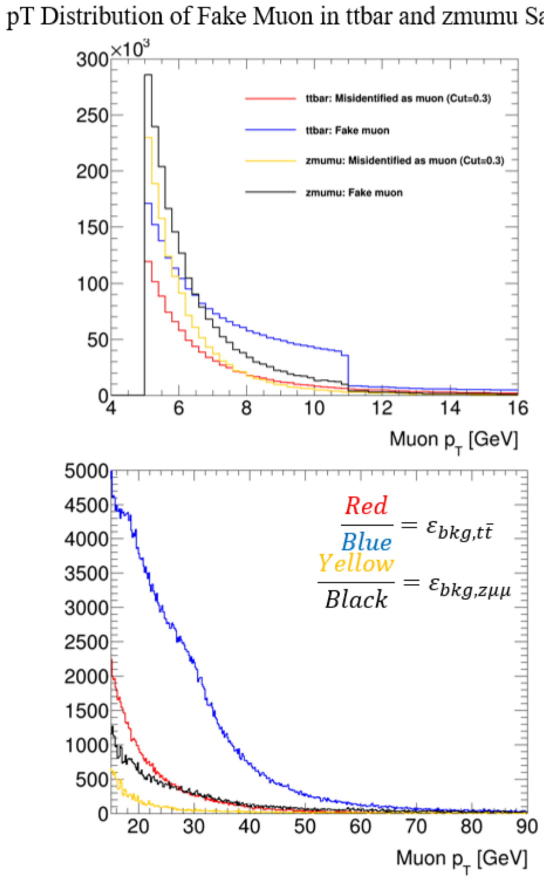

Fig. 15. The pT distribution of the number of fake muon in ttbar and zmumu sample. As well as the amount of fake muon that is misidentified as muon by CaloMuonScore when the cut value is constant 0.3. Note that the distribution of low pT (top) and high pT (bottom) are shown separately because the distribution at low pT have a much larger scale than that of high pT.

C. High pT Fluctuation

The causes of the high pT fluctuation can be known by looking at the pT distribution of fake muon in ttbar and zmumu sample (Figure 15). At high pT region, there are only a few hundred muons in each pT bin and among them, only a few tens of them are misidentified. So, there are not enough statistics to provide a meaningful background efficiency. That is why the fluctuation in high pT is less when a larger pT bin is considered. When a larger pT bin is considered, more muon is grouped in the same bin and making it statistically a meaningful background efficiency.

To confirm the effect is statistical fluctuation, a convergence test is performed by obtaining the cut value curve when only a fraction of the events are run. The result is shown in Figure 16. Even when only running 10% of the event, the cut value curve at low pT stays the same. However, the curve at high pT changes a lot even when 75% of the event is run. This shows that the fluctuation of high pT is caused by an inadequate amount of sample.

The most direct solution of this would be using a sample with more high pT particle. However the sample is generated by Monte Carlo Simulation and it is very computational costly.

So, a alternative solution is to increase the bin size for high pT in ε_{bkg} curve when calculating the interception point. The cut value curve obtained using a bin of 2GeV at high pT is shown in Figure 17. So, the number of particles in each bin will be double and the curve is much smoother. The final cut value curve will be using the cut value curve from 1GeV pT bin for pT<22 and using the cut value curve from 2GeV pT bin for pT>22.

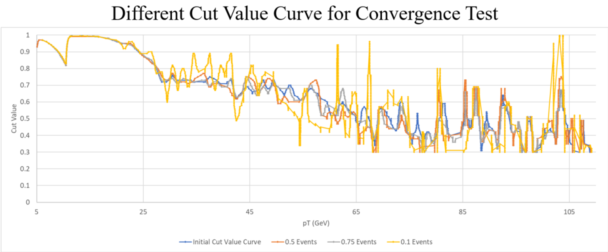

Fig. 16. Different cut value curve obtained from the ttbar sample when only running a fraction of events.

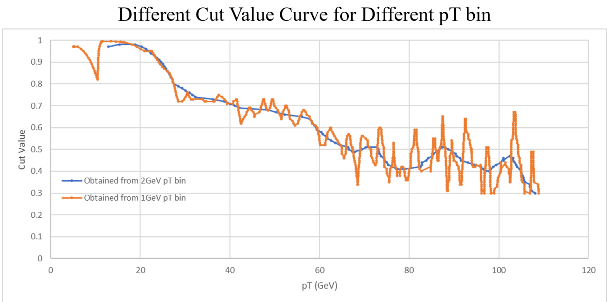

Fig. 17. Different cut value curve obtained from the ttbar sample when different pT bin size of ε_{bkg} curve is used in calculating interception.

IV. CONCLUSION

Using the old 3rd-order polynomial cut value curve for analysis is not ideal as it may introduce inconsistent background efficiency and causes bias. Instead, a program is created to obtain a cut value curve which will give a constant background efficiency depending on the background event. So, the cut value curve of different events can be precalculated based on the pT distribution of fake muon such that different cut value curves can be chosen depending on the origin of the suspected muon. And a flat background efficiency performance of CaloMuonScore can be achieved.

Future work will include find out what causes the huge drop of fake muon at 11GeV in the ttbar and zmumu sameple (Figure 15). Also, the code need to be modify so that the bin size will change according to the number of fake muon in order to smooth the high pT region. And also how to achieve this in actual event when the number of fake muon is not known. Lastly, is to study the change in signal efficiency to know is it worth to keep the background efficiency at 6.8%. Maybe using a larger background efficiency will give a huge increase in the signal efficiency.

ACKNOWLEDGMENT

This work will not be possible without Tong Pan support, such as providing reading materials and package as the bases of this work. Also, special thanks to Luis for supervising my work and CERN for giving me such great opportunities and valuable experiences.

REFERENCES

[1] R. Wölker, "Search for Heavy ZZ Resonances in the 4 Final State with the ATLAS Detector and Identifying Muons in the ATLAS Calorimeter using Machine Learning", *Proefschrift, University of Oxford*, 2021.

[2] CERN, "ATLAS muon spectrometer: Technical Design Report", *Technical Design Report ATLAS*, 1997.

[3] G. O. Sanz, "Muon Identification in the ATLAS Calorimeters", *Proefschrift, Radboud University of Nijmegen*, 2009.