



BELLE2-NOTE-PL-2023-003
April 29, 2023

Development of FastBDT Classifiers to Suppress Beam Background Clusters and Fake Photons

The Belle II Collaboration

We present an overview of the methodology used to develop `FastBDT` classifiers that suppress beam background clusters and fake photons, with the aim of using these classifiers to improve the signal-background separation power of the residual calorimeter energy E_{ECL} . The feature selection, hyperparameter tuning, and training of the classifiers is performed using photons from simulated $B^0\bar{B}^0$ events. The application of these classifiers to the $B^* \rightarrow D^*\ell\nu$ decay is shown, with E_{ECL} distributions before and after classifier cuts also included. These E_{ECL} distributions also provide a comparison between simulated events and a subset of Belle II data amounting to 25.4 fb^{-1} of integrated luminosity.

1. DATA SAMPLES

Signal photons, beam background clusters and fake photons are all selected from photon candidates reconstructed from Monte-Carlo (MC) simulated $B^0\bar{B}^0$ events with simulated background overlay [1]. All photon candidates are required to have a cluster energy greater than 0.05 GeV, a polar angle θ between 17° and 150° , and the sum of crystal weights in the cluster greater than 1.5. Signal photons are 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 is used. 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. For clusters passing this step, the total energy in the cluster due to particles descended from the e^+e^- collision is summed. Clusters where this sum ≥ 0.053 GeV are classified as fake photons while clusters where this sum ≤ 0.025 GeV are classified as beam background clusters. 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. The number of photons that remain for each simulated sample following the aforementioned cuts is 98,000 for beam background clusters and 420,000 for fake photons. To avoid class imbalances, equal sized samples of signal photons were used in either case for the development of the classifiers.

2. CLASSIFIER FEATURES AND FEATURE SELECTION

Features were selected from a pool of cluster-based variables that demonstrated some level of separation between signal photons and beam background clusters or fake photons. Only features with high importance scores as calculated by the `FastBDT` [2] were selected. These scores were obtained by training dedicated `FastBDT` classifiers using the entire feature pool and the default hyperparameter values [2]. Redundant features were identified using correlation matrices and were removed. For the beam background classifier, the features selected were: the energy (*clusterE*), polar angle (*clusterTheta*) and timing (*clusterTiming*) of the cluster; the output of an external classifier that uses eleven of the Zernike moments of a cluster to identify the shape of energy distributions as belonging to hadronic particles or photons (*clusterZernikeMVA*); the output of an external classifier that identifies clusters as either electromagnetic showers or hadronic showers (*clusterPulseShapeDiscriminationMVA*) [3]. For the fake photon classifier, the features selected were: all the features for the beam background classifier; the lateral energy distribution of the cluster (*clusterLAT*) [4] and the distance between the cluster and its nearest track (*minC2TDist*). Distributions of the three most important features across both of the classifiers are given in Figure 1.

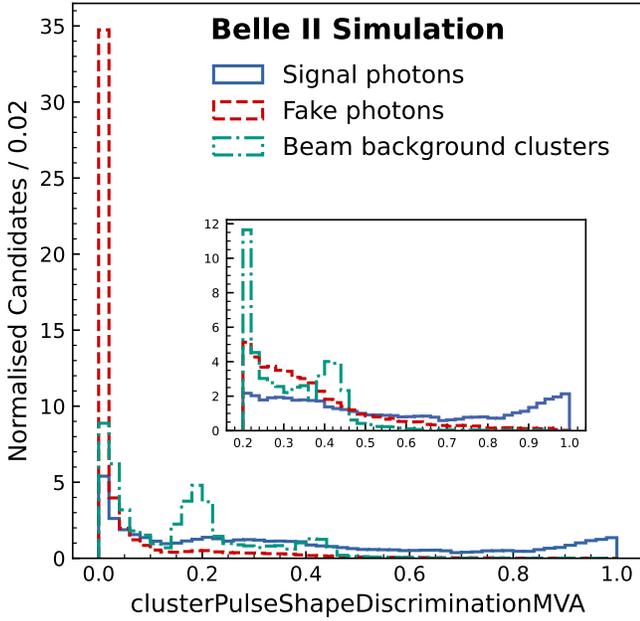
3. HYPERPARAMETER TUNING AND TEST RESULTS

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 done for the following values: the number of trees from 100 \rightarrow 1000 in steps of 100; the maximum depth from 1 \rightarrow 6 in steps of 1. The holdout procedure was used which involved dividing the photon samples into a training, validation and test subset using the ratio 60:20:20 for each class. For each set of hyperparameter values, the classifier is trained

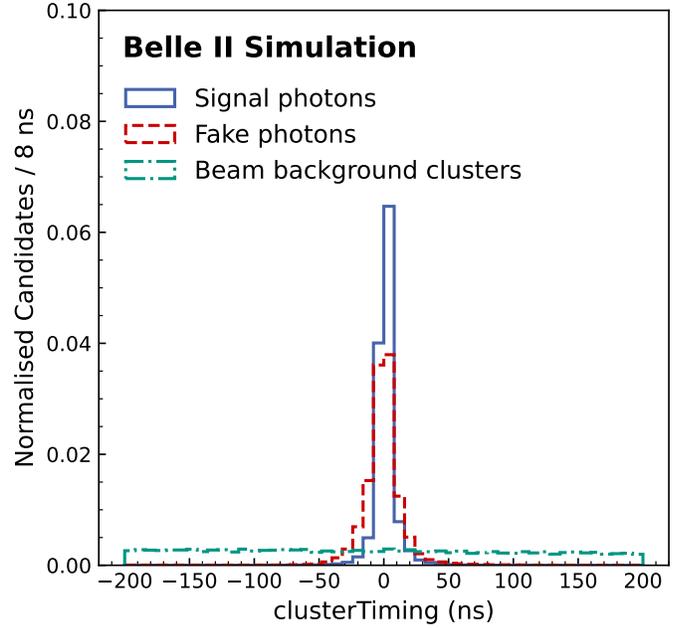
using the training subset, and tested on the validation subset. The log-loss [5] values for the training and validation subsets are compared to check for overfitting. In summary, the hyperparameters chosen for the fake photon (beam background) classifier were: number of trees = 300 (100), and maximum depth = 3 (3). Plots of the fake photon classifier log-loss scores for training and validation are shown in Figure 2. A final training of each classifier was done on the training and validation subsets combined using the optimal hyperparameters before a final test on the test subset. Distributions of the classifier output probability for class 1 after being applied to the test subset is given in Figure 3. The area under the curve (AUC) for the receiver operating characteristic (ROC) curves 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.

4. APPLICATION TO $B^0 \rightarrow D^{*-}\ell^+\nu$

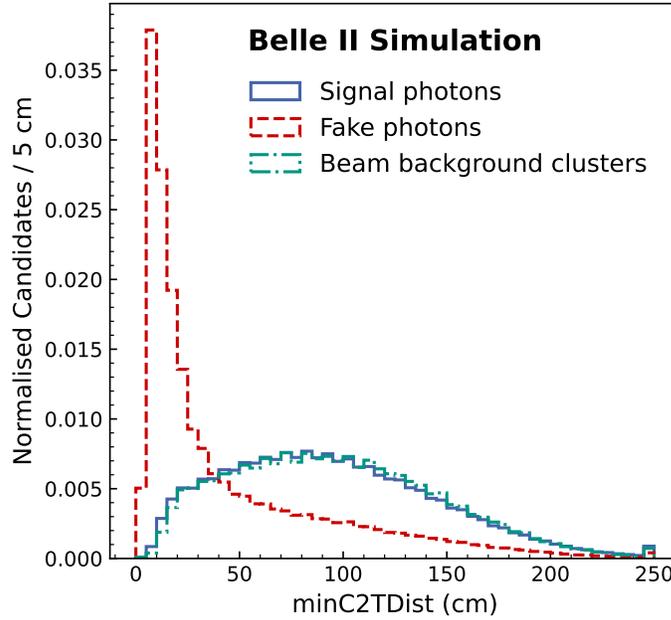
Reconstruction of $B^0 \rightarrow D^{*-}\ell^+\nu$ events begins with collision events being passed through the Full Event Interpretation (FEI) algorithm for B_{tag} reconstruction[6]. Events that pass the FEI have the following selections applied: at least 3 charged tracks with $|z_0| < 2.0$ cm and $|d_0| < 0.5$ cm, and with a minimum transverse momentum p_T of 100 MeV/c, are present (where z_0 and d_0 are the z coordinate and distance in the r - ϕ plane respectively to the point of closest approach to the interaction point); at least 3 neutral clusters with cluster energy greater than 0.1 GeV and a polar angle between 17° and 150° are present; the total visible energy of all the tracks and clusters in the events is greater than 4 GeV; the total energy deposited in the electromagnetic calorimeter is between 2 and 7 GeV; the ratio of the second to zeroth Fox-Wolfram moment is less than 0.3 [7]. B_{tag} candidates are also required to have an FEI signal probability > 0.001 . The B_{tag} is required to have $M_{\text{bc}} > 5.27$ GeV/ c^2 and $-0.15 < \Delta E < 0.1$ GeV. Here, $M_{\text{bc}} = \sqrt{E_{\text{CM}}^2 - |\vec{p}_{B_{\text{tag}}}|^2}$ where E_{CM} is half the total collision energy and $\vec{p}_{B_{\text{tag}}}$ is the momentum of the B_{tag} candidate in the centre of mass frame. Furthermore, $\Delta E = E_{B_{\text{tag}}} - E_{\text{CM}}$, with $E_{B_{\text{tag}}}$ denoting the centre of mass frame energy of the B_{tag} candidate. On the B_{sig} side, all tracks must also fulfill the same quality criteria as described above and, except for the slow pion π_s daughter produced in the $D^{*+} \rightarrow D^0\pi_s^+$ decay, must have at least one hit in the central drift chamber. The D^0 meson is reconstructed in the decay channels $D^0 \rightarrow K^-\pi^+$, $K^-\pi^+\pi^0$, $K^-\pi^+\pi^-\pi^+$, $K_s^0\pi^+\pi^-$. Kaon candidates are required to satisfy particle identification criteria for a kaon with likelihood ratio greater than 0.5, while pion candidates, except for the π_s , are required to satisfy a pion particle identification likelihood ratio greater than 0.1. π^0 daughters of the D^0 must satisfy the mass requirement $0.124 < M_{\pi^0} < 0.145$ GeV/ c^2 and are reconstructed via $\pi^0 \rightarrow \gamma\gamma$, where the photons are required to satisfy the energy threshold of $E > 0.05$ GeV. Only D^0 candidates with mass between 1.86 and 1.878 GeV are kept, while for D^* candidates, the mass difference, defined as $\Delta m = m_{D^*} - m_D$, is allowed to range between $[0.142, 0.150]$ GeV/ c^2 . Electrons and muons are identified as tracks with a particle identification likelihood greater than 0.9. Events with additional tracks after the $\Upsilon(4S)$ reconstruction are excluded. If an event still has more than one $\Upsilon(4S)$ candidate per event, the $\Upsilon(4S)$ candidate reconstructed using the B_{tag} with the highest FEI signal probability is chosen. Particles produced in the collision event that are not included in the $\Upsilon(4S)$ reconstruction are collectively labelled as the rest of event (ROE). The E_{ECL} distributions are obtained from the total energy sum of all electromagnetic calorimeter clusters in the ROE and are shown in Figure 4.



(a)

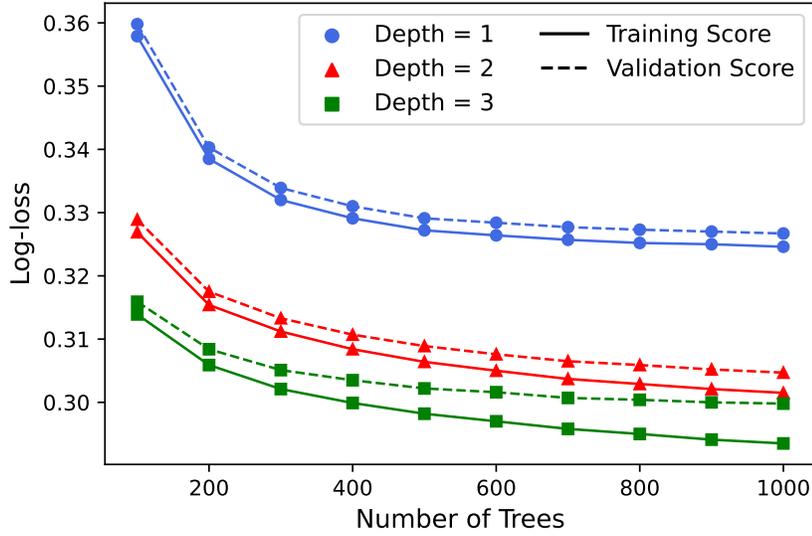


(b)

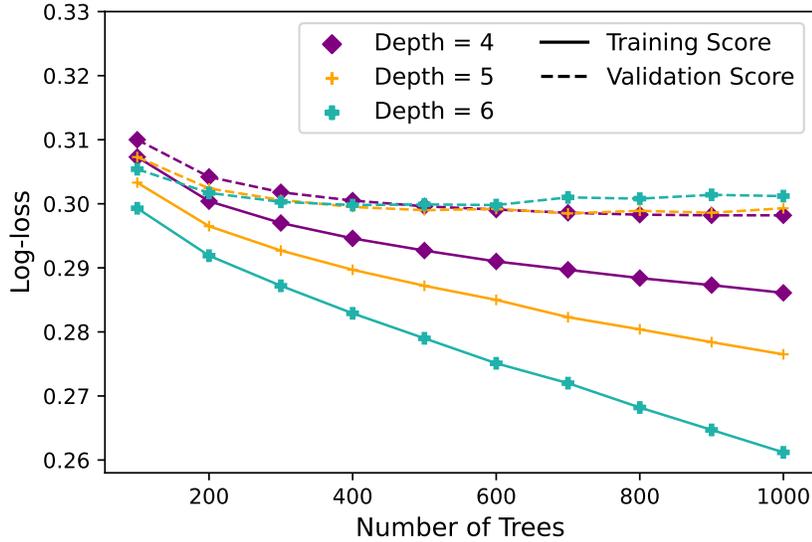


(c)

FIG. 1: Distributions for (a) $clusterPulseShapeDiscriminationMVA$, (b) $clusterTiming$ and (c) $minC2TDist$. The features shown in (a) and (b) were the two most important for the beam background classifier, while features in (a) and (c) were the two most important for the fake photon classifier. All distributions are individually normalised to 1. An inset for (a) is provided to show the distributions in the tail region [0.2, 1]. The photons used for the distributions are reconstructed from $B^0\bar{B}^0$ MC events with simulated background overlay.

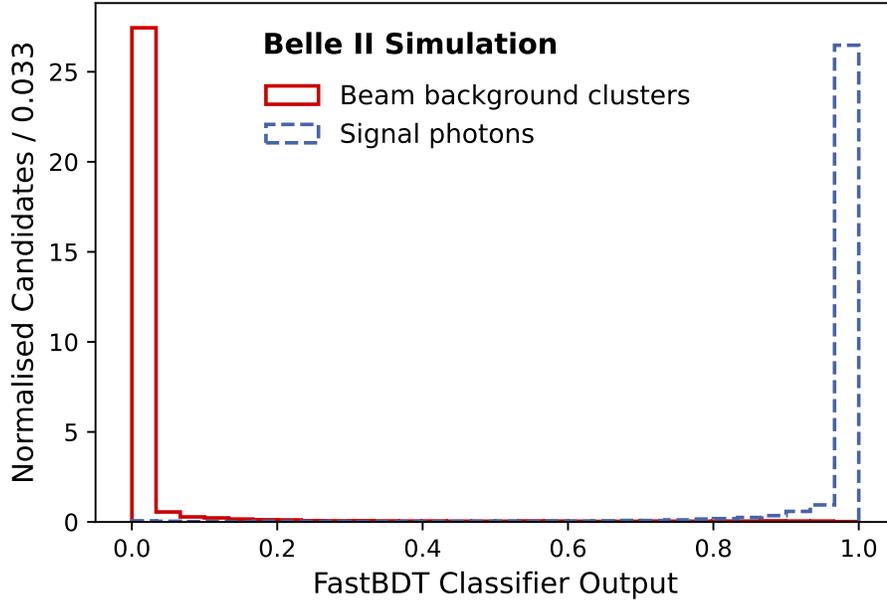


(a)

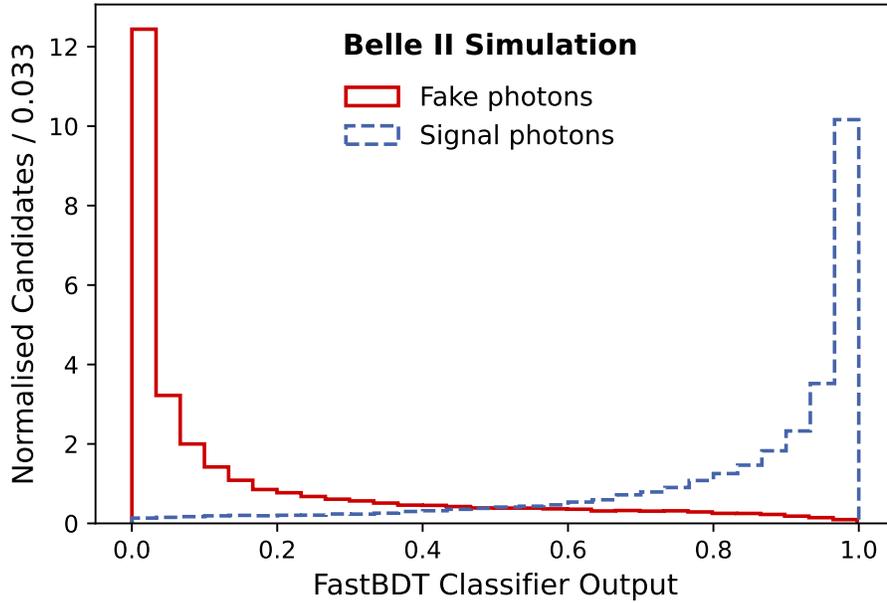


(b)

FIG. 2: The log-loss scores obtained during the training and validation of the fake photon classifier for each hyperparameter setting. All other `FastBDT` hyperparameters were set to their default values. Results for maximum depths 1-3 are shown in (a) while (b) shows the results for maximum depths 4-6. Based on these log-loss scores, the number of trees and maximum depth chosen for the fake photon classifier were 300 and 3 respectively.



(a)



(b)

FIG. 3: Distributions of the classifier outputs for the beam background classifier in (a) and the fake photon classifier in (b). The classifier output represents the probability of a photon being class 1 i.e. a signal photon. The classifiers used for these distributions have their hyperparameters set to the optimal values, and have undergone a final training on the training and validation subsets combined, with the distributions obtained by applying the classifiers to the test subset. All distributions are independently normalised to 1.

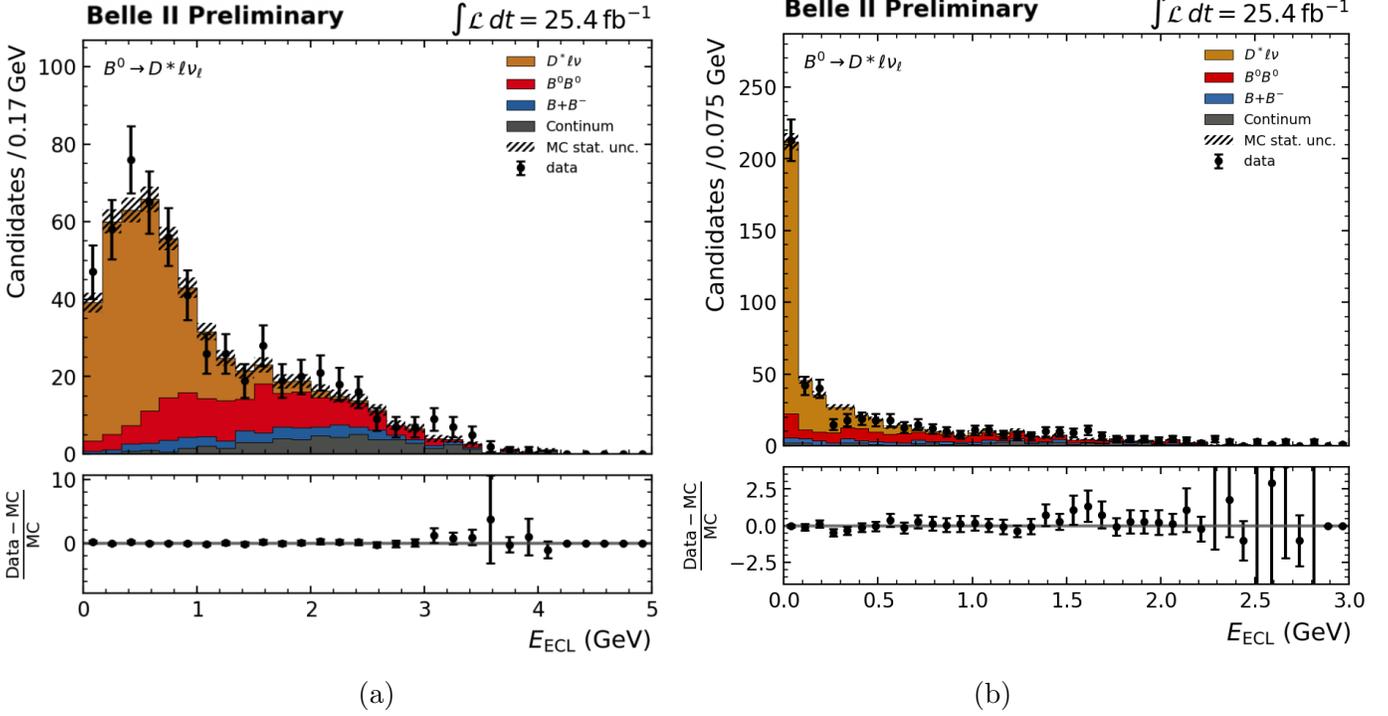


FIG. 4: E_{ECL} distributions for $B^0 \rightarrow D^{*-}\ell^+\nu$ events reconstructed from a small subset of Belle II data and simulated samples. The simulated samples are decomposed into the true $B^0 \rightarrow D^{*-}\ell^+\nu$ events in bronze, combinatorial background (incorrectly reconstructed $B^0\bar{B}^0$ events) in red, B^+B^- background in blue, and continuum background in grey. Data points are denoted with black dots. The subplot in (a) shows the E_{ECL} distribution without any classifier cuts, while the (b) shows the distribution under the following cuts: beam background classifier > 0.6 and fake photon classifier > 0.7 . To quantify the 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 + 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 and estimated using $\sigma_S/\sqrt{2N}$ where N is the number of events in the fit. A fit to the E_{ECL} distribution in (a) gives $S/\sigma_S = 4.10$ while a fit to (b) $S/\sigma_S = 10.08$.

-
- [1] A. Natochii, T. E. Browder, L. Cao, K. Kojima, D. Liventsev, F. Meier, K. R. Nakamura, H. Nakayama, C. Niebuhr, A. Novosel, G. Rizzo, S. Y. Ryu, L. Santelj, X. D. Shi, S. Stefkova, H. Tanigawa, N. Taniguchi, S. E. Vahsen, L. Vitale, and Z. Wang, *Beam background expectations for Belle II at SuperKEKB*, 2022. [arXiv:2203.05731](https://arxiv.org/abs/2203.05731) [hep-ex].
- [2] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, [arXiv: 1609.06119](https://arxiv.org/abs/1609.06119) (2016) .
- [3] S. Longo et al., *CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron-positron colliders*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **982** (2020) 164562.
- [4] D. Brown, J. Ilic, and G. Mohanty, *Extracting longitudinal shower development information from crystal calorimetry plus tracking*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **592** (2008) no. 3, 254–260.
- [5] I. J. Good, *Rational Decisions*, Journal of the Royal Statistical Society. Series B (Methodological) **14** (1952) no. 1, 107–114. <http://www.jstor.org/stable/2984087>.
- [6] T. Keck and others, *The Full Event Interpretation.*, Computing and Software for Big Science **3** (2019) no. 6, .
- [7] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in $e^+ e^-$ Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978) 1581.