Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC

W. Adam, R. Frühwirth
Institute for High-Energy Physics, Austrian Academy of Sciences, Vienna, Austria

A. Strandlie
Gjøvik University College, Gjøvik, Norway

T. Todorov
IReS, Strasbourg, France and CERN

Abstract

The bremsstrahlung energy loss distribution of electrons propagating in matter is highly non-Gaussian. Because the Kalman filter relies solely on Gaussian probability density functions, it is not necessarily the optimal reconstruction algorithm for electron tracks. A Gaussian-sum filter (GSF) algorithm for electron reconstruction in the CMS tracker has therefore been developed and implemented. The basic idea is to model the bremsstrahlung energy loss distribution by a Gaussian mixture rather than by a single Gaussian. It is shown that the GSF is able to improve the momentum resolution of electrons compared to the standard Kalman filter. The momentum resolution and the quality of the error estimate are studied both with a fast simulation, modelling the radiative energy loss in a simplified detector, and the full CMS tracker simulation.
1 Introduction

Modern tracking detectors based on semiconductor technologies contain larger amounts of material than gaseous detectors, due in part to the detector elements themselves, and in part to additional material required for on-sensor electronics, power distribution, cooling and mechanical support. A precise modelling of material effects in track reconstruction is therefore necessary to obtain the best estimates of the track parameters. Such material effects are particularly relevant for the reconstruction of electrons, which, in addition to ionisation energy loss and multiple Coulomb scattering, suffer from large energy losses due to bremsstrahlung.

The baseline for track reconstruction in the CMS tracker is the Kalman filter [1], which was introduced to the high-energy physics community in the DELPHI experiment at the LEP collider at CERN. The Kalman filter is a recursive formulation of the least-squares method of fitting a set of measurements to a track model. It alternates between propagation steps and update steps. In the propagation step, the track state is extrapolated to the next observation layer, and material effects are added to the covariance matrix. In the update step, the extrapolated state is combined with the observation. Due to its recursive character, the Kalman filter is well suited for doing track finding, also called pattern recognition, and track fitting concurrently [2]. A brief description of the CMS tracker and an outline of the track reconstruction using the Kalman filter is presented in Section 2.

Since the Kalman filter is a linear least-squares estimator, the optimal treatment of radiative energy loss is in this context to correct the momentum by the mean value of the energy loss and to increase the variance of the momentum by adding the variance of the energy loss distribution during each propagation step. This procedure ensures unbiased estimates of the track parameters and of the associated uncertainties [3]. Also due to the virtue of being a linear least-squares estimator, the Kalman filter is optimal when all probability densities encountered during the track reconstruction procedure are Gaussian, but not necessarily in other cases.

The implicit approximation of the energy loss distribution by a single Gaussian in the Kalman filter is quite crude. It is therefore plausible that a non-linear estimator which takes the actual shape of the distribution into account would yield better results. A non-linear generalisation of the Kalman filter (KF), the Gaussian-sum filter (GSF) [4, 5], has therefore been implemented in the reconstruction software of the CMS tracker [6]. The algorithm requires that all probability densities involved in track reconstruction are described by Gaussian mixtures. The basic idea of the present work is to approximate the energy loss distribution by a Gaussian mixture. We have chosen the simple and well-known model of energy loss that is due to Bethe and Heitler [7]. The model and the GSF algorithm are described in some more detail in Section 3.

The algorithm has been implemented as an alternative track reconstructor in the CMS reconstruction program. It has been validated using a simplified simulation and evaluated on tracks from a full simulation of electrons in the CMS tracker. The results are presented in Section 4. Our final conclusions are summed up in Section 5.

2 Track reconstruction in the CMS tracker

The CMS tracker [8] consists of a central “barrel” part and two endcaps. The barrel is the relevant part for the tracks used in this study. It is composed of silicon pixel and microstrip detectors, which are arranged in cylindrical layers.

The three barrel pixel layers are situated at distances of 4 cm, 7 cm and 11 cm from the beam line. In the configuration used for this study the pixels are square with a pitch of 150 $\mu$m. The microstrip sensors are arranged in ten layers at radii ranging from about 25 cm to 110 cm. The strips are parallel to the beam line. The pitch varies from about 100 $\mu$m to 180 $\mu$m. Layers one, two, five and six are equipped with double-sided modules with a stereo angle of 100 mrad. A schematic view of the barrel part of the CMS tracker is shown in Fig. 1.

Most of the support structure is concentrated on the layers, close to the sensors. Hence, for reconstruction purposes material effects have only been taken into account at the position of the sensitive elements. The amount of material attributed to each layer was based on the full detector description and varied between 2% and 5% of a radiation length at normal incidence. Taking into account the inclination of tracks with respect to the sensor plane, the effective amount of material crossed by the particle can be significantly larger. The parametrisations of the Bethe-Heitler model in Ref. [9] cover the relevant range of effective path lengths.

The default strategy for track reconstruction in CMS starts out with generating track segments or seeds by grouping together measurements in the pixel layers close to the beam pipe. The seeds are used as a starting point for a combinatorial Kalman filter track finding procedure [10], in which a tree of possible track candidates is generated for each seed. The surviving track candidates are submitted to a Kalman smoothing procedure, which starts at the
the extrapolated state is combined with the observation.

forwards filter – and a backwards filter proceeding in the opposite direction.

is increased by the variance of the energy loss distribution at each intersection with a surface. In the update step, the extrapolated state is combined with the observation.

The smoother works by calculating a weighted mean of the results of the filter proceeding outwards – the so-called forward filter – and a backward filter proceeding in the opposite direction.

The GSF algorithm for electrons

The energy loss of electrons due to bremsstrahlung has been described by Bethe and Heitler [7]. In the Bethe-Heitler model, the probability density function (PDF), \( f(z) \), of the energy loss of an electron is

\[
   f(z) = \frac{[-\ln z]^{c-1}}{\Gamma(c)}, \quad c = t / \ln 2, \tag{1}
\]

where \( t \) is the path length in the material (in units of radiation length) and \( z \) is the fraction of energy remaining after the material is traversed. The distribution is assumed to be independent of the energy of the incoming particle. It is shown in Fig. 2 for different values of the path length.

A key component in the application of the GSF to electrons is the approximation of the energy loss distribution in Eq. (1) by a Gaussian mixture. The different components of the mixture model different degrees of hardness of the bremsstrahlung in the layer under consideration.

Results of a detailed study on the determination of the mixture parameters can be found in Ref. [9]. Gaussian
mixtures with four, five or six components have been obtained by minimising either of the following two distances

\[ D_{\text{CDF}} = \int_{-\infty}^{\infty} |F(z) - G(z)| \, dz, \]  
\[ D_{\text{KL}} = \int_{-\infty}^{\infty} \ln \left( \frac{f(z)}{g(z)} \right) f(z) \, dz, \]  

where \( f(z) \) and \( F(z) \) are the PDF and cumulative distribution function (CDF) of the model distribution and \( g(z) \) and \( G(z) \) are the PDF and CDF of the Gaussian mixture, respectively. \( D_{\text{KL}} \) is the so-called Kullback-Leibler distance between the model distribution and the mixture. Hereafter, \( n \)-component mixtures obtained by minimising \( D_{\text{CDF}} \) are called CDF\(_n\)-mixtures, whereas the mixtures obtained by minimising \( D_{\text{KL}} \) are called KL\(_n\)-mixtures. The minimisations in Ref. [9] have been done for discrete values of \( t \), ranging from 0.02 to 0.20. The resulting mixture parameters have been used to fit fifth-degree polynomials as functions of \( t \). This allows for a fast generation of the mixture during reconstruction, based on the effective thickness of a layer as deduced from the incident angle of the track.

In the GSF, the distributions of all state vectors are allowed to be Gaussian mixtures, i.e. weighted sums of Gaussians, instead of single Gaussians as in the Kalman filter. Still, the basic structure of the GSF is very similar to the standard Kalman filter, having the same alternating sequence of propagation and update steps. In each propagation step, a weighted mixture of state vectors and corresponding covariance matrices is propagated to the nearest surface, yielding a predicted mixture. This propagation is done independently for each component and does not affect the weights.

The effects of radiative energy loss on a surface are described by a Gaussian mixture that is convoluted with the predicted state, which is usually itself a Gaussian mixture. Measurements are included by invoking Bayes’ theorem. The state vector and covariance matrix of each component are updated with the measurement as in the Kalman filter. The compatibility of the measurement with the individual components is reflected in the resulting posterior weights associated to the components. As the weights depend on the observations, the GSF is no longer a linear filter. For more details concerning the Gaussian-sum filter in general, the reader is referred to the literature [4, 5].

The convolution of the mixture describing the predicted state with the mixture modelling the energy loss results in a multiplication of the number of components. This procedure quickly leads to a combinatorial explosion. Hence every realistic implementation of the GSF must repeatedly reduce the number of components describing the state of the track. As little information as possible should be lost in this step.

The strategy adopted in this study was the merging of components into clusters, based on the Kullback-Leibler distance as defined in Eq. (3). Clusters of components were replaced by a single Gaussian of equal mean and variance, while the weights of the cluster members were summed to obtain the weight of the new component.
4 Results from simulated tracks in the CMS tracker

Single electron tracks have been simulated in the barrel region of the CMS tracker, for absolute values of the pseudo-rapidity, \( \eta \), less than 1.0 and for different transverse momenta, \( p_T \). The vertices were Gaussian distributed around the origin, with standard deviations of 15 \( \mu \)m in the transverse dimensions and 5.3 cm along the beam axis.

Only tracks crossing at least eight layers have been retained. To separate the effects of pattern recognition from the fit itself, the default combinatorial Kalman filter track finding procedure has been bypassed, and the reconstructed hits have instead been chosen by their association with the simulated tracks.

Each track was fitted with a Gaussian-sum filter and a Kalman filter. Since the best estimate of the track parameters is needed close to the production vertex, both fits proceeded from the outermost to the innermost hit. GSF and KF results were compared at a virtual plane corresponding to the point of closest approach to the average vertex position in the transverse plane – the transverse impact point (TIP). On this surface – as on any other one used during the fit – the trajectory is described by the ratio of charge to momentum (\( q/p \)), two orthogonal co-ordinates on the surface and the corresponding direction tangents w.r.t. the normal to the surface.

The result of the GSF contains more information than estimated parameters and their variances: it provides a, generally non-Gaussian, probability density function. This is illustrated in Fig. 3, which shows an example of the estimated PDF of \( q/p \) for one single track, where the KF result is shown as a single Gaussian. Many applications still require a result in the form of a single parameter vector and its covariance matrix. Therefore residuals and derived quantities shown in the following sections have been calculated using the first two moments of the GSF mixtures.

![Figure 3: Estimated \( q/p \) residual distribution for a single track with \( p_T = 10 \text{ GeV/c} \) and \( \eta = 0.44 \). The 12-component GSF estimate is shown as solid histogram with the dotted lines representing individual components. The non-Gaussian nature of the mixture is clearly visible. The solid (dashed) line represents the first two moments of the KF (GSF) estimate in the form of a single Gaussian.](image)

4.1 Simplified simulation

The intrinsic performance of the algorithm was tested using a simplified simulation, matching the assumptions made for reconstruction. In this simulation, a constant magnetic field of 4T parallel to the beam axis was used. Position and amount of material used in the simulation matched exactly the values used in reconstruction. At each detector surface, energy loss was simulated according to the Bethe-Heitler model. No other material effects were taken into account. Reconstructed hits were generated from the simulated intersection points by Gaussian smearing with the nominal detector resolution.
Figure 4 shows residuals and pull quantities for the momentum-related part of the state vector at the TIP surface. For the GSF, a six-component CDF-mixture has been used to approximate the energy loss. During the fit, the number of components was limited to twelve.

Several versions of the GSF, using different mixtures for the energy loss and different merging strategies, have been compared based on the width of the normalised residuals of the reconstructed momentum. A value \( Q(x) \) was defined as the half-width of the symmetric interval covering a fraction \( x \) of the residuals. Figure 5 shows the variation of \( Q \) for the equivalent of \( \pm 1\sigma \) and \( \pm 2\sigma \) coverage. The GSF yields better results than the KF in all tested cases, and CDF-mixtures are superior to KL-mixtures in the modelling of the energy loss. The differences between CDF-mixtures with four, five and six components are small, as is the gain for keeping more than twelve components in the fit.

Three merging strategies based on clustering were tested for \( p_t = 10 \, \text{GeV}/c \), using CDF_6-mixtures and limiting the number of components of the track state to twelve:

1. The pair with the smallest distance was combined into a single component. The distances were recalculated and the combination step was repeated until the required number of components was reached.
2. Pairs were combined iteratively as in the previous algorithm, but the distances were only calculated once for the initial set of components. In the subsequent steps distances to combinations of components were approximated by a weighted sum of the individual distances.
3. The component with the largest weight was combined with the one closest to it. This step was repeated with the remaining components. If all components had been used the procedure was restarted with the newly combined pairs.

The procedure described in item 2 was chosen as the standard solution. It was about twice faster than the first one at the same momentum resolution of \( Q(68\%) = 0.069 \). The third algorithm, which proceeds in the order of weights, was marginally worse (\( Q(68\%) = 0.071 \)).

An alternative and simpler approach is the selection of the components with the highest weights. This is computationally light, but does not conserve the first two moments of the mixture, unless all deselected components are combined into a single one and added to the new mixture. Even in this case the selection by weight is inferior to clustering by distance (\( Q(68\%) = 0.092 \)).

Figure 6 shows the evolution of the results with increasing transverse momentum. The GSF residuals have a narrower core, whereas the difference is less pronounced when the tails are included. For high transverse momenta the GSF results approach those obtained by the KF.
Figure 5: Half-widths of the symmetric intervals covering 68% (left) and 95% (right) of the normalised momentum residuals as a function of the maximum number of components kept during the fit. The corresponding values for the KF are also shown, together with the estimated uncertainties on $Q$.

Figure 6: Half-widths of the symmetric intervals covering 68% (left) and 95% (right) of the residuals in momentum as a function of $p_t$, for a maximum of twelve components kept during the fit. The round markers and asterisks show the result for a Gaussian-sum filter with a CDF-mixture and a Kalman filter, respectively.

The GSF provides an estimate for the distribution of the residuals. Since pull quantities are not adequate to test this feature, a more general approach has been chosen: on a track-by-track basis, the probability transform at the TIP surface, i.e. the probability of the mixture below the true value, was computed. For the KF, a single Gaussian distribution was used. The distributions of the probability transforms for $q/p$, shown in Fig. 7, should be flat if the true values were distributed according to the mixtures. The GSF estimate is not perfect, but provides a good estimate of non-Gaussian tails for each track.

4.2 Detailed simulation

The Gaussian-sum and the Kalman filter have also been tested on tracks produced by the full CMS simulation program [11], which uses a detailed description of the spatial distribution of material in the tracker, a realistic map of the magnetic field and takes into account all relevant types of interaction of electrons with matter. Intersection
Figure 7: Distributions of the probability transforms for $q/p$ for the KF (open histograms) and GSF (hatched histograms). Left: distributions for two values of the transverse momentum. Right: distributions for the KF and for two versions of the GSF with different limitations on the number of components.

points with the sensitive volumes were reconstructed based on a detailed modelling of the detector response. In the reconstruction uncertainties due to multiple scattering were taken into account in addition to the modelling of the energy loss described in the previous sections.

The evolution of the momentum resolution as a function of $p_t$ is shown in the left part of Fig. 8. The GSF reconstructs the majority of the tracks with better precision than the KF. At high momenta, however, the $Q(95\%)$ values are worse than those of the KF, indicating higher tails. This qualitative difference to the results obtained with a simplified simulation is mainly due to non-Gaussian tails in the reconstructed positions in the innermost layers. The GSF is sensitive to these outliers, as they will not only distort the updated trajectory parameters, but also increase the weight of components far from the true values. This is confirmed by the resolution plots shown in the right part of Fig. 8. They were obtained under the same conditions as above except for the reconstructed positions in the sensitive elements, which were created from the simulated positions by a Gaussian smearing. Ultimately this problem could be solved using outlier detection or, more natural in the context of the GSF, by an appropriate parametrisation of the measured positions by a Gaussian mixture.

Figure 9 shows the pull quantities and the distribution of the probability transform for $q/p$ w.r.t. to the estimated mixture for a CDF$_6$-mixture and $p_t = 10$ GeV/$c$. As in the simplified simulation, the GSF estimate is clearly superior to the description by a single Gaussian distribution.

Radiation in the innermost layers of the detector cannot be detected by the fit, leading to irreducible tails in the residual distributions for the momentum, transverse impact parameter and azimuthal direction. This effect can be partially compensated by including a vertex constraint – a procedure which will be used in many realistic applications. For this study the prior knowledge on the distribution of generated vertices has been used, a conservative assumption compared to the precision expected for the reconstruction of primary vertices in high energy proton-proton collisions. Figure 10 shows the distributions of the momentum residuals with and without the vertex constraint for a CDF$_6$-mixture and a $p_t$ of 10 GeV/$c$. Using the constraint renders distributions less skew, with modes closer to zero and a reduction of the population of tracks in the tails. The resolution, measured in terms of $Q(68\%)$, improves by 35% for the GSF and 20% for the KF. This indicates the adaptive power of the GSF, which will select components based on the information contained in the vertex constraint.

The fully simulated tracks have also been used to measure the CPU time consumption of the Gaussian-sum filter. No attempt had been made to optimise the implementation of the GSF in terms of execution time. Figure 11 shows the average time spent in different components of the GSF as a function of the maximum number of components kept, normalised to the time used by the KF. It is dominated by the calculation of distances between components in the process of merging. The GSF is significantly slower than the KF, but since it would only be applied to preselected electron candidates, the CPU time consumption is no limiting factor for its use in event reconstruction.
Figure 8: Half-widths of the symmetric intervals covering 68\% (dashed lines) and 95\% (solid lines) of the normalised momentum residuals as a function of $p_t$. Reconstructed hit positions were obtained based on detailed detector simulation (left) or by Gaussian smearing of the simulated intersection points (right). GSF results are indicated by circular symbols. The KF results are shown with the estimated uncertainties on $Q$.

Figure 9: Left: Pull quantities for $q/p$ for the GSF (solid histogram) and the KF (open histogram). The values for mean and RMS shown in the figure were obtained using truncation at the histogram limits. Right: Distributions of the probability transform for $q/p$. 
Figure 10: Normalised momentum residuals without (left) and with (right) the vertex constraint at $p_t = 10$ GeV/$c$. GSF (KF) results are shown as solid (open) histograms.

Figure 11: Average elapsed time in GSF components, normalised to the average time for KF track reconstruction. Circles: full GSF fit (includes all other components). Downward pointing triangles: component merging. Upward pointing triangles: propagation and estimation of material effects. Squares: combination of predicted state and measurement (including weight calculation). Asterisks: Calculation of the weights of the updated state.

5 Conclusion

The Gaussian-sum filter has been implemented in the CMS reconstruction program. It has been validated with electron tracks from a simplified simulation in which the energy loss distribution (Bethe-Heitler model) and the exact amount and location of material are known to the reconstruction program. It has been shown that the quality of the momentum estimate depends on the number of components kept during reconstruction and on the number of components in the mixture modeling the energy loss distribution. Mixtures obtained by minimising the distance of the cumulative distribution function to the one according to Bethe and Heitler are superior to those obtained using the Kullback-Leibler distance. A comparison with the best linear unbiased estimator, the Kalman filter, shows a clear improvement of the momentum resolution. At high momenta the results approach those of the Kalman filter.
The same algorithm has been used to reconstruct electron tracks from full detector simulation. As above, a perfect
pattern recognition has been assumed, but neither the exact energy loss distribution nor the precise amount and
location of material were known to the reconstruction program. The core of the residual distribution obtained from
the GSF is still significantly narrower than for the KF, and the resulting mixture is a good approximation of the
distribution of the true momentum. This indicates that perfect knowledge of the amount and location of the material
is not essential for the performance of the GSF. The number of tracks in the tails of the residual distribution was
found to be higher than for the KF, but this is mainly because of a higher sensitivity to tails in the distribution of
the measured positions. Within the framework of the GSF, a parameterisation of the measurement uncertainty in
form of a Gaussian mixture would allow to include this effect in a natural way.

More systematic studies with electrons from the full simulation are clearly needed. In particular, the effect of the
pattern recognition needs to be evaluated. While the pattern recognition may well introduce a bias into the track
sample, we have no reason to assume that this will affect the performance of the Gaussian-sum filter as compared
to the Kalman filter. One can therefore conclude that in electron reconstruction the Gaussian-sum filter yields an
improvement in precision as compared to the Kalman filter.

Acknowledgements

We would like to acknowledge the many collaborators contributing to the reconstruction software of CMS, which
served as a basis for the this study. We would also like to thank the CMS editors and referees for their comments
and suggestions.

References