Internship report

PSA Master 2

Timing Studies in Belle II Vertex Detector

Yitong Liu
Supervisor: **Jérôme Baudot**
Hubert Curien Multi-disciplinary Institute (IPHC)

20 Feb. 2023 - 13 June 2023

Abstract

In this master thesis, we will study the strategies related to the timing information at Belle-II to cope with the high-luminosity condition. We will first study how the reduction of signal waveform samples will affect the performance of two different algorithms. Then, we will present the initial time delay study on a prototype CMOS sensor developed at Strasbourg.

Contents

Introduction

Belle II is an particle collider experiment at the SuperKEKB machine in Tsukuba, Japan. The experiment observes e^+ e^- collisions at the center of mass of energy of the Υ (4S), where a large of B mesons are produced in order to make precise measurements to reveal beyond-standard-model physics [1]. The SuperKEKB + Belle II program is built on the the success of the previous KEKB + Belle project, but with a luminosity increased between one to two orders of magnitude. The luminosity measures the amount of collisions happening per unit of cross section and time, and the high luminosity is the key to discover new physics events. For this reason, the detector of Belle II is designed to reach the peak luminosity of $6 \times 10^{35} cm^{-2} s^{-1}$ and to record an integrated luminosity of 50 ab^{-1} data in the next decade. Belle II collects its first data in 2019 as shown in the operation plan in Figure 2, and reached the highest instantaneous luminosity in the world, $4.7 \times 10^{34} cm^{-2} s^{-1}$, in June 2022 during the first phase run I from $2019 - 2022$. The plan is to increase this instantaneous luminosity by up to a factor of 6 in the next 5 years. Currently, Belle II is under a one-year (LS1) shutdown for a detector upgrade, and the Run II $(2023 - 2027)$ after this upgrade will be expected to record a dataset equivalent or close to 10 ab^{-1}. Another upgrade LS2 of Belle II sub-detectors is foreseen to happen around 2027 [2].

The high luminosity will lead to challenging data-taking conditions for the detector. The peak luminosity rise after LS2 will greatly increase beam-induced backgrounds. This will result in the Silicon Vertex Detector (SVD), using strip detectors, approaching the occupancy limit around 3%. The 3% occupancy will saturate the bandwidth of current acquisition system which

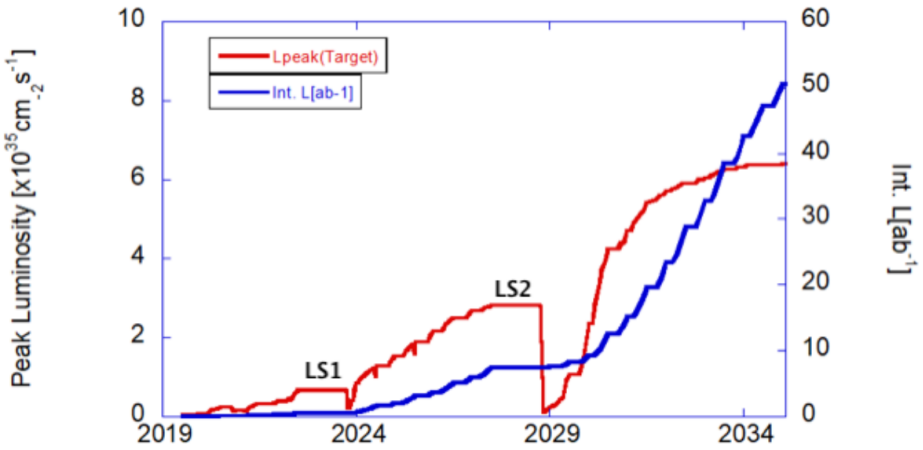

Figure 2: Operation plan as presented at the June 2022 Belle II general meeting.[3]

may cause the problem that only partial data is collected [3]. High background condition will also cause difficulties in distinguishing signals from the background. To cope with these problems anticipated with the high luminosity condition, two major strategies are proposed. One is to reduce the bandwidth of the data by changing the data format to a lighter version (from 6-sample to 3-sample acquisition). The other one is to replace the current sensors equipping the vertex detector (VXD) with high granularity monolithic CMOS pixel sensors to drastically reduce the occupancy, which will alleviate both the data throughput and background issues. This master thesis presents the preliminary study on comparing 3-sample and 6-sample acquisition modes using different algorithms as well as studying the timing information in CMOS pixel sensors.

Acknowledgements

I would like to thank Professor Jérôme Baudot for giving me the opportunity to explore new topics in the physics research, and IPHC to give me the financial funding for my master thesis. For the SVD part, I would especially thank Christian Finck for his help on the code debugging and explanations on data monitor system at Belle-II, and Varghese Babu for his help on ROC curve study. For the CMOS sensor part, I would particularly thank Ajit Kumar for his cooperation on helping me with the MIMOSIS-Fast test, and Andrei Dorokhov's kindness and patience in helping us to find the optimal DAQ values and explaining things, Matthieu Specht and Mathieu Goffe for setting up the lab equipment. I am grateful to everyone's generous help in the Belle-II and PICSEL group, also the generous and beautiful Strasbourg sunshine during the last few weeks of my internship.

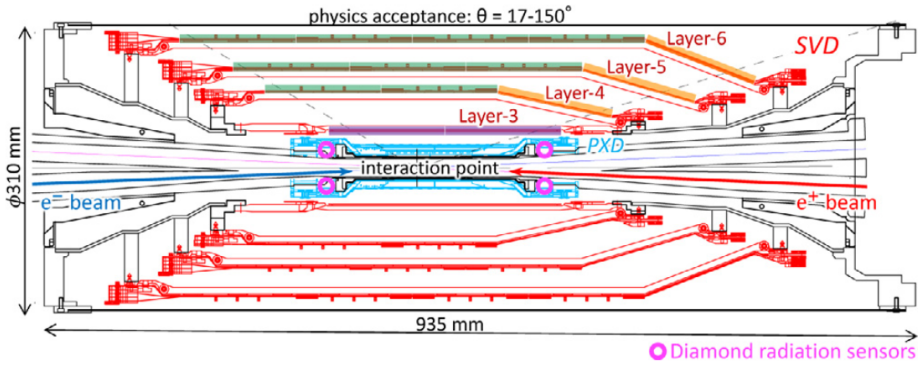

Figure 3: Schematic of the Belle-II vertex detector (VXD), the PXD layer is blue, and the four outer layers with red marks are the Silicon Vertex Detector (SVD).[2]

1 Belle II SVD

The Silicon Vertex Detector (SVD) is the outer part of the innermost Belle II detector, Vertex Detector (VXD). The two innermost layers, Layer-1 and Layer-2 (in blue) are the Pixel Detector (PXD) using the DEPFET technology. The Silicon Vertex Detector (SVD) is the outer four layers, Layer-3 to Layer-6 (in red) of the VXD (Figure 3). The SVD is important for track extrapolation from the main Belle-II tracking device (a drift chamber) to the PXD. It is also important for track reconstruction for low momentum charged particles and decay vertex reconstructions [2].

The current SVD is made of the double-sided silicon strip detectors (DSSDs) for precise tracking information while maintaining material budget. The strip width is down to 50μm with orthogonal p^+ and n^+ strips (referring to thin implants with higher doping than the 300m thick low-doped sensitive volume) on the two sides of the sensor. When a charged particle crossing the bulk of the sensor, it generates typically 24000 electron-hole pairs, and the holes will drift towards the p^+ strip while the electrons to the n^+ strip. The induced current by electrons and holes will be further measured in the readout electronics [1].

However, the problem of using strip detector is that the number of readout channels is low and generate high occupancy. The anticipated SVD upgrade into pixel sensors is explained in Section 3.

1.1 Data acquisition system (DAQ)

The bunch crossing at SuperKEKB is up to every 4ns, but for most of the cases, either no collision happens or no interesting results happened during the collisions. Keeping record of the results from each collision would be wasteful and impossible due to the bandwidth restriction. Therefore, the process of triggering and filtering events is crucial before storing the data.

With the increasing luminosity, the trigger rate is expected to approach 20kHz, and the Belle II data acquisition (DAQ) needs to handle 30kHz trigger rate [4]. The simplified diagram of the Belle II DAQ system is shown in Figure 4. For the topic of this study, we will only focus on the SVD DAQ chain. The trigger and clock signals are sent to the detector front-end electronics (FEE) boards. After receiving the trigger, the digitized signal output from the FEE board will be transferred to a unified readout board, COmmon Pipelined Platform for Electronics Readout (COPPER). Then, the data will be further collected by a readout PC and sent to High Level

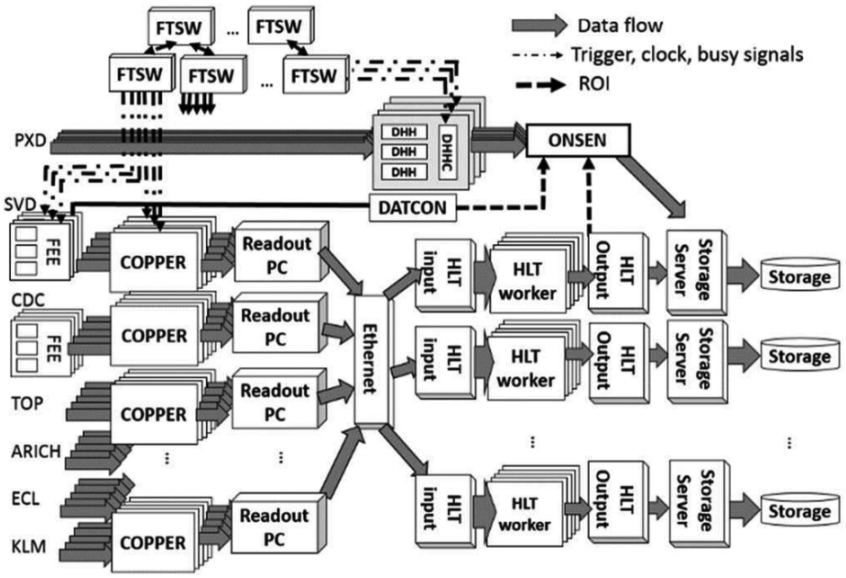

Figure 4: Data Acquisition System for the Belle II[4]

Trigger (HLT) via Ethernet. In the HLT step, the events will be fully reconstructed and further selected down to a factor of 5 compared to the initial trigger.

The front-end Application-Specific Integrated Circuits (ASIC), the APV25, is an important part in the DAQ system. It samples the data with the multi-peak mode shown in Figure 5. For the SVD, the chip samples the signal waveform with a 32 MHz sampling clock and stores each sample in a ring buffer. After a trigger is decided, the chip will go back in time to search for the samples stored in the buffer that match the trigger time [4]. Currently, the SVD is using 6-sample readout mode that the APV25 reads out 6 consecutive samples stored in the buffer to cover the peak of the signal. With the increasing luminosity condition in the future, this may cause a problem that too many strips are fired at the same period of time which leads to bandwidth saturation issue. Therefore, three/six-mixed acquisition mode from APV25 readout is proposed. Depending on the timing precision of the Level-1 trigger signal, the chip will read out either 6-sample or 3-sample mode. The 3 successive samples will be chosen from the 6-sample that contains the maximum information about the signal pulse. If a precise trigger timing is available, the 3-sample mode will be used, otherwise, the 6-sample readout mode will be used [2].

1.2 SVD cluster time reconstruction

From the APV 25 readout, 6-sample or 3-sample, we can extract useful physics parameters from the signal waveform for further data monitoring. The cluster time, measuring the difference between the hit time and the trigger signal, is important for background rejections. When a charged particle transverses the strip detector, the clusters will be formed in SVD strips. All the strips within the cluster will be summed sample by sample. Then, the reconstruction algorithms will extract information from these summed samples to calculate the cluster time. There are multiple ways of doing cluster time reconstruction in SVD. In this master thesis, we will focus

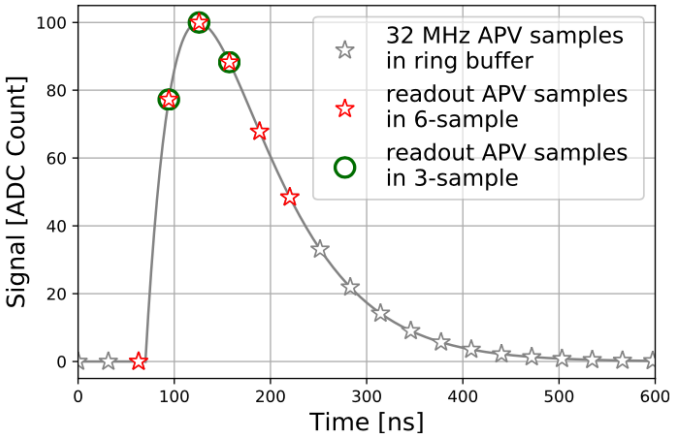

Figure 5: Example of sampling in "multi-peak" mode of the APV25. The gray curve is the signal waveform. The grey stars represent the sampled signal stored in the ring buffer using 32 MHz sampling clock. The red stars are the 6 consecutive samples read out, and the red stars with green circle are the 3 consecutive samples chosen from the 6-sample mode with a relative shift[2]

on the two main algorithms named CoG3 and ELS3, both applied to the 6-sample and 3-sample acquisition modes.

CoG3 uses center of gravity to calculate the cluster time. It chooses 3 best summed consecutive samples that give the highest sum. And then, the cluster time is calculated as:

$$t_{cluster} = 31.44ns \frac{\sum_{n=0}^{n<3} n a_n}{\sum_{n=0}^{n<3} a_n} \tag{1}$$

where $31.44ns$ is the sampling period of the APV readout chip and a_n is the amplitude of the nth sample [5].

For the ELS3 algorithm, the way of choosing the 3 best summed consecutive samples is the same as in CoG3. But instead of using center of gravity algorithm, it uses an analytical equation (shown in equ.2) which is the estimation of the time from a chi-squared fit of the theoretical CR-RC waveform, to calculate the cluster time [5]:

$$t_{cluster} = -31.44ns \frac{2e_\tau^4 + \omega e_\tau^2}{1 - e_\tau^4 - \omega(2 + e_\tau^2)} \tag{2}$$

1.2.1 6-sample mode

The CoG3 6-sample is currently chosen as the default algorithm. However, some recent studies have shown that ELS3 might give a better performance for the future high luminosity condition. The initial cluster time result on comparing the 6-sample using CoG3 and ELS3 algorithms is shown in Figure 6. All the tests shown in this master thesis are using the data at Belle II on low-background condition collected in 2021 (Experiment 21, Run 234). The x-axis is the cluster time (ns) is with respect to the trigger time. The signal distribution is peaked at around zero since the trigger time is tuned in favor for of the signal. The background arrival time is not synchronous to the trigger time. Therefore, the waveform for the real background hits are not

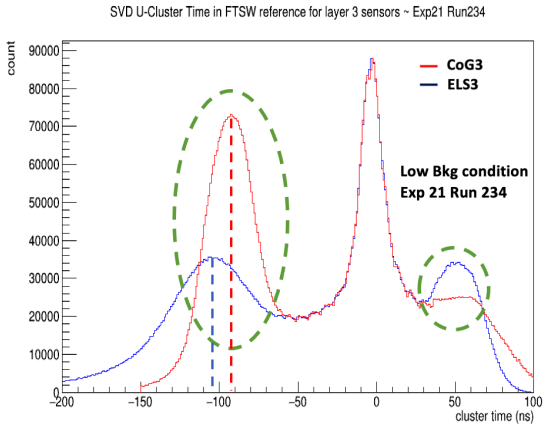

Figure 6: The 6-sample DAQ: CoG3 vs. ELS3 tested on low background condition data collected in 2021. The red line is the 6-sample readout mode using GoG3 algorithm for cluster time reconstruction, the blue line using ELS3 algorithm. The two green circles represent the beam-induced background peaked at the positive and negative sides.

like in Figure 5, but can be a cut-off anywhere earlier or later compared to this waveform. This distorted time estimation will lead to the peaks in background hits as in Figure 6. In the cluster time plot, the blue curve corresponds to the ELS3 algorithm with 6-sample readout mode, and the red curve is the CoG3 algorithm with 6-sample. For both algorithms, the signal distribution has similar mean (around zero) and width. For the background hits on the negative side, the CoG3 has a higher peak that centered around -90ns with a short tail, while ELS3 gives a lower peak centered around -105 ns, but with a longer tail. For the background in the positive side, the CoG3 gives a flat distribution while ELS3 gives a peak around 50 ns. Just from the cluster time graph of the two algorithms, it is hard to determine which algorithm gives a better signal-background separation and a better performance. An advanced study on the receiver operation characteristic (ROC) curve on signal efficiency, background rejection and purity comparison with different time window cuts is shown in the Section 2.

1.2.2 3-sample emulation

Since Belle II does not have real data collected with only 3-sample readout mode, we need to emulate the 3-sample case from the real data collected using the 6-sample DAQ. The emulation process is shown in Figure 7 [6]. The first event in the 3-sample mode is chosen from the 6-sample event by doing a relative shift, and the other two are chosen successively following the first one. The current relative shift is set to be 7 ns because these chosen 3-samples will best cover the peak of the signal and also best match the 6-sample calibration.

Going from 6 to 3-sample, a worse performance in signal and background separation is expected, since we are extracting less information from the pulse distribution. The comparison between 6-sample and 3-sample mode using GoG3 algorithm is shown in Figure 8. Compared to the 6-sample case, the peak of the background increase from -90ns to -45ns, and the tail of the background merges with the signal distribution centered at zero which creates a higher peak around zero. The background in the positive side completely merges with the signal distribution for the 3-sample case. It clearly shows that it is more difficult to separate signal from the background using 3-sample emulation because the background distribution shifts towards the signal distribution.

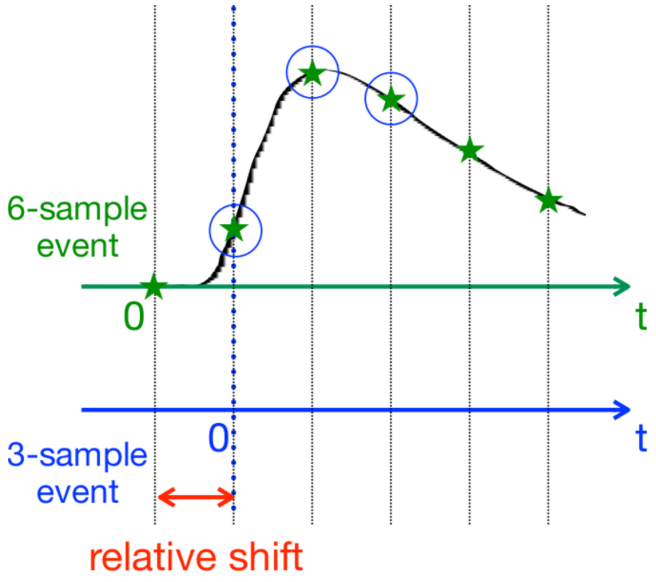

Figure 7: 3-sample emulation from 6-sample DAQ. Green stars are the 6-sample. Green stars with blue circles are the 3-sample chosen from the 6-sample with a relative shift[6]

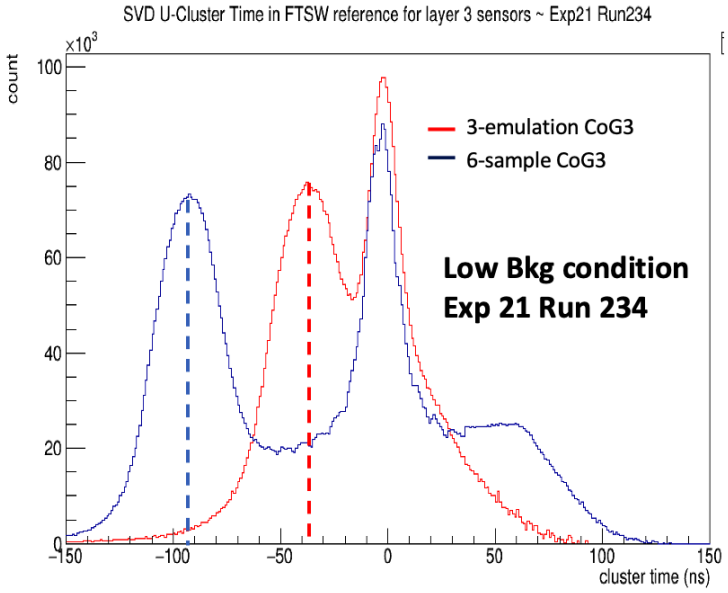

Figure 8: 3-sample emulation vs. 6-sample using CoG3 algorithm. Red curve is the CoG3 with 3-sample emulation mode; Blue curve is the CoG3 with 6-sample mode

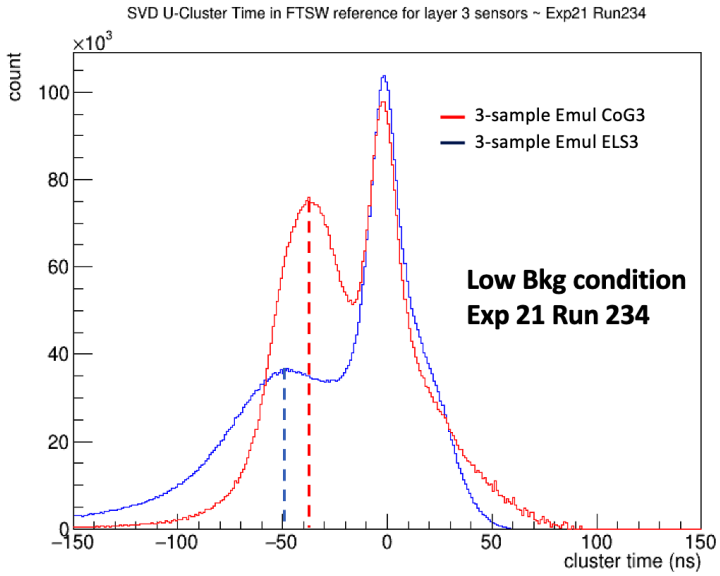

Figure 9: 3-sample emulation result on comparing CoG3 and ELS3. Red curve: 3-sample emulation using CoG3; Blue curve: 3-sample emulation using ELS3.

The comparison between CoG3 and ELS3 with 3-sample mode is shown in Figure 9. The ELS3 3-sample emulation also shifts the background distribution towards the signal distribution compared with the ELS3 6-sample distribution in Figure 6. The peak of the background distribution increases form -105ns to -50ns. From the two distributions, CoG3 and ELS3 using 3-sample emulation, it is hard to determine which algorithm gives a better signal background separation with the 3-sample mode. A further study is also needed as indicated by the result from the 6-sample.

2 Efficiency-purity ROC curve study

The Receiver Operating Characteristic (ROC) curve is often used to compare the algorithm performance quantitatively with different threshold cuts. In order to compare the background rejection between CoG3 and ELS3, 6-sample and 3-sample emulation, we need to do a different time window cut and compare the performance of different situations by constructing the ROC curve. In the ROC curve test, four variables, S_{total}, B_{total}, S_{sel} and B_{sel} are needed to obtain from the cluster plot:

S_{total} = Total number of signal hits without any constraints
B_{total} = Total number of background hits without any constraints
S_{sel} = Number of signal hits selected given a certain time cut
B_{sel} = Number of background hits selected given a certain time cut

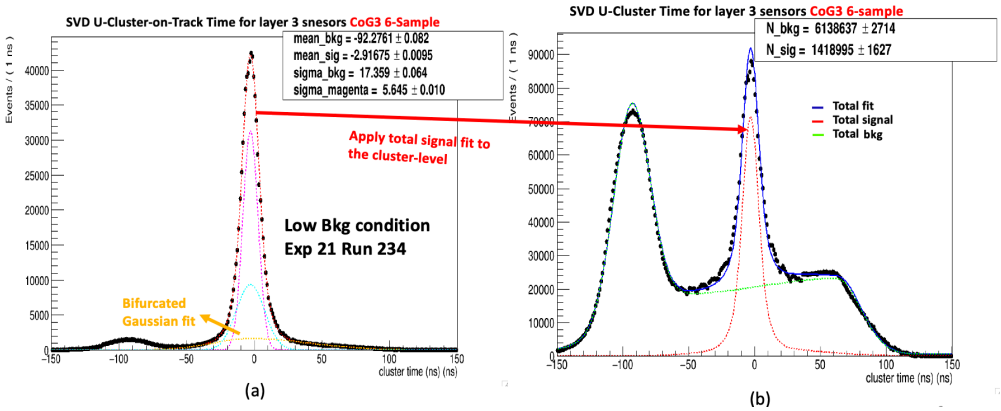

Figure 10: The total Gaussian fit to the 6-sample CoG3 in the cluster-time plot. The plot (a) determines the Gaussian fit to the signal distribution. The black dots are the real data; the magenta and blue lines are two Gaussian fits; the yellow line is the asymmetric Gaussian fit; the red line is the final fit to the signal distribution, the sum of all three Gaussian fits. The total signal fit shape parameters will be fixed in the cluster time plot (b). The fit to the background distribution is separated into two parts, one with the negative peak and another one with the positive peak. Each background is also fitted with one Gaussian and one asymmetric Gaussian fit. The summed background fit is shown as the green curve in plot (b). The overall fit, signal plus background, is shown as the blue curve in plot (b).

Then from these four variables, we can construct different figures of merit to construct the ROC curve. In this study, we will focus on the following three parameters:

$$
\begin{aligned}
S_{efficiency} &= \frac{S_{sel}}{S_{total}} \\
B_{rejection} &= 1 - \frac{B_{sel}}{B_{total}} \\
Purity &= \frac{S_{sel}}{S_{sel} + B_{sel}}
\end{aligned}
\tag{3}
$$

The $S_{efficiency}$ gives the percentage of the signal events that are selected within a specific time constraint over the total signal events. The $B_{rejection}$ gives the percentage of background events that are not selected within a specific time constraint. The purity measures the fraction of signal events over total events within a time cut. Then we can build $S_{efficiency}$ vs. $B_{rejection}$ and *purity* vs. $S_{efficiency}$ curves with different time cuts to compare the performance of different algorithms and sampling modes.

2.1 Gaussian fit to the cluster time plot

The values of the variables, S_{total}, B_{total}, S_{sel} and B_{sel}, cannot be obtained directly from the real data collected at Belle II. In order to estimate these values, we need to fit the signal and background distributions separately in the cluster time plot. The Gaussian fits to 6-sample using CoG3 algorithm is shown in Figure 10. In order to simplify the fitting process, we need to first fix the shape parameters for the signal distribution. The Gaussian fits to the signal curve is obtained from the track-level cluster time plot (shown in Figure 10 (a)). The events in the track-level are from the clusters associated to a track. These cluster have passed some quality criteria, so the majority is signal, with very low background probability. The overall signal fit (red line) uses two Gaussian and one asymmetric Gaussian fit. For each Gaussian fit,

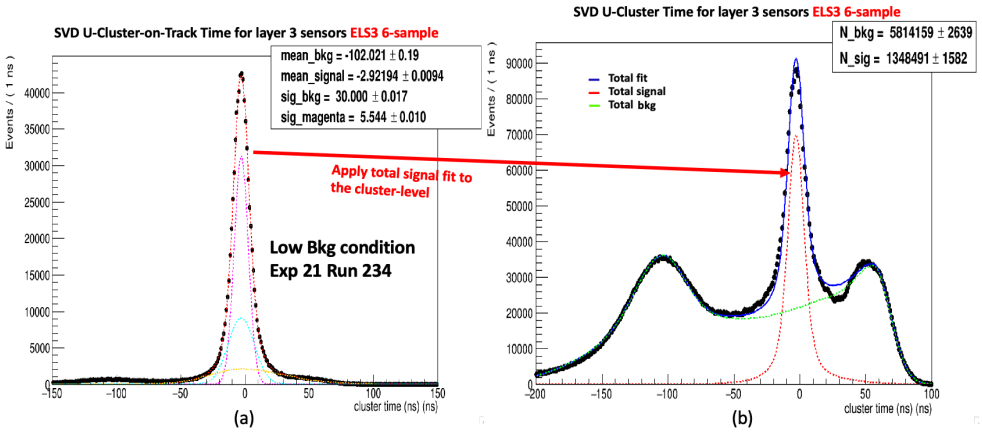

Figure 11: The total Gaussian fit to the 6-sample using ELS3 algorithm in the cluster-time plot.

there are three free parameters, mean, sigma and the fraction of this Gaussian fit, and four for asymmetric Gaussian, with the sigma divided into left and right parts. Then, we can fix the shape parameters, mean and width, for the signal distribution in the general cluster time plot (plot (b)). With the fixed signal distribution parameter shape, we constructed the background fit (green curve) shown in Figure 10 (b). Since there are two peaks for the background distribution, one on the negative side and another one on the positive side, we will fit each part separately. Each of the background distribution is fitted with one Gaussian and one asymmetric Gaussian. The overall background fit is shown as the Green curve in plot (b). Combining the signal fit (red curve) and the background fit (green curve), the final fit model to the cluster time distribution on the cluster-level is shown as the blue curve in plot (b).

The total Gaussian fit to the cluster-level plot for the other case, 6-sample using ELS3 (Figure 11), follows the same procedure as described in the 6-sample using CoG3 case, but tuning the means and widths of the Gaussian distributions to match the distribution. The 3-sample emulation using CoG3 and ELS3 (Figure 12 and 13) also follow the same procedure. From the total fit to the signal and background distribution in cluster time plot, we will get the total number of signal and background automatically from the Gaussian fits which is shown in the box on the right corner in each plot (b).

2.2 ROC curve performance with general time window cut

After getting the total fit to the distribution in the cluster time plot, we can now perform different time cut to get the different variable values mentioned in equation (3). The Figure 14 (a) shows the $S_{efficiency}$ vs. *time window* curve with a general time window cut from 0 to 150 ns to see the general evolution of the signal efficiency. The $S_{efficiency}$ increases as the time window increases and approaches to 1 since the signal distribution is Gaussian. Different colors of dots represent different test cases. The signal efficiency is defined as the ratio of $S_{selected}/S_{total}$. If the time window (x-axis) equals to 20 ns, the $S_{selected}$ will be the number of signal events that is within ± 10 ns, and the S_{total} is the total signal event taken from the total signal distribution. For each window cut, the ideal is to maximize the $S_{efficiency}$ to avoid signal loss. The error bar of the $S_{efficiency}$ is computed through the error propagation law and will not be mentioned in this test because we are testing over 7 million events, and the error bar is too small to show the significance (around 0.01%). All four cases, 6-sample and 3-sample emulation using CoG3 and

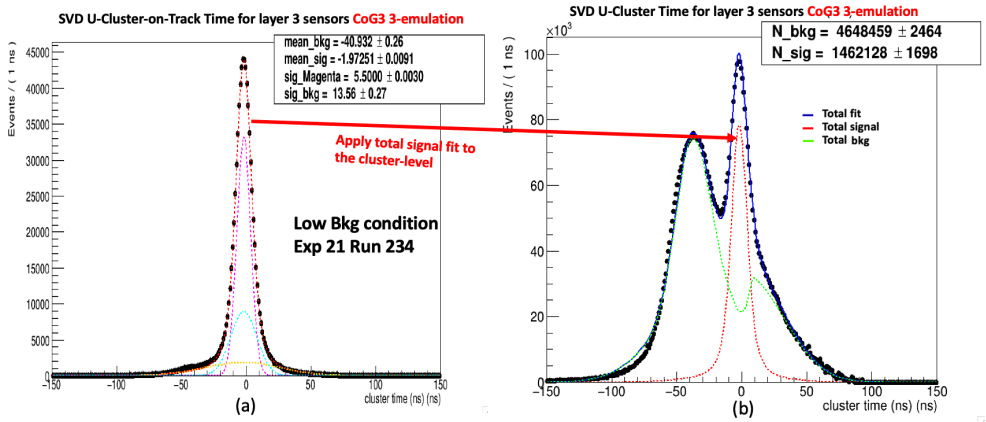

Figure 12: The total Gaussian fit to the 3-sample emulation using CoG3 algorithm in the cluster-time plot

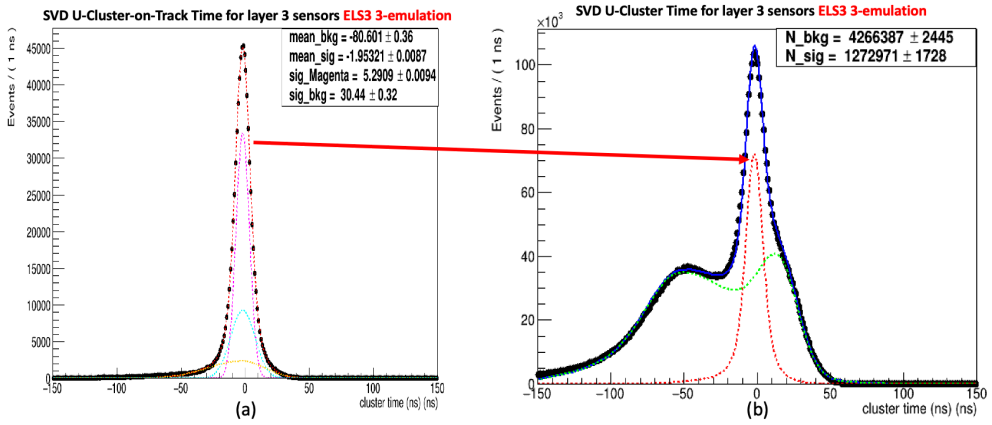

Figure 13: The total Gaussian fit to the 3-sample emulation using ELS3 in the cluster-time plot

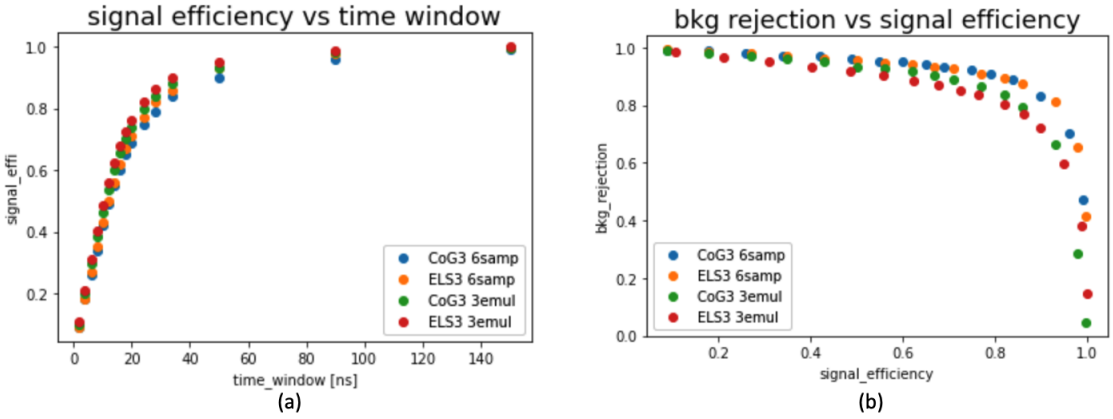

Figure 14: Comparison of four different case 6-sample and 3-sample emulation using CoG3 and ELS3 algorithms for (a) signal efficiency vs. time window (b) background rejection vs. signal efficiency. Different colors of dots represent different cases.

ELS3 algorithms, perform the same in signal efficiency (Figure 14 (a)). No significant difference is observed with different time cuts. This result shows that the mean and the width of the signal distribution are not affected by using different algorithms or sampling modes.

The background rejection power at each signal efficiency level is also important to study the performance. The background rejection decreases as signal efficiency increases, because the time window also increased and more backgrounds are included. The equation (3) indicates that higher $B_{rejection}$ means higher distinguishing power between signal and background. Therefore, the ideal case would be keeping high $S_{efficiency}$ and $B_{rejection}$ for each time window. The Figure 14 (b) shows the $B_{rejection}$ vs. $S_{efficiency}$ curve for the four different situations. The four cases perform about the same at low $S_{efficiency}$ (below 30%). As $S_{efficiency}$ goes higher, the difference in background rejection increases. For $S_{efficiency} > 70\%$, the plot clearly shows that the 6-sampling mode has a higher $B_{rejection}$ than the 3-sample emulation, and no big difference between 6-sample CoG3 and ELS3. For the 3-sample, the ELS3 performs worse than CoG3 at low $S_{efficiency}$, but seems to catch up with the CoG3 at $S_{efficiency} > 80\%$.

The next ROC curve study is *purity* vs. $S_{efficiency}$. The purity also decreases as $S_{efficiency}$ increases with more background events included, but the signal events will not change much after a certain time window. The higher purity of the dataset will be useful for the later physics analysis, but keeping a higher $S_{efficiency}$ will also be essential. As shown in Figure 15, the 3-sample emulation using ELS3 has a significant low purity at low $S_{efficiency}$, while the other three gives about the same purity. However, the 3-sample ELS3 seems to catch up the performance of other cases at high $S_{efficiency}(> 90\%)$. No significant difference between 6-sample CoG3 and ELS3 is observed.

2.3 ROC curve performance with signal efficiency $> 80\%$

The physics analysis usually focuses on the high $S_{efficiency}$ to eliminate the background disruption. A typical choice is $S_{efficiency} > 80\%$. In this section, we will only plot the events that has $S_{efficiency}$ greater than 80% and up to 100%. For different cases, the time cut that gives a $S_{efficiency} = 80\%$ is slightly different, around 28 ns time window ($\pm$14 ns). For the 100% $S_{efficiency}$, the time window for 6-sample is 140ns ($\pm$70 ns), and 84ns ($\pm$42 ns) for 3-sample emulation.

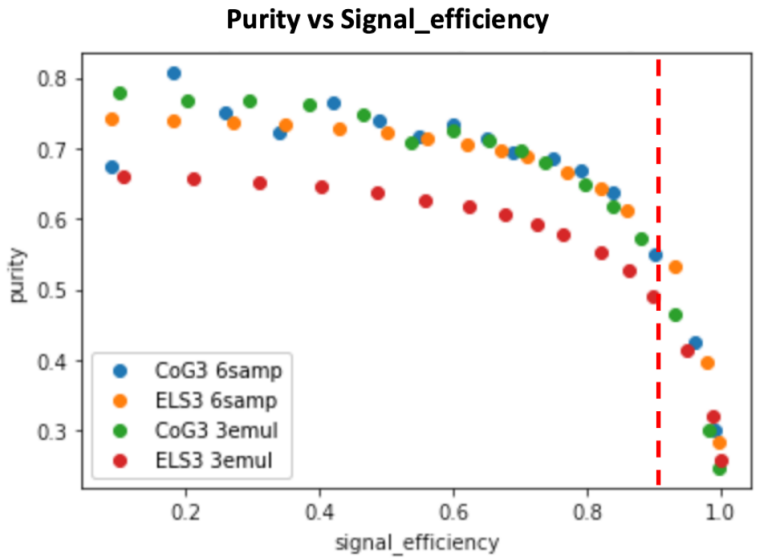

Figure 15: The purity vs. signal efficiency for four different cases

The Figure 16 (a) shows the background rejection vs. signal efficiency curve with the cut $S_{efficiency} >= 80\%$. The 3-sample emulation (green and red) curves gives lower $B_{rejection}$ at each $S_{efficiency}$ compared to the 6-sample DAQ (orange and blue) curves. This is already expected from the cluster time plot in Figure 8. However, this is only the emulation case, the real 3-sample DAQ will have a better time precision, so might still be able to catch up with the 6-sample performance. For the 6-sample, ELS3 gives about the same $B_{rejection}$ power as CoG3 at $S_{efficiency} < 90\%$, but starts to perform better after 90%. At around $S_{efficiency} = 95\%$, the ELS3 is 6% higher than the CoG3 in background rejection. For the 3-sample emulation case, the CoG3 performs better than ELS3 for the range: $80\% < S_{efficiency} < 95\%$. At high $S_{efficiency}$ ($> 95\%$), ELS3 performs slightly better than CoG3 in $B_{rejection}$, about 4% higher at 97.6%. The same behavior is also shown in the purity vs. $S_{efficiency}$ plot (Figure 16 (b)). The 6-sample ELS3 performs marginally better than the 6-sample CoG3 in purity at high $S_{efficiency}$ ($> 90\%$), about 1% at 95% signal efficiency. The 3-sample emulation ELS3 also performs slightly better than 3-sample emulation CoG3 in purity at high $S_{efficiency}$ ($> 95\%$), about 1% at 97.6% signal efficiency.

From the ROC curve study, it is shown that for both sampling mode, 6 and 3, the ELS3 performs better than CoG3, higher background rejection and purity, at high $S_{efficiency}$ ($> 95\%$). However, the difference is marginal, no absolute conclusion can be drawn in this initial study on low background data.

2.4 Purity vs signal efficiency with increasing background

With the increasing luminosity situation in Belle-II in the future, there will be more events "seen" by the detector, but the majority would be backgrounds, and the signal events would stay relatively the same. Therefore, the purity of the data is a big concern. This study tries to mimic the high-background situation by increasing the number of background events artificially in the purity equation (equation 3) and studies how the amount of background events will affect

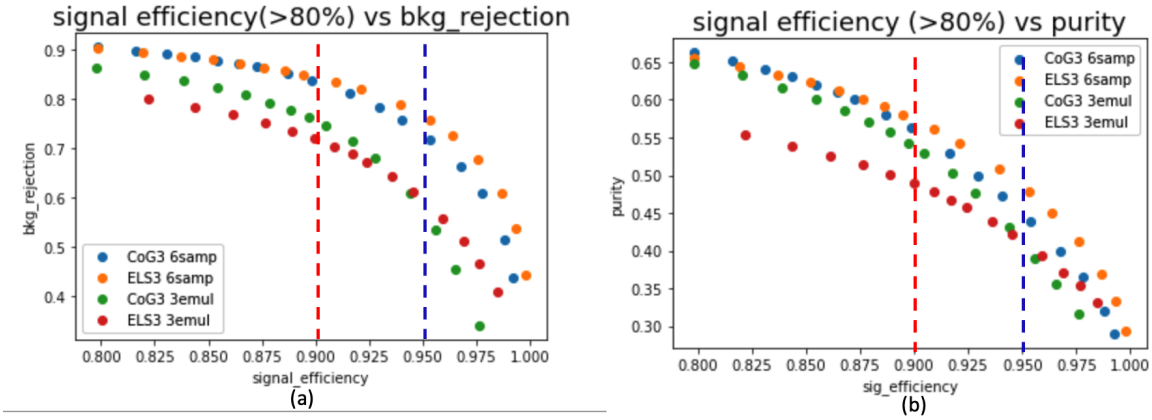

Figure 16: Signal efficiency above 80% for (a) background rejection vs. signal efficiency (b) purity vs. signal efficiency

the purity.

The Figure 17 reflects how the purity would change when we have 3, 5, 8 and 10 times more background situations. It is done on the 6-sample CoG3 case. To mimic the high background situation, we can just add a coefficient, A, in front of the B_{sel} in the purity calculation. The new calculation is: $purity = S_{sel}/(S_{sel} + A * B_{sel})$, where in the plot, A equals to 1 for the original curve (purple). In this purity calculation, S_{sel} stays the same for different coefficients, and the calculation can be simplified as $constant/(constant + A * B_{sel})$ with $A * B_{sel}$ as the only variable. As the $A * B_{sel}$ is way higher than the S_{sel}, we would expect that the purity decreases as $\frac{1}{x}$ distribution that the purity decreases slower and slower as the background increases higher and higher. The Figure 18 shows the trend of the decrement in purity with respect to the increment in background from 1 time to 10 times at 80% signal efficiency. The orange line is the $constant/(constant + x)$ fit to the blue dots with $constant = 2.25$. This results might be used in further tracking performance studies where one can embed in the real data a fraction of background hits, corresponding to the various expected purity estimated here.

3 VTX upgrade

An upgrade of the Belle II vertex Detector (VXD) is proposed to improve the detector performance with the high luminosity condition anticipated after the second upgrade around 2027. The proposal is to replace the whole VXD with a new vertex detector (VTX), composed with 5-layers of Monolithic Active Pixel Sensors (MAPS) based on the CMOS technology [1]. The VTX has much finer space and time granularity than the current sensor used for VXD. The occupancy level with VTX is expected to be below $1e^{-4}$ in any background level conditions, which is improved significantly compared to the current VXD operating occupancy, 1 to $4e^{-2}$ [7]. In the past, the problem with the MAPS is that it has a sequential readout chain which is slow. The new version of MAPS for VTX uses integrated circuits in each pixel that many process can happen at the same time. Instead of the sequential readout, it will be data-driven readout which will be much faster. The proposed MAPS has a 25 to 100ns integration time. The shorter integration time from the pixel sensor can reduce the occupancy and improve the background rejection. In this study, we studied the time integration in another CMOS chip, MIMOSIS-Fast, developed at Strasbourg.

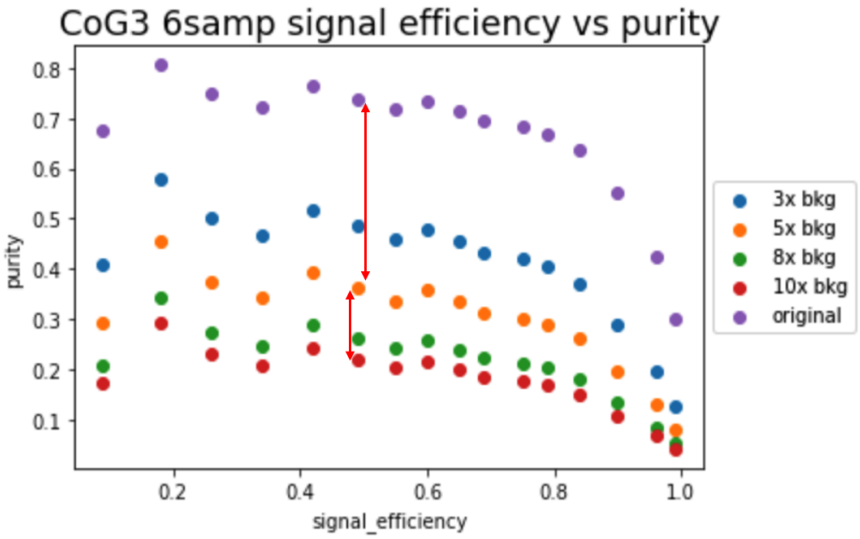

Figure 17: Purity vs. signal efficiency plot for 6-sample using CoG3 with increasing background events.

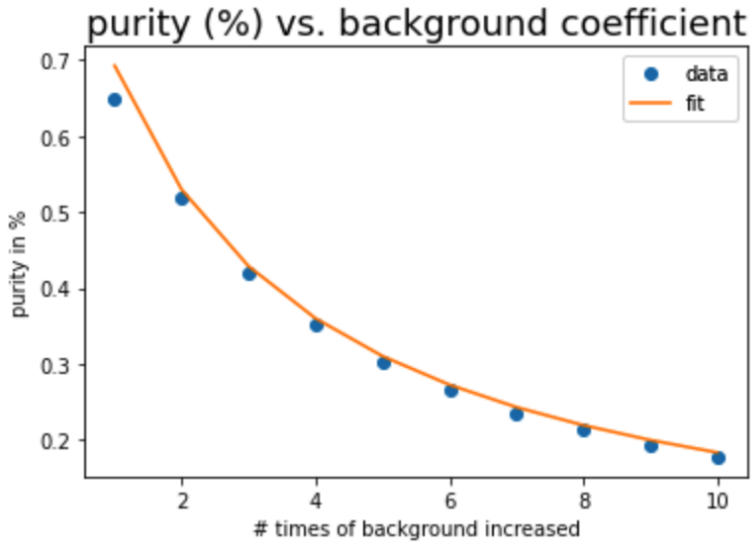

Figure 18: The purity vs different times (1 to 10 times)increased selected background at 80% signal efficiency. The orange curve is the $constant/(constant + x)$ fit to the blue dots where constant equals to 2.25.

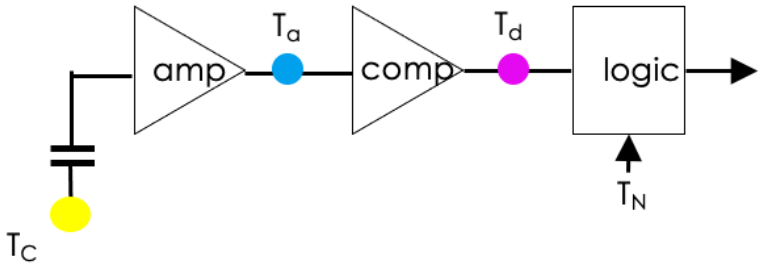

Figure 19: The signal propagation in each pixel in a CMOS sensor: T_c is the collision time; T_a is the analogue pulse passes the detection threshold; T_d is the digital pixel output switch from 0 to 1; T_d is the time when frame N starts)

3.1 MIMOSIS-Fast

The MIMOSIS-Fast is based on MIMOSIS0 which has a long integration time, approximately 5 μs [8]. The MIMOSIS-Fast is designed to have a shorter integration time. It has 504 rows and 32 columns with DC and AC pixels. The readout diagram of each pixel is shown in Figure 19. When a signal arrives, the sensor converts the light into charges. The charges are collected and accumulated inside each pixel. They are further transformed into voltage and amplified. Then, the amplified waveform goes through the comparator. The comparator prints out digital output 0 or 1 depending on if the amplified pulse is above the threshold or not. If the digital output is 1, it will further go through the logic at the matrix periphery which combines with the frame information to define which frame this signal is in to determine the arrival time. In this scenario, the frame counter is a periodic signal that determines the arrival time of the hit. There is no specific time for the digital converter to estimate the precise hit time. Consequently, the timing precision is given by the frame period.

3.1.1 Time delay test

The time delay is important for accurately defining the frame number associated to an incoming signal. Before going into the comparator, amplified waveforms with different amplitudes will have different rising slopes, so the time of reaching the threshold will be different (Figure 20 (a)). The difference between the starting point of a pulse and the time of reaching the threshold is defined as time delay (plot (b)). As the signal height increases, the time delay will approach a constant, t_0 [9]. If the time delay is too long, the incoming signal might be considered as in the next frame which will mess up with the true arrival time of the signal. Therefore, the time delay should be much shorter than the frame period, about one order of magnitude lower, otherwise the timing information provided by the frame identification will be invalid.

The MIMOSIS-Fast injects charges into each pixel by using capacitors. The Figure 20 shows the schematic of getting the injected charge. The capacitor can be connected to the pulsing voltage, V_{PL} or V_{PH}. The VPulseHigh (V_{PH}) and VPulseLow (V_{PL}) are two eight-bit on-chip DACS that generates the voltages. They are set in the software, *MIMOSIS0 Device Configuration*, which sets the value into the chip. The injected voltage is first set to V_{PL} with V_{Bias} set to the working voltage of a diode. Once the charge is injected, the injection voltage will be set to

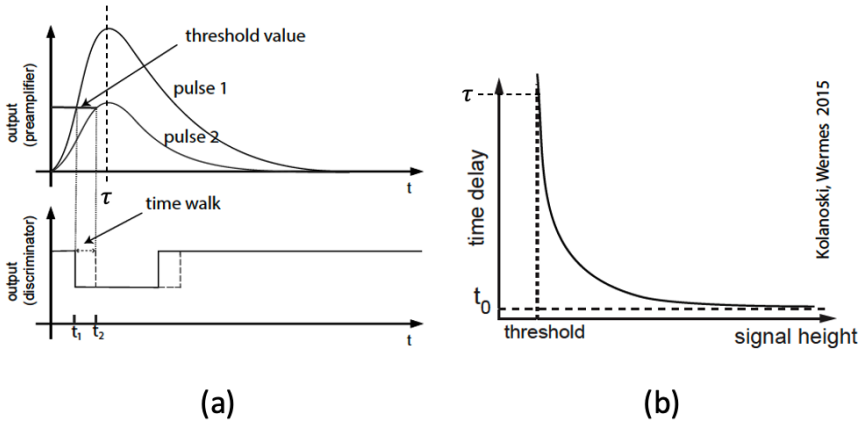

Figure 20: The procedure to get time delay plot. (a) upper plot is the analogue output and the lower plot is the digital output. (b) The time delay plot with varying signal height [9]

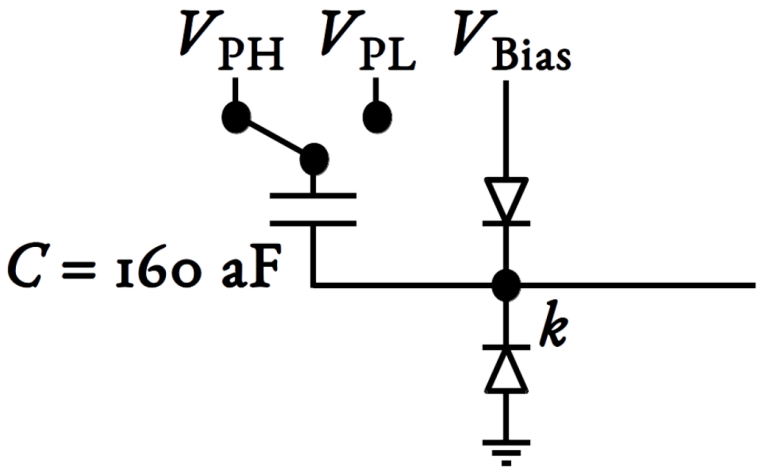

Figure 21: Schematic sketch of the charge injection system in MIMOSIS-Fast

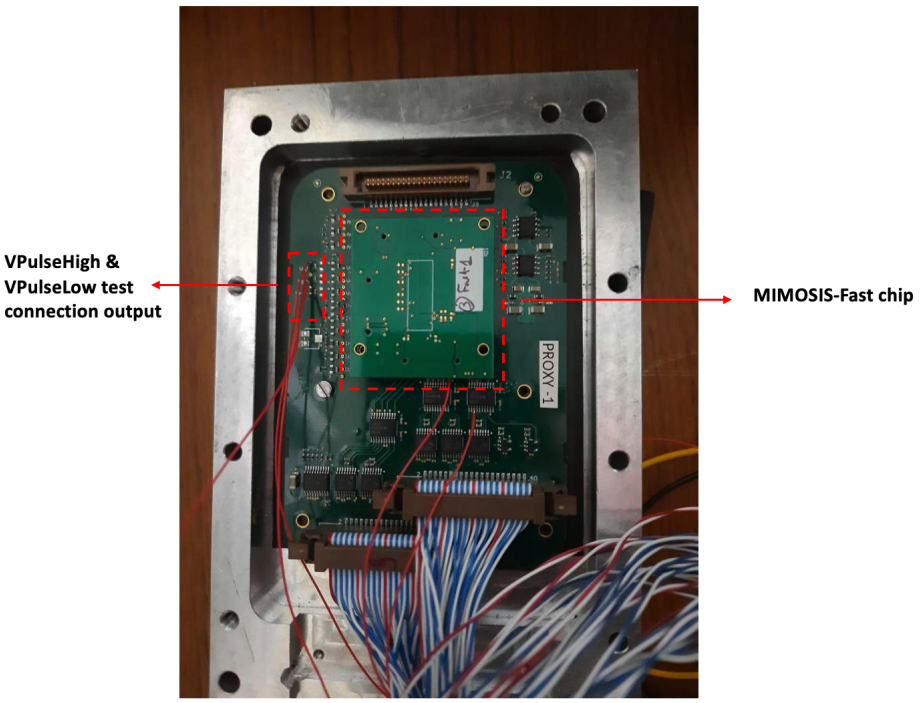

Figure 22: The set-up of the proximity board in the MIMOSIS-Fast study. The proximity board has different connections for test outputs, VPulseHigh and VpulseLow outputs are marked in the plot. The MIMOSIS-Fast chip is located at the center of the proximity board.

V_{PH}. The quick change of voltages induces the charge to the node k with $Q = C \cdot (V_{PH} - V_{PL})$. The capacitance is set to be 160aF so that 1mV voltage difference corresponds to 1e charge generated on the electrodes [8]. Using the ammeter to measure the VPulseHigh and VPulseLow values to the ground directly from the proximity board (Figure 22), where the MIMOSIS-Fast chip is positioned. We found that the voltage value is off by 0.384 V when the DAC value is at zero.

To measure the time delay, we used the oscilloscope to visualize the injected pulse (trigger), analogue and digital output (Figure 23). The colors of the curves match the colors of the output points shown in Figure 19. The oscilloscope is triggered on the rising edge of the injected pulse (yellow line). The magenta line and blue line show the accumulated view of the digital and analogue outputs separately. The pulse injection in MIMOSIS-Fast is set to be once per every eight frames. The optimal configuration is found by tuning different parameters and observing the results from the oscilloscope to maximize the analogue pulse height, but minimize the analogue pulse width at the same time. To define the time delay, we used the built-in time delay function in the oscilloscope which measures the time difference between the trigger (pulse injection) to 50% of digital pulse height.

The final time delay plot is shown in Figure 24. The test uses a -1V reverse bias voltage with a varying injected charge scan from 132 e to 1397 e charges. In this plot, we showed the test on three different types of pixels, and each dot is an average value of 20 samples. The DC pixel 28 (blue dots) has both analogue and digital output test connections, while the DC pixels from

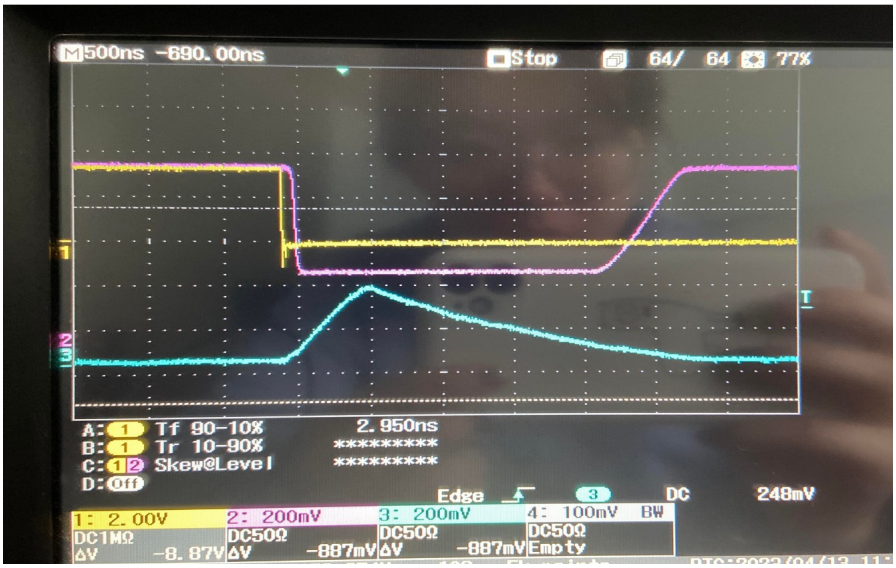

Figure 23: The output of the analogue, digital and injected pulse from MIMOSIS-Fast shown on the oscilloscope. yellow line: injected pulse; magenta line: digital output; blue line: analogue output. The width of each grid on the screen is 500ns.

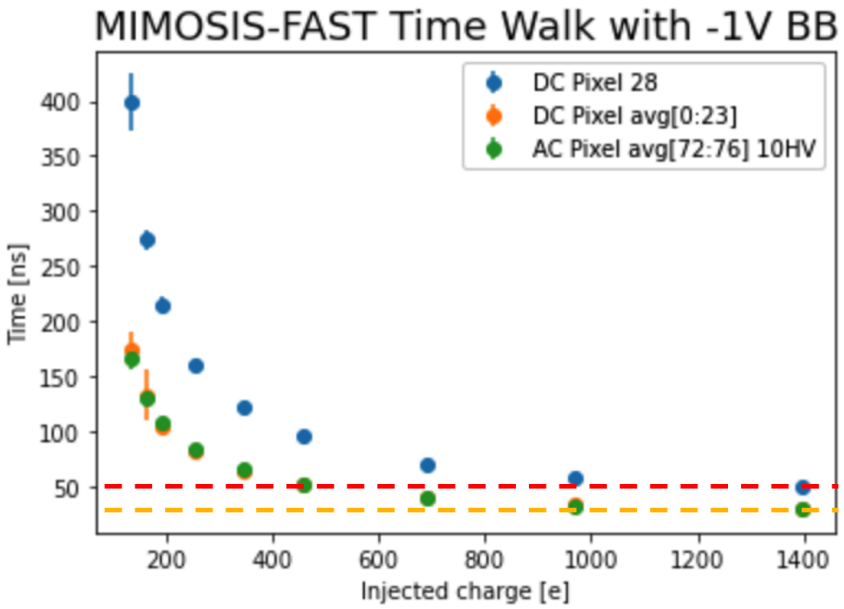

Figure 24: The time delay plot for the MIMOSIS-Fast

0 to 23 (orange dots) only have the digital one. The AC is connected with a 10 HV, and AC pixels from 72 to 76 (green dots) only have the digital output test connection. At high injected charges, the time delay for DC pixel 28 is approaching 50ns while the average time delay for DC pixels (0 to 23) and the AC pixels (72 to 76) is approaching 30ns. Without the analogue output test connection, the time delay is improved by around 20ns at high injected charge since there is no extra capacitance at the analogue output test connection. In the real detector, the pixels will have neither analogue nor digital output test connections, so the time delay will be even faster. The initial study on MIMOSIS-Fast shows that both DC and AC pixels have a short time delay, down to 30ns. Therefore, we could use the frame period down to 300ns to 500ns. There is no big difference between DC and AC because the high voltage applied is not high enough to show the difference in our test.

4 Conclusion

This master thesis investigates timing strategies on the software and hardware at Belle II in response to the anticipated high-luminosity situation in the future. The high-luminosity will bring more background hits in the detector and create high occupancy. On the software part, reducing the readout samples by two can solve bandwidth problem. However, the simple cut on the 6-sample to 3-sample readout mode does not seem to work very well, lower purity and background rejection at high signal efficiency. A more complex selection strategy based on machine learning algorithm is currently under development at KEK. This strategy uses both the time and charge on each strip within the cluster to cut down the readout sample numbers. On the hardware part, replacing current sensors in the vertex detector with all-pixel designed sensors using the CMOS technology can further reduce the occupancy by providing low granularity in space and time. A prototype CMOS sensor, MIMOSIS-Fast, shows that the time delay can be down to 30 ns with around 1400e injected charges. However, this study only uses the injected charge which will only test the electronics behavior. To test the ionisation charge collection time, the next step would be using a laser with a precisely defined arrival time.

References

[1] L. Massaccesi. *Performance study of the SVD detector of Belle II and future upgrades.* https://docs.belle2.org/record/2759. Nov. 2021.

[2] Y. Uematsu et al. *The Silicon Vertex Detector of the Belle II experiment.* https://doi.org/10.1016/j.nima.2022.166688. Apr. 2022.

[3] J. Baudot et al. *IN2P3 contributions to an all-layer monolithic pixel vertex detector for the Belle II upgrade.* Belle II document to IN2P3 Scientific Council. Oct. 2022.

[4] S. Yamada et al. *Data Acquisition System for the Belle II Experiment.* doi: 10.1109/TNS.2015.242471. June 2015.

[5] Belle II Software Group. *SVD:SVD reconstruction.* basf2 Development Documentation.

[6] S. Mondal. *Test SVD EventT0 Performance in 3/6 mixed Sample.* BIISVD-448: Belle II SVD Meeting. Jan. 2023.

[7] M. Babeluk et al. *CMOS MAPS upgrade for the Belle II Vertex Detector.* https://doi.org/10.1016/j.nima.2023.168015. Mar. 2023.

[8] M. Deveaux et al. *Observations on MIMOSIS-0, the first dedicated CPS prototype for the CBM MVD.* https://doi.org/10.1016/j.nima.2019.162653. Apr. 2020.

[9] H. Kolanoski N. Wermes. *Particle Detectors.* Oxford University Press. Aug. 2020.