Journal of
Instrumentation

SISSA

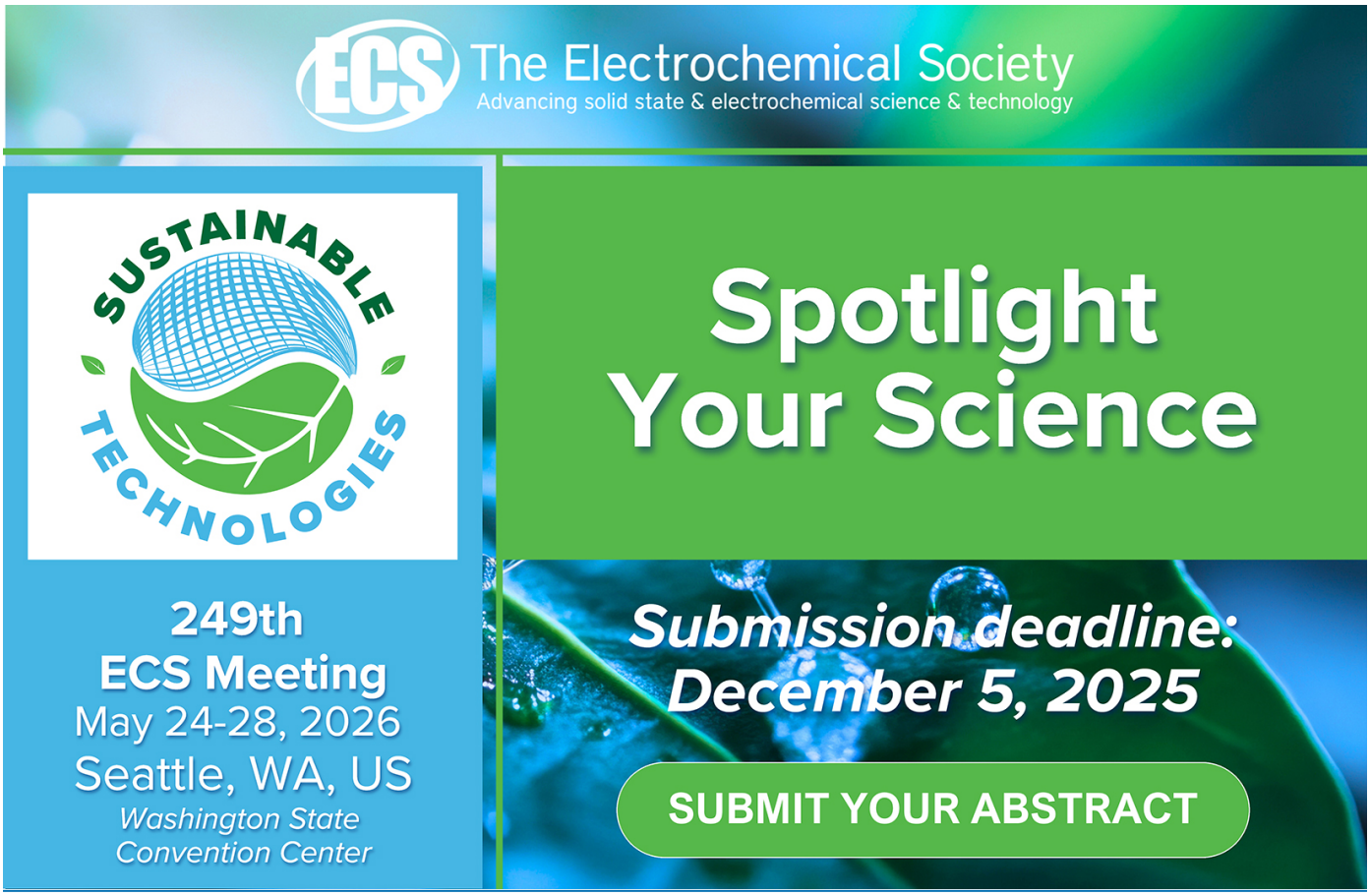
A matched filter based approach for high-energy estimation in calorimetry

To cite this article: B.S. Peralva *et al* 2021 *JINST* **16** P02016

View the article online for updates and enhancements.

You may also like

- Radiation hardness studies of neutron irradiated CMOS sensors fabricated in the ams H18 high voltage process
M. Fernández García, C. Gallrapp, M. Moll et al.
- Energy landscapes and their relation to thermodynamic phase transitions
Michael Kastner
- Two-time distribution functions in the Gaussian model of randomly forced Burgers turbulence
Victor Dotsenko



This content was downloaded from IP address 200.131.56.50 on 15/10/2025 at 16:08

Published by IOP Publishing for Sissa Medialab

Received: *September 9, 2020*
Accepted: *December 21, 2020*
Published: *February 10, 2021*

A matched filter based approach for high-energy estimation in calorimetry

B.S. Peralva,[a,1] **J.M. Seixas,**[b] **L.M.A. Filho**[c] **and A.S. Cerqueira**[c]

[a] *Computational Modeling Graduate Program, Rio de Janeiro State University, Nova Friburgo, RJ, Brazil*
[b] *Signal Processig Lab, COPPE-Poli, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil*
[c] *Electrical Enginnering Graduate Program, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil*

E-mail: bernardo.peralva@uerj.br

Abstract: In high-energy calorimeters, a crucial task is the reconstruction of the energy deposited by particle interactions. Standard techniques used in modern calorimeters rely on the energy estimation to select the signal with relevant information. This work presents a new approach, which performs the signal detection against noise as a first step, followed by the energy estimation task. The method is fully based on the Matched Filter (MF) theory, which is known to produce the optimum detection efficiency with respect to the signal-to-noise ratio. Furthermore, the MF output can be calibrated to estimate the signal amplitude and, thus, the energy. The proposed method is compared to different optimum filtering algorithms, which are currently being used for energy reconstruction in modern calorimeter systems. The results from simulated data show that the proposed method achieves better performance in terms of both signal detection efficiency and estimation error.

Keywords: Analysis and statistical methods; Calorimeters; Data reduction methods; Digital signal processing (DSP)

[1]Corresponding author.

© 2021 IOP Publishing Ltd and Sissa Medialab
https://doi.org/10.1088/1748-0221/16/02/P02016

2021 JINST 16 P02016

Contents

2021 JINST 16 P02016

1 Introduction

Modern particle physics experiments rely strongly on calorimeter systems. Calorimeters provide information on the total amount of detected energy, geometric imbalances in energy (missing transverse energy) and both single particle and jet identifications. Moreover, this information can be available fast enough to offer powerful online event selection capabilities [1].

From a structural point of view, a calorimeter is a block of matter in which particles get absorbed. In order to extract relevant particle properties, practical calorimeter designs for collider experiments comprise up to hundreds of thousand cells. Due to such fine-grained segmentation, not all channels will have relevant information for an incoming event. Therefore, efficient energy reconstruction algorithms and triggering systems are at most importance.

When the particle is absorbed within the calorimeter, its properties can be reconstructed from the signals produced by all the cells that were involved in the shower development process [2]. The energy deposited in each cell is measured individually, through its own readout electronics. For the data acquisition system point of view, each cell is a separate readout channel. Eventually, the calorimeter cells are combined for triggering purposes into towers of different sizes [3, 4].

Typically, the readout channels comprise an electronic shaping circuit [5]. Such pre-processing step aims at providing a fixed pulse shape that has the amplitude proportional to the deposited energy within the cell. Therefore, by estimating the signal amplitude of the received signal, the cell energy can be found. Ideally, the pulse parameters are deterministic, except for the amplitude.

Actually, fluctuations, such as electronic noise and pulse parameters (time of arrival and deformities), are present and can be measured by using test beam or calibration data taking [6, 7]. Although the effect of some of these sources of uncertainties can be minimized, they unavoidably deteriorate the energy resolution. Choosing the energy estimation algorithm involves overall estimation performance and implementation feasibility (in case of online operation), in order to meet the memory and processing time constraints that are often imposed by a trigger system [8, 9].

The standard methods to estimate the amplitude of the readout signals are based on the minimization of the estimator variance [10–14], which are referred to as optimum filtering. Therefore, they achieve optimum performance in cases where the shape of the pulse of interest is invariant and the additive background noise is Gaussian. It is worth mentioning that many channels are likely to have low energy deposition or even just noise for a given event and the correct cell selection impacts on the reconstruction of the physics objects during offline analysis. Thus, cells sampling only noise must be suppressed from event reconstruction.

In this paper, a Matched Filter (MF) based approach for energy reconstruction is developed. The MF technique is known to be the optimum signal detector (in terms of signal-to-noise ratio, SNR) for cases where the received signal is corrupted from additive noise [15]. An approximation of this technique was previously developed for electron/pion separation using the spaghetti calorimeter [16] and has recently been applied in a muon trigger assisted from hadronic calorimetry [17]. Unlike standard energy reconstruction methods in calorimetry, the MF approach is not limited to Gaussian random processes. Therefore, different noise sources can be incorporated in its design, leading to a more robust estimation method. The proposed MF approach treats the detection of incoming calorimeter signals as a general random process detection problem, which allows to properly handle signal fluctuations. For the cases where computation resources are limited (envisaging online operation, for instance), and the calorimeter shaped signals are sufficiently stable, a deterministic signal approximation for the MF design may be considered.

The first goal of the proposed technique is the selection of the calorimeter cells that should be used for event reconstruction. The MF approach handles this detection problem as a signal against noise hypothesis testing. Additionally, as a result of the signal modeling, it turns out that the MF output can be calibrated in terms of the signal amplitude and thus be also used to estimate the deposited energy for cells that passed the detection condition. While classical methods perform the signal detection based on the amplitude estimates (using an energy threshold), the proposed approach initially solves the detection problem in an optimum way and then derives the amplitude estimation for cells with relevant physics information.

2021 JINST 16 P02016

The rest of the paper is organized as follows. The next section introduces typical digital energy reconstruction methods applied in modern calorimeters. Section 3 presents the MF theory, as well as the proposed approach for energy estimation. Section 4 presents the simulation data used for performance evaluation. Results are given in terms of both signal detection efficiency and energy estimation error. Simulation data considered a low SNR condition for assessing the signal detection capability, as the presence of small energy levels in a given cell might be masked by the noise. In order to evaluate the energy estimation in practical operation, a large dynamic range data set was also produced. A comparison between MF and the optimum filter reconstruction methods is also carried out. Conclusions are derived in section 5.

2 Energy reconstruction in segmented calorimetry

This section introduces the basic concepts in the context of the energy estimation in modern calorimeters. Firstly, the signal conditioning of the collected analog pulse is presented. Then, the digital treatment to extract the parameters of interest is outlined.

2.1 Signal processing

The information from a particle interaction with the calorimeter is provided in the form of electrical signals. In order to extract the relevant information, such as the signal amplitude and timing, the readout pulse must be processed. The amplitude of the shaped pulse (unipolar or bipolar) is supposed to be proportional to the deposited energy within the cell [18–20]. Additionally, the width of the shaped pulse is defined in order to minimize the signal pile-up on the tails of preceding pulses (in the case of high-event rate experiments). The stability of such pulses may also be affected by electronics aging or miscalibration, which can cause phase shifts and deformation of the original shaped pulse.

The fast analog signals are converted into digital samples with a given sampling frequency. A window of certain number of samples representing the entire pulse (or partially) is read out for every event. These samples are used to estimate the parameters of interest such as the amplitude and time of the incoming signals for a given readout channel.

2.2 Energy reconstruction

The calorimeter signal is usually modeled from a first-order approximation [10]:

$$r_k = A(h_k - \tau \dot{h}_k) + n_k + ped \qquad k = 0, 1, 2, \ldots, N-1 \tag{2.1}$$

where r_k represents the received digital signal at sample k and N is the total number of samples within the acquisition window. The amplitude A is the parameter to be estimated, while n_k is the noise. Here, n_k comprises the readout channel noise, which is typically modeled by a Gaussian distribution [21, 22]. The parameters h_k and $\dot{h}_k$ are, respectively, the reference pulse shape and its time derivative (linear approximation for the pulse phase), while the parameter τ is the signal phase. The variable ped corresponds to the signal pedestal and it is a constant value added to the analog signal just before its analog-to-digital conversion. Therefore, we assume n_k as a zero-mean random process.

2021 JINST 16 P02016

As the noise is often assumed stationary and modeled by a Gaussian distribution, variance minimization techniques are extensively employed for energy reconstruction since they perform close to the optimal operation. It should be stressed that signal shape fluctuations also introduce uncertainties to the final energy estimation, but they are not taken into account in standard energy estimator designs.

Most algorithms for energy reconstruction in calorimeter systems are based on optimum filtering [10, 11]. The signal amplitude $\hat{A}$ is estimated through a linear combination of the N samples:

$$\hat{A} = \sum_{k=0}^{N-1} w_k r_k \tag{2.2}$$

The weights w_k are obtained from the front-end pulse shape and the noise covariance matrix. The procedure aims at minimizing the variance of the amplitude distribution. Thus, these algorithms are optimal for deterministic signals corrupted from Gaussian noise. The correct weights are computed by minimizing the effect of the noise in the amplitude reconstruction.

For a linear and unbiased estimator, it is required that the expected value of $\hat{A}$ to be A ($E[\hat{A}] = A$). Therefore, equation (2.3) can be derived from substituting equation (2.2) into equation (2.1) for an Optimal Filter (OF).

$$E[\hat{A}] = \sum_{k=0}^{N-1} (A w_k h_k - A\tau w_k \dot{h}_k + w_k ped) \tag{2.3}$$

The following constraints can be deduced in order to reach the condition $E[\hat{A}] = A$:

$$\sum_{k=0}^{N-1} w_k h_k = 1 \tag{2.4}$$

$$\sum_{k=0}^{N-1} w_k \dot{h}_k = 0 \tag{2.5}$$

$$\sum_{k=0}^{N-1} w_k = 0 \tag{2.6}$$

These constraints are added to the minimization procedure and they make the estimator immune against pedestal and phase shifts. On the other hand, the imposition of these constraints may increase the estimator variance. Thus, for example, in cases where the electronic pedestal value can accurately be measured and subtracted from the incoming digitized signals, the constraint expressed in equation (2.6) may be removed, increasing the number of degrees of freedom of the optimization procedure.

Regardless which constraints are included for the optimization procedure, the variance of the estimator is given by,

$$var(\hat{A}_{OF}) = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} w_k w_j C_{kj} = \mathbf{w}^t \mathbf{C} \mathbf{w} \tag{2.7}$$

where $\mathbf{C}$ corresponds to the noise covariance matrix and $\mathbf{w}$ is the weight vector. The superscript t represents the transpose operator. Hence, to find the optimal weights, equation (2.7) is often

2021 JINST 16 P02016

minimized subjected to the constraints in equations (2.4), (2.5) and (2.6) using Lagrange multipliers [23]. Most of the modern calorimeters use the above described algorithm to reconstruct the signal amplitude. However, each algorithm is slightly modified from one system to another to estimate the final amplitude.

For the sake of performance comparisons, three different OF algorithms are considered in section 4. The first one, which is referred to here as Method 1, does not use the pedestal constraint (equation (2.6)), but, instead, it applies a baseline value previously computed and stored in a database. Therefore, the parameter ped can be subtracted from every incoming received time samples of $\mathbf{r}$, producing:

$$\mathbf{r'} = \mathbf{r} - ped \tag{2.8}$$

The second one, called Method 2, does not use the equation (2.6) in its optimization procedure either. Similarly to Method 1, it subtracts the baseline value from the incoming time samples according to equation (2.8). Therefore, since Method 1 and Method 2 do not use the baseline information in the optimization procedures, both minimization processes lead to the same weights. However, unlikely for Method 1, Method 2 estimates the baseline (ped value) on an event-by-event basis, using some pre-samples right before the physics pulse, instead of using specific runs for this purpose, as performed in Method 1. Finally, the Method 3 uses the pedestal constraint in its design (it does not use any estimate of the baseline value), leading to a different set of weights.

3 The matched filter method

This section develops the MF approach for signal detection applied on calorimeter signals. In the end of the section, a relationship between the MF output and the signal amplitude is derived, aiming at providing the reconstructed energy value.

3.1 Signal detection

Unlike the methods described in section 2.2, where the calorimeter response time samples h_k are taken from a deterministic function, the MF method models the calorimeter response as a random process [24] G that incorporates the pulse uncertatinties, such as pulse phase shifts and deformities. Therefore, the process realizations correspond to realistic calorimeter response samples in which phase shifts and deformities are present. Hence, the signal description represented by equation (2.1) may be stated as:

$$r_k = Ag_k + n_k + ped \qquad k = 0, 1, 2, \ldots, N-1 \tag{3.1}$$

where both g_k and n_k correspond to realizations of the time samples from the random processes associated to signal and noise, respectively. The parameter A is the incoming signal amplitude, which is to be estimated (see subsection 3.2).

The general detection problem can be described as an hypothesis testing applied to the received signal, considering the presence of additive noise [15] and $k = 0, 1, 2, \ldots, N-1$:

$$\begin{aligned} H_0 &: r_k = n_k + ped \\ H_1 &: r_k = Ag_k + n_k + ped \end{aligned} \tag{3.2}$$

where H_0 represents the null-signal hypothesis, while H_1 corresponds to the full-signal hypothesis.

2021 JINST 16 P02016

By considering **r'** (see equation (2.8)) a given outcome at the receiver input as it was mentioned before (the parameter ped can be computed beforehand) and, based on the Neyman-Pearson design approach [15], the relationship that maximizes the detection efficiency with respect to SNR is given by the likelihood ratio test [25]:

$$\Lambda(\mathbf{r'}) = \frac{p(\mathbf{r'}|H_1)}{p(\mathbf{r'}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{3.3}$$

The terms $p(\mathbf{r'}|H_1)$ and $p(\mathbf{r'}|H_0)$ are the probability density functions for full-signal hypothesis H_1 and null-signal hypothesis H_0, respectively, given the data vector **r'** was received. The parameter γ is a detection threshold. The detection system decides in favor of hypothesis H_1 if the likelihood ratio of the received signal is greater than (or equal to) the detection threshold, and otherwise in favor of H_0. In such a way, the likelihood ratio maximizes the detection probability and minimizes the error detection probability.

The required likelihood functions are unknown for the majority of real detection problems. Therefore, the detection algorithm should estimate both $p(\mathbf{r'}|H_1)$ and $p(\mathbf{r'}|H_0)$ from a set of known data. Additionally, if the noise samples are correlated, a pre-processing step (whitening filter) is required prior to MF operation, aiming at decorrelating the noise samples and allowing the likelihood ratio test to profit from orthogonal statistical information [15]. In this sense, the received input signal **r'** after the whitening filter is represented as **r''**. In case the noise samples are uncorrelated, the whitening operation may be suppressed since it leads to the identity matrix. It is worth mentioning that the MF design can be developed for any calorimeter response shape.

3.1.1 Stochastic design

Assume the random process G is Gaussian and independent of the additive noise process. As the signal to be detected is a random process, it can be expressed into the Karhunen-Loeve (KL) series [26] in order to access its uncorrelated components. Then, a MF can be designed for matching each series component and resulting projections weighted for signal decision, according to their signal energy description (optimizing SNR) [27]. This can be viewed as matching the detector to the Gaussian part of the incoming signals. In addition, the shaper circuit may produce Gaussian shaped signals according to the experiment requirements [28], which improves the matching and makes the filter to operate close to optimum.

Typically, KL representation is performed through Principal Component Analysis (PCA) [29]. The process G is mapped onto the principal directions, which are ranked according to the signal variance they carry. The eigenvalues (λ_k) and eigenvectors ($\mathbf{\Phi}_k$) of the covariance matrix of G give, respectively, the represented variance and mapping vector directions. After some manipulations, the likelihood ratio test defined in equation (3.3) may be expressed in two terms: deterministic (I_D) and stochastic (I_R) [27]:

$$\Lambda(\mathbf{r''}) = I_D + I_R \tag{3.4}$$

$$I_D = \mathbf{m}\left[\sum_{k=0}^{N-1} \frac{1}{\lambda_k + N_0/2} \mathbf{\Phi}_k \mathbf{\Phi}_k^t\right] (\mathbf{r''})^t \tag{3.5}$$

$$I_R = \frac{1}{N_0}\mathbf{r''}\left[\sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + N_0/2} \mathbf{\Phi}_k \mathbf{\Phi}_k^t\right] (\mathbf{r''})^t \tag{3.6}$$

2021 JINST 16 P02016

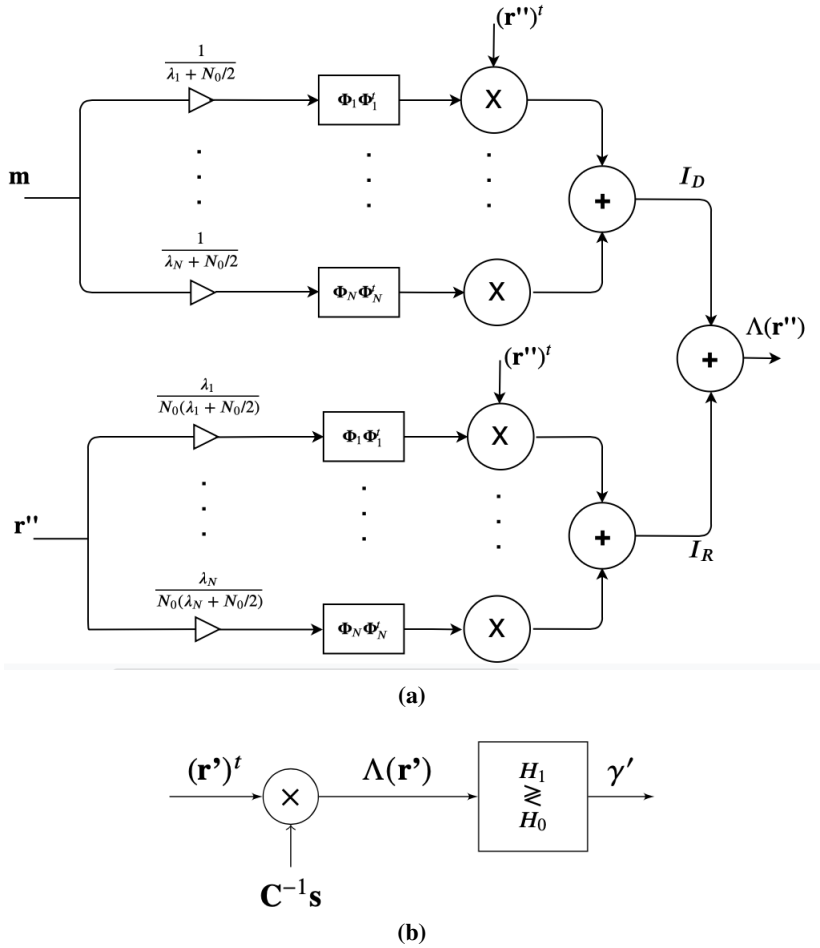


Figure 1. Block diagrams for stochastic (a) and deterministic (b) MF hypothesis testings.

where **m** corresponds to the normalized mean vector (unitary amplitude) of the random signal process (G) after the whitening operation, and $N_0/2$ is the white noise spectral density. Figure 1a exhibits the block diagram for the MF signal processing chain and how the I_D and I_R terms are obtained. In order to evaluate its matching with respect to each signal component, the whitened received signal is projected onto the principal components, which reconstruct the signal random process in terms of orthogonal directions. The mean signal is also projected onto the principal components for allowing the deterministic projection to be added to such random signal part, so that the received signal is properly compared to the detection threshold.

It is known that PCA allows for data compaction, as the less energetic components may be discarded [29]. Thus, the filter design may be simplified when only $M(M < N)$ components expressing signal variance above a given threshold are retained. Online applications may benefit from such compact MF design.

The MF approach is developed using the probability density function of the received signals. Therefore, it can be designed for any additive noise model. As from equation (3.3), no Gaussian assumption is made and non-Gaussian random noise processes could also be supported. This non-Gaussian condition may arise, for instance, when there are contributions from non-electronic

2021 JINST 16 P02016

sources, such as the pile-up signal from previous events that appear in the same readout window [30]. In case of noise non-gaussianity, one could still apply the same MF strategy (see figure 1a), which relies on the KL transformation. Thus, the Gaussian part of the noise would be mapped onto the principal components for evaluating the filter components and producing the corresponding response, which would be suboptimal as the nonlinear statistics were not fully handled. On the other hand, full access to the high-order statistics available from the non-Gaussian noise probability density function could be accomplished by replacing the KL transformation by means of a BSS (Blind Source Separation) [31] algorithm, for instance, and the implementation complexity would not be increased.

3.1.2 Deterministic approximation

Here, it is assumed that the signal of interest may be considered as deterministic. This approximation may be valid for calorimeter systems where the shaper circuit provides a fixed and stable pulse at its output. In this case, the received signal represented by equation (2.1) may be now rewritten as

$$r_k = As_k + n_k + ped \qquad k = 0, 1, 2, \ldots, N-1 \tag{3.7}$$

where $s_k = h_k - \tau \dot{h}_k$ represents the normalized (unitary amplitude) samples from the approximated calorimeter response function, assumed as deterministic, and n_k remains as the samples from the Gaussian random noise process. The signal time samples (s_k) can be computed from an average process using a high SNR data set, incorporating both phase and deformities uncertainties from the actual random signal process.

Thus, the noise second-order statistics determines the fluctuations on the received signals. The hypothesis test becomes:

$$\begin{aligned} H_0 &: r'_k = n_k \\ H_1 &: r'_k = As_k + n_k \end{aligned} \tag{3.8}$$

Hence, the corresponding distributions under H_0 and H_1 can be computed as:

$$p(H_1|\mathbf{r'}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp\left(-\frac{1}{2}(\mathbf{r'} - A\mathbf{s})^t \mathbf{C}^{-1} (\mathbf{r'} - A\mathbf{s})\right) \tag{3.9}$$

$$p(H_0|\mathbf{r'}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp\left(-\frac{1}{2}\mathbf{r'}^t \mathbf{C}^{-1} \mathbf{r'}\right) \tag{3.10}$$

The detection problem simplifies significantly, as the likelihood ratio test becomes:

$$\Lambda(\mathbf{r'}) = \frac{\exp\left(-\frac{(\mathbf{r'}-A\mathbf{s})^t \mathbf{C}^{-1} (\mathbf{r'}-A\mathbf{s})}{2}\right)}{\exp\left(-\frac{\mathbf{r'}^t \mathbf{C}^{-1} \mathbf{r'}}{2}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \phi \tag{3.11}$$

and the decision is given by:

$$\mathbf{r'}^t \mathbf{C}^{-1} \mathbf{s} \underset{H_0}{\overset{H_1}{\gtrless}} \phi' \tag{3.12}$$

where ϕ' represents the output of the MF deterministic approximation.

2021 JINST 16 P02016

This is known as the MF for detecting a deterministic signal corrupted from Gaussian noise [15]. The operations to achieve the hypothesis testing output are shown in figure 1b. For normalized white Gaussian noise, the covariance matrix $\mathbf{C}$ is the identity matrix and can be suppressed. As a result, the MF weights simply become the normalized averaged signal $\mathbf{s}$ itself. For the case where the Gaussian noise is correlated, the inverse of the covariance matrix has a pre-whitening function [32] and the elements of the vector $\mathbf{C}^{-1}\mathbf{s}$ become the MF coefficients.

3.2 Amplitude estimation from MF output

Both stochastic and deterministic outputs may be used to assess the signal amplitude (energy), as exploited next. It is worth to remind that the estimation step is applied only for calorimeter cells for which a given received signal is detected (above threshold) from the detection step.

3.2.1 Stochastic design

In the presence of a signal of interest (H_1 hypothesis), the stochastic MF operation can be summarized by the following equation:

$$I_D + I_R = \mathbf{b}_1(\mathbf{r''})^t + b_2(\mathbf{r''})^t(\mathbf{r''}) = \theta \tag{3.13}$$

where $\theta = \Lambda(\mathbf{r''})$ and the elements $\mathbf{b}_1$ and b_2 correspond to:

$$\mathbf{b}_1 = \mathbf{m}\left[\sum_{k=0}^{N-1} \frac{1}{\lambda_k + N_0/2}\mathbf{\Phi}_k\mathbf{\Phi}_k^t\right] \tag{3.14}$$

$$b_2 = \frac{1}{N_0}\sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + N_0/2}\mathbf{\Phi}_k\mathbf{\Phi}_k^t \tag{3.15}$$

Since the signal amplitude corresponds to the information to be recovered, the received signal can be represented as the normalized mean vector $\mathbf{m}$ scaled by an amplitude A, or $E[\mathbf{r''}] = A\mathbf{m}$, which leads to the following equation:

$$(\mathbf{b}_1 A\mathbf{m})^t + (b_2 A\mathbf{m})(A\mathbf{m})^t = \theta \tag{3.16}$$

Hence, since $\mathbf{m}\mathbf{m}^t = 1$, the stochastic amplitude estimation can then be found by solving equation (3.16) for the variable A:

$$\hat{A}_{MF(s)} = \frac{-\mathbf{b}_1\mathbf{m}^t + \sqrt{((\mathbf{b}_1\mathbf{m}^t)^2) + 4(b_2\theta)}}{2b_2} \tag{3.17}$$

It should be stressed that only the positive part of the square root function is used, as the amplitude estimation requires that $A \geq 0$. As a result, unlike for OF approaches, the stochastic MF method is particularly robust against the undesired negative energy estimation.

3.2.2 Deterministic approximation

For the deterministic approximation, the received signal is represented according to equation (3.7). Since the noise time samples n_k account for the noise, $E[n_k] = 0$, as:

$$E[\mathbf{r'}] = A\mathbf{s} \tag{3.18}$$

2021 JINST 16 P02016

where **s** is the averaged normalized pulse (unitary amplitude). Thus, the detection operation for the deterministic MF approximation (equation (3.12)) can be written as:

$$E[(\mathbf{r'})^t \mathbf{C}^{-1}\mathbf{s}] = (A\mathbf{s})^t \mathbf{C}^{-1}\mathbf{s} \tag{3.19}$$

and by solving this equation for the variable A, it leads to the expression shown by equation (3.20), which is the deterministic approximated estimation of the input pulse amplitude. The noise accounts for the estimation error.

$$\hat{A}_{MF(d)} = \frac{(\mathbf{r'})^t \mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^t \mathbf{C}^{-1}\mathbf{s}} \tag{3.20}$$

The numerator $(\mathbf{r'})^t \mathbf{C}^{-1}\mathbf{s}$ is the usual MF detector, whereas the denominator $\mathbf{s}^t \mathbf{C}^{-1}\mathbf{s}$ can be seen as a constant that normalizes the MF operation in order to recover the original signal amplitude.

4 Simulation results

The efficiency achieved by each method was evaluated in terms of both signal detection and energy estimation for a single calorimeter readout channel. Data come from a toy Monte Carlo simulation. Two data sets were built: one comprising low SNR signals aiming at studying the signal detection efficiency, which may be critical for low amplitude signals corrupted from noise, and another one considering high SNR conditions from which the energy estimation performance is evaluated for a large energy range. From the latter, the practical operation (signal detection followed by energy estimation) was emulated. As Method 1 and Method 2 come from the same optimization procedure, they have the same set of coefficients. Whenever differences in performance are not significant for both methods, only Method 1 is shown in plots, in order to improve the visibility.

4.1 Low SNR data set

For signal detection analysis, separate data sets were formed from noise only and signal without noise samples comprising 200,000 observations each. In a real application, where usually pure signal samples (without noise) are not available, high SNR signals corrupted with noise may be used to build such signal data set. The noise process is modeled as zero-mean Gaussian where correlation between noise samples is admitted up to the 50% level. The noise variance was set to 30 MeV, which is the noise level found in modern electromagnetic and hadronic calorimeter systems [12, 33]. The noise set was used to estimate the pedestal value, the noise covariance matrix **C** for the MF whitening transformation, the noise spectral density and OF optimization procedures. For MF designs and Method 1, the pedestal value was estimated by taking the average of the first time sample of the incoming signals of the noise development set. For Method 2, a larger acquisition window is used, from which the three additional samples were acquired as a pre-trigger (before the signal pulse) for estimating the pedestal through an averaging process.

For the pure signal data set, a 180 ns slightly asymmetric unipolar pulse, sampled at 40 MHz, was synthetized. This pulse shape strategy has been chosen in different modern calorimeters [18–20]. Therefore, seven time samples were enough to acquire the entire pulse in an acquisition window. A signal phase shift was simulated using a zero-mean Gaussian distribution with standard deviation of 12.5 ns (half of the bunch crossing), in order to add time uncertainties to the readout signal. In

2021 JINST 16 P02016

addition to that, a small pulse deformity was introduced to the simulation on each digitized sample, also using a zero-mean Gaussian distribution, but with standard deviation of 2% with respect to the associated time sample index from the reference pulse shape. The energy of the signal was randomly taken from an exponential distribution with mean value equals to 120 MeV.

In order to assess the statistical fluctuations, a cross-validation method was carried out, where the signal and noise data sets were randomly shuffled separately before being split into development and test sets [34]. A 10-fold cross-validation was performed, where both signal and noise data sets were equally split into 10 subsets. A single subset was separated from each data set (noise and signal) in order to form the test set for efficiency evaluation, while the remaining subsets formed the development set. For performance evaluation the test sets were used for building the received signals. For this, the signal observations from the pure signal test set were corrupted by adding the realizations of the noise random process taken from the noise test set. The signal development set was used to perform PCA and extract signal decomposition into its eigenvectors. In the cross-validation procedure, each subset was acting once as the testing data. For quoting the different performance measures, the average and RMS values obtained from the ten-fold cross-validation procedure were taken as the performance estimates and the corresponding error bars, respectively.

4.2 High SNR data set

Aiming at estimating the cell energy along a large dynamic range, a data set containing 200, 000 high SNR events was built. In this data set, the received signals were simulated from signal realizations with additive noise. The parameters used for simulating both the signal pulses and the noise were as from section 4.1. However, the signal energy values were taken from an exponential distribution whose mean value was 3 GeV [1, figure 5.58, page 390].

4.3 Design method and performance measures

In order to describe the noise second-order statistics, the noise covariance matrix was estimated through a robust approach [35] using the noise development set. As for the MF designs, the noise whitening transformation $\mathbf{W}$ was obtained from [32]:

$$\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\mathbf{V}^{t} \tag{4.1}$$

where matrices $\mathbf{D}$ and $\mathbf{V}$ contain the eigenvalues (diagonal) and unit-norm eigenvectors of the noise covariance matrix, respectively. The noise cross-talk $n_{ct}(\%)$ was used as a measure of the whitening quality. It was computed from the nondiagonal noise covariance matrix elements before and after applying the whitening transformation [36]. Since the noise covariance matrix is symmetric, only its upper triangular block was considered:

$$n_{ct}(\%) = \frac{\sum_{i=1}^{7}\sum_{j=i+1}^{7}\mathbf{C}(i,j)}{\sum_{i=1}^{7}\sum_{j=i}^{7}\mathbf{C}(i,j)} \times 100 \tag{4.2}$$

Concerning signal detection, one can assess the number of principal components that account for most of the process variance and those that may be discarded from the accumulated energy

2021 JINST 16 P02016

in the eigenvectors of the signal covariance matrix [29]. The Receiver Operating Characteristics (ROC) [15, 37] and the area under the ROC curve (Area Under the Curve (AUC)) [38] were both employed as measures for comparing detection capabilities. The ROC curve shows the signal detection probability (P_d) against the false alarm (F_a) rate (noise misdetected as signal) for a binary detection system. The P_d for a given energy threshold was estimated from the number of signals correctly detected (above threshold) divided by the total number of received signals. The F_a rate was computed from the number of noise events misclassified as signals divided by the total number of noise events under evaluation. An AUC value equals to one represents a perfect test. A performance measure that provides an operation point for balanced signal detection for both P_d and F_a is achieved through the sum-product (SP) index, which was computed for binary detection according to [39]:

$$SP = \sqrt{\frac{P_d + (1 - F_a)}{2}\sqrt{P_d(1 - F_a)}} \tag{4.3}$$

The SP value was computed for every pair of P_d and F_a values from the ROC curve, and its largest value indicated the optimal threshold value to be used. The larger the SP value is, the better trade-off between P_d and F_a is achieved.

The amplitude estimation performance was evaluated from the filters designed from the signal detection phase by measuring the estimation error and analyzing the error uniformity over a large energy range (high SNR data). The estimation error was computed by measuring how far the estimate values deviate from the reference values (true energy values taken from simulation). One way of performing such measurement is to subtract each energy estimate from its associated reference value. The characteristics of the estimation error distributions may be measured in terms of kurtosis [40], in order to explore high-order statistics. Moreover, the relative error (in percent) described by:

$$RE(\%) = \frac{A_{\text{estimated}} - A_{\text{truth}}}{A_{\text{truth}}} \times 100 \tag{4.4}$$

as a function of the true signal phase was also explored. The parameters $A_{estimated}$ and A_{truth} correspond to the amplitude estimated by the algorithms and the true amplitude value known from the simulation, respectively. Additionally, a linearity evaluation was performed to show the estimation error dependence as a function of the energy.

Another interesting point in energy estimation performance is the quality of the event reconstruction. Here, the quality factor (QF) measures how well the input pulse (time samples) is reconstructed [41]. If the noise is Gaussian, the error between the input samples and the respective estimates leads to a χ^2 distribution:

$$QF = \frac{1}{\hat{A}}\sqrt{\sum_{k=0}^{N-1}\frac{((x_k + ped) - r_k)^2}{N}} \tag{4.5}$$

where the elements x_k and r_k represent the reconstructed and received samples (vector $\mathbf{r}$), for a given time sample k, respectively. The ped parameter corresponds to the baseline estimation performed by each method. In practice, apart from selecting good reconstructed signals, the quality factor may also be used to flag corrupted data [42].

An emulation of practical operation was also performed. For OF methods, the detection step was based on a threshold applied on the estimated energy value. For the proposed MF approaches,

2021 JINST 16 P02016

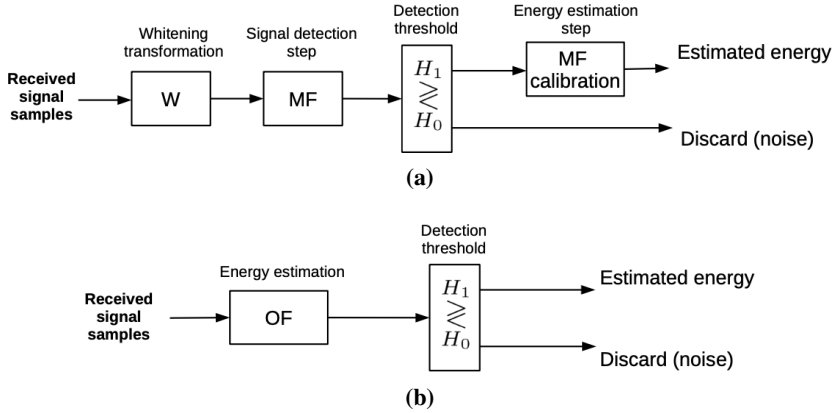


Figure 2. Emulation for calorimeter signal detection and energy estimation strategies from (a) MF designs and (b) standard referenced methods.

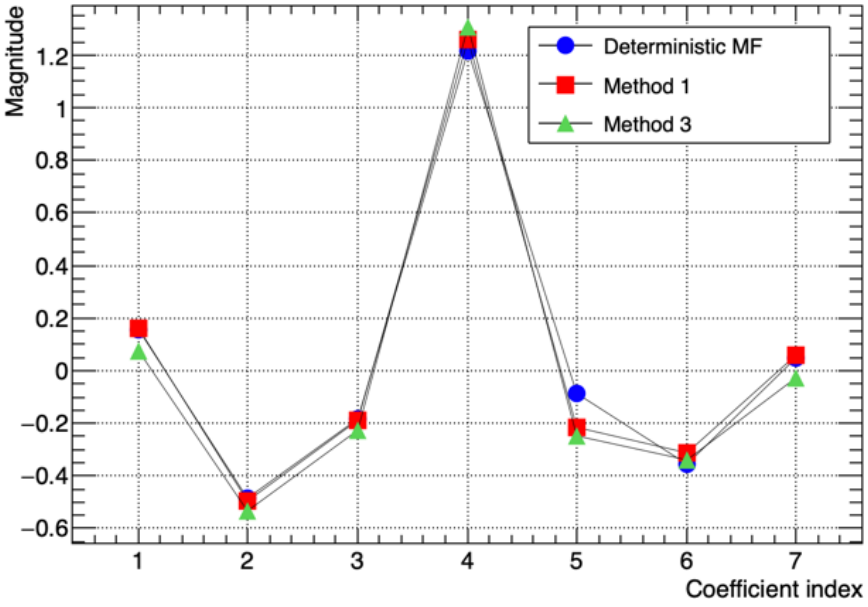
the calorimeter cells were assigned to either signal or noise according to the likelihood ratio test that maximized the SNR. For the received signals that were approved by the likelihood ratio test, the energy was recovered. Otherwise, the signal was discarded as noise. Figure 2 shows the emulation strategies adopted from MF based techniques and standard methods.

4.3.1 Filter coefficients

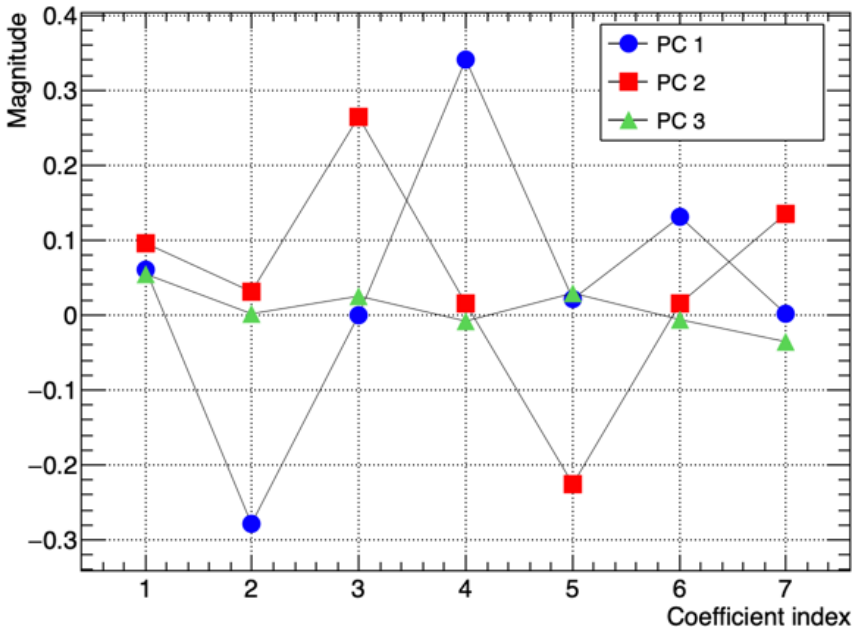
For the filter designs, apart from the noise covariance information used for the MF whitening processing step besides the OF optimization procedures, the information concerning the expected signal pulse shape is needed. In the deterministic approaches, this information comes from a normalized representative version of the signal of interest. For the OF techniques, a normalized pulse acquired from several acquisitions of the readout shaper circuit output was carried out through a high-rate sampling process, whereas for the deterministic MF a normalized average pulse was computed from the signal development data set, but only signals above 100 MeV (approximately three times the noise standard deviation) were considered to avoid noise contamination. Additionally, in order to have an accurate estimation of the original pulse shape to be matched, pulse average was formed from time aligned pulses. For this, the maximum pulse first-order derivatives were made to coincide in time. Figure 3a shows the filter coefficients computed for the deterministic approaches, which are quite similar. It is worth noticing that the fifth sample is considered slightly more relevant in the weighted signal processing by the deterministic MF approximation.

For guiding the PCA, figure 4 shows the accumulated variance according to the number of principal components (PCs) to be considered for MF design. It can be seen that three components retained roughly 94% of the signal variance. Therefore, in case data are mapped onto the principal components and dimension reduction is exploited, one may expect that the likelihood ratio test efficiency will not be degraded significantly if only three components are used. In case of online operation, the number of components to be retained impacts on the computational effort needed to compute the likelihood ratio test. As an example, table 1 summarizes the number of multiplications required to compute the likelihood ratio test by the different methods, including the whitening operation. The use of three instead of seven components produces a considerable reduction in the number of multiplications needed. Additionally, the approximated deterministic MF presents a

2021 JINST 16 P02016



(a)



(b)

Figure 3. Coefficients for amplitude estimation of the deterministic methods (a) and the three most energetic principal components (PCs) used for the stochastic MF (b).

2021 JINST 16 P02016

competitive option with respect to OF methods regarding the computational cost, being attractive for online operation where the resources may be limited.

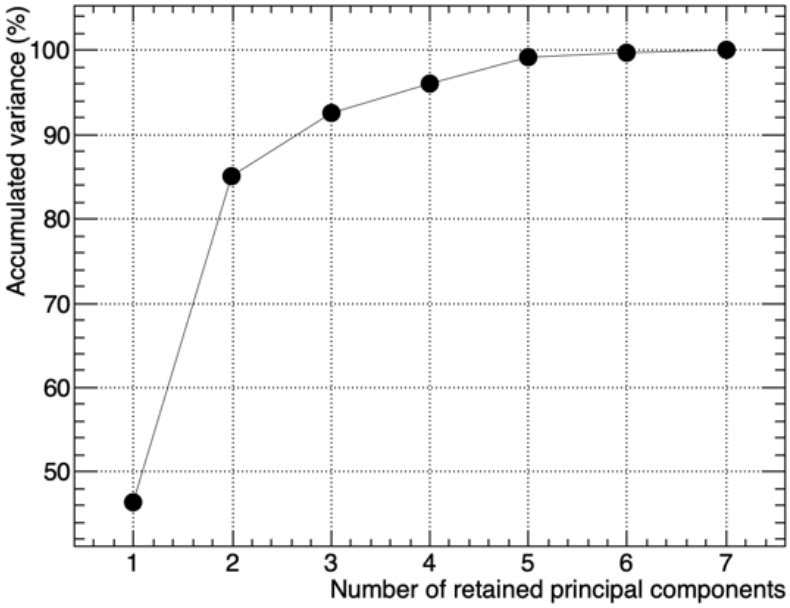


Figure 4. Accumulated variance on PCs for signal description.

Table 1. Number of multiplications required by each method, for which the stochastic design is evaluated per number of retained principal components (PCs).

Method	**Number of multiplications**
Stochastic MF (7 PCs)	98
Stochastic MF (3 PCs)	42
Stochastic MF (2 PCs)	28
Deterministic MF	7
OF methods	7

Concerning the stochastic MF coefficients, the random signal process is represented through its principal components. Figure 3b shows the three most energetic components weighted by their associated eigenvalues (projected signal variance). It should be noticed that the second PC pointed out that the third signal sample is worth to be weighted significantly, although this was not revealed by the deterministic approaches.

4.4 Performance comparisons

The cross-talk in the noise samples before and after applying the whitening transformation to the noise test set was 54.73% and 0.23%, respectively. This points out that the noise cross-talk is considerably suppressed after whitening.

2021 JINST 16 P02016

4.4.1 Signal detection

Figure 5 shows the ROC curves for the methods considered in this work applying the low SNR data set. The data points correspond to the averaged value taken from the 10-fold cross-validation process. The stochastic MF presented the best detection performance among the considered methods, due to the SNR improvement. Particularly, since the deterministic MF approximation computed its pulse representative through an average process that partially incorporated the signal uncertainties, it represented the superior average envelope for the performance of the deterministic strategies. On the other hand, as expected, Method 3 represented the inferior average performance envelope, as a consequence of its additional constraint imposed to the optimization procedure, which increases its variance. In order to have a flavour on how much was gained in terms of signal detection ability, both the best OF method (Method 1) and the MF outputs are displayed in figure 6. To this end, the stochastic MF output was calibrated to recover the energy information (see equation (3.17)). It can be seen that the MF output for noise events (hypothesis H_0) is considerably more concentrated, so that low-amplitude signals can be better reconstructed.

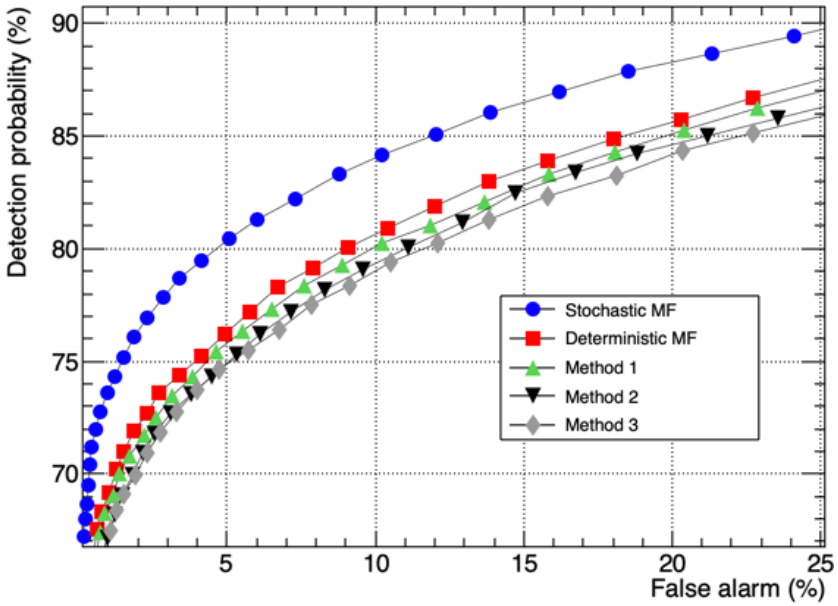


Figure 5. ROC curves for the evaluated methods.

Detection efficiencies from the cross-validation procedure are shown in table 2. It can be seen that the stochastic MF (using all principal components) surpassed all the other techniques (better signal detection capability). This is explained by the fact that the stochastic approach takes into account the statistics from the random signal process, while the other methods assume the signal of interest as purely deterministic or provide a deterministic approximation of the random process. The reconstructed energy performance for the test set is also shown in table 2.

The OF methods 2 and 3 presented similar but slightly lower detection efficiency values, which may be explained by the way these methods deal with the signal baseline. Method 3 makes use of a constraint to become immune against pedestal fluctuations and, therefore, it increases its amplitude estimation variance. The other OF approaches do not take into account such constraint and rely on

2021 JINST 16 P02016

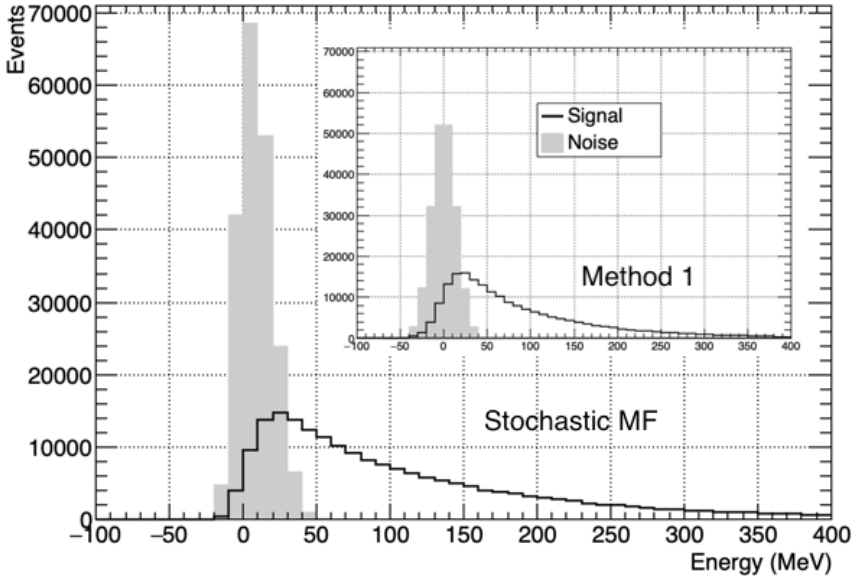


Figure 6. The stochastic MF output. The output for the OF Method 1 is also shown embedded to the figure.

Table 2. Detection performance in terms of AUC, detection probability (P_d) for 10% false alarm (F_a), the maximum SP index and reconstructed energy performance.

Method	**AUC (%)**	**P_d (%) for 10% F_a**	**SP (%) (max.)**	**Estimation error (MeV)**
Stochastic MF	92.39 ± 0.19	83.13 ± 0.46	75.52 ± 0.32	-2.30 ± 18.21
Deterministic MF	90.16 ± 0.22	79.75 ± 0.36	72.41 ± 0.30	-21.81 ± 49.40
Method 1	89.79 ± 0.18	79.10 ± 0.42	71.94 ± 0.25	-25.63 ± 51.77
Method 2	89.48 ± 0.18	78.14 ± 0.43	71.28 ± 0.36	-25.63 ± 51.98
Method 3	89.18 ± 0.24	78.09 ± 0.46	70.91 ± 0.30	-27.03 ± 54.53

a value computed apart. However, computing the pedestal value from a dedicated data taking, with a large number of observations (Method 1), instead of event-by-event basis (Method 2), revealed to be a slightly better option.

The signal detection efficiencies found for the deterministic MF and Method 1 were very similar. Although these methods come from different approaches (likelihood ratio test for MF and variance minimization for Method 1), they make similar assumptions about the signal characteristics. While Method 1 uses a constraint to compensate for phase fluctuations (equation (2.5)), the deterministic MF incorporates the pulse fluctuations (phase plus deformities) in the averaged pulse (**s**) used in equation (3.7). Moreover, both methods use the pedestal computed from the noise development set and assume the signal of interest as deterministic.

Concerning the number of PCs to be used in the MF design, figure 7 shows the ROC curves considering a different number of PCs. It can be noted that the random signal process was roughly described by the most energetic component (1 PC) and, thus, the likelihood ratio test performed suboptimally. However, when the stochastic was designed from the two most energetic components

2021 JINST 16 P02016

onwards, the detection performance became close to optimal (when all seven components are used). Table 3 summarizes such comparison and shows how much was gained as more components were retained. It should be stressed that only two components were enough for achieving a better detection efficiency with respect to the OF designs.

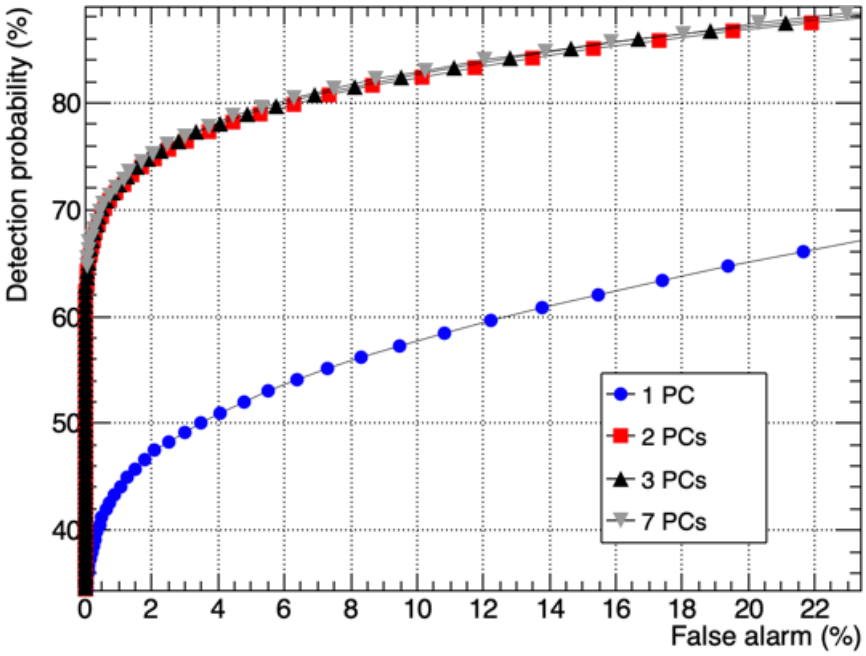


Figure 7. ROC curves for stochastic MF design when considering an increasing number of PCs.

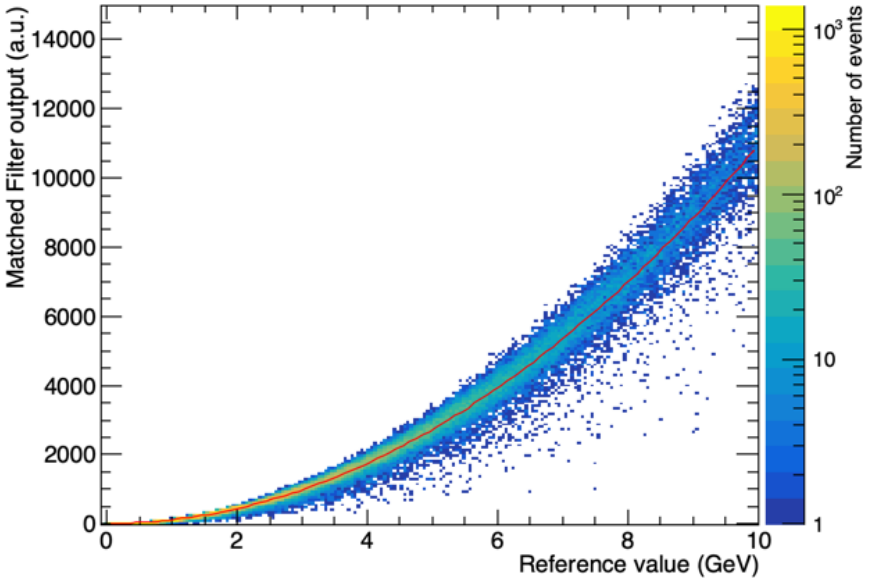
Table 3. Detection capability in terms of AUC and the detection probability (P_d) for 10% false alarm probability, as the number of retained principal components (PCs) was increased.

Number of PCs	AUC (%)	P_d (%) for 10% F_a
1 component	78.63 ± 0.36	59.92 ± 0.48
2 components	91.70 ± 0.24	82.89 ± 0.29
3 components	92.01 ± 0.09	83.11 ± 0.16
7 components	92.39 ± 0.19	83.13 ± 0.46

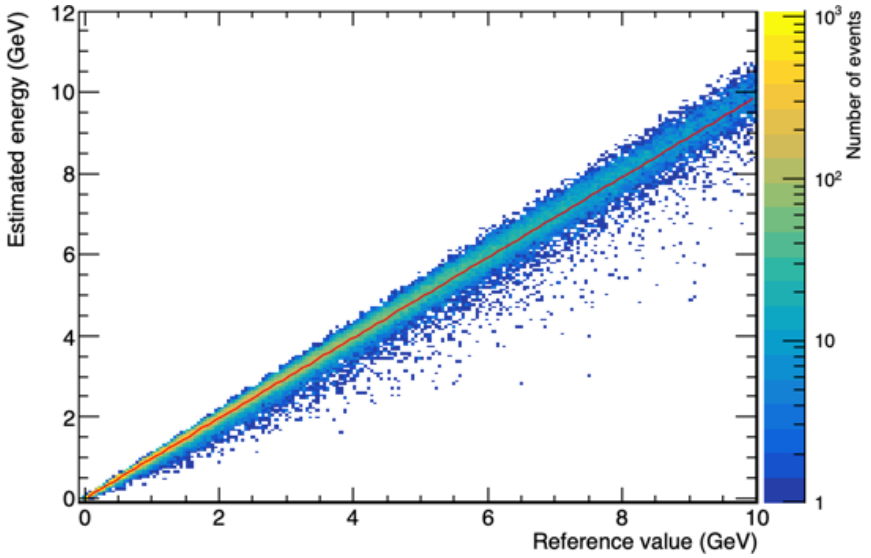
4.4.2 Energy estimation

Figure 8 shows the correlation between the reference value and the stochastic MF output. From figure 8a, it can be noticed the quadratic component introduced by the I_R term (see equation (3.6)). A second-order polynomial fitting was performed, achieving a χ^2 test of 152.80 for 155 degrees of freedom, and 53.57% probability of the data points to be explained by the fitted curve. By applying the calibration step (see equation (3.17)), the amplitude (energy) was recovered, as shown in figure 8b. A linear fit was then performed and both linear and angular coefficients are shown in table 4, considering all methods. The angular coefficients from the fittings revealed that the

2021 JINST 16 P02016



(a)



(b)

Figure 8. Performance for energy estimation: correlation between the reference amplitude value (truth energy from simulation) and the stochastic MF output before (a) and after calibration for amplitude recovery (b). The fitting curves are also shown in both plots.

OF methods significantly underestimated the energy with the MF deterministic approximation performing slightly better.

Regarding the energy estimation accuracy, figures 9a and 9b show, respectively, the estimated energy spectrum ratio with respect to the reference energy distribution, and the estimation error for each method. It is noticeable that the stochastic MF design estimates were closer to the truth values in higher energies and such design approach produced the narrowest negative error for lower energies. In addition, the deterministic MF approximation proved to be the best performance

2021 JINST 16 P02016

Table 4. Linear fitting parameters computed from the correlation between the reference values (truth energy from simulation) and estimation from each of the considered methods.

Method	Linear coef.	Angular coef.
Stochastic MF	-0.004	0.990
Deterministic MF	-0.007	0.831
Method 1	-0.009	0.803
Method 2	-0.009	0.803
Method 3	-0.011	0.797

envelope for the deterministic approaches in both (low and high) energy ranges, mainly due pulse uncertainties to be partially absorbed through the normalized average signal construction. As for the linearity, the relative errors for a large energy range are shown in figure 10 for the stochastic MF filter. It can be noticed that the proposed technique is not biased since its estimates deviated uniformly from the reference values.

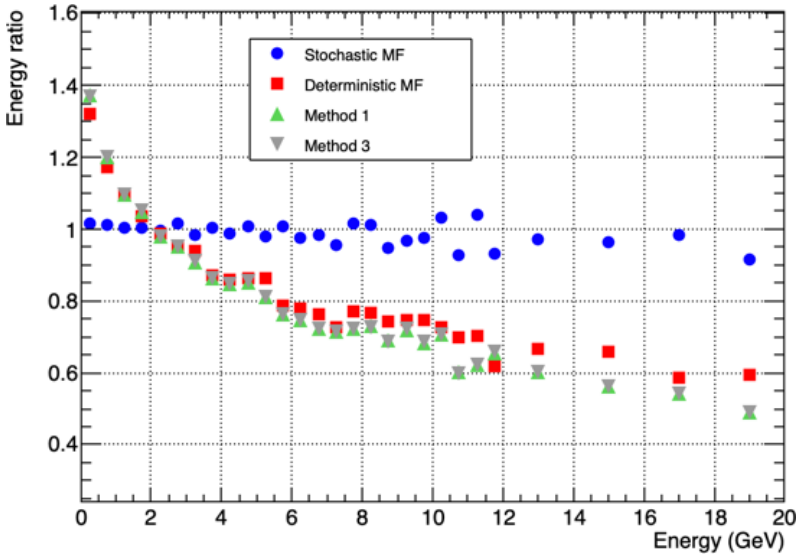
Table 5 summarizes the RMS and kurtosis values computed from the distribution for each method. The stochastic MF shows a higher value of kurtosis due to its sharpness (super-Gaussian) characteristics, whereas a perfect Gaussian distribution has a kurtosis value equal to zero. With respect to the RMS of the estimation error, it was considerably smaller for the stochastic MF, showing how resilient the stochastic design is to the signal uncertainties such as phase and pulse shape deformities. On the other hand, Method 3 presented the worst performance (larger error dispersion) in terms of energy estimation error due to its additional constraint associated with the pedestal parameter estimation (see equation (2.6)).

Table 5. Statistics from estimation error distributions.

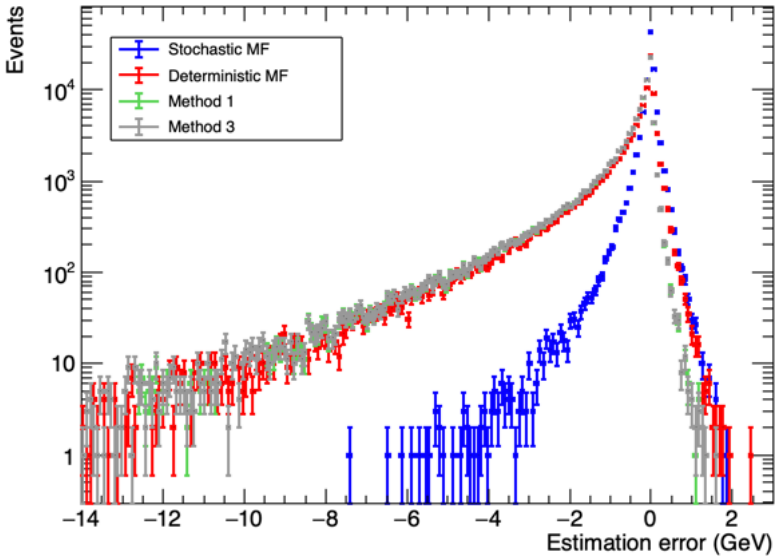
Method	RMS	Kurtosis
Stochastic MF	0.27	69.67
Deterministic MF	1.32	19.69
Method 1	1.38	18.13
Method 2	1.38	17.23
Method 3	1.44	18.13

Figure 11 shows that deterministic methods suffered from phase deviation. The mean of the relative error (in percent) is plotted as a function of the received signal phase. It is shown that the stochastic MF was capable of estimating the amplitude with an error close to zero for signal phase deviations up to half of the bunch crossing period (±12.5 ns). For the remaining methods, the efficiency depended significantly from the signal phase, as the amplitude was underestimated as the phase absolute value increased. This underestimation effect was associated with the negative tails shown in figure 9b. As both Method 1 and Method 2 achieved similar performance with respect to phase shift, it might be concluded that their pedestal computation choices produced roughly the same deterioration impact on the estimation efficiency, while adding the pedestal constraint in the

2021 JINST 16 P02016



(a)



(b)

Figure 9. Energy estimation performance: (a) reconstructed energy spectrum ratio with respect to the reference and (b) estimation error distribution.

optimization process (Method 3) provoked a slightly worse estimation error. Anyhow, a function describing the relative error may be estimated and applied in order to mitigate this effect [6].

Considering the signal reconstruction quality, figure 12a shows the QF as a function of the signal energy. It can be noted that the stochastic MF approach systematically presented smaller QF values up to approximately 12 GeV, which pointed out a better signal reconstruction characteristic. For such an energy range, the MF deterministic approximation was the performance envelope for the deterministic methods. For energies above 12 GeV, the statistics uncertainties increased significantly as the number of events decreased. Additionally, the QF performance is preserved within the ±12.5 ns range as shown in figure 12b.

2021 JINST 16 P02016

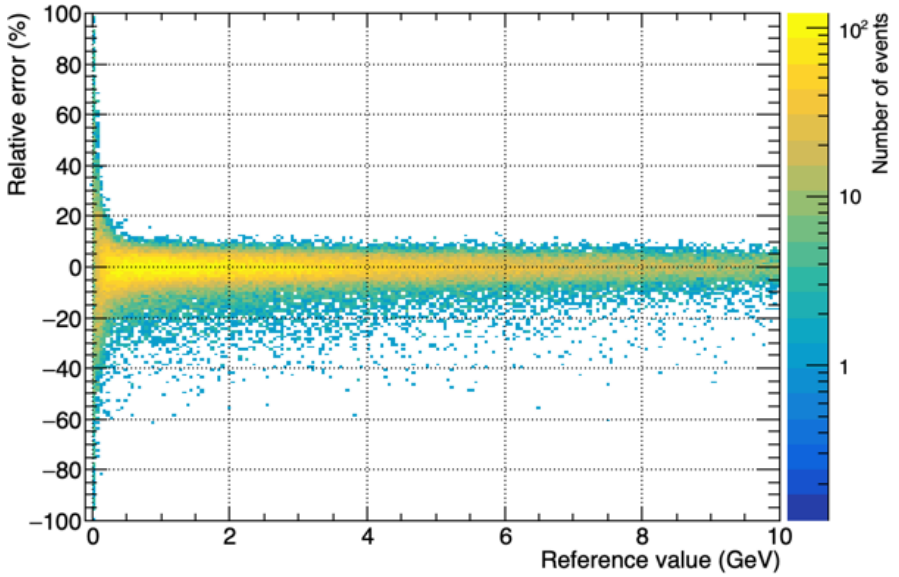


Figure 10. The relative error of the stochastic MF.

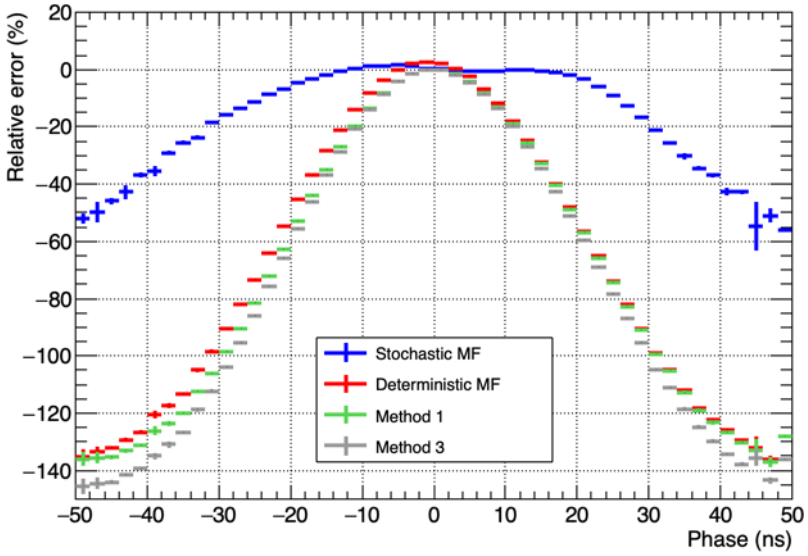
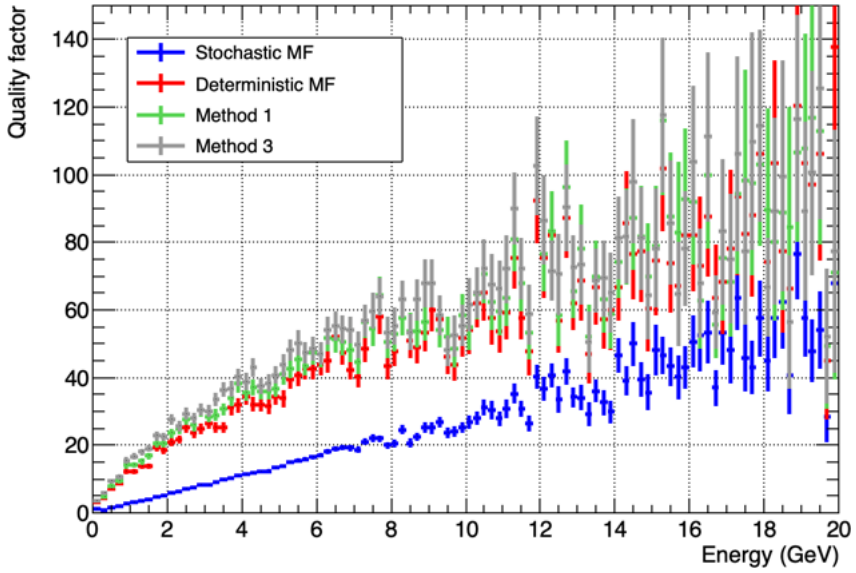


Figure 11. The relative error profile as a function of the signal phase deviation.

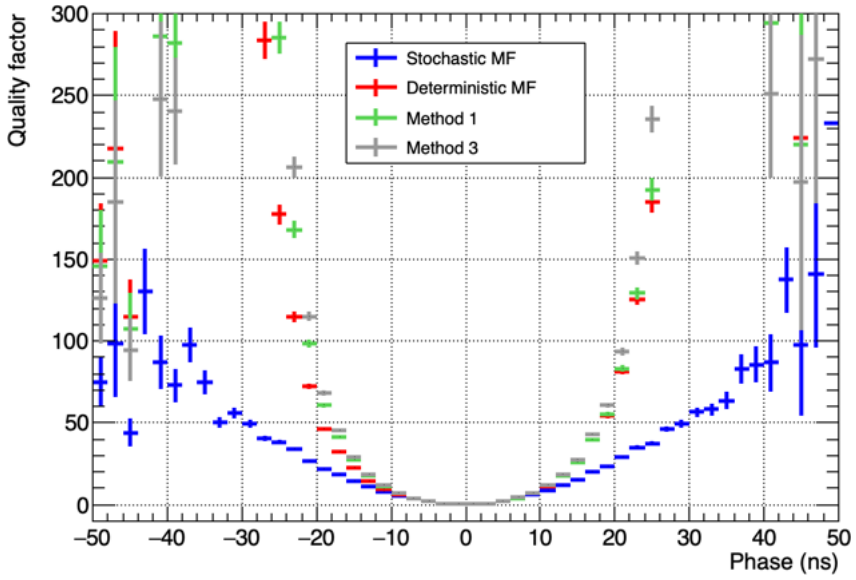
4.4.3 Operation emulation

The OF methods use the information from the estimated energy to apply an energy threshold in order to select only signals with relevant information, whereas the MF approaches detect the signals before computing the signal amplitude (energy). In order to evaluate how this MF processing sequence impacts on a practical calorimeter operation, in which only part of the cells actually participates on the signal reconstruction step, the fraction of the signals that are detected with relevant information as a function of energy was evaluated. For this, the high SNR data set was used. The smaller the fraction of discarded signals is, the better the efficiency is with this respect.

2021 JINST 16 P02016



(a)



(b)

Figure 12. The quality factor evolution as a function of the signal energy (a) and the signal phase (b).

In figure 13, the energy cut values were chosen according to the maximum SP value operation point. For better visualization, only signals up to 1 GeV were considered. It could be seen that the stochastic MF was capable of preserving more signals than any of the deterministic methods along the energy spectrum. The first three data points in this figure corresponded up to three RMS values from the noise. Thus, the stochastic MF retained signals practically up to three noise sigma values, whereas the deterministic approaches rejected around 7% of the incoming signals. Once more, the MF deterministic approximation acted as the performance envelope for the deterministic methods.

2021 JINST 16 P02016

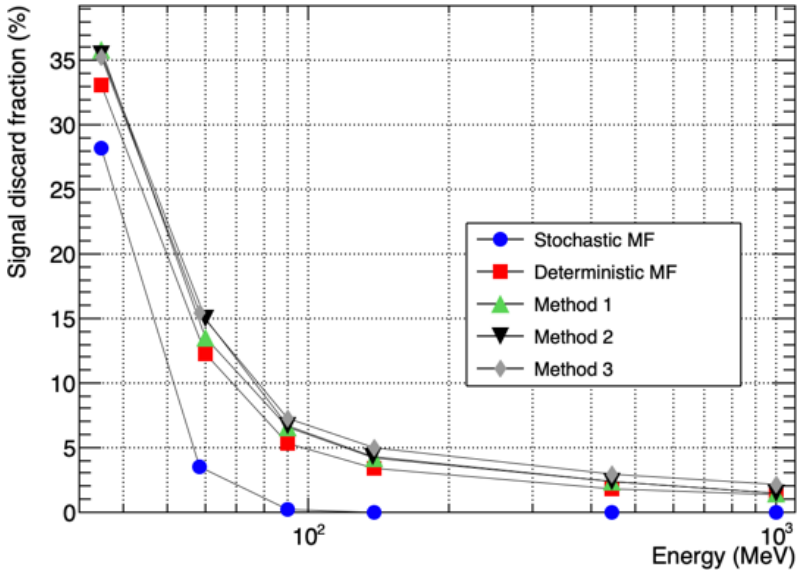


Figure 13. The signal discard ratio as a function of the readout signal energy. The stochastic MF is designed considering full PCA reconstruction (from 7 PCs).

5 Conclusions

In this paper, a MF-based approach was proposed for both calorimeter signal detection and estimation. The filter design was developed considering an hypothesis testing for detecting the signal of interest as a random process and it was also shown that the MF output can be used to recover the signal amplitude (energy). Envisaging online operation, a deterministic approximation for the MF design was also developed. Both approaches were applied to simulated signals corrupted from noise and their efficiencies were compared to typical methods currently used in modern calorimetry. The MF approaches are developed from minimizing the detection error probability through the maximization of the likelihood ratio, whereas the reference standard for energy estimation in calorimetry is based on a noise variance minimization strategy.

Results show that for online applications, where limited computation resource conditions may be imposed, the deterministic approximation of the MF method is attractive, as it requires the same computational effort when compared to the OF approach. Furthermore, it represents the highest performance bound for the deterministic approaches, being slightly better with respect to any OF method in terms of signal detection efficiency, energy estimation error, and signal reconstruction from the quality factor. Depending on the computational power available, a compact version of the stochastic MF (using 3 principal components, for instance) should also be considered due to its even better response. On the other hand, for offline operation, in conditions where the signal phase is an issue, the full stochastic MF provides a considerable better performance improvement when compared to the deterministic approaches. These improvements have a positive impact on the selection of the cells to be considered for event reconstruction. Particularly, particles emerging from the high-energy collisions and interacting with the calorimeter system may be reconstructed from topological clustering algorithms using the energy information from individual readout channels and, thus, removing the calorimeter cells with insignificant signals [43].

2021 JINST 16 P02016

In spite of the gaussianity assumed for the noise process, the MF weights can be re-designed and tested for other noise models, by solving the likelihood ratio analytically or numerically. Although the stochastic MF requires more computational resources, PCA allows to considerably reduce the computational cost. In the case of fast online operations [44, 45], where a simple FIR filter is usually adopted, the signal detection and amplitude estimation may be performed by the deterministic MF approximation.

Acknowledgments

We are thankful to CNPq, CAPES, RENAFAE, FAPERJ and FAPEMIG (Brazil), and European Union (E-planet project) for their support to this work.

References

[1] R. Wigmans, *Calorimetry: Energy Measurement in Particle Physics*, 2nd Edition, Oxford University Press (2017) [ISBN: 9780198786351].

[2] K. Kleinknecht, *Detector for Particle Radiation*, 2nd Edition, Cambridge University Press (1998) [ISBN: 9780521648547].

[3] M. Abolins et al., *Design and Implementation of the New D0 Level-1 Calorimeter Trigger*, *Nucl. Instrum. Meth. A* **584** (2008) 75 [arXiv:0709.3750].

[4] CMS collaboration, *The CMS trigger system*, 2017 *JINST* **12** P01020 [arXiv:1609.02366].

[5] G.F. Knoll, *Radiation Detection and Measurement*, 4th Edition, John Wiley and Sons (2010) [ISBN: 9788126567720].

[6] ATLAS collaboration, *Operation and performance of the ATLAS Tile Calorimeter in Run 1*, *Eur. Phys. J. C* **78** (2018) 987 [arXiv:1806.02129].

[7] LHCf collaboration, *Calibration of LHCf calorimeters for photon measurement by CERN SPS test beam*, *Nucl. Instrum. Meth. A* **671** (2012) 129.

[8] ALICE collaboration, *Online Calibration of the TPC Drift Time in the ALICE High Level Trigger*, *IEEE Trans. Nucl. Sci.* **64** (2017) 1263 [arXiv:1712.09423].

[9] LHCb collaboration, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, 2019 *JINST* **14** P04013 [arXiv:1812.10790].

[10] G. Bertuccio, E. Gatti, M. Sampietro, P. Rehak and S. Rescia, *Sampling and optimum data processing of detector signals*, *Nucl. Instrum. Meth. A* **322** (1992) 271.

[11] W.E. Cleland and E.G. Stern, *Signal processing considerations for liquid ionization calorimeters in a high rate environment*, *Nucl. Instrum. Meth. A* **338** (1994) 467.

[12] P. Adzic et al., *Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter*, *Eur. Phys. J. C* **46S1** (2006) 23.

[13] E. Fullana et al., *Digital signal reconstruction in the ATLAS hadronic tile calorimeter*, *IEEE Trans. Nucl. Sci.* **53** (2006) 2139.

[14] A. Shahinyan et al., *An Electromagnetic Calorimeter for the JLab Real Compton Scattering Experiment*, *Nucl. Instrum. Meth. A* **643** (2011) 17 [arXiv:0704.1830].

2021 JINST 16 P02016

[15] H.V. Trees and K. Bell, *Detection, Estimation, and Modulation Theory, Part I*, 2nd Edition, John Wiley and Sons (2013) [ISBN: 978-0-470-54296-5].

[16] J. Seixas, L. Caloba and L. Pinto, *Fast particle discriminator based on matched filters*, in *Proceedings of Midwest Symposium on Circuits and Systems*, Ames, U.S.A. (1996), vol. 2, pp. 843–846 [DOI].

[17] T. Ciodaro, J.M. de Seixas and A. Cerqueira, *Use of Hadronic Calorimetry Information in the ATLAS Level-1 Muon Trigger*, *IEEE Trans. Nucl. Sci.* **61** (2014) 1047.

[18] A. Caldwell et al., *Design and implementation of a high precision readout system for the ZEUS calorimeter*, *Nucl. Instrum. Meth. A* **321** (1992) 356.

[19] H. Muller et al., *Front-end electronics for PWO-based PHOS calorimeter of ALICE*, *Nucl. Instrum. Meth. A* **567** (2006) 264.

[20] M. Tylmad, *Pulse shapes for signal reconstruction in the atlas tile calorimeter*, in *Proceedings of 16th IEEE-NPSS Real Time Conference*, Beijing, China (2009), pp. 543–547 [DOI].

[21] G. Drake, D. Frei, S.R. Hahn, C.A. Nelson, S.A. Segler and W. Steurmer, *The Upgraded CDF front end electronics for calorimetry*, *IEEE Trans. Nucl. Sci.* **39** (1992) 1281.

[22] ATLAS LAr collaboration, *Electronic calibration of the ATLAS LAr calorimeter and commissioning with cosmic muon signals*, *J. Phys. Conf. Ser.* **160** (2009) 012050.

[23] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, 1st Edition, Academic Press (1982) [DOI].

[24] A. Papoulis and S.U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th Edition, McGraw-Hill Europe (2002) [ISBN: 978-0071226615].

[25] G. Turin, *An introduction to matched filters*, *IRE Trans. Inf. Theory* **6** (1960) 311.

[26] H. Stark and J. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, 2nd Edition, Prentice-Hall (1994) [ISBN: 978-0137287918].

[27] H.V. Trees, *Detection, Estimation and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*, 1st Edition, Wiley-Interscience (2001) [ISBN: 978-0-471-10793-4].

[28] P. Nicholson, *Nuclear Electronics*, 1st Edition, John Wiley and Sons (1974) [ISBN: 978-0471636977].

[29] I.T. Jolliffe and J. Cadima, *Principal component analysis: a review and recent developments*, *Phil. Trans. A. Math. Phys. Eng. Sci.* **374** (2016) 20150202.

[30] E. Kowalski, *Nuclear Electronics*, 1st Edition, Springer (1970) [ISBN: 978-3-642-87663-9].

[31] P. Comon and C. Jutten eds., *Handbook of Blind Source Separation*, Academic Press (2010) [ISBN: 978-0-12-374726-6].

[32] A. Kessy, A. Lewin and K. Strimmer, *Optimal Whitening and Decorrelation*, *Am. Stat.* **72** (2018) 309.

[33] ATLAS collaboration, *ATLAS Tile calorimeter calibration and monitoring systems*, *IEEE Trans. Nucl. Sci.* **66** (2019) 1228 [arXiv:1806.09156].

[34] R. Kohavi, *A study of cross-validation and bootstrap for accuracy estimation and model selection*, in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Quebec, Canada (1995), vol. 2, pp. 1137–1143.

[35] M. Hubert, M. Debruyne and P.J. Rousseeuw, *Minimum covariance determinant and extensions*, *WIREs Comput. Stat.* **10** (2018) e1421.

[36] Z. Weiping, W. Yi, L. Yuanjing and C. Jianping, *A new design of MRPC to reduce noise and crosstalk*, 2014 *JINST* **9** C07007.

2021 JINST 16 P02016

[37] N.A. Obuchowski and J.A. Bullen, *Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine*, *Phys. Med. Biol.* **63** (2018) 07TR01.

[38] T. Fawcett, *An introduction to roc analysis*, *Pattern Recogn. Lett.* **27** (2006) 861.

[39] E.F. Simas Filho, J.M. Seixas and L.P. Caloba, *Optimized calorimeter signal compaction for an independent component based ATLAS electron/jet second-level trigger*, *PoS* **ACAT08** (2008) 102.

[40] P. Peebles, *Probability, Random Variables, and Random Signal Principles*, 4th Edition, McGraw-Hill (2000) [ISBN: 978-0073660073].

[41] M. Delmastro, *Quality factor analysis and optimization of digital filtering signal reconstruction for liquid ionization calorimeters*, *Nucl. Instrum. Meth. A* **600** (2009) 545 [arXiv:0812.3486].

[42] J.M. Seixas, *Quality Factor for the Hadronic Calorimeter in High Luminosity Conditions*, *J. Phys. Conf. Ser.* **608** (2015) 012044.

[43] ATLAS collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490 [arXiv:1603.02934].

[44] ALICE collaboration, *Real-time data processing in the ALICE High Level Trigger at the LHC*, *Comput. Phys. Commun.* **242** (2019) 25 [arXiv:1812.08036].

[45] R. Oishi, *LHC-ATLAS Phase-1 upgrade: firmware validation for real time digital processing for new trigger readout system of the Liquid Argon calorimeter*, 2020 *JINST* **15** C05013.

2021 JINST 16 P02016