INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A deterministic annealing PMHT algorithm with an application to particle tracking

Are Strandlie — Josiane Zerubia

N° 3711

Juin 1999

THÈME 3

Rapport de recherche

ISSN 0249-6399

A deterministic annealing PMHT algorithm with an application to particle tracking

Are Strandlie* †, Josiane Zerubia

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 3711 — Juin 1999 — 20 pages

Abstract: We introduce the Probabilistic Multi-Hypothesis Tracking (PMHT) algorithm for particle tracking in high-energy physics detectors. This algorithm has been developed recently for tracking multiple targets in clutter, and it is based on maximum likelihood estimation by aid of the EM algorithm. The resulting algorithm basically consists of running several iterated and coupled Kalman filters and smoothers in parallel. It is similar to the Elastic Arms algorithm, but it possesses the additional feature of being able to take process noise into account, as for instance multiple Coulomb scattering. Herein, we review its basic properties and derive a generalized version of the algorithm by including a deterministic annealing scheme. Further developments of the algorithm in order to improve the performance are also discussed. In particular, we propose to modify the hit-to-track assignment probabilities in order to obtain competition between hits in the same detector layer. Finally, we present results of an implementation of the algorithm on simulated tracks from the ATLAS Inner Detector Transition Radiation Tracker (TRT).

Key-words: EM algorithm, deterministic annealing, PMHT algorithm, particle tracking, particle physics.

* Department of Physics, University of Oslo, P. O. Box 1048 Blindern, 0316 Oslo, Norway and Department of Electrical Engineering and Science, Gjøvik College, Service box 191, 2810 Gjøvik, Norway

† We would like to thank Dr. Rudolf Frühwirth for many valuable comments on the manuscript. One of us (AS) is also grateful to the Norwegian Research Council for partial funding of a half-year stay at INRIA.

Algorithme PMHT avec recuit déterministe. Application au suivi de particules

Résumé : Nous considérons l'algorithme PMHT pour suivre la trajectoire de particules dans des détecteurs utilisés en physique des hautes énergies. Cet algorithme a récemment été développé pour suivre des cibles multiples dans un environnement encombré. Il est fondé sur l'estimateur du maximum de vraisemblance, et s'appuie sur un algorithme de type EM. L'algorithme résultant correspond à l'utilisation en parallèle de plusieurs filtres de Kalman itératifs couplés. Il est proche de l'algorithme EA, mais il est de plus capable de prendre en compte le bruit associé au processus, comme par exemple la diffusion de Coulomb multiple. Dans ce rapport, nous présentons les propriétés classiques d'un tel algorithme et proposons une généralisation incluant un recuit déterministe. Nous proposons également plusieurs modifications améliorant les performances de cet algorithme. En particulier, nous avons modifié les probabilités reliant les événements élémentaires aux trajectoires afin d'obtenir une compétition entre ces événements dans une même couche du détecteur. Enfin, nous présentons des résultats obtenus sur des simulations réalisées à partir du détecteur ATLAS (TRT).

Mots-clés : algorithme EM, recuit déterministe, algorithme PMHT, suivi de particules, physique des particules

Contents

1 Introduction

Developments in track finding and fitting during the last one and a half decades can roughly be divided into two separate categories. The first category consists of methods or algorithms based on the now widely used Kalman filter [1] or variations thereof. The Kalman filter performs a least-squares fit of the data in a track candidate to a given track model. One of the main advantages of the application of this filter compared to a global least-squares fit is that process noise, as for instance multiple Coulomb scattering, can be taken into account locally, i.e. there are no long-range correlations between the observations. In addition, by supplementing the filter with a smoother optimal estimates of the track parameters can be evaluated anywhere along the track. However, the application of the Kalman filter requires that the pattern recognition problem, i.e. the measurement-to-track assignment procedure, has been completely resolved in advance. If not, one can propose to make a list of all possible combinations of points inside a track candidate that constitute a valid track and run the filter on all these combinations. In the end, one picks the combination with the least value of the chisquare statistic as the fitted track. One major disadvantage with this approach is that the number of combinations can get very large when the density of measurements is high, as will be the case in future experiments at the Large Hadron Collider (LHC) at CERN.

More recently, there have been attempts to apply the Gaussian-sum filter (GSF) to problems related to tracking. The GSF takes the form of several Kalman filters running in parallel and is therefore a relatively straightforward extension of the original Kalman filter. It has turned out to be useful in situations where one encounters long-tailed or non-Gaussian measurement noise [2], and it has also successfully solved the problem of simultaneous track finding and fitting to data coming from a detector with ambiguous measurements [3]. The major problem with the GSF in the latter case is its lack of robustness towards noise, i.e. its behaviour in the case where none of the measurements in some detector layer is coming from the track to be fitted. In its original formulation, the GSF will always believe that at least one out of possibly several measurements in a layer originates from the correct track, thereby inducing a bias in the estimates of the track parameters. This can in principle be cured, but the computational cost will rapidly increase.

The most recent development in the class of filters, the Deterministic Annealing Filter (DAF) [4], has been constructed in an attempt to overcome the shortcomings of the GSF. It is an iterated Kalman filter with reweighted observations and is an example of an EM algorithm [5], well-known in statistics. The filter is by construction robust, and the deterministic annealing scheme included efficiently prevents the algorithm from ending up in a local maximum on the likelihood surface. It is equivalent to the Elastic Arms algorithm [6] formulated for single track candidates, but it has the additional feature of, if necessary, including process noise. Moreover, it does not require the minimization of a non-quadratic energy function, which is generally a non-trivial task.

The second category of developments consists of global tracking algorithms. Examples of these are the Elastic Tracking algorithm by Gyulassy and Harlander [7] and the previously mentioned Elastic Arms algorithm. The main advantage of these methods compared to the different filter algorithms described above is the fact that they can handle situations where a hit possibly can belong to several tracks. This might occur for instance in the reconstruction of very narrow jets. The filter methods are by construction working on single track candidates, so they lack at the present stage the flexibility provided by the global algorithms. On the other hand, the global algorithms do in general not exhibit the same statistical rigour as the filters. It is therefore desirable to construct a formalism which is able to extract the best properties of both the filters and the global methods and combine these.

In this work, we introduce a novel algorithm for track finding and fitting in particle detectors: the Probabilistic Multi-Hypothesis Tracking (PMHT) algorithm. This algorithm has recently been developed for tracking multiple targets in clutter [8]. It is based on the maximization of a likelihood structure by aid of the EM algorithm, and it fits the data of a collision event to a hypothesized number of tracks. The algorithm takes the form of several iterated and coupled Kalman filters and smoothers working in parallel. Thus, the PMHT algorithm exhibits the same power of global tracking as the Elastic Arms algorithm. Furthermore, because the backbone of the PMHT is Kalman filters, it is as statistically rigorous as the filter methods described above. In order to make the algorithm work even better, we propose some modifications to the original formulation of the PMHT. One of these is a generalization of the algorithm to include a deterministic annealing scheme, and we will show that this gives significant improvements in the accuracy of the estimates of the track parameters. With these modifications, the PMHT is in the single-track case and under the assumption of Gaussian noise equivalent to the DAF. The PMHT can therefore also be regarded as a multi-track generalization of the DAF.

The paper is organized in the following way. In Section 2 we review the basic properties of the PMHT algorithm. We introduce the modifications and generalizations of the algorithm in Section 3. In Section 4 we bring results from experiments performed on simulated data from the ATLAS Inner Detector TRT. The paper is concluded in Section 5 with a summary of the main results and a brief outlook to future research.

2 Review of basic properties of the PMHT

Detailed descriptions of all aspects of the PMHT algorithm exist in the literature [8]. To our knowledge, however, the application of the algorithm to tracking problems in high-energy physics detectors is novel, and we therefore assume little or no knowledge of it by the readers of this paper. Therefore we will herein give a review of the basic features of the algorithm.

The scenario is a given collection of measurements from a particle detector, and we want to fit M tracks to these measurements in an optimal manner. The output of the algorithm is the estimated state vectors of the tracks. The number M of tracks has to be hypothesized beforehand, as well as the initial values of the track parameters. A plausible way of tackling the problem in practice would be to first apply some kind of heuristic pattern recognition procedure, for instance a Hough transform, to define regions of interest in the detector and give initial estimates of the track parameters. Depending on the complexity of the problem. i.e. the density of measurements and tracks, the PMHT algorithm can then be used either on each track candidate independently or on several neighbouring candidates simultaneously. If one chooses to treat each track candidate independently, the algorithm is used to discern the true track points from noise hits, hits from other tracks and possibly also mirror hits, if the hits arise from a detector with ambiguous measurements. This case is the one to be considered in the simulation experiments of this work. If several track candidates are taken into account at the same time, the algorithm is used also to decide which track the measurements are assigned to.

The PMHT algorithm is derived by first constructing a probabilistic likelihood function. This likelihood is a function of the measurements $\{\boldsymbol{m}_{ik}\}$ in the collision event and the state vectors $\{\boldsymbol{x}_{km}\}$. Here $\boldsymbol{m}_{ik}$ denotes measurement i $(i = 1, \ldots, n_k)$ in layer k $(k = 1, \ldots, K)$, and $\boldsymbol{x}_{km}$ represents the state vector of track m $(m = 1, \ldots, M)$ in layer k. We assume that the detector can be represented as a collection of shells or layers and that all measurements therefore come from these layers. The likelihood is also a function of the *assignment variables* $\{k_{ik}\}$ $(k_{ik} = 1, \ldots, M)$ associated with the measurements. For instance, $k_{ik} = m$ means that measurement $\boldsymbol{m}_{ik}$ is assigned to track m. It is one of the crucial features of the PMHT algorithm that these assignment variables are modelled as stochastic variables and that the parameters describing their distributions should either be estimated by the algorithm or be given in some other way. The PMHT algorithm requires thus by construction no "hard" measurement-to-track assignments. It should be noted that with respect to this feature, the PMHT differs from the original formulation of the Elastic Arms algorithm. In the Elastic Arms algorithm the assignment variables are stochastic as long as the temperature is different from zero, but the algorithm is formulated in such a way that the $T \to 0$ limit always should be taken in the end. The temperature is introduced merely as a tool to avoid ending up in a local extremum of the energy function during the search for the global minimum. It has recently been shown [4], however, that the most accurate estimates for the Elastic Arms algorithm are not found in the limit $T \to 0$, but rather at a temperature which is related to the variance of the measurement error. This indicates that an approach based on hard assignments is suboptimal.

The construction of the likelihood function is based on a number of independence assumptions. Firstly, the measurements are assumed independent, conditioned on the state vectors and the assignment variables. Secondly, the state vectors in layer k are assumed independent, conditioned on the state vectors in layer $k-1$. Finally, all assignment variables are assumed independent. This gives the fundamental likelihood structure of the PMHT:

$$\begin{aligned} P(\boldsymbol{M},\boldsymbol{X},\boldsymbol{K}) &= \left\{\prod_{\nu=1}^{M}\phi_\nu(\boldsymbol{x}_{0\nu})\right\}\prod_{k=1}^{K}\left\{\left[\prod_{s=1}^{M}\phi_s(\boldsymbol{x}_{ks}|\boldsymbol{x}_{k-1,s})\right]\right. \\ &\quad \left.\cdot\prod_{i=1}^{n_k}\left[\pi_{km}\zeta_m(\boldsymbol{m}_{ik}|\boldsymbol{x}_{km})|_{m=k_{ik}}\right]\right\}. \qquad (1) \end{aligned}$$

Here $\phi_\nu(\boldsymbol{x}_{0\nu})$ is the *a priori* probability density function of state vector ν, $\phi_s(\boldsymbol{x}_{ks}|\boldsymbol{x}_{k-1,s})$ is the probability density function of state vector s in layer k, conditioned on state vector s in layer $k-1$, n_k is the number of points in layer k, $\pi_{km} = P(k_{ik} = m)$ is the prior probability that a measurement in layer k originates from track m, and $\zeta_m(\boldsymbol{m}_{ik}|\boldsymbol{x}_{km})$ is the probability density function of measurement $\boldsymbol{m}_{ik}$, conditioned on state vector $\boldsymbol{x}_{km}$. The quantities $\boldsymbol{M}$, $\boldsymbol{X}$ and $\boldsymbol{K}$ denote the collection of all measurements, state vectors and assignment variables, respectively. The prior probabilities $\boldsymbol{\Pi} = \{\pi_{km}\}$ can be estimated by the algorithm, but other assumptions are also possible.

One possible way to proceed would now be to calculate the posterior probabilities of the state vectors and the assignment variables, given the data, and find the maximum of this function with respect to the state vectors and the assignment variables. Such an approach would yield the *maximum a posteriori* estimate of these quantities. However, this would require a complete enumeration of all possible configurations of the assignment variables and is therefore computationally unfeasible. The approach of the PMHT is to consider the assignment variables as *missing data* and obtain the estimates as the maximum of the marginal probability density function $P(\boldsymbol{M},\boldsymbol{X})$

with respect to the state vectors and possibly the prior probabilities. This can effectively be done by aid of the EM algorithm. In fact, the convergence theorem by Dempster et al. [5] guarantees that in this case we will find at least a local maximum of the marginal probability density function. It can be noted that the estimates are found without knowledge of the probabilities of any particular configuration of the assignment variables. The effect of these different configurations is accumulated in the marginal. This aspect of the PMHT is very similar to the marginalization that is found in the Elastic Arms algorithm.
The EM algorithm consists of an expectation part and a maximization part. The expectation is done by defining a function

$$Q(\boldsymbol{X}, \boldsymbol{\Pi} | \boldsymbol{X'}, \boldsymbol{\Pi'}) = \sum_{\{\boldsymbol{K}\}} \log P(\boldsymbol{M}, \boldsymbol{X}, \boldsymbol{K}; \boldsymbol{\Pi}) \cdot P(\boldsymbol{K} | \boldsymbol{M}, \boldsymbol{X'}; \boldsymbol{\Pi'}), \tag{2}$$

where the sum is over all configurations of the assignment variables. We have herein included the dependence on the prior probabilities explicitely in the expressions. The primed quantities are supposed to be fixed during one EM step, while the others are variable. The measurements $\boldsymbol{M}$ are of course fixed and viewed as constant quantities in all operations. The probability density of the assignment variables given all the other quantities can be shown to be

$$P(\boldsymbol{K} | \boldsymbol{M}, \boldsymbol{X}; \boldsymbol{\Pi}) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} w_{iks}|_{s=k_{ik}}, \tag{3}$$

with

$$w_{iks} = \frac{\pi_{ks} \zeta_s(\boldsymbol{m}_{ik} | \boldsymbol{x}_{ks})}{\sum_{m=1}^{M} \pi_{km} \zeta_m(\boldsymbol{m}_{ik} | \boldsymbol{x}_{km})}. \tag{4}$$

The quantity w_{iks} is to be interpreted as the probability that measurement $\boldsymbol{m}_{ik}$ is assigned to track s, conditioned on the state vectors $\boldsymbol{X}$ and the collection of measurements $\boldsymbol{M}$. Due to the independence assumptions stated earlier this probability does not depend on the locations of any other measurements in the same layer.
By writing out the probability density functions explicitely one can obtain an analytic expression of the Q function of Equation (2). It can be written

$$Q = \sum_{k=1}^{K} Q_{k,\Pi} + \sum_{m=1}^{M} Q_{m,\mathbf{X}}, \tag{5}$$

with

$$\begin{aligned}
Q_{k,\Pi} &= \sum_{i=1}^{n_k} \sum_{m=1}^{M} w'_{ikm} \log \pi_{km}, && (6)\\
Q_{m,\mathbf{X}} &= \log \phi_m(\boldsymbol{x}_{0m}) + \sum_{k=1}^{K} \Big\{ \log \phi_m(\boldsymbol{x}_{km}|\boldsymbol{x}_{k-1,m}) \\
&\quad + \sum_{i=1}^{n_k} w'_{ikm} \log \zeta_m(\boldsymbol{m}_{ik}|\boldsymbol{x}_{km}) \Big\}, && (7)
\end{aligned}$$

where the primes on the assignment probabilities denote that they are a function of $\boldsymbol{X}'$ and $\boldsymbol{\Pi}'$.
The second part of the EM step is to maximize this Q function with respect to the parameters $\boldsymbol{\Pi}$ and the state vectors $\boldsymbol{X}$, and due to the structure of Equation (5) we see that the maximization problem reduces to independent maximizations of the $Q_{k,\Pi}$'s and the $Q_{m,\mathbf{X}}$'s, respectively. For $Q_{k,\Pi}$ this task can be solved by aid of Lagrangian multipliers, and the result is

$$\pi_{km} = \frac{1}{n_k} \sum_{i=1}^{n_k} w'_{ikm}. \tag{8}$$

The same task for $Q_{m,\mathbf{X}}$ is more complicated. However, the following relation can be relatively straightforwardly derived from Equation (7):

$$\exp(Q_{m,\mathbf{X}}) \propto \phi_m(\boldsymbol{x}_{0m}) \prod_{k=1}^{K} \left\{ \phi_m(\boldsymbol{x}_{km}|\boldsymbol{x}_{k-1,m}) \varphi\left(\widetilde{\boldsymbol{m}}_{km}; \boldsymbol{H}_k \boldsymbol{x}_{km}, \widetilde{\boldsymbol{V}}_{km}\right) \right\}, \tag{9}$$

with *effective* measurements and covariance matrices defined by

$$\begin{aligned}
\widetilde{\boldsymbol{m}}_{km} &= \frac{1}{\sum_{j=1}^{n_k} w'_{jkm}} \sum_{i=1}^{n_k} w'_{ikm} \boldsymbol{m}_{ik}, && (10)\\
\widetilde{\boldsymbol{V}}_{km} &= \frac{\boldsymbol{V}_k}{\sum_{j=1}^{n_k} w'_{jkm}}. && (11)
\end{aligned}$$

We have now made the assumption that the measurement probability density function is Gaussian: $\varphi(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{V})$ represents a Gaussian with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$. If we further assume that the track model probability density

function ϕ also is Gaussian, Equation (9) can be seen to be the probability density function of a standard Kalman filter. The only difference is that the measurements and the covariance matrices are replaced by the effective measurements and the effective covariance matrices defined above. Therefore, the output of the Kalman filter and smoother is exactly the state vectors maximizing Equation (9). This completes the maximization part of the EM step.
The PMHT algorithm will then be as follows:

1. Initialize state vectors and prior probabilities.

2. Repeat until convergence

   (a) Update the assignment probabilities according to Equation (4), using the prior probabilities and the state vectors from the previous iteration.

   (b) Update the prior probabilities, the effective measurements and the effective covariance matrices according to Equations (8), (10) and (11), respectively, using the updated assignment probabilities.

   (c) Update the state vectors by running M Kalman filters and smoothers with the updated effective measurements and effective covariance matrices as input.

3. Store state vectors.

In summary, the PMHT is a global tracking algorithm which in the case of Gaussian measurement errors and process noise works by iteratively running M coupled Kalman filters and smoothers in parallel. Inclusion of process noise is straightforwardly done in the same way as for the standard Kalman filter. In the Gaussian case the algorithm does not require the optimization of a non-quadratic objective function. However, the formalism is not restricted to Gaussian noise only. From Equation (7) it can be seen that any probability density function of process noise and measurement noise can be taken into account. In general, in order to obtain the state estimates one will then have to apply numerical optimization techniques. The computational cost will of course be larger in the general case than in the Gaussian case.

3 A generalization of the PMHT algorithm using a Deterministic Annealing EM technique

The EM algorithm is indeed a very powerful tool to obtain good estimates in situations where the data is incomplete. In our case the assignment variables can be viewed as the missing data, i.e. we do not have the information about the correct hit-to-track assignments. Nevertheless, it is known that the standard EM algorithm has some unwanted properties. One of these is the fact that the EM algorithm is guaranteed to converge only to a local maximum of the likelihood function. If this function has many peaks, the performance of the algorithm will become very sensitive to the initialization procedure. In other words, the starting values of the quantities to be estimated have to be very close to the global maximum to obtain the optimal performance.
Several modifications of the EM algorithm have been proposed in order to cope with this problem. One of these is the Stochastic EM algorithm (SEM) [9]. Instead of calculating the full Q function of Equation (2), which might very well be computationally intractable, it draws samples of the assignment variables $\boldsymbol{K}$ according to the distribution $P(\boldsymbol{K}|\boldsymbol{M}, \boldsymbol{X'}; \boldsymbol{\Pi'})$ and calculates the corresponding maximum likelihood estimate of the complete data likelihood function $P(\boldsymbol{M}, \boldsymbol{X}, \boldsymbol{K}; \boldsymbol{\Pi})$. It can be shown that for each SEM iteration there is a non-zero probability of choosing an updated value of the estimated parameters which decreases the incomplete data likelihood function. In contrast to the standard EM algorithm, the SEM will therefore in general not be trapped in the closest local maximum. A possible disadvantage of this method is that for each iteration one has to sample a probability distribution, which in a computer implementation means that a random number generator has to be called maybe a large number of times. The computational cost of this method might therefore be quite large.
In this paper, we propose a generalization of the PMHT algorithm based on a recent development: the Deterministic Annealing EM algorithm (DAEM) [10]. The DAEM has also been constructed in an attempt to overcome the problem of ending up in a local maximum during the search. It basically works by smoothing out the likelihood surface at high temperatures, leading the search into the correct region of parameter space as the temperature gets lower. The final step of the DAEM is always equivalent to the corresponding standard EM algorithm. The PMHT with the Deterministic Annealing EM algorithm corresponds in the linear-Gaussian case to a very simple extension of the standard PMHT: one simply defines a series of temperatures or measurement errors and runs the PMHT to convergence at each of

the temperatures sequentially. This can be seen by realizing that the DAEM uses a modified $P(\boldsymbol{K}|\boldsymbol{M},\boldsymbol{X};\boldsymbol{\Pi})$ (denoted P_β) during the first iterations to calculate the Q function of Equation (2):

$$P_\beta(\boldsymbol{K}|\boldsymbol{M},\boldsymbol{X};\boldsymbol{\Pi}) = \frac{P(\boldsymbol{M},\boldsymbol{X},\boldsymbol{K};\boldsymbol{\Pi})^\beta}{\sum_{\{\boldsymbol{K}\}} P(\boldsymbol{M},\boldsymbol{X},\boldsymbol{K};\boldsymbol{\Pi})^\beta}, \tag{12}$$

where $0 < \beta \leq 1$ and $\beta = 1/T$, T denoting the temperature. For $\beta = 1$, P_β is the same as $P(\boldsymbol{K}|\boldsymbol{M},\boldsymbol{X};\boldsymbol{\Pi})$. The output of the PMHT at one temperature serves as input to the PMHT at the next one. One should start with a large value of the measurement error in the expressions for the assignment probabilities (see Equation (4)), successively decrease this during the iterations and obtain the final estimates at the nominal value of the measurement error. It can be noted that with this approach no random numbers have to be generated. In the next section, we will show that the generalization of the PMHT including such a deterministic annealing scheme leads to significant improvements in the accuracy of the estimated parameters.

There have also been earlier attempts [11] to modify the PMHT by increasing the measurement error during the first iterations, but apparently without any great success. According to the DAEM, however, this is the correct thing to do when the noise can be described by a Gaussian probability density function. We would also like to point out that the DAEM algorithm is not restricted to this assumption. Equation (12) is valid for any probability density, and the DAEM thus gives a well-defined prescription of how to formulate a deterministic annealing PMHT also in the case of non-Gaussian noise. The approach of this work is therefore more general than what has been presented earlier.

As mentioned earlier, due to the independence assumptions applied in the construction of the PMHT likelihood structure, the probability of assigning a hit to a specific track does not depend on the locations of the other measurements in the same layer. In this work we propose to modify these assignment probabilities in order to obtain competition between hits in the same layer. We will therefore use weights with the same structure as those in the DAF and the Elastic Arms algorithm of earlier work [4]:

$$w_{iks} = \frac{\varphi\left(\boldsymbol{m}_{ik};\boldsymbol{H}_k\boldsymbol{x}_{ks},\boldsymbol{V}_k\right)}{n_k \cdot \varphi\left(\boldsymbol{\lambda};\boldsymbol{0},\boldsymbol{V}_k\right) + \sum_{j=1}^{n_k}\varphi\left(\boldsymbol{m}_{jk};\boldsymbol{H}_k\boldsymbol{x}_{ks},\boldsymbol{V}_k\right)}. \tag{13}$$

The weights have been stated for the case of single tracks, since it is this case we will consider in the simulation experiments of the next section. It is not entirely obvious how the weights should look in the multi-track case. In fact, this is the subject of an ongoing study, and the topic will be addressed in a later paper. The quantity $\boldsymbol{\lambda}$ defines a cutoff in the same sense as for the Elastic Arms algorithm, and it can be seen that the prior probabilities do not appear in the expression of the weights. This is due to the fact that there is no reason to believe beforehand that any track deposits energy in the detector more often than any other one. The prior probabilities should therefore be equal and, hence, cancel out. The simulations will be used to determine a suitable magnitude for $\boldsymbol{\lambda}$.

Because of the modification of the weights the algorithm also has to be slightly modified. The expectation part of one basic EM step will now be to update the assignment probabilities using the state vectors from the previous iteration. From these updated assignment weights the effective measurements and covariance matrices are recalculated. The maximization part is to update the state vectors by using the updated effective measurements and covariance matrices as input to the Kalman filters and smoothers. One major point with the modified weights is that they now depend on the locations of all measurements in a layer. It will be demonstrated in the next section that this also leads to improvements in the accuracy of the estimated track parameters compared to the standard case of Equation (4).

With the modifications described above, the PMHT is in the single-track case identical to the DAF. It is easy to see that the filter and covariance matrix updates of the PMHT, using the effective measurements and effective covariance matrices given in Equations (10) and (11), are totally equivalent to the corresponding updates of the DAF, as stated in [4]. The full PMHT is therefore, under the assumption of Gaussian noise, a multi-track generalization of the DAF. In addition, the algorithm can be formulated to handle any kind of process noise and measurement errors. This also makes it more general than the Elastic Arms algorithm, because the latter is optimal only in the case of Gaussian measurement errors and negligible process noise.

4 Simulation experiments of tracks from the ATLAS Inner Detector TRT

This section presents results from simulation experiments of single tracks from the barrel part of the ATLAS Inner Detector TRT. A description of all the details of this detector can be found in [12]. The TRT consists of drift tubes, the measurements are therefore ambiguous. The information of each observation consists of the radius of the layer, the polar angle Φ of the centre of the straw, the drift distance, the sign of z and the track label or KINE number of the particle causing the hit. Since our

measurements are two-dimensional, three parameters are needed to uniquely define a track.
The data sample consists of 9800 “perfect” tracks, i.e. no material effects have been taken into account. However, a measurement error of 250 μm has been simulated. The correct solution of the left-right ambiguity is known and used in the analysis of the results, but not during the reconstruction phase. The PMHT requires an initial guess of the track parameters, and this is provided by a least-squares fit of all points in a track candidate to a straight line in the (R, Φ)-projection. In this projection, all tracks with transverse momenta p_T larger than about 2 GeV/c are approximately straight. Since all our simulated tracks have $p_T > 1$ GeV/c, fitting to a straight line is justified. The accuracy of the track parameters is assessed by the generalized variance, i.e. the determinant of the covariance matrix of the residuals of the estimated track parameters with respect to the true values. All generalized variances are given relatively to the generalized variance obtained by a least-squares fit to only the track points, the mirror hits being turned off. This corresponds to the case of a complete knowledge of all hit-to-track assignments and therefore sets a lower limit on the spread of the estimates of the track parameters.
We will first investigate the effect of the adjusted assignment probabilities, as given in Equation (13), on the accuracy of the estimates of the track parameters. This has been done by running the algorithm on track candidates consisting of track points together with their corresponding mirror hits. A plot of the relative generalized variances in two cases - adjusted assignment weights and standard assignment weights - as a function of the cutoff λ, in terms of standard deviations of the measurement error, is shown in Figure 1. We have also included a cutoff term in the standard weights. The structure of these is therefore similar to the adjusted weights given in Equation (13), the only difference being that the sum in the denominator consists of the one term with index i. An annealing schedule has been included and is the same in both cases. Since the standard weights do not exhibit any competition between a hit and its corresponding mirror hit, the optimal cutoff has to be a compromise between not including too many mirror points and not excluding too many true track points. For the adjusted weights, however, the competition implies that the mirror points usually are given a low weight. The cut should therefore be as loose as possible in order not to lose any good points, and the simulation results indeed confirm this picture. The optimal V_{rel} with the standard weights is 1.85 times larger than the optimal V_{rel} with the adjusted weights, so the gain in accuracy with the adjusted weights is significant.
The results of another experiment are shown in Figure 2, and here noise has been

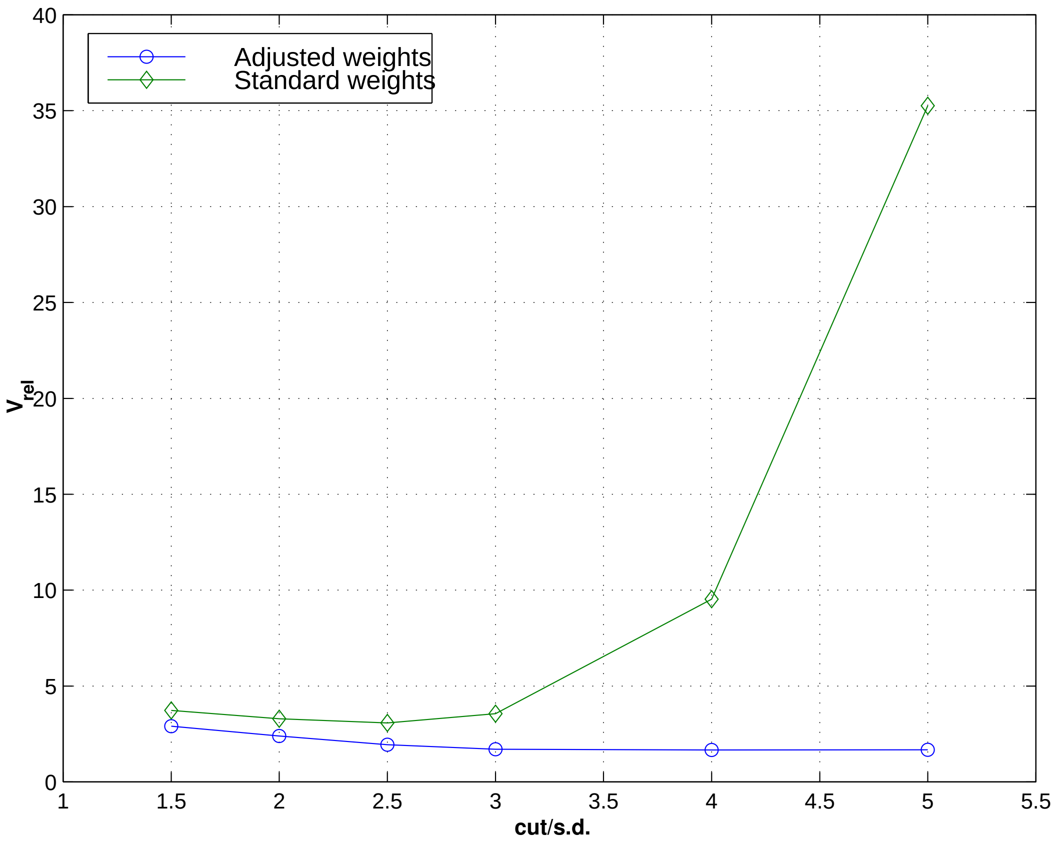

Figure 1: Relative generalized variance for tracks with mirror hits but without noise hits, as a function of the cutoff divided by the standard deviation of the measurement error.

included. This has been done by replacing the correct drift distance by a random one with 10 % probability, which means that for some observations both the hit and the mirror hit are wrongly positioned. Again the relative generalized variances have been plotted as a function of the cutoff in terms of standard deviations of the measurement error. Since some of the hits now are pure noise, the optimal choice of the cutoff for the case of adjusted weights also has to be a compromise of the same type as the one above mentioned. At the optimal cutoff values the generalized variance for the case of standard weights is 1.58 times larger than for the case of adjusted weights.

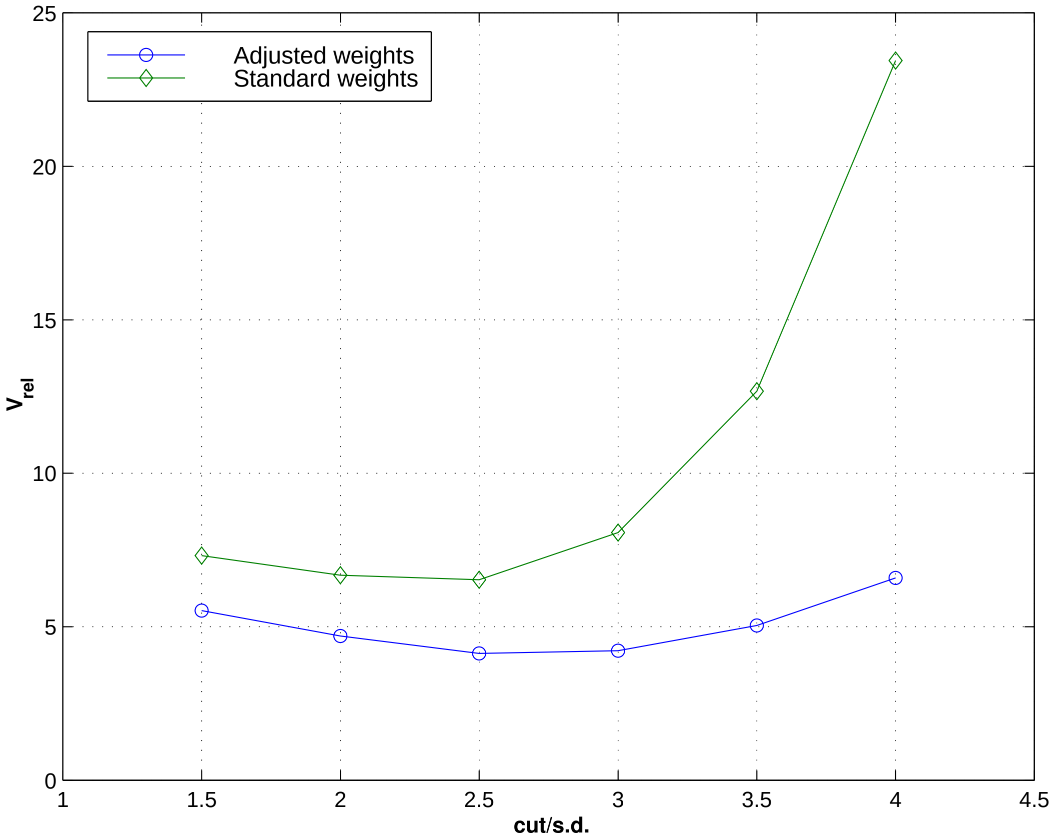

Figure 2: Relative generalized variance for tracks with mirror hits and noise hits, as a function of the cutoff divided by the standard deviation of the measurement error. The noise level is 10 %.

We will then proceed to investigate the influence of the annealing on the accuracy of the estimates. The algorithm therefore has been run both with and without annealing on tracks with a noise level of 10 %. The annealing schedule adopted foresees three different values of the measurement error. At each of the two largest values of the measurement error only one pass of the filter and the smoother has been performed, but in order to make the algorithm converge, four iterations have been allowed at the final, nominal value of the measurement error. In the case of no annealing we have made six EM iterations, making the computing time the same as in the annealing case. Figure 3 illustrates the difference in behaviour for one specific track. The plot is shown in the (R, Φ)-projection. It is here seen how the plain PMHT obviously

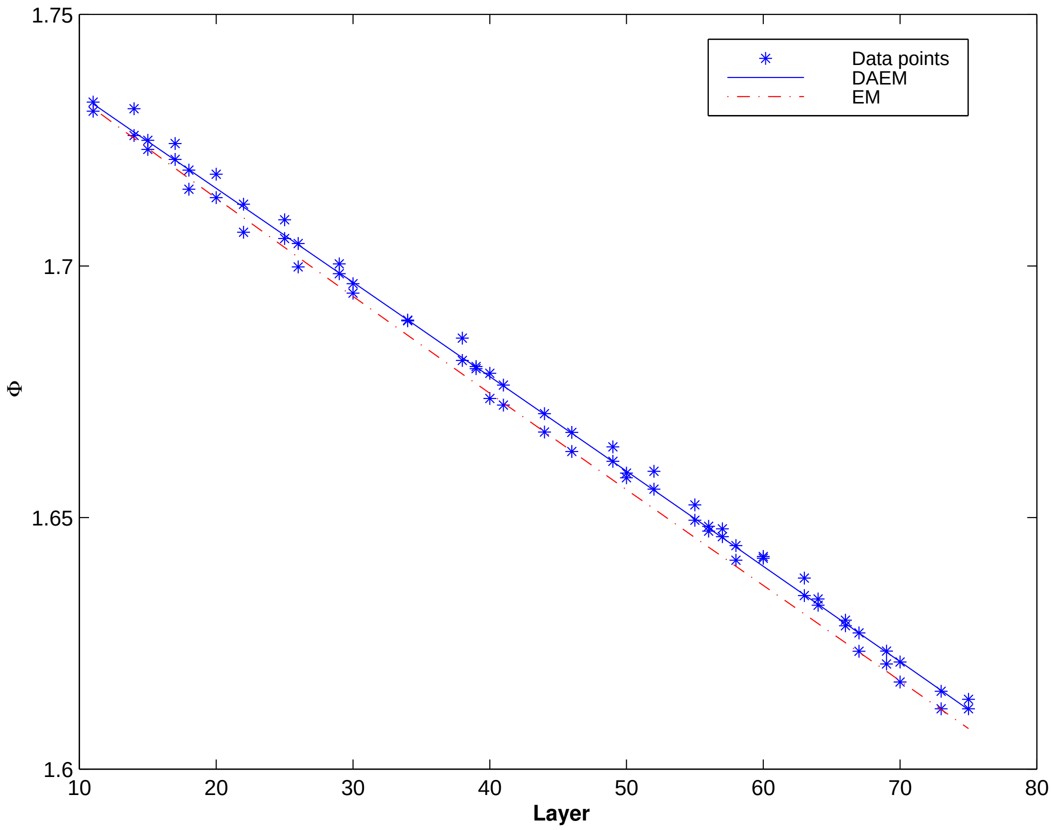

Figure 3: Fitted tracks for the PMHT algorithm with and without deterministic annealing.

ends up in a local maximum of the likelihood function during the EM steps, due to a non-perfect initialization of the track parameters. The PMHT with annealing, however, is able to recover from this and seems to have found the global maximum. The annealing thus makes the algorithm much less sensitive to the quality of the initialization. In the general case this is a very attractive feature, since one often will encounter problems where a good initialization is difficult to achieve.
Figure 4 shows the different generalized variances again as a function of the cutoff. Notice that we have used a semi-logarithmic scale. The lowermost curve in this plot is of course exactly the same as the lowermost curve in Figure 2. Without annealing, the accuracy improves as the cutoff increases. This means that with a loose cut, the algorithm more rarely tends to end up in a local maximum of the

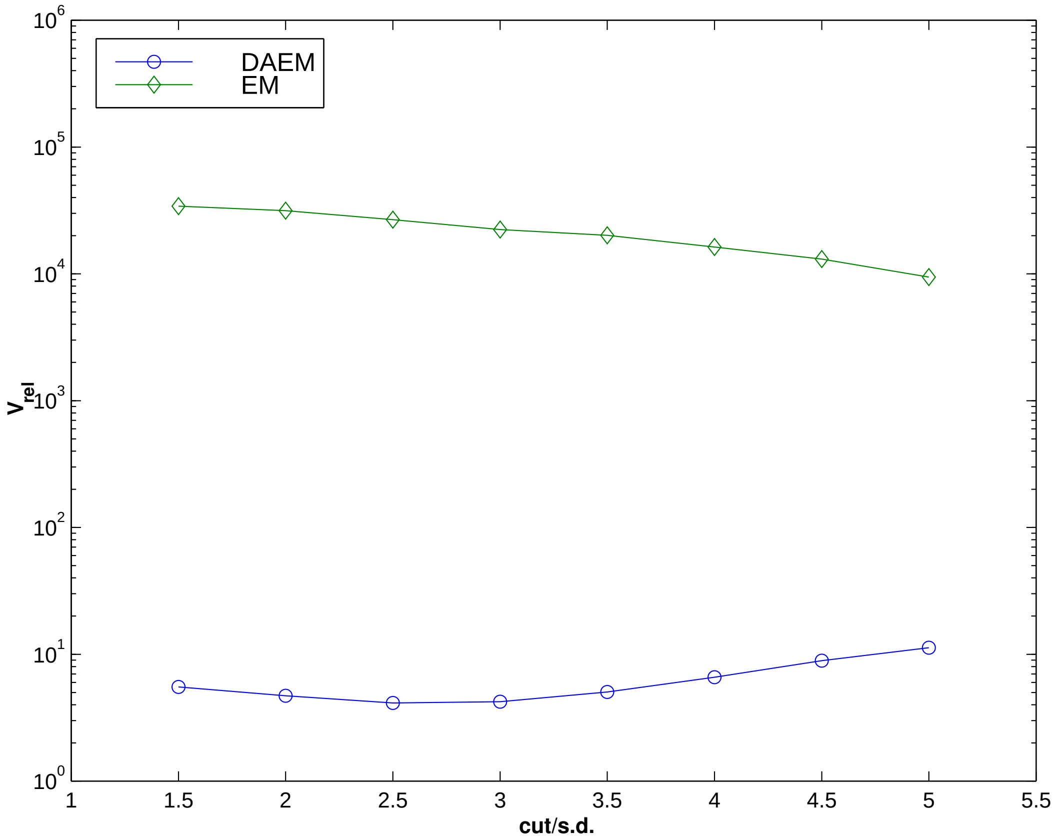

Figure 4: Relative generalized variance for tracks with mirror hits and noise hits, as a function of the cutoff divided by the standard deviation of the measurement error. The noise level is 10 %.

likelihood function. Nevertheless, note that the generalized variance in the optimal cutoff region is worse for the plain PMHT by more than three orders of magnitude. There is a vast improvement in accuracy due to the annealing in this case.
In order to check out the robustness of the PMHT we have also run the algorithm on track candidates with a noise level of 20 %. The best result of V_{rel} is now around 11.0, and this value is about 2.7 times larger than the corresponding value at 10 % noise. Note that on average this implies an increase in the standard deviations of the estimated track parameters of about 18 % only.

5 Conclusions and outlook

In this paper, we have introduced the PMHT algorithm for particle tracking in high-energy physics detectors. It is a global tracking algorithm with similarities to the Elastic Arms algorithm, but with the additional feature of including material effects in the form of process noise in the formalism. We have proposed two new features of the PMHT algorithm in this work. One of them is to adjust the assignment probabilities in order to obtain competition between hits in the same detector layer. The other one is a generalization of the algorithm to include a deterministic annealing scheme. By means of simulation experiments of tracks from the ATLAS Inner Detector TRT, both of these proposals have been shown to significantly improve the accuracy of the estimated track parameters.

With the improvements included, the PMHT is in the single-track case equivalent to the DAF. The full PMHT is therefore, under the assumption of Gaussian noise, a multi-track generalization of the DAF. In addition, the PMHT can be formulated to handle any kind of assumptions regarding the functional forms of the noise probability density functions. The PMHT with deterministic annealing is well-defined also in this very general situation. In contrast, the Elastic Arms algorithm always does a global least-squares fit in the low-temperature limit, regardless of the structure of the underlying probability densities, and is therefore optimal only in the case of Gaussian measurement errors and negligible process noise.

Even though the PMHT has shown to perform very well indeed on single track candidates, the full power of the algorithm will not be exhibited until one considers the case of possibly several tracks competing for the same hit. This might happen for instance in very narrow jets. It would therefore be very interesting to see how the PMHT performs under such conditions. It is not obvious how to formulate the assignment weights in the multi-track case. These topics are currently under investigation, and the outcome of the study will be presented in a subsequent paper.

References

[1] R. Frühwirth, Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods* A262 (1987) 444.

[2] R. Frühwirth, Track fitting with non-Gaussian noise. *Computer Physics Communications* 100 (1997) 1.

[3] R. Frühwirth and A. Strandlie, Track finding and fitting with the Gaussian-sum filter. *Proceedings of the Conference on Computing in High Energy Physics (CHEP'98)*, Chicago, September 1998.

[4] R. Frühwirth and A. Strandlie, Track fitting with ambiguities and noise: a study of elastic tracking and non-linear filters. *To appear in: Computer Physics Communications*.

[5] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society B* 39 (1977) 1.

[6] M. Ohlsson, C. Peterson and A. Yuille, Track finding with deformable templates — the elastic arms approach. *Computer Physics Communications* 71 (1992) 77.

[7] M. Gyulassy and M. Harlander, Elastic tracking and neural network algorithms for complex pattern recognition. *Computer Physics Communications* 66 (1991) 31.

[8] R.L. Streit and T.E. Luginbuhl, *Probabilistic Multi-Hypothesis Tracking.* NUWC-NPT Technical Report 10 428, Naval Undersea Warfare Center, Newport, Rhode Island, 1995.

[9] G. Celeux and J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly* 2 (1985) 73.

[10] N. Ueda and R. Nakano, Deterministic annealing EM algorithm. *Neural Networks* 11 (1998) 271.

[11] P. Willett, Y. Ruan and R. Streit, A variety of PMHTs. *Proceedings of the workshop "Probabilistic methods in multi-target tracking"*, ENST, Paris, 1998.

[12] *ATLAS Inner Detector Technical Design Report.* CERN/LHCC 97-16, CERN, Geneva, 1997.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399