Suppressing Beam Background and Fake Photons at Belle II using Machine Learning

Priyanka Cheema[1],*

[1]School of Physics, University of Sydney, Australia

Abstract. The Belle II experiment situated at the SuperKEKB energy-asymmetric e^+e^- collider began operation in 2019. It has since recorded half of the data collected by its predecessor, and reached a world record instantaneous luminosity of 4.7×10^{34} cm^{-2}s^{-1}. For distinguishing decays with missing energy from background events at Belle II, the residual calorimeter energy measured by the electromagnetic calorimeter is an important quantity. Ideally, calorimeter clusters due to beam backgrounds and fake photons should be excluded when the residual calorimeter energy is calculated, so identifying them during the analysis process is key. We present two new boosted decision tree classifiers that have been trained to identify such clusters at Belle II and distinguish them from real photons originating from collision events at the interaction point. We provide results from their application to the $B \rightarrow D^* \ell \nu$ decay mode, and show that the distribution of residual calorimeter energy for signal events is significantly improved.

1 Introduction

The Belle II detector is situated at SuperKEKB which is an asymmetric e^+e^- collider at the KEK laboratory in Tsukuba, Japan [1] [2]. Collisions started in 2019, and since then 428 fb^{-1} of data has been recorded. SuperKEKB is built to collide positrons and electrons at energies of 4 and 7 GeV respectively. These energies are tuned to ensure that the centre of mass energy $\sqrt{s}$ matches the $\Upsilon(4S)$ rest mass of 10.58 GeV/c^2 [2] [3]. Once created at the interaction point, the $\Upsilon(4S)$ decays almost exclusively into a pair of B mesons (denoted as $B\bar{B}$). The Belle II detector is used to detect particles emitted from these collisions, with the sub-detectors designed to achieve a cylindrical geometry that yields a near-4π solid angle coverage. The Belle II experiment is particularly useful at identifying B decays with missing energy-momentum signatures due to (1) the relatively clean environment and (2) the known initial state provided by e^+e^- collisions. These features allow the energy-momentum of neutrinos or other invisible final-state particles to be reconstructed, and in principle allow for a complete reconstruction of each event. A by-product of this reconstruction is the *residual calorimeter energy*, E_{ECL}, which is a key variable for many B decay analyses. This variable is used for signal extraction and background suppression due to its strong signal-background separation power. The quantity refers to the total energy sum of electromagnetic calorimeter clusters unaccounted for following the reconstruction of a collision event. For correctly reconstructed signal events, E_{ECL} peaks at 0 GeV [4], while for background events, the E_{ECL}

*e-mail: pche3675@uni.sydney.edu.au

distribution broadens and peaks at higher values. Within the landscape of current Belle II analyses, E_{ECL} is vital for several high priority measurements such as $R(D^{(*)})$, the $|V_{ub}|$ and $|V_{cb}|$ elements of the CKM matrix, and the branching fractions of $B \to K\nu\bar{\nu}$ and $B \to \tau\nu_\tau$. The use of a quantity like E_{ECL} is not specific to B Factories, and can be easily extended to any detector that can achieve full reconstruction through an electromagnetic calorimeter with close to 4π coverage such as BES-III, KLOE or SND.

To increase the separation power of E_{ECL}, ensuring a strong peak at 0 GeV for the signal distribution is imperative. In practice this is difficult as contributions from energy deposits in the electromagnetic calorimeter created by beam background processes, or calorimeter energy deposits that are split into multiple clusters during the reconstruction process, also referred to as fake photons, will contribute to E_{ECL}. Such contributions will affect properly reconstructed signal events, making the distribution appear more background-like, i.e. they acquire a broad nonzero peak. This dilutes the separation power of E_{ECL} and degrades the precision of measurements with missing energy at Belle II. It is therefore imperative that these beam background clusters and fake photons are identified and suppressed to allow for a "clean" signal E_{ECL} distribution.

2 The Belle II Detector

The Belle II detector consists of cylindrical layers that comprise a magnetic spectrometer, electromagnetic calorimeter and muon detector. A superconducting solenoid provides a 1.5 T magnetic field that envelops the former two layers. A complete description of the Belle II detector and its sub-detectors is given in [1], with a diagram of the detector provided in figure 1. A detailed discussion of the electromagnetic calorimeter, from which E_{ECL} is derived, is given in Sect. 2.1.
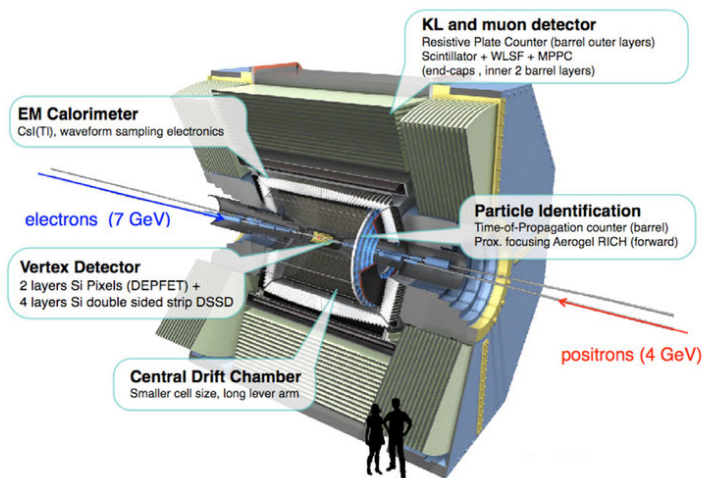

Figure 1: A diagram of Belle II and its various sub-detectors. Credit: Belle II Collaboration.

2.1 Electromagnetic Calorimeter

The electromagnetic calorimeter covers an acceptance region of 12.4°–155.1° in the polar angle with respect to the beam pipe, with two 1° gaps between the barrel and endcaps lo-

cated at 31.8° and 129.7°. It comprises 8736 CsI(Tl) crystals, with 6624 of these crystals located in the barrel, and 2112 located in the endcaps. The thallium doping in the crystal structure helps to increase the light output of the crystals. Each crystal has a trapezoidal geometry with a length of 30 cm and a square cross-section averaging 6 x 6 cm². At the rear of each crystal, two photodiodies 10 x 20 mm² in size are attached for scintillation light detection. Preamplifiers provide independent output signals from the photodioides of each crystal to an external shaper board where the signal is digitised by field-programmable gate arrays (FPGA). From this digitised signal, the amplitude and the time of the waveform can be extracted, and for crystals with energy above 50 MeV, the waveforms are recorded offline for pulse shape analysis [5].

Adjacent crystals with significant energy are grouped to form a cluster using a clustering algorithm [2]. During reconstruction, tracks of charged final state particles are extrapolated to the electromagnetic calorimeter, and matched to corresponding clusters, with all unmatched clusters reconstructed as photon candidates [6]. A limitation of the clustering algorithm is the potential to reconstruct fake photons. This arises when a particle leaves an energy deposit in multiple calorimeter crystals that are then incorrectly divided into multiple clusters, some of which remain unmatched and are thus reconstructed as photons. An example of where this occurs is for low momentum particles with a long trajectory at a shallow angle to the surface of the calorimeter. For such particles, the extended chain of crystals with deposited energy can lead to the incorrect subdivision of crystals into multiple clusters, or matching failures between the track and corresponding cluster. Hadronic showers are another common source of fake photons due to their irregularly shaped and dispersed energy deposits.

The aim of this study is to distinguish calorimeter clusters created by true photons from the collision event (hereafter referred to as signal photons), as opposed to clusters reconstructed as fake photons, or caused by beam background processes such as beam-gas, Touschek or radiative Bhabha scattering [7].

3 `FastBDT` Architecture

A multivariate analysis approach was chosen for the identification of fake photons and beam background clusters to allow for the information from various cluster-related variables to be used in combination. The classifier chosen was the `FastBDT` as it has been specifically optimised to handle the substantial amount and type of data that exists in particle physics [8]. The architecture of the `FastBDT` classifier consists of a cache-optimised stochastic gradient boosted decision tree [8]. Such architectures are robust against overfitting, and help ensure minimal correlations between the weak learners which maximizes performance when the individual learners are aggregated. The main hyperparameters for the architecture are the number of trees in the ensemble, the maximum depth of each weak learner, and the shrinkage parameter for regularisation. Hyperparameter tuning was performed for the number of trees and maximum depth as the `FastBDT` classifier shows great variability in performance over a range of values for these hyperparameters [8]. Shrinkage was only tuned if overfitting is observed. Built into to the architecture is an estimation of the importance of each feature used in the classifier, which is calculated using the total information gain achieved by each feature across all weak learners in the ensemble. To compare the performance between classifiers during hyperparameter tuning, the binary cross-entropy loss function (also called log-loss) was used. This is defined as

$$L_{\log} = -(y \log(p) + (1 - y) \log(1 - p)) \tag{1}$$

where $y \in \{0, 1\}$ is the true label of a given sample, and p is the probability of the given sample being in class 1 as determined by the classifier [9]. The smaller the log-loss, the better

the predictions made by the model. In addition to the log-loss, the area under the receiver operating characteristic curves (AUC) was also used as a measure of classifier performance. The higher the AUC, the more separated the classes are, with an AUC of 1 indicating that a perfect separation has been achieved by the classifier.

4 Data Samples

Signal, beam background and fake photons were all selected from photon candidates reconstructed from simulated $B^0\overline{B}^0$ events with simulated beam background processes overlaid across all events. A detailed description of the background simulation can be found in [7]. All considered photon candidates were required have a minimum cluster energy of 0.05 GeV and a polar angle between 17° and 150° to match the acceptance of the central drift chamber [1]. Additionally, the sum of crystal weights in the photon cluster was required to be more than 1.5. For non-overlapping clusters, this sum equals the total number of crystals in the cluster. For overlapping clusters, a division of the crystal energy occurs between the clusters leading to a non-integer value of the crystal weight sum. Clusters were classified using information from the Monte Carlo generators used to create the simulated data. The Monte Carlo generators retain information about the energy deposited by individual particles in individual crystals, and then calorimeter clusters. This allows clusters whose energy is predominantly from a single particle to be *matched* to that particle in the simulation. Signal photons were defined using a cut which ensures the photon has been correctly reconstructed from the collision event. To select beam background and fake photon clusters, the following procedure was used:

1. Any cluster that is matched to a particle descended from the fundamental e^+e^- collision (e.g. B– or D–decay daughters, $e^+e^- \rightarrow q\bar{q}$ fragmentation products, initial- and final-state radiation) is rejected.

2. For clusters passing this step, the total energy in the cluster due to particles descended from the e^+e^- collision is summed.

3. Clusters where this sum ≥ 0.053 GeV are classified as fake photons.

4. Clusters where this sum ≤ 0.025 GeV are classified as beam background clusters.

The thresholds in steps 3 and 4 were determined by studying distributions from various Monte Carlo samples, with and without hadronic tracks, and with and without overlaid beam background. In the case of both the beam background and fake photon classifiers, class 1 refers to signal photons, while class 0 signifies either beam background clusters or fake photons. For the photons that remain for each simulated sample following the aforementioned cuts, holdout was implemented so that feature selection and hyperparameter tuning could be performed without biasing the test data. The ratio of the split between the training, validation and test data was 60:20:20.

5 Classifier Features and Hyperparameter Selection

Feature selection for the beam background and fake photon classifiers was performed by first collating a pool of cluster-related variables that demonstrated some level of separation between signal photons and beam background clusters or fake photons. From this pool, features that demonstrated a high importance score were selected. Redundant features were identified through both their importance scores and correlation matrices. A list of features

that were selected for the beam background and fake photon classifiers are given below. For the beam background classifier, features 1–5 were used. The fake photon classifier uses all the listed features. Distributions of the features 1, 4 and 5 are presented in figure 3 as they are the most important across the two classifiers.

1. The estimated energy of the photon which produced the cluster.

2. The polar angle (with respect to the beam pipe) of the cluster.

3. The output of a separate classifier that uses eleven of the Zernike moments of a cluster to identify the shape of energy distributions as belonging to hadronic particles (class 0) or photons (class 1).

4. The output of a separate classifier that uses pulse shape information from activated electromagnetic calorimeter crystals to distinguish between hadronic showers (class 0) and electromagnetic showers (class 1) [5].

5. The difference in timing $t - t_0$ between the time of the collision at the interaction point t_0 and the cluster time t

6. The lateral energy distribution of the cluster

$$ S = \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^{n} w_i E_i r_i^2} $$

where $r_0 \approx 5$ cm, and for crystals sorted in decreasing order of energy (with E_0 the highest), the w_i and E_i are the weight and energy respectively of crystal i, and r_i is the distance from crystal i to a projection of the cluster onto a plane perpendicular to the photon trajectory [10]

7. The distance (in cm) between the cluster and its closest particle track (this is measured using the point of closest approach for the particle track extrapolated through the electromagnetic calorimeter)

Following feature selection, hyperparameter tuning for both classifiers was performed for the number of trees and the maximum depth of each weak learner. A simple grid search was conducted for the following values: number of trees $\in [100, 1000]$ in steps of 100; maximum depth $\in [1, 6]$ in steps of 1. The log-loss values for each pairing was compared between the training and validation data sets to check for overfitting. The shrinkage parameter was decreased for the value pairings where overfitting was observed, with only minimal improvements seen. The hyperparameters chosen were number of trees = 100, and maximum depth = 3 for the beam background classifier, and number of trees = 300, and maximum depth = 3 for the fake photon classifier. For each classifier, a shrinkage value of 0.1 was used.

6 Results and Application To $B^0 \rightarrow D^{*-} \ell^+ \nu$

A final training of each classifier was done on the training and validation data sets combined, and using the optimal hyperparameters found in Sect. 5. The AUC obtained from the final training (test) of each classifier was 0.998 (0.998) for the beam background classifier, and 0.943 (0.944) for the fake photon classifier. The log-loss scores obtained from the final training (test) of each classifier was 0.0557 (0.0560) for the beam background classifier, and 0.290 (0.291) for the fake photon classifier. Distributions of the classifier output following the final
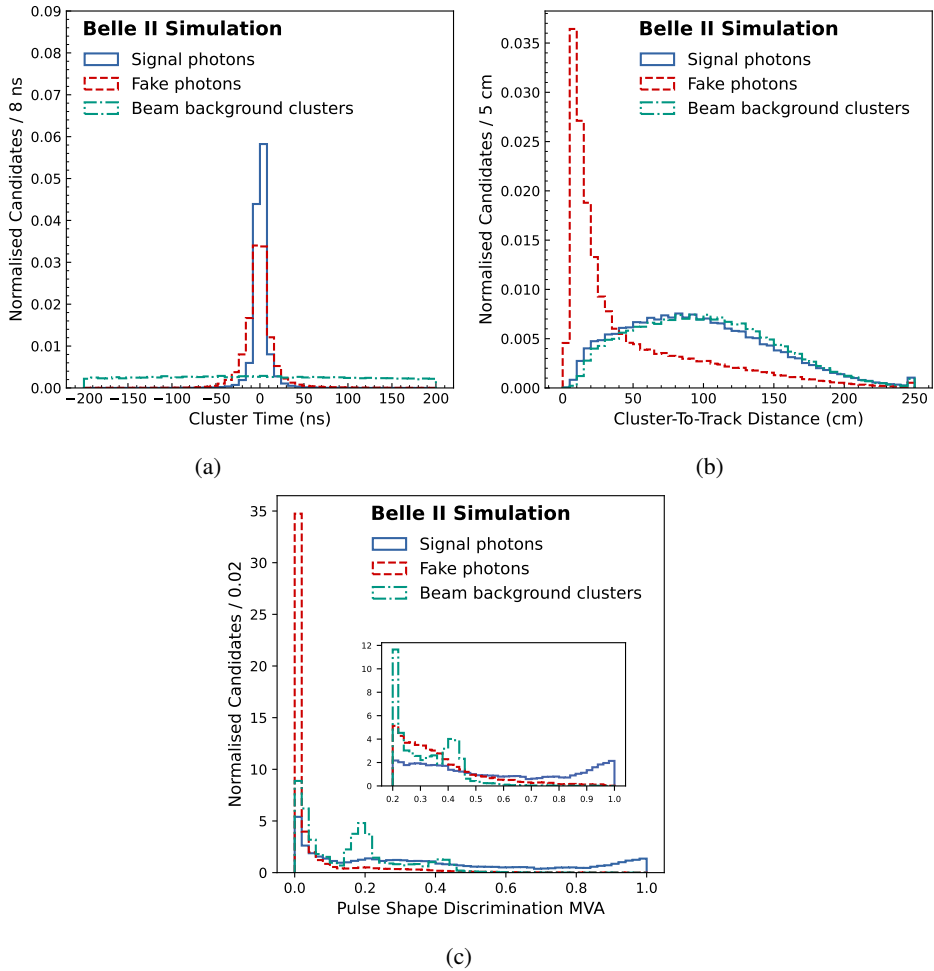
Figure 2: Distributions of the three most important features for signal photons, fake photons and beam background clusters. The cluster time and output of the pulse shape discrimination MVA are the most important for the beam background classifier, while the latter and the cluster-to-track distance are the two most important for the fake photon classifier. All distributions are normalised to 1.

training and application to the test data is provided in figure 3. To determine the impact of the beam background and fake photon classifiers on E_{ECL}, the classifiers were applied to a $B^0 \rightarrow D^{*-}\ell^+\nu$ ($\ell = e, \mu$) analysis. Simulated samples of $\Upsilon(4S)$ events overlaid with simulated beam background processes are used. The classifier output cuts tested were an output greater than 0.6 for the beam background classifier, and an output greater than 0.7 for the fake photon classifier. Distributions for E_{ECL} before and after these classifier cuts are given in figure 4.

As can be readily seen from figure 4a, true $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ events do not show a peak at 0 GeV despite being properly reconstructed. This is largely due to the presence of beam background and fake photon clusters, and the broadness of the distribution limits the separation power between signal and background events, thus decreasing the power of E_{ECL} for background rejection. The resulting impact of the classifiers on the E_{ECL} distribution is shown
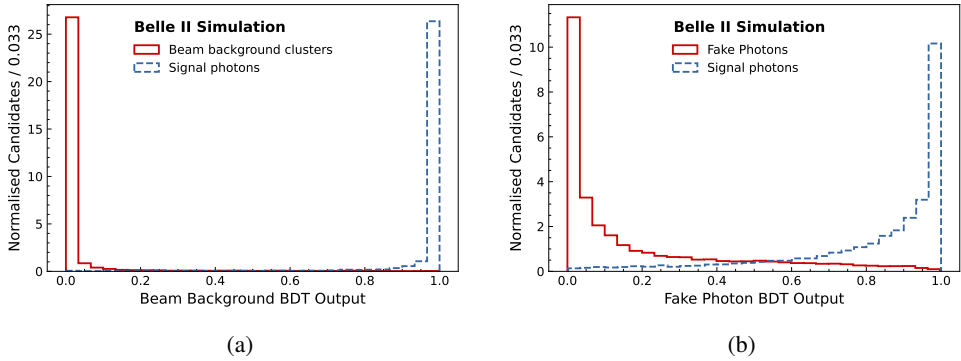
Figure 3: Distributions of the classifier outputs for the beam background classifier in (a) and the fake photon classifier in (b). All distributions are independently normalised to 1.
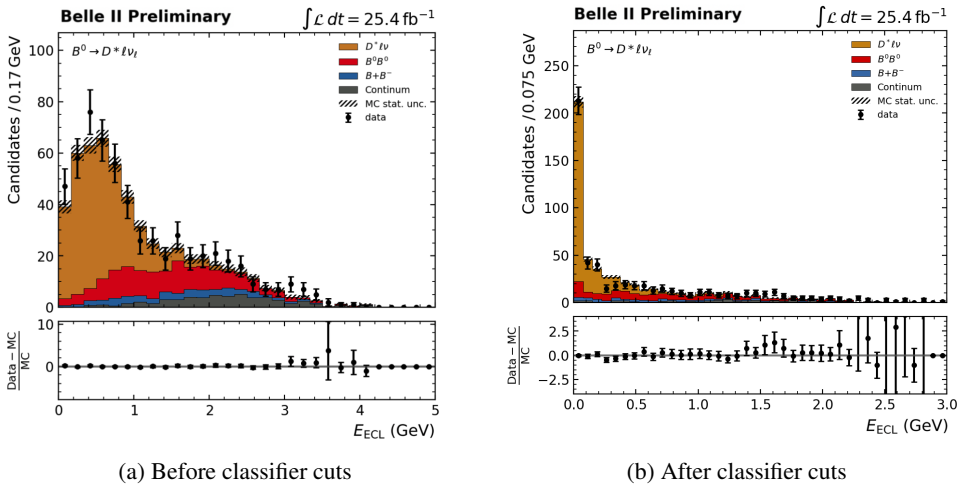

(a) Before classifier cuts

(b) After classifier cuts

Figure 4: The E_{ECL} distributions for $B^0 \rightarrow D^{*-}\ell\nu_\ell$ events after a cut of greater than 0.6 on the output of the beam background classifier, and a cut of greater than 0.7 on the output of the fake photon classifier. The simulated samples are decomposed into the true $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ events, combinatorial background (incorrectly reconstructed $B^0\overline{B}^0$ events), B^+B^- background and continuum background.

in Figure 4b. As can be seen, usage of the classifiers produces a signal distribution that is significantly shifted to lower E_{ECL} values, with a strong peak at $E_{\text{ECL}} \approx 0$ which is the desired performance. The result is an E_{ECL} distribution that has an improved discrimination between signal and background events due to the difference in shape. To quantify this improvement, the signal yield S from data for $B^0 \rightarrow D^{*-}\ell^+\nu$ is determined via a single toy fit to the $E_{\text{ECL}} < 0.8$ GeV region. For the fit, only two templates are considered: signal, and background (which includes combinatorial, B^+B^- and continuum). The signal significance S/σ_S is calculated where σ_S is the uncertainty of the fitted yield. A fit to the E_{ECL} distribution in figure 4a gives $S/\sigma_S = 4.10$ while a fit to the distribution in figure 4b gives $S/\sigma_S = 10.08$,

which demonstrates a clear improvement in the shape of E_{ECL} for signal-background separation, and the resulting increased sensitivity following the usage of the classifiers.

7 Conclusion

In these proceedings, new `FastBDT` classifiers built to identify beam background clusters and fake photons, and separate them from signal photons, were presented. Various features containing information on the timing, energy distribution, scintillation pulse shape and location of clusters in the electromagnetic calorimeter were considered. For each classifier, feature selection followed by hyperparameter tuning using holdout was performed. The result of this were high performing classifiers, that gave high AUC scores and strong separation between beam background clusters, fake photons and signal photons. Upon applying the classifiers to the $B^0 \to D^{*-}\ell^+\nu_\ell$ analysis, substantial improvements were seen in the distributions of residual calorimeter energy E_{ECL}. In particular, clear peaks at $E_{\mathrm{ECL}} \approx 0$ are seen for signal, and the result is an improved E_{ECL} that provides an even greater signal-background separation. Due to this, multiple missing energy decay analyses at Belle II are successfully employing these classifiers to significantly reduce their backgrounds, and improve their signal significance. Additional studies have demonstrated that these classifiers are robust against higher beam background levels, which is important as SuperKEKB looks to ramp up its instantaneous luminosity in the coming years.

References

[1] T. Abe et al., *Belle II Technical Design Report* (2010), `https://arxiv.org/abs/1011.0352`

[2] W. Altmannshofer et al. (Belle-II), PTEP **2019**, 123C01 (2019), [Erratum: PTEP 2020, 029201 (2020)], `1808.10567`

[3] R.L. Workman et al. (Particle Data Group), PTEP **2022**, 083C01 (2022)

[4] Ed. A.J. Bevan, B. Golob, Th. Mannel, S. Prell and B.D. Yabsley, Eur. Phys. J. C **74**, 3026 (2014)

[5] S. Longo et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **982**, 164562 (2020)

[6] *Belle II Analysis Software Framework (basf2)*, `https://doi.org/10.5281/zenodo.5574115`

[7] A. Natochii et al., *Beam background expectations for Belle II at SuperKEKB*, in *Snowmass 2021* (2022), `2203.05731`

[8] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification* (2016), `https://arxiv.org/abs/1609.06119`

[9] I.J. Good, Journal of the Royal Statistical Society. Series B (Methodological) **14**, 107 (1952)

[10] D. Brown, J. Ilic, G. Mohanty, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **592**, 254 (2008)