

Article

Probability Distributions for Track Fitting and the Advanced Lucky Model

Gregorio Landi ^{1,*} and Giovanni E. Landi ²

¹ Dipartimento di Fisica e Astronomia, Università di Firenze Largo E. Fermi 2 (Arcetri), 50125 Firenze, Italy

² ArchonVR S.a.g.l., Via Cisieri 3, 6900 Lugano, Switzerland

* Correspondence: gregorio.land@unifi.it

Abstract: Probability distributions for the center of gravity are fundamental tools for track fitting. The center of gravity is a widespread algorithm for position reconstruction in tracker detectors for particle physics. Its standard use is always accompanied by an easy guess (Gaussian) for the probability distribution of the positioning errors. This incorrect assumption degrades the results of the fit. The explicit error forms evident Cauchy–Agnesi tails that render the use of variance minimizations problematic. Therefore, it is important to report probability distributions for some combinations of random variables essential for track fitting: $x = \xi / (\xi + \mu)$, $y = (\xi - \mu) / [2(\xi + \mu)]$, $w = \xi / \mu$, $x = \theta(x_3 - x_1)(-x_3) / (x_3 + x_2) + \theta(x_1 - x_3)x_1 / (x_1 + x_2)$ and $x = (x_1 - x_3) / (x_1 + x_2 + x_3)$. The first two are partial forms of the two strip center of gravity. The fourth is the complete two strip center of gravity, and the fifth is a partial form of the three strip center of gravity. For the complexity of the forth equation, only approximate expressions of the probability are allowed. Analytical expressions are calculated assuming ξ , μ , x_1 , x_2 and x_3 independent Gaussian random variables. The analytical form of the probability for the two strip center of gravity allows one to construct an approximate proof for the lucky model of our previous paper. This proof also suggests how to complete the lucky model by its absence of a scaling constant, relevant to combine different detector types. This advanced lucky model (the super-lucky model) can be directly used in trackers composed of non-identical detectors. The construction of the super-lucky model is very simple. Simulations with this upgraded tool also show resolution improvements for a combination of two types of very different detectors, near to the resolutions of the schematic model.

Keywords: center of gravity; weighted average; probability density functions; position reconstruction; micro-strip detectors; track fitting; least squares method; lucky model



Citation: Landi, G.; Landi, G.E. Probability Distributions for Track Fitting and the Advanced Lucky Model. *Physics* **2022**, *4*, 1026–1049. <https://doi.org/10.3390/physics4030068>

Received: 11 May 2022

Accepted: 4 July 2022

Published: 6 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The top resolutions in track fitting, discussed in [1–3], require special types of probability density functions (PDFs) for their realizations. The maximum likelihood, used to obtain the top resolutions, implies the handling of products of the observation PDFs (likelihood), one PDF for each observation of the fit. Such PDFs must have analytical expressions, appositely tuned on the statistical properties of the used signals. In our case, signals were released by minimum ionizing particles (MIPs) in silicon micro-strip detectors. In fact, a careful observation of the simulated data evidences the impossibility to reproduce the scatter-plots with a single variance (homoscedasticity). The homoscedasticity of the observations implies immense simplification of the fitting problem: the variances of the observations disappear from the equations for the least-squares. Those equations can be applied to any type of fitting problem. The homoscedasticity is a fundamental assumption of mathematical statistics, as repeatedly recalled in [4]. The failure of this key point of mathematical statistics imposes a completely different strategy. The properties of the observations (hits) must be carefully studied to extract the essential information to achieve the fit optimality [5]. The scatter-plots, which provided a hint to these studies, were those

of [6], showing samples of center of gravity (COG) of MIP signals as a function of the MIP impact points, (ϵ).

To explore the relevance of additional pieces of information, those simulations were used to produce seven very rough approximations of the PDFs as functions of ϵ for a fixed interval of COG values. Those rough PDFs were used in a maximum likelihood search of fitted parameters of MIP straight tracks. The evident improvement of the parameter distributions convinced us about the importance of the additional pieces of information. A feature was particularly convincing: the unexpected richness of exact values, absent in the standard fits (homoscedastic least squares). This feature was clearly evident as narrow peaks in all our simulations. However, to effectively use those hints, the rough PDFs need to be replaced with accurate forms such as those derived in the following and used in [1–3]. In addition, we demonstrated in Refs. [5,7] that the usual fitting methods are non-optimal just for the neglect of the variance differences of the hits. Those studies proved that the variances of the fitted parameters, given by the standard fits, are always greater than the variances obtained by weighted fits with the hit variances as weights. The demonstrations are perfectly consistent with our early empirical results.

The aim of this study is to complete the previously employed methods giving the derivations and the explicit expressions of the PDFs essential to obtain our previous results. The implied PDFs are those calculated for the center of gravity (COG) algorithm. The COG algorithm is an easy and efficient positioning algorithm that is widely used in particle physics and in many other technological problems. The generic COG definition is $\sum_j E_j X_j / \sum_j E_j$, where $\{E_j\}$ are the signals of a cluster inserted in the COG and $\{X_j\}$ denote their positions. The different selections of the signals, inserted in the COG expression, generate positioning algorithms with different analytical and statistical properties. Our special attention is directed to the two strip COG (COG₂) for its minimal noise and usefulness for tracks at orthogonal incidence (direction with the worst resolution). The COG₂ is computed with the signals of the leading strip (seed) and the maximum of the two contiguous strips. The corresponding PDF has a typical gap, the explanation of this feature and examples are reported in [8]. It is just this typical feature that renders very complicated the calculation and very long the equation of the PDF for the COG₂. Nevertheless, the COG₂ PDF is extensively used in the simulations of [2,3] with very large improvements for the resolutions of the track parameters (up to a factor three for the most favorable cases). Although our attention is focalized on COG₂, other COG PDFs are calculated as well, and few of them are frequently used in our approaches. However, the COG PDFs are only a part of the probabilities for track fitting, the other part is the insertion of the functional dependence on ϵ . Completed with the particle impact point (ϵ), the PDF can take into account the signal-to-noise ratio of each strip and correct the COG systematic errors. The insertion of the ϵ functional dependence requires the exploration and filtering of special types of random processes and the availability of large sets of homogeneous data [2]. Further details about the handling of these types of random processes will be discussed elsewhere.

As a non-trivial by product, the analytical expression of the COG₂ PDF also allows us to discuss an approximate demonstration for the lucky model [1]. The same equation also consents to an extension of the lucky model to trackers formed with non-identical detectors (a declared limitation for the lucky model in [1]).

In Section 2, the necessity to go beyond the standard (homoscedastic) least squares method is illustrated and the simplest forms of COG PDFs are reported. In the following, the least squares method, which uses the equations for homoscedastic systems, is called “standard least squares.” Sections 3 and 4 are devoted to the complete COG₂ PDF and the PDF for the three strip COG. Applications of these results are reported in Section 5, with proofs and extensions of the lucky model. Section 6 contains the conclusions. Appendix A gives a derivation of a partial COG₂ PDF from the cumulative probability distribution and Appendix B reports on a better approximation of the COG₂ PDF. These results are obtained

with the extended use of MATHEMATICA [9] and verified in many realistic cases with MATLAB [10] simulations.

2. Definition of the Problem

The peculiar non-uniformity (heteroscedasticity) of the point distributions in the scatter plots of COG simulations are easily observed, (as in [2,6] and in the figures shown in this study). In [2], this heteroscedasticity is demonstrated with the definition of an effective variance for each hit and with the distributions of the samples of these values. The plots of those distributions substantially differ from a horizontal line, the obvious result for an identical variance for all hits (homoscedasticity). Therefore, the hypothesis of homoscedasticity for the positioning errors must be ruled out in favor of more realistic assumptions. In fact, the variance reduction, introduced by the distinction of good or excellent hits (heteroscedasticity) on a track containing two or more hits of such types, is easily grasped. This is the reason for the evident richness of the exact values of the heteroscedastic fits, absent in the standard least squares. The distributions of the hit effective standard deviations also show trends surprisingly similar to the COG histograms. This similarity consents the definition of a sub-optimal model (the lucky model [1]) for the hit variances that introduce good improvements in the fitting results, without the complications of the full calculation.

The corrections of the hit properties, due to the differences of the detector technologies along the lines of Ref. [11], are small steps in the right direction but absolutely insufficient. Experimental indications about the differences of the hit resolutions are reported in [12]. However, the Landau distribution of the charge released by a MIP is another well known experimental fact, adding further differences to the hit properties and definitively suppressing the homoscedasticity. Instead, the maximum likelihood method involves all the available information contained in the data, and it is able to obtain drastic improvements of the track parameters also in presence of outliers [2]. This ability to handle outliers is a consequence of the tails of PDFs. As previously discussed, another important feature is introduced by the different quality of hit PDFs. Two goods (or excellent) hits suffice for a good (or excellent) straight track fit as in [2], or three of them for momentum reconstruction in a known homogeneous magnetic field as in [3]. The probability of good (or excellent) hits grows proportionally to the number N of hits (detecting layers) per track [1,5,7]. Therefore, the pool of the track parameters is enriched at this rate ($\approx$ linear in N) for a range of N -values of interest for tracker physics. Instead, the standard least squares, not accounting for the hit differences, grows as $\approx \sqrt{N}$. This is a very slow growth compared with the linear one, implying a waste of tracker resolution and an increase of tracker complexity.

In spite of the proofs of the maximum likelihood as the best fitting method, intrinsic difficulties limit its use. For very complex trackers, the full machinery of the maximum likelihood could be beyond reasonable computing resources. Although the schematic model [2] substitutes the maximum likelihood method with a weighted least squares, the computing of the effective hit variances also requires large processing time. The lucky model of Ref. [1] and its advanced version can be an easy substitute. Evidently, each departing step from the maximum likelihood adds a loss of resolution. It must be remembered that the schematic model, the lucky model, and its advanced form are ineffective with the outliers.

The required analytical expressions of the PDFs have the general functional forms $P(\epsilon, x_{g_n})$, where, x_{g_n} is a generic COG with n -strips. The conditional probabilities, $P(\epsilon|x_{g_n})$ and $P(x_{g_n}|\epsilon)$, are connected with the marginal probabilities, $P(x_{g_n})$ and $P(\epsilon)$ [13]:

$$P(\epsilon|x_{g_n})P(x_{g_n}) = P(\epsilon, x_{g_n}) = P(x_{g_n}|\epsilon)P(\epsilon).$$

The constant probability of the impact point ϵ is assured by

$$P(\epsilon) = \int_{-\infty}^{+\infty} P(\epsilon, x_{g_n}) dx_{g_n} = 1, \quad (1)$$

and it is consistent with the assumption of uniformity and normalization on a strip.

The Kolmogorov axioms [13] attribute to the cumulative probability distribution a fundamental role to calculate the PDF. The cumulative probability distribution for a continuous case is given by integrations on the appropriate portion of the plane, or the space, as required by the geometry of the problem. Differentiating the cumulative probability distribution produces the PDF. This method becomes extremely long with complicated algorithms. However, our first approach was modeled on the ratio of two random variables of [13], and we used the cumulative distributions for all the considered PDFs, from the simplest to the most complex one. This very long set of integrals is quite tedious, to be reported in a paper. This is the principal reason for the delay of this report. Here, a different approach is utilized, very direct and flexible, with the use of Dirac δ -functions and Heaviside θ -functions, operating directly on the PDFs of the strip noise; an assay was given in [3]. This method is a variance of the Fermi “golden rule 1” that is used for the cross-section calculations (or diffusion probabilities), and it recovers the results of our geometric approaches. However, Gauss [14] gives similar generic expressions as obtained after the δ -integration, but we found that old fundamental paper only recently. To underline the consistency with the geometric approach, the first part of the COG₂ PDF is obtained with the cumulative probability distribution in Appendix B.

The random signals of our interest are modeled on the charges released on the strips by the hitting MIPs. The signals are corrupted by additive random noises produced by the rest of the acquisition system. The data are at their final elaboration procedure (calibration, pedestal, common noise suppression, etc.) and are ready to be used in a positioning algorithm of any type. The stream of primary charges, released by a MIP, diffuses in the detector toward a cluster of strips. The charges, collected by a strip, are correlated with those, collected by the cluster. The distributions of the charges in the cluster depend, among other parameters such as particle direction and total released charges, on the MIP impact point. Hence, the ε dependence of $P(\varepsilon, x_{g_n})$ must be contained in the strip signal a_i . The signals a_i are considered as parameters, and the PDF will be expressed in the form $P(\{a_i\}, x_{g_n})$, where $\{a_1\}$ is the set of strips inserted in x_{g_n} . The variable x_{g_n} is abandoned for a simpler $x_{g_n} \rightarrow x$. The strip size is always taken to be one, and this is the length scale of the system (all the implied lengths are divided by the strip widths). The parameters, $\{a_i\}$, can be expressed in any dimensional units consistent with those of the noise $\{\sigma_i\}$ (we use directly the ADC (analog-to-digital converter) counts.) The variable x turns out to be a pure number as the PDFs.

Each strip has its own random additive noise uncorrelated with that of any other strip. In strips without MIP signals, the strip noise is well reproduced with a Gaussian PDF. Thus, the PDFs for the signal plus noise of the strip i become:

$$P_i(z) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(z - a_i)^2}{2\sigma_i^2}\right], \quad i = 1, 2, \dots, n. \quad (2)$$

The Gaussian mean values $\{a_i\}$ are the (noiseless) charges, collected by the strips, and are positive numbers (we assume to have signals from real particles). The parameters $\{\sigma_i\}$ are the standard deviations of the additive zero-average Gaussian noise.

2.1. Probability for the Ratio $x = \xi/(\xi + \mu)$

The first explored PDF is the distribution of the random values of x with $x = \xi/(\xi + \mu)$. This expression has the structure of a COG with the origin of the reference system in the center of the strip 2 (μ random variable) and another signal on the right strip 1 (ξ random variable). This form of COG is the right part of the full COG₂ algorithm. Due to its limitation to only two random variable $\{\xi, \mu\}$, it is a first step toward more complex PDFs.

The derivation of the PDF for $P_{xg_2R}(x)$ with the cumulative distribution is illustrated in Appendix B. However, this PDF can be rapidly obtained with this method:

$$P_{xg_2R}(x) = \int_{-\infty}^{+\infty} \delta\left(x - \frac{\xi}{\xi + \mu}\right) P_1(\xi) P_2(\mu) d\xi d\mu. \quad (3)$$

Integrating the δ -function with $z = \xi + \mu$:

$$\int_{-\infty}^{+\infty} F(z) \delta\left(x \pm \frac{\xi}{z}\right) dz = F\left(\mp \frac{\xi}{x}\right) \frac{|\xi|}{x^2}, \quad (4)$$

the expression for $P_{xg_2R}(x)$ is easily obtained:

$$P_{xg_2R}(x) = \frac{1}{x^2} \int_{-\infty}^{+\infty} P_1(\xi) P_2\left(\frac{1-x}{x}\xi\right) |\xi| d\xi. \quad (5)$$

The heavy tails of a Cauchy–Agnesi-like distribution are evident; Equation (5) shows $1/x^2$ behavior for $x \rightarrow +\infty$, and the factor $(1-x)/x$ goes to -1 . In this limit, the integral is convergent and different from zero. Identically for $x \rightarrow -\infty$. The singularity for $x = 0$ is apparent because one integrand goes to zero for $x \rightarrow 0$ and can be removed with the coordinate transformation $z/x = \xi$. However, it is useful to save the $1/x^2$ factor to remember the divergence of the variance for $P_{xg_2R}(x)$. The Gaussian integral is analytic for any x and a_i and has the following form:

$$P_{xg_2R}(x) = \frac{a_2(1-x)\sigma_1^2 + a_1x\sigma_2^2}{\sqrt{2\pi}[(1-x)^2\sigma_1^2 + x^2\sigma_2^2]^{3/2}} \exp\left[-\left(\frac{a_1}{a_1+a_2} - x\right)^2 \frac{(a_1+a_2)^2}{2(\sigma_1^2(1-x)^2 + x^2\sigma_2^2)}\right] \\ \times \operatorname{erf}\left[\frac{a_2(1-x)\sigma_1^2 + a_1x\sigma_2^2}{\sqrt{2}\sigma_1\sigma_2\sqrt{(1-x)^2\sigma_1^2 + x^2\sigma_2^2}}\right] + \exp\left(-\frac{a_1^2}{2\sigma_1^2} - \frac{a_2^2}{2\sigma_2^2}\right) \frac{\sigma_1\sigma_2}{\pi[(1-x)^2\sigma_1^2 + x^2\sigma_2^2]}. \quad (6)$$

The expression of the $P_{xg_2R}(x)$ shows some aspects that will be found often in the following. One can easily recognize, in Equation (6), part of the PDF reported in [3]. Equation (6) has a maximum for $x \approx a_1/(a_1 + a_2)$. The maximum is the noiseless COG for this combination of variables and tends to eliminate the COG systematic error [8]. Around the maximum, $P_{xg_2R}(x)$ looks similar to a Gaussian. However, the exponential differs from a Gaussian for large x . The modulating term of the exponential is connected to the signal to noise ratio of the two strips. The positivity of the PDF is granted by the term $A \operatorname{erf}(A)$, which, for a not too small A , converges rapidly to $|A|$. Around zero, $A \operatorname{erf}(A)$ is a continuous differentiable function and differs from $|A|$ essentially for the cusp at $A = 0$ of $|A|$. The range of the differences with respect to $|A|$ are of the order of σ_1 (or some weighted average with σ_2). This range is expected to be negligible, if the hit-detection algorithm works well and discards almost all the fake hits (with $a_i = 0$). Thus, very often we will substitute $A \operatorname{erf}(A)$ with $|A|$.

The last term will be indicated as “Cauchy-like” term; it is very similar, but not identical, to a Cauchy–Agnesi PDF. This term survives also for $a_1 = a_2 = 0$ and could be a probability of fake hits. It assures the strict positivity of the PDF. For $a_i \neq 0$, the Cauchy-like term is heavily suppressed by the negative exponents, quadratic in the strip signal-to-noise ratio.

The validity of this PDF is limited to one side of the COG₂ algorithm. The track reconstruction requires a rigid connection to the local reference system, which is naturally locked in the seed-strip center. Therefore, it is important to conserve a difference from the left strip, the central strip, and the right strip. The track impact point, ε , can also be located outside the seed strip.

Another PDF, which composes the complete COG₂ PDF, contains the random variable β , the signal collected by strip 3 positioned to the left of strip 2. This PDF will be indicated as $P_{xg_2L}(x)$. As for $P_{xg_2R}(x)$, the strip 2 is the seed of the strip cluster. As always, the

origin of the reference system is the center of strip 2. Now, x is given by the combination of random variables $-\beta/(\beta + \mu)$. The function $P_{xg2L}(x)$ is obtained from Equation (6) with the substitution $a_1 \rightarrow a_3$, $\sigma_1 \rightarrow \sigma_3$ and $x \rightarrow -x$. We report here $P_{xg2L}(x)$; often in the following, terms of this type will be indicated with the substitutions needed for their construction:

$$P_{xg2L}(x) = \frac{a_2(1+x)\sigma_3^2 - a_3x\sigma_2^2}{\sqrt{2\pi}[(1+x)^2\sigma_3^2 + x^2\sigma_2^2]^{3/2}} \exp\left[-\left(\frac{a_3}{a_3+a_2} + x\right)^2 \frac{(a_3+a_2)^2}{2(\sigma_3^2(1+x)^2 + x^2\sigma_2^2)}\right] \\ \times \operatorname{erf}\left[\frac{(1+x)a_2\sigma_3^2 - a_3x\sigma_2^2}{\sqrt{2}\sigma_3\sigma_2\sqrt{(1+x)^2\sigma_3^2 + x^2\sigma_2^2}}\right] + \exp\left[-\frac{a_3^2}{2\sigma_3^2} - \frac{a_2^2}{2\sigma_2^2}\right] \frac{\sigma_3\sigma_2}{\pi[(1+x)^2\sigma_3^2 + x^2\sigma_2^2]}. \quad (7)$$

The small x approximation is now:

$$P_{xg2L}(x) = \frac{|a_2|}{\sqrt{2\pi}} \frac{\exp\left[-\left(x + \frac{a_3}{a_3+a_2}\right)^2 \frac{(a_3+a_2)^2}{2(\sigma_3^2(1+x)^2)}\right]}{\sigma_3(1+x)^2}.$$

The approximation of $P_2[z(-1-x)/x]$ as a Dirac δ -function in the integration of Equation (5) does not reproduce the Cauchy-like term. The factor $(1+x)$ is retained, being contained in the argument of the δ -function. Its role is essential to obtain the maximum of $P_{xg2L}(x)$ in the expected position $-a_3/(a_3 + a_2)$ of its noiseless COG.

2.2. Probability Distribution for $y = \frac{\xi - \mu}{2(\xi + \mu)}$

Another type of COG₂ algorithm is of frequent use, for example, in [15]. The main difference of this combination of random variables, from previous expressions, is a translation with respect to the standard reference system (centered in the middle of the strip 2). Now, the reference system is centered on the right border of the strip 2. This COG₂ algorithm has the following form:

$$y = x - \frac{1}{2} = \frac{\xi - \mu}{2(\xi + \mu)}. \quad (8)$$

Although this is another direct transformation of Equation (6), we report its general form and the case of Gaussian PDF for completeness:

$$P_G(y) = P_{xg2R}\left(y + \frac{1}{2}\right) = \int_{-\infty}^{+\infty} P_1(\xi)P_2\left(\frac{1-2y}{1+2y}\xi\right) \frac{|\xi|}{(y+1/2)^2} d\xi. \quad (9)$$

In the form of $P_G(y)$, we directly use the substitution of $A \operatorname{erf}(A)$ with $|A|$. In any case, $A \operatorname{erf}(A)$ is easily obtained from Equation (6):

$$P_G(y) = \frac{4|a_2(1-2y)\sigma_1^2 + a_1(1+2y)\sigma_2^2|}{\sqrt{2\pi}[(1-2y)^2\sigma_1^2 + (1+2y)^2\sigma_2^2]^{3/2}} \exp\left[-\left(\frac{a_1-a_2}{2(a_1+a_2)} - y\right)^2\right] \\ \times \frac{2(a_1+a_2)^2}{(\sigma_1^2(1-2y)^2 + (1+2y)^2\sigma_2^2)} \\ + \exp\left[-\frac{a_1^2}{2\sigma_1^2} - \frac{a_2^2}{2\sigma_2^2}\right] \frac{4\sigma_1\sigma_2}{\pi[(1-2y)^2\sigma_1^2 + (1+2y)^2\sigma_2^2]}. \quad (10)$$

With a similar transformation, the PDF for $y = x + 1/2$ can be obtained; here, the reference system is centered on the left border of strip 2 with strip 3. A discussion of the variance of y for small errors is given in [16], although the variance is an ill defined parameter due to the Cauchy–Agnesi-like tails of the PDF. In this case, the results depend on the assumptions introduced.

These PDFs have simple analytical forms; they are defined in reference systems that depend on the signal in the second strip. Their use in maximum likelihood search, could imply heavy complications in the exploration of the likelihood surface. In fact, to allow for the maximum to be outside the two strips of the PDF, another function must be introduced with a different reference system.

2.3. Probability Distribution for $w = \xi/\mu$

As a final use of Equation (5), we apply it to obtain the PDF for the ratio of random variables $w = \xi/\eta$. Now it reads:

$$x = \frac{\xi}{\xi + \mu}, \quad w = \xi/\mu, \quad x = \frac{w}{1+w}, \quad P_{\xi/\mu}(w) = P_{x_{g2R}}\left(\frac{w}{1+w}\right) \frac{1}{(1+w)^2}. \quad (11)$$

The integral expression of $P_{\xi/\mu}(w)$ becomes:

$$P_{\xi/\mu}(w) = \frac{1}{w^2} \int_{-\infty}^{+\infty} P_1(z) P_2\left(\frac{z}{w}\right) |z| dz, \quad (12)$$

and transforming Equation (6) in w , as indicated, the $P_{\xi/\mu}(w)$ for Gaussian PDFs becomes:

$$P_{\xi/\eta}(w) = \frac{a_2\sigma_1^2 + a_1w\sigma_2^2}{\sqrt{2\pi}(\sigma_1^2 + w^2\sigma_2^2)^{3/2}} \exp\left[-\left(\frac{a_1}{a_2} - w\right)^2 \frac{a_2^2}{2(\sigma_1^2 + w^2\sigma_2^2)}\right] \\ \times \operatorname{erf}\left[\frac{a_2\sigma_1^2 + a_1w\sigma_2^2}{\sqrt{2}\sigma_1\sigma_2\sqrt{\sigma_1^2 + w^2\sigma_2^2}}\right] + \exp\left[-\frac{a_1^2}{2\sigma_1^2} - \frac{a_2^2}{2\sigma_2^2}\right] \frac{\sigma_1\sigma_2}{\pi(\sigma_1^2 + w^2\sigma_2^2)}. \quad (13)$$

Now, the maximum of the first term is moved to be around a_1/a_2 . For $a_1 = a_2 = 0$, the last term, the only non-zero, coincides with that reported in [13].

3. The PDF of the Complete COG₂ Algorithm

To obtain the PDF for the COG₂ algorithm, we have to define in detail this algorithm. As previously recalled, the signals of the three strips must be simultaneously accounted for: the strip with the maximum signal (strip 2, the seed strip) and the two lateral strips (strip 1 to the right and strip 3 to the left). Any signal value in these two strips is accepted, as well as any signal value below the threshold for the insertion of a strip signal in a cluster by the cluster detection algorithm (as in [17]). Around strip 2, the strip with the maximum signal is selected between the two strips 1 and 3. Due to the smallest number of strips, this COG₂ has a very favorable signal-to-noise ratio. It is the natural selection for orthogonal incidence on strip detectors with strip widths near to the average lateral drift of the primary charges [6].

3.1. The Definition of the COG₂ Algorithm

The definition of the COG₂ algorithm can be condensed into the following equation [3]:

$$x_{g2} = \frac{x_1}{x_1 + x_2} \theta(x_1 - x_3) - \frac{x_3}{x_3 + x_2} \theta(x_3 - x_1). \quad (14)$$

where x_1, x_2, x_3 are the random signals in the three strips and $\theta(z)$ is the Heaviside θ -function: $\theta(x) = 0$ for $x \leq 0$ and $\theta(x) = 1$ elsewhere. The two θ -functions select the lateral strip with the highest signal. No condition is imposed on the strip 2, although it has some constraints for its role of seed strip. This choice eliminates inessential complications and saves the normalization of the PDF.

Our aim is to reproduce the gap for $x_{g2} \approx 0$, which is typical of the histograms of the COG₂ algorithm. This gap is given by the impossibility (or lower probability) to have $x_{g2} \approx 0$ if the charge drift populates one or both the two lateral strips. The gap

grows rapidly with an increase of these two charges. The noise and our selection allow forbidden values.

The η -algorithm of Ref. [18] uses a slightly different definition. The term $-x_3/(x_2 + x_3)$ of Equation (14) is modified in $x_2/(x_2 + x_3)$. In this way, the values of η are contained in the interval $\{0 \vdash 1\}$, and the gap for $x_{g2} \approx 0$ is spread at the borders of this interval.

The constraints of Equation (14), on the three random signal, $\{x_1, x_2, x_3\}$, are inserted in the integral for the PDF of $P_{x_{g2}}(x)$. The integral expression is given by (with the usual substitution of x_{g2} as x):

$$P_{x_{g2}}(x) = \int_{-\infty}^{+\infty} dx_1 dx_2 dx_3 P_1(x_1) P_2(x_2) P_3(x_3) \left[\delta\left(x - \frac{x_1}{x_1 + x_2}\right) \theta(x_1 - x_3) + \delta\left(x + \frac{x_3}{x_3 + x_2}\right) \theta(x_3 - x_1) \right]. \quad (15)$$

The normalization of $P_{x_{g2}}(x)$ can be immediately proved with a direct x -integration. The other integrals are executed with the following transformations: $x_1 = \xi$, $x_1 + x_2 = z_1$, $x_3 = \beta$, and $x_3 + x_2 = z_2$. The Jacobian of each couple of transformations is one. The integrals on z_1 and z_2 of the two δ -functions can be performed with the rule of Equation (4). The general form of $P_{x_{g2}}(x)$ for any type of signal PDF $\{P_1, P_2, P_3\}$ becomes:

$$P_{x_{g2}}(x) = \frac{1}{x^2} \left[\int_{-\infty}^{+\infty} d\xi P_1(\xi) P_2\left(\xi \frac{1-x}{x}\right) |\xi| \int_{-\infty}^{\xi} d\beta P_3(\beta) + \int_{-\infty}^{+\infty} d\beta P_3(\beta) P_2\left(\beta \frac{-1-x}{x}\right) |\beta| \int_{-\infty}^{\beta} d\xi P_1(\xi) \right]. \quad (16)$$

The Gaussian PDFs of Equation (2), inserted in Equation (16), allow for the explicit expression of the two integrals on $P_3(\beta)$ and $P_1(\xi)$ with the appropriate erf-functions. Indicating the remaining integration variable as z , Equation (16) becomes:

$$P_{x_{g2}}(x) = \frac{1}{2\pi\sigma_1\sigma_2x^2} \times \int_{-\infty}^{+\infty} dz |z| \left\{ \exp\left[-\frac{(z-a_1)^2}{2\sigma_1^2} - \frac{\left(\frac{(1-x)z}{x} - a_2\right)^2}{2\sigma_2^2}\right] \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{a_3-z}{\sqrt{2}\sigma_3}\right)\right] + \frac{\sigma_1}{\sigma_3} \exp\left[-\frac{(z-a_3)^2}{2\sigma_3^2} - \frac{\left(\frac{(-1-x)z}{x} - a_2\right)^2}{2\sigma_2^2}\right] \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{a_1-z}{\sqrt{2}\sigma_1}\right)\right] \right\}. \quad (17)$$

The combination of erf-functions and $|z|$ render an analytical integration of Equation (17) impossible. The serial development of the erf-function and its successive integration term by term is too cumbersome to be of practical use. (A generalization of the erf-function to the two dimensions is missing). Thus, we have to explore analytical approximations apt to be useful in a maximum likelihood search.

3.2. Small- $|x|$ Approximation

The small $|x|$ approximation is one of the easiest ways to handle Equation (17). The function $P_2(z(1-x)/x)$ can be transformed to approximate a Dirac δ -function for small $|x|$:

$$\frac{\exp\left[-\left(\frac{1-x}{x}z - a_2\right)^2 \frac{1}{2\sigma_2^2}\right]}{\sqrt{2\pi}\sigma_2} = \frac{\exp\left[-\left(\frac{z}{a_2} - \frac{x}{1-x}\right)^2 \frac{(1-x)^2 a_2^2}{2x^2\sigma_2^2}\right]}{\frac{\sqrt{2\pi}\sigma_2|x|}{a_2|(1-x)|}} \frac{|x|}{a_2|(1-x)|} \approx \frac{|x|}{a_2|(1-x)|} \delta\left(\zeta - \frac{x}{1-x}\right) \quad \zeta = z/a_2. \quad (18)$$

The effective standard deviation of the Gaussian is $\sigma_2|x|/(a_2|1-x|)$; this term, for $|x| \rightarrow 0$, produces the convergence of the Gaussian toward a Dirac δ -function. The term $|1-x|$ is useful to obtain the combination $a_1/(a_1+a_2)$ in the exponent of the Gaussian-like function. A similar transformation can be applied to $P_2[z(-1-x)/x]$; the integration on ζ is now immediate, and the small $|x|$ probability P_{xg_2} becomes [3]:

$$P_{xg_2}(x) = \frac{|a_2|}{2\sqrt{2\pi}} \left\{ \frac{\exp\left[-\left(x - \frac{a_1}{a_1+a_2}\right)^2 \frac{(a_1+a_2)^2}{2(\sigma_1^2(1-x)^2)}\right] \left(1 - \operatorname{erf}\left[\left(\frac{a_3}{a_3+a_2} - x\right) \frac{a_2+a_3}{\sqrt{2(1-x)\sigma_3}}\right]\right)}{\sigma_1(1-x)^2} + \frac{\exp\left[-\left(x + \frac{a_3}{a_3+a_2}\right)^2 \frac{(a_3+a_2)^2}{2(\sigma_3^2(1+x)^2)}\right] \left(1 - \operatorname{erf}\left[\left(\frac{a_1}{a_1+a_2} + x\right) \frac{a_1+a_2}{\sqrt{2(1+x)\sigma_1}}\right]\right)}{\sigma_3(1+x)^2} \right\}. \quad (19)$$

The term a_2 is a positive constant (the noiseless charge of the seed strip) and the absolute value can be eliminated, but for future developments it is better to remember its presence. It is easy to recognize in Equation (19) the two maxima of Equations (6) and (7), the noiseless position of the two branches of the COG₂ algorithm. The main difference is due to the two $[1 - \operatorname{erf}(-z)]/2$ functions, these types of functions are similar to continuous (and derivable) Heaviside θ -functions. They interpolate, in a very smooth way, the two sides of the PDF. Two different simulated distributions are reported in [2,3] and compared with Equation (19); the first one did not contain the Landau fluctuations, while the second one contained the approximate Landau fluctuations. At orthogonal incidence, the Landau fluctuations are well described by the fluctuations of the total collected charges.

The approximation of Equation (19) reproduces, in a reasonable way, the COG₂ PDF for non-small- x too. Actually, the real useful range of x is $|x| \leq 0.5$, and the factor that is supposed small is $|x|\sigma_2/a_2$. However, the constant a_2 is connected to the seed of the cluster, and it has a high probability to be larger than a few times σ_2 . The a_2 noisy detected part, x_2 , must assure a reasonable detection efficiency of the hit. Certainly, Equation (19) becomes of no use around $x = \pm 1$. Anyway, improved approximations are always useful, given that the probability $P_{xg_2}(x)$ has to be applied to a large set of experimental configurations. The conceptual incompleteness of Equation (19) is the lack of the normalization. The normalization assures a constant probability of the impact point (Equation (1)), but its lack is not a real limitation for a practical use of Equation (19).

3.3. A Better Approximation for $P_{xg_2}(x)$

A more accurate approximation for $P_{xg_2}(x)$ can be obtained retaining the small- x approximation for the two erf-function of Equation (17) and integrating on z the remaining parts. Now, the two integrals have analytical forms, one identical to Equation (6) (a part a factor 1/2) and the other to Equation (7). This approximation also saves the normalization, obviously within the precision of a numerical integration of a heavy-tail PDF. It must be recalled that normalization is the only converging integral for all these PDFs.

As with simplification described above, we substitute $A \operatorname{erf}(A)$ for $|A|$. In any case, the expressions of the erf-functions are those of Equations (6) and (7). Additionally, the Cauchy-like terms are neglected. They are very small. For example, the seed charge in some experiments is selected to be around $6 \sim 7\sigma$; thus, the term $\exp(-a_2^2/2\sigma_2^2)$ could be around $\exp(-18) \approx 10^{-8}$. In some special conditions, these terms might be useful for the outlier suppressions [2] that depend on the PDF tails. As previously stated, these terms assure the strict positivity of $P_{xg_2}(x)$ for $a_1 = a_2 = a_3 = 0$. However, we did not insert those terms in our track reconstructions.

$$P_{xg_2}(x) = \frac{|a_2(1-x)\sigma_1^2 + a_1x\sigma_2^2|}{2\sqrt{2\pi}[(1-x)^2\sigma_1^2 + x^2\sigma_2^2]^{3/2}} \exp\left[-\left(\frac{a_1}{a_1+a_2} - x\right)^2 \frac{(a_1+a_2)^2}{2(\sigma_1^2(1-x)^2 + x^2\sigma_2^2)}\right] \quad (20)$$

$$\times \left\{1 - \operatorname{erf}\left[\left(\frac{a_3}{a_3+a_2} - x\right) \frac{a_2+a_3}{\sqrt{2}(1-x)\sigma_3}\right]\right\} + (a_1 \leftrightarrow a_3, \sigma_1 \leftrightarrow \sigma_3, x \rightarrow -x).$$

Although Equation (20) represents a better approximation, compared to Equation (19), it shows appreciable deviation from the numerical integral of Equation (17) in some extreme cases. For example, for tracks with large inclination, the combination of parameters $a_1/(a_1+a_2)$ are quite close to $a_3/(a_3+a_2)$ and slightly lower than 0.5. In this case, the two maxima are widely separated, and the $P_{xg_2}(x)$ of Equation (20) shows discrepancies, compared to the numerical integration of Equation (17). These discrepancies are absent, or heavily reduced, in the longer approximation, given in Appendix B.

Some properties of this PDF and its extensions, which are useful in track reconstruction, are illustrated with a set of simulations in Section 5.

4. Simplified form of the Three Strip COG

To test the accuracy of the functions $\{a_i(\varepsilon)\}$, the reconstruction of the two-strip and three-strip COG (COG₂ and COG₃) histograms were extensively used in [2]. For this task, the COG₃ PDF was also essential. To simplify, we leave out of this discussion the complications of the full form of the COG₃ PDF and their gaps at the strip borders as discussed in [8], limiting the demonstration to the easiest COG₃ form. This incomplete PDF is useful in all the cases when the border gaps are very small (near orthogonal incidence):

$$P_{xg_3}(x) = \int_{-\infty}^{+\infty} dx_1 dx_2 dx_3 P_1(x_1) P_2(x_2) P_3(x_3) \delta\left(x - \frac{x_1 - x_3}{x_1 + x_2 + x_3}\right). \quad (21)$$

Again, the normalization of $P_{xg_3}(x)$ is easily verified. The substitution of variables, $\xi = (x_1 - x_3)$, $z = (x_1 + x_2 + x_3)$, and $\beta = x_2$, simplifies the Dirac δ -function integration. The Jacobian of the substitution equals 1/2. Integrating over ξ the Dirac δ -function, the remaining double integral has the following form:

$$P_{xg_3}(x) = \frac{1}{2} \int_{-\infty}^{+\infty} dz d\beta |z| P_1\left(\frac{z(1+x) - \beta}{2}\right) P_2(\beta) P_3\left(\frac{z(1-x) - \beta}{2}\right). \quad (22)$$

The integration in β is a convolution of Gaussian PDFs and provides another Gaussian. Due to the $|z|$, the integral on z produces the term of the form $A \operatorname{erf}(A)$ that, as above, we approximate as $|A|$. The introduction of the auxiliary constants X_3 and E_3 further simplifies the form of $P_{xg_3}(x)$. The Cauchy term, indicated with $P_{xg_3}^C(x)$, is the first one discussed and it is:

$$P_{xg_3}^C(x) = \exp\left\{-\frac{E_3^2 [\sigma_1^2(X_3-1)^2 + \sigma_2^2(X_3)^2 + \sigma_3^2(X_3+1)^2]}{2(\sigma_1^2\sigma_2^2 + 4\sigma_1^2\sigma_3^2 + \sigma_3^2\sigma_2^2)}\right\} \quad (23)$$

$$\times \frac{\sqrt{\sigma_1^2\sigma_2^2 + 4\sigma_1^2\sigma_3^2 + \sigma_3^2\sigma_2^2}}{\pi[(1-x)^2\sigma_1^2 + x^2\sigma_2^2 + (1+x)^2\sigma_3^2]},$$

$$X_3 = (a_1 - a_3)/(a_1 + a_2 + a_3), \quad E_3 = a_1 + a_2 + a_3.$$

The term $P_{xg_3}^C(x)$ for $\sigma_1 = \sigma_2 = \sigma_3$, as it is often, has a very simple form:

$$P_{xg_3}^C(x) = \exp\left[-\frac{E_3^2 (X_3^2 + 2/3)}{4\sigma_1^2}\right] \frac{\sqrt{2/3}}{\pi(x^2 + 2/3)}.$$

This term also survives for $E_3 = 0$ and becomes an exact Cauchy–Agnesi PDF. The main term $P_{xg_3}(x)$ is:

$$P_{xg_3}(x) = \exp \left\{ -(X_3 - x)^2 \frac{E_3^2}{2[(1-x)^2\sigma_1^2 + x^2\sigma_2^2 + (1+x)^2\sigma_3^2]} \right\} \times \frac{|E_3 [(1-X_3)(1-x)\sigma_1^2 + X_3x\sigma_2^2 + (1+X_3)(1+x)\sigma_3^2]|}{\sqrt{2\pi} [(1-x)^2\sigma_1^2 + x^2\sigma_2^2 + (1+x)^2\sigma_3^2]^{3/2}}, \quad (24)$$

the approximation of $A \operatorname{erf}(A)$ as $|A|$ has no observable differences in our realistic simulations. Anyway, the erf-function is:

$$\operatorname{erf} \left\{ \frac{E_3 [(1-X_3)(1-x)\sigma_1^2 + X_3x\sigma_2^2 + (1+X_3)(1+x)\sigma_3^2]}{[2(\sigma_1^2(1-x)^2 + x^2\sigma_2^2 + (1+x)^2\sigma_3^2)(\sigma_2^2\sigma_3^2 + \sigma_1^2(\sigma_2^2 + 4\sigma_3^2))]^{1/2}} \right\}. \quad (25)$$

The upper part of the fraction in the erf-argument in Equation (25) coincides with the corresponding term in the absolute value of Equation (24). The numerator of the erf-argument represents rather peculiar recurrent form (in the main term (24)); the σ_i^2 of each strip is multiplied by the noiseless COG₃, referred to in the position of that strip and identically for the running x . A similar feature can also be seen in the COG₂ PDF.

A complete form for the COG₃ algorithm must account for the gap at the strip borders. This happens when the signal distribution is larger than two strips, and the values of $x \approx |0.5|$ are suppressed in the COG₃ creating gaps at the strip borders. The suppression increases rapidly as the (average) signal distribution grows beyond the two-strip size. Near to the strip borders, the noise can increase the signal collected by the nearby strip (as the strip 1), becoming the seed of another three-strip cluster. In this case, the COG₃ algorithm must operate with the signal triplet $\{x_2, x_1, x_4\}$, where x_4 is the signal of the strip to the right of the strip 1. Thus, the algorithm becomes:

$$x_{g_3} = \left(\frac{x_1 - x_3}{x_1 + x_2 + x_3} \right) \theta(x_2 - x_1) + \left(\frac{x_4 - x_2}{x_1 + x_2 + x_4} + 1 \right) \theta(x_1 - x_2). \quad (26)$$

The two sides of Equation (26) are defined in the identical reference system centered on strip 2. A similar expression is required for the gap at the other border of the strip 2. The details of these extensions of the COG₃ algorithm are reported in [19] (and references therein.) These are the instruments to move toward large angles of incidence. However, the selection of the COG algorithm depends on the sizes and shapes of the average signal distributions. Such parameters must be extracted from the data with the equations from [2].

5. Applications

5.1. Plots and Simulations with Equation (20)

A set of simulations highlight some peculiar properties of the PDF of Equation (20). The other PDFs are of minor use at the orthogonal incidence. The effective use of Equation (20) in track reconstruction, and similarly for all the other PDFs, requires the completion with the functional dependence from the hit impact point. The method for this completion is discussed in [2], with the construction of the functions $a_j(\epsilon)$. Unlike their definition in the previous equations, the $a_j(\epsilon)$ are constructed normalized to one. For tuning them to each hit, the parameter E_0 scales the a_j to the appropriate values; E_0 is the noiseless total charge collected by the three strips of each hit. The functions $a_j(\epsilon)$ are obtained by large averages of hit positioning given by the η -algorithm [6]. The data of the test-beam [20] were used for those constructions. The unbiasedness of the η -algorithm and the special averages [2] give to the a_j realistic functional dependencies from the hit impact point ϵ . The finiteness of the averages leave small artifacts, but they have no detectable effects on the distributions of the fitted parameters.

A simple simulation can be used to verify the consistency of Equation (20) and to illustrate the weak gap present in a distribution of simulated x_{g2} in the scatter-plot of Figure 1. The data of the simulation are generated with the function `randn` of MATLAB and with the equations $x_i = \sigma_1 \text{randn}(1, N) + a_i * E_0$ and inserted in Equation (14) to calculate a large sample of x_{g2} . The values for a_i ($a_1 = 0.01$, $a_2 = 0.91$, $a_3 = 0.08$, $E_0 = 150$ ADC counts, and $\sigma_1 = 8$ ADC counts) and σ_i are extracted from [2] for orthogonal incidence on the two types of silicon detectors studied there. The type of detector considered in Figures 1 and 2 is of normal type with a high noise of 8 ADC counts, one of the two types studied in [2,3]. To simplify the simulations, all the σ_i are identical to the most probable noise of the strips. The noisy strips are very few, and it is a useless complication to explore different σ_i . The left side of Figure 2 reports the overlaps of the analytical PDF and the PDF obtained from the simulated data (empirical PDF).

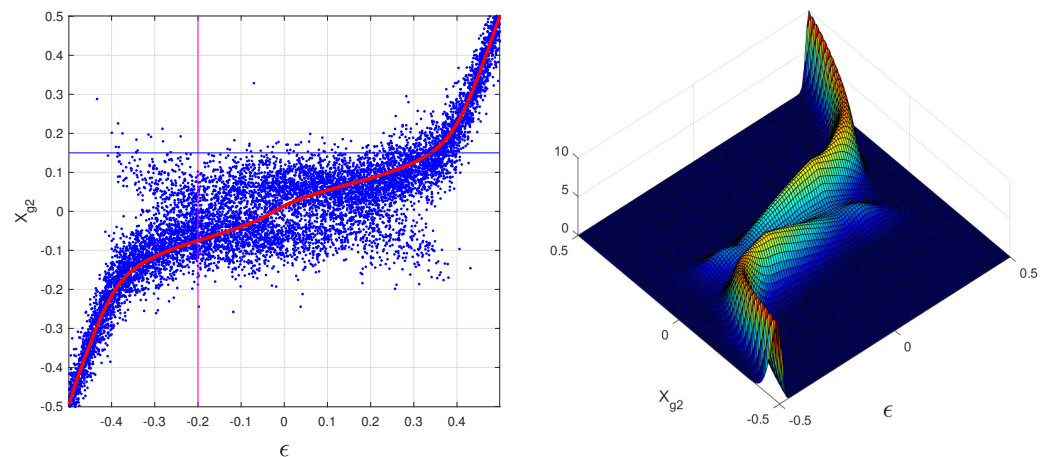


Figure 1. **Left:** scatter-plot of a sample of simulated x_{g2} data as a function of the impact point ϵ ; the red line is the η -algorithm. The magenta line and the blue line are for the following figures. **Right:** a two-dimensional (2D) illustration of the PDF for x_{g2} in function of the impact point ϵ for $E_0 = 150$ ADC counts. See text for details.

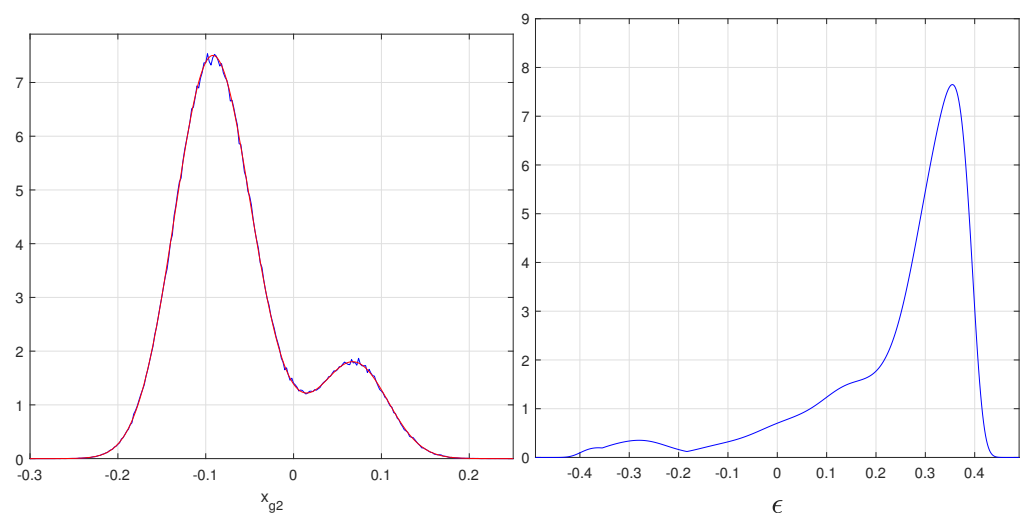


Figure 2. **Left:** empirical PDFs of x_{g2} (blue line), compared with Equation (20) (red line), for a model of silicon detector [2] and an impact point $\epsilon = -0.2$; all the $\sigma_i = 8$ ADC counts, and $a_1 = 0.01 E_0$, $a_2 = 0.91 E_0$, and $a_3 = 0.08 E_0$ are the charges, collected by the three strips; here, E_0 , the total charge of the three strips, is 150 ADC counts (the magenta line of Figure 1). **Right:** the probability distribution for $x_{g2} = 0.15$, the blue line in the scatter-plot of Figure 1.

The probability decreases between the principal and secondary maximum of Figure 2, originating a similar reduction along the magenta line in the scatter-plot of Figure 1. The

secondary maximum is produced by the noise that promotes the minority noiseless signal to become the greater one. The signal clusters with a lower total charge show larger gaps.

The right side of Figure 2 shows the probability distribution of the hit impact points for a given value of the two strip COG x_{g2} . $x_{g2} = 0.15$ is the constant value of x_{g2} , the blue line in the scatter-plot of Figure 1. The integral of this distribution provides the probability for this value of x_{g2} for 150 ADC counts. The integration is limited to a two-strip width, centered on the maximum of the distribution. In this way, the histogram of x_{g2} is analytically reconstructed and reproduces well the histogram of the simulated data [2]. Products of probability distributions, similar to that of Figure 2, are used to find the maximum likelihood of a set of hits for a track. Figure 2 also shows the presence of outliers in the tail of the distribution. These outliers are difficult to handle because they are masked as good hits in the schematic model. Instead, the maximum likelihood is able to avoid their disturbances in the fits. In [2], we illustrate one of the worst outliers and how the maximum likelihood finds almost the exact parameters of the fitted track. We accumulated many similar simulated events with excellent maximum likelihood reconstructions and the worst for schematic model and standard fit reconstructions.

Figures 3 and 4 report plots similar to Figures 1 and 2 for the floating strip detector [2,20]. These plots illustrate the large differences of this excellent detector type with very low noise and the floating strips, which are able to distribute the incoming signal to the nearby strips. The gap for $\text{COG}_2 \approx 0$ is a real gap without data in a scatter-plot with a moderate number of events. The effects of the low noise (of 4 ADC counts) are quite evident in the distributions.

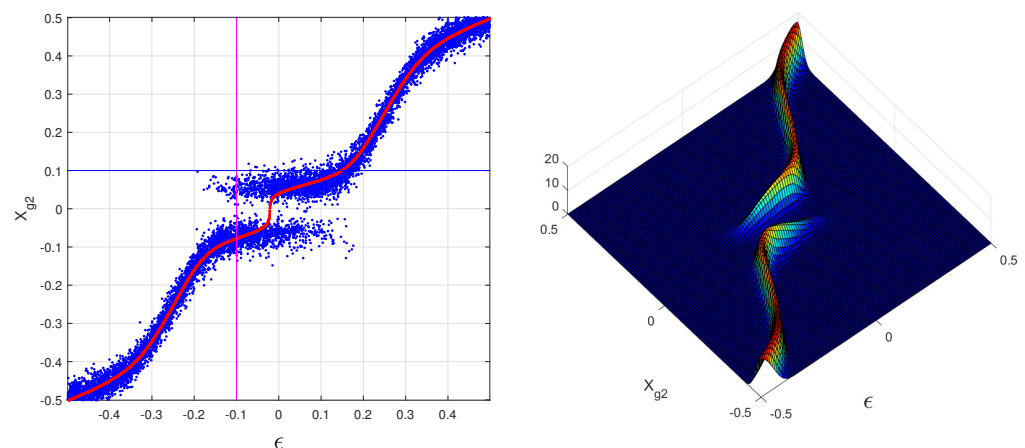


Figure 3. Same as Figure 1 but for floating strip detectors. **Left:** scatter-plot of x_{g2} data as a function of ϵ . **Right:** 2D PDF for x_{g2} as a function of ϵ .

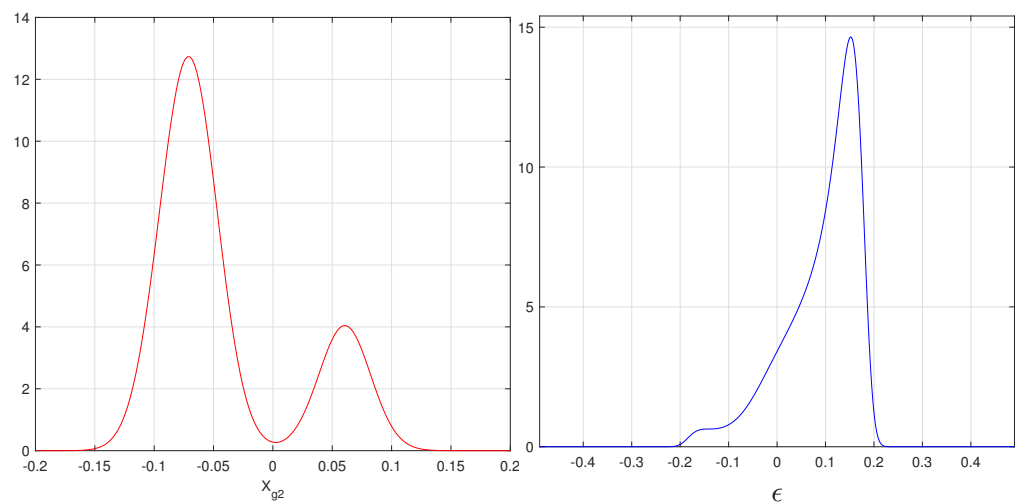


Figure 4. Floating strip detector. **Left:** analytical PDFs of x_{g2} for $\epsilon = -0.1$ and $E_0 = 150$ ADC counts. **Right:** the probability distribution for $x_{g2} = 0.1$, the blue line in the scatter-plot of Figure 3.

5.2. The Lucky Model

The availability of the analytical equations for the COG PDFs allows for a direct discussion about the lucky model of Ref. [1]. The structure of [1] had to heavily deviate from its aims to complete this discussion. As sketchily illustrated in [1], this sub-optimal model is suggested by the similarity of the trends of the effective standard deviations for the schematic model and the trends of the x_{g2} histograms (similarly for the COG₃ histograms). Those effective standard deviations were obtained from the variances of functions of ε , two of them illustrated in the right side of Figures 2 and 4, as explained in [2,3]. The scatter-plots of these effective standard deviations clearly show the trends of the x_{g2} histograms. In [2], we gave an approximate motivation of this non-obvious correlation. We can complete those motivations with more details. The reasoning of [2] supposes (as always) a uniform population of events on a strip; thus, for regions with large effective variances, a relative larger fraction of events ends up with the corresponding x_{g2} -values. Instead, for small effective variances, a relative lower fraction of events ends up with these x_{g2} -values. Assuming the correctness of this explanation, a test of its effectiveness in improving a fit is a natural output. Although the result of this “lucky” test of [1] was successful, a detailed supporting proof is essential for confident use. This proof was a pledge of reference [1].

Actually, the construction of this proof requires an analysis of Equation (20) and the left sides of Figures 2 and 4, unavailable tools in [1].

Equation (20) shows the $P_{x_{g2}}(x)$ as formed by two bumps (approximately Gaussian) whose maxima follows the noiseless function $x_{g2}(\varepsilon)$. Beyond the center of the strip, the last part of each bump decreases rapidly as illustrated in the two-dimensional 2D PDF of Figures 1 and 3. As discussed in [6], the η -algorithm produces functions that show strong similarity with the noiseless $x_{g2}(\varepsilon)$, in the relevant parts of the two bumps. Therefore, we can use the functions $x_{g2}(\eta)$, of the η -algorithm, in place of the exact noiseless $x_{g2}(\varepsilon)$, and to follow the path of the bump maxima and the two corresponding positions at the half-maximum. The distance between these two positions is the full-width-at-half-maximum of the $P_{x_{g2}}(x)$ in the direction of the hit impact point. Each of these two points follows paths parallel to the noiseless $x_{g2}(\varepsilon)$, one above and the other below the path followed by the bump maximum. Again, we approximate these paths with the $x_{g2}(\eta)$ -function. The two sides of Figure 5 (out of scale for a better illustration) present these paths for the two types of detectors considered in [2,3]. The vertical segments are the full-width-at-half-maximum of $P_{x_{g2}}(x)$ and are defined as $\sigma_{x_{g2}}$. The positioning errors σ_η of a generic x_{g2}^0 , possible realizations of their COG₂, are the horizontal black lines. The amplitudes of the horizontal segments are the positioning errors relevant for the weighted least squares.

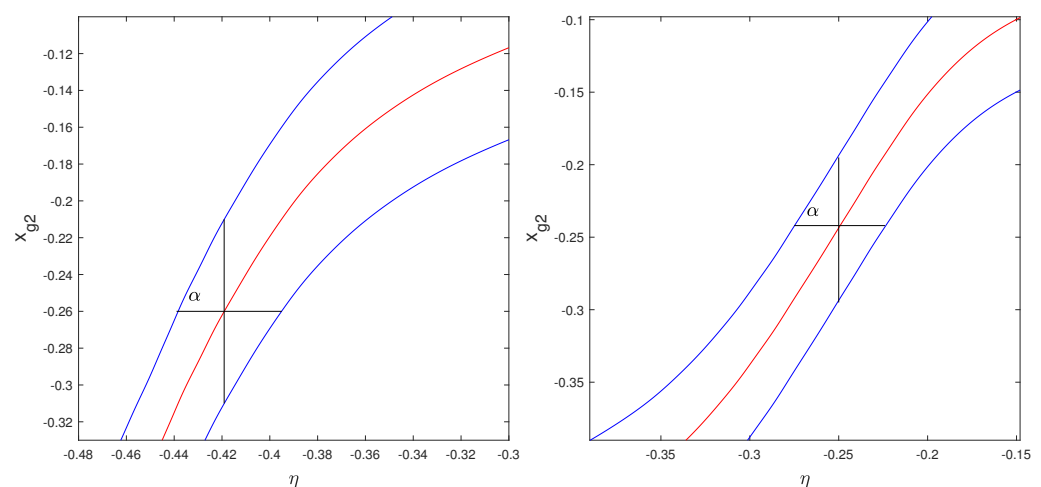


Figure 5. Relation between the amplitude of the x_{g2} error of Equation (20) and the positioning error of a given value of COG₂ due to an observation for normal strip detector (left) and for floating strip detector (right). See text for details.

The blue lines in Figure 5 are curved lines and only approximatively describe a triangle, while the ratio of $\sigma_{x_{g2}}$ and σ_η can be estimated as

$$\tan(\alpha) \approx \frac{\sigma_{x_{g2}}}{\sigma_\eta}, \quad \frac{1}{\tan(\alpha)} = \frac{d\eta}{dx_{g2}} \Big|_{x_{g2}^0} = \Gamma(x_{g2}^0), \quad \sigma_\eta \approx \frac{\sigma_{x_{g2}}}{\tan(\alpha)} = \sigma_{x_{g2}} \Gamma(x_{g2}^0), \quad (27)$$

where $\Gamma(x_{g2}^0)$ is the amplitude of the normalized histogram of COG₂ for the value x_{g2}^0 . Actually, the starting point of the η -algorithm of [6] is the differential equation $d\eta/dx_{g2} = \Gamma(x_{g2})$. Neglecting the differences of $\sigma_{x_{g2}}$, the η -algorithm gives all the elements for a good fit; the corrections of the COG₂ systematic errors and the weight for track fitting in an array of identical detector layers. In this case, the expressions of the parameter estimators of the fit are independent from the (assumed) constant $\sigma_{x_{g2}}$. Therefore, the values of $\Gamma(x_{g2})$ can be used directly as weights of the observations in the weighted least squares of a track.

In the presence of large data gaps, (i.e., the absence of data) in the histogram of COG₂, the function $x_{g2}(\eta)$ acquires discontinuities that complicate the plots of Figure 5, as discussed in [8]. Anyhow, the presence of large gaps in the COG histograms should be avoided for the excessive loss of information. It is preferable to use the COG-algorithms with more strips.

Advanced form for the Lucky Model: The Super-Lucky Model

The precedent discussion is a justification of the simple lucky model of [1]. We utilized this model with a set of identical detectors, directly inserting the amplitudes of the COG₂ histograms in place of effective standard deviations of the observations. The scaling factor, ($\sigma(x_{g2}) = \text{constant}$) of this approximate guess, is simultaneously present in the numerator and denominator of the expressions of the fitted parameters and simplifies. This simplification is impossible for arrays with detectors of different types. Hence, the (assumed) constant values of the scaling factors ($\sigma_{x_{g2}}$) become relevant to tuning the amplitudes of different histograms with the properties of the different detectors. The full calculation of some effective standard deviations of the schematic model looks unavoidable. However, the demonstration of Equation (27) recalls one's attention to the possible variations of $\sigma_{x_{g2}}$. Let us suppose that the principal bumps of Equation (20) are close to the form of Gaussian PDFs. We can push this assumption forward and extract from Equation (20) approximate forms for $\sigma_{x_{g2}}$ (abandoning the full-width-at-half-maximum in favor of the usual standard deviation):

$$\sigma_{x_{g2}}^R = \frac{\sqrt{\sigma_1^2(1 - |X^R|)^2 + \sigma_2^2(X^R)^2}}{a_1 + a_2}, \quad X^R = \frac{a_1}{a_1 + a_2} \quad (28)$$

$$\sigma_{x_{g2}}^L = \frac{\sqrt{\sigma_3^2(1 - |X^L|)^2 + \sigma_2^2(X^L)^2}}{a_3 + a_2}, \quad X^L = -\frac{a_3}{a_3 + a_2}.$$

Actually, a careful substitution of X^R and X^L in place of x in the two branches of Equation (20), is able to transform the two bumps in two Gaussian PDFs. Anyhow, even these approximate forms are outside our reach; a_1 , a_2 , and a_3 are the noiseless signals released by the MIP. Only their noisy version is available. Similarly for X^R and X^L , they represent the noiseless COG₂, the ratio of noiseless signals. The only well defined parameters are σ_1 , σ_2 , and σ_3 , the strip noises, calculated at the initialization stage of the strip detectors. However, we can try to combine the noisy data to see what might happen (we were lucky with the COG₂ histogram in Ref. [1]). The COG₂ algorithm is described in Equation (14) with the definition of x_{g2} ; therefore, the approximate (super-lucky) Σ_{sup} could be:

$$\Sigma_{\text{sup}} = \frac{\sqrt{\sigma_1^2(1 - |x^R|)^2 + \sigma_2^2(x^R)^2}}{x_1 + x_2} \theta(x_1 - x_3) + \frac{\sqrt{\sigma_3^2(1 - |x^L|)^2 + \sigma_2^2(x^L)^2}}{x_3 + x_2} \theta(x_3 - x_1), \quad (29)$$

$$x^R = \frac{x_1}{x_1 + x_2}, \quad x^L = -\frac{x_3}{x_3 + x_2}, \quad \sigma_\eta = \Sigma_{\text{sup}}(x_{g2}) \Gamma(x_{g2}).$$

The insertion of Equation (29) into the calculation of the effective standard deviations of the observations can be compared with those of the schematic model for the two widely different detector types of our simulations. We also tested the use of the soft cut-offs with the erf-functions of Equation (20) in place of the sharp cut-offs of Equation (29), but this insertion adds complications without any visible effect on the simulations.

The upper plots of Figure 6 illustrates the relations of the effective standard deviations from Refs. [2,3] for the two schematic models and the corresponding amplitudes of the lucky model. The lower plots of Figure 6 show the surprisingly good overlaps of the guessed weights of Equation (29) for the advanced lucky model (hereafter, the super-lucky model) with the two schematic models of [2,3]. Additionally, the limitation to an identical set of detectors is solved by the super-lucky model.

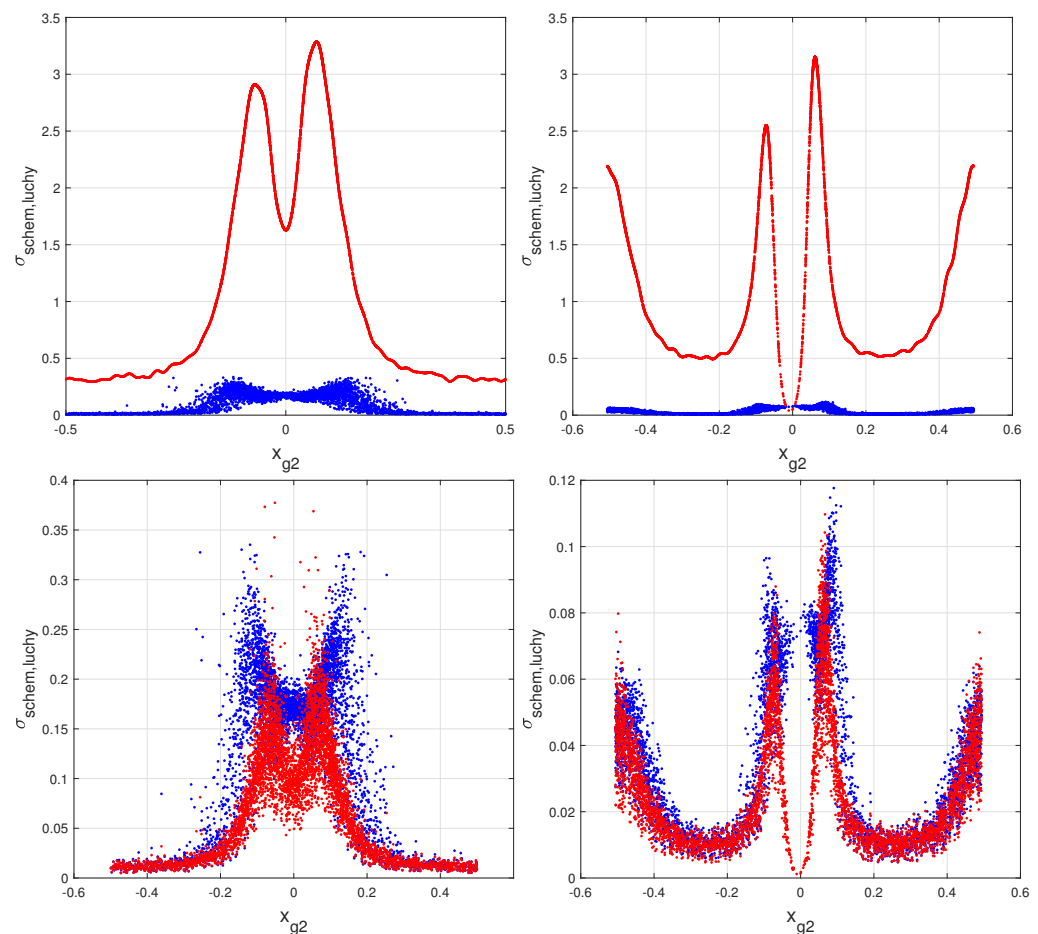


Figure 6. Top: blue dots represent the schematic model, red dots the COG₂-histogram for normal strip detector (left) and floating strip detector (right). Bottom: blue dots represent the schematic model, red dots the $\sigma_\eta(x_{g2})$ of Equation (29) for normal strip detector (left) and for floating strip detector (right). See text for details.

Figure 7 shows the quality of the track reconstructions produced by the super-lucky model that is quite close to that of the schematic model, which is substantially better than the simple lucky model (not plotted for a better readability). Similarly to that in [1], Figure 7 reports the empirical PDFs for the fitted directions (the tangents of small angles)

of 150,000 straight tracks at orthogonal incidence on a set of parallel and equidistant detector layers. To clearly visualize the differences of the parameter distributions, the first distributions with two detector layers are centered on zero (as it must be); the other distributions with $N = 3, N = 4, \dots$, up to $N = 13$, are shifted by $N-2$ identical steps. We always start from two detector layers as a check of the methods.

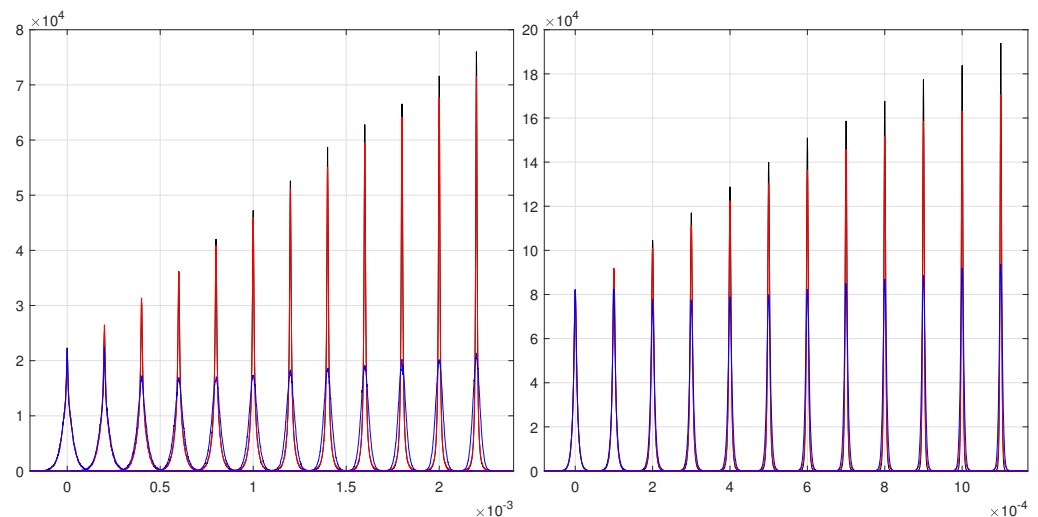


Figure 7. Comparison of the schematic model (black lines), the super-lucky model (red lines), and the standard model (blue lines) for the track direction reconstructions with 2, 3, ..., 13 layers for normal strip detectors (**left**) and for floating strip detectors (**right**). See text for details.

An interesting comparison is the relations of the maxima of the distributions of Figure 7 and the standard deviations of those distributions. We have to recall the possible complications with the standard deviation for systems with PDFs having tails similar to the Cauchy–Agnesi PDFs. As in Refs. [1,5], we report $1/\sqrt{2\pi S_d^2}$, where S_d is the standard deviation of one of the empirical PDFs in Figure 7. For the Gaussian PDF, this ratio coincides with the maximum of the PDF. As expected by the Cauchy–Agnesi tails of the PDFs, we observe large distances from the maxima in these plots. The maxima of the standard least squares are the nearest.

The results of the lucky model are not reported to avoid an excessive complication in the plots. However, the simulations with the lucky model for the normal strip detector are reported in Ref. [1]. The results of the lucky model for the floating strip detector are appreciably lower ($\approx 12\%$) than the results of the super-lucky model.

5.3. The Super-Lucky Model for the Combination of Two Very Different Detector Types

The super-lucky model was studied just for the application of the lucky model to trackers with different types of detector layers. Thus, we simulated a set of trackers starting with two detector layers, one for each type, and added via alternation a floating strip detector and a normal detector. Figures 7 and 8 show the large differences in resolution of these two types of detectors composing this new system; the floating strip detectors with a low noise of four ADC counts and the normal detectors with a noise of eight ADC counts. The floating strips add further improvement in resolution because they spread the signal in the nearby strips.

The left sides of Figures 9 and 10 illustrate the similarity in the resolution of the super-lucky model and the schematic model for this non-homogeneous set of trackers. For comparison, in the right sides of Figures 9 and 10, we also report the simulation with the simple lucky model. The improvements of the super-lucky model are evident. Despite its inconsistency in this case, the simple lucky model shows a substantial increase of resolution compared with the standard least-square. The rough pieces of information contained in the model are able to enrich the parameter distributions of the exact values. Instead, the

standard deviations of the parameters of this simple model are close enough (although better) to those of the standard least-squares, signaling large tails given by tracks without good hits. The super-lucky model has standard deviations appreciably better than the lucky model; thus, the added corrections (29) recover part of the tails. However, the effects due to the large tails of the distributions can be attenuated by the selection of tracks with two or more good hits.

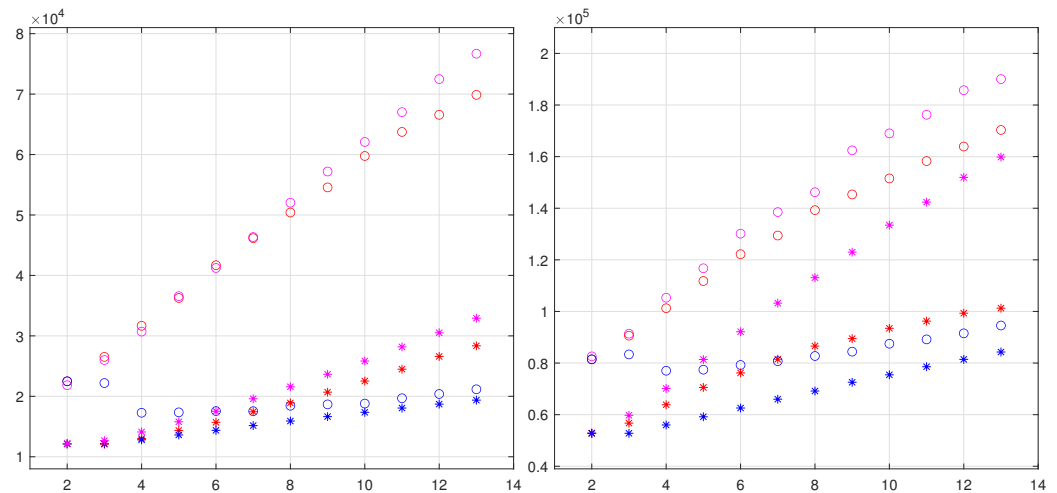


Figure 8. Comparisons of maxima and $1/\sqrt{2\pi S_d^2}$, S_d is the standard deviation of one of the three distributions for the normal strip detectors (**left**) and for the floating strips detectors (**right**). The magenta markers represent the schematic model, the red markers the super-lucky model, and the blue markers the standard fit. The circles indicate the maxima, and the asterisks the ratio of the standard deviations. See text for details.

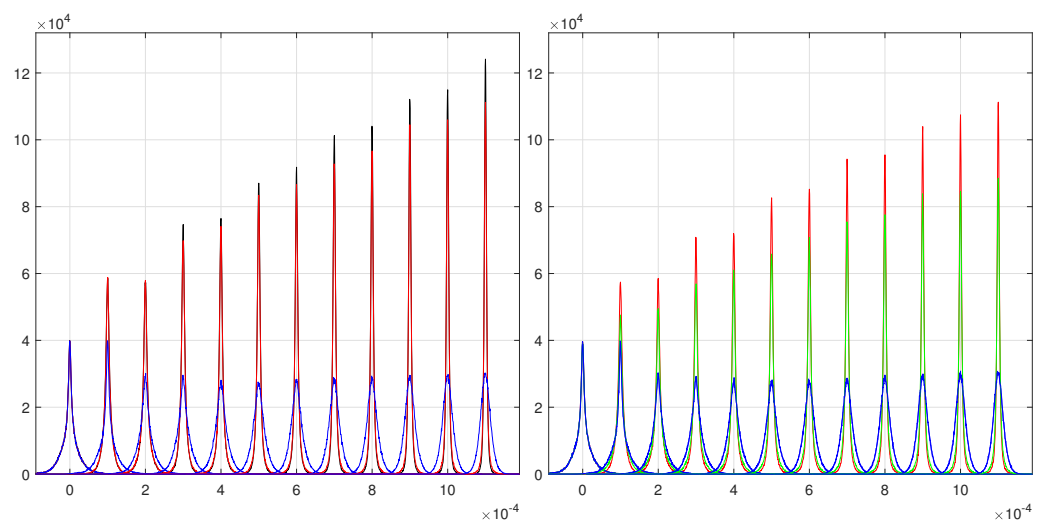


Figure 9. **Left:** comparison of the schematic model (black lines), super-lucky model (red lines), and the standard model (blue lines) for the track direction reconstructions half detector layers of floating strip detector and half with normal detector layers. **Right:** the same half-and-half detector arrays. The green lines represent the simpler lucky model, and other lines are those of the left plot.

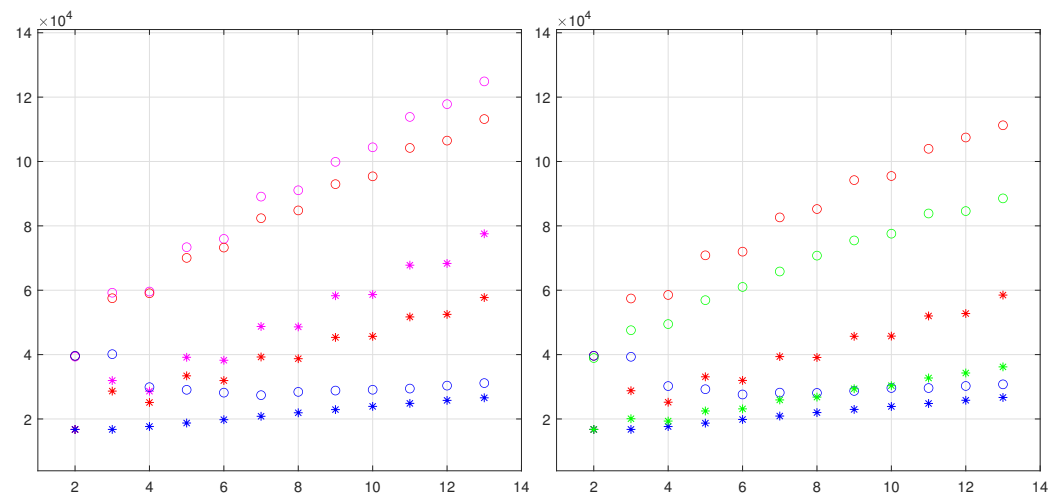


Figure 10. Mixed detector quality. Comparison of maxima and $1/\sqrt{2\pi S_d^2}$ for the normal strip case (left) and for the comparison with the simpler lucky model shown by green circles (right). The magenta markers represent the schematic model, the red markers the super-lucky model, and the blue markers the standard fit. The circles indicate the maxima, and the asterisks the ratio of the standard deviations. See text for details.

In this set of trackers, the growth of the maxima are dominated by the floating strip detectors; the addition of a layer of this type of detectors shows an evident increase of the maxima of the distributions. Instead, the addition of a layer of normal detectors has a negligible effect. The usual linear growth shown in the two sides of Figure 8 transforms into the step-like growth in Figures 9 and 10. Additionally, the standard least-square fit shows a drastic reduction of the maxima and a general deterioration in comparison with the simpler analysis of the parts with good detector types. All other models also show improvements for the addition of low-quality detectors. In [5], the resolutions for these non-Gaussian distributions are defined by the maxima of the distributions: the higher the maximum, the better the resolution. The method of fitting a Gaussian in the core of the distributions (as, e.g., in [12,21]) gives too high values for these very narrow distributions. The demonstrations from Refs. [5,7] prove that the standard least squares model is never optimum outside the homoscedastic systems. The results of the previous two suboptimal models enforce the power of those demonstrations. Actually, any deviation from homoscedasticity, as well as with weak correlations with the true variances for the observations, is able to improve the fit resolution beyond the results of the standard least squares.

These results are exclusively obtained with the hit position, given by the η -algorithm, as it is shown in the two sides of Figure 5. The use of the positions, given by the COG₂-algorithm (as often done), totally suppresses the goodness of these results, and the parameter distributions are lower than those of the standard least-squares fits, (even these substantially lower than those given by the η -algorithm).

5.4. Further Discussions and Comparisons for the Simulations and Data

For a happy coincidence, we found a very effective approximation of the schematic model. When we started to write Equation (27), we had no idea of those further upgradings. They were suggested by the environment in which the equation was written. Surprising enough were the overlaps among the effective standard deviations for excellent hits; they are illustrated in the low part of Figure 6. Further checks also showed that the differences of these parameters, for the same hit, are around $\approx 10\%$, 20% for many thousands of excellent hits. This is a very improbable random event; thus, a mathematical explanation has to be found. We have to recall that the construction of the effective variances in the schematic model contains an arbitrariness to be fixed. Equation (11) of [3] reports a detailed description of this construction. The probability distributions in the integrals are

similar to those at the right side of Figures 2 and 4, very different from Gaussian PDFs for the COG₂ around zero. Instead, in the regions of excellent hits for the floating strip detectors, the simulations show a strong similarity with the Gaussian PDFs for the core of the distributions. The tails are surely different (Cauchy–Agnesi type) and had to be cut to avoid divergences. The effective variances showed a sensible dependency from the cut position. Our decision about the cut was essentially “aesthetical”; the effective variances had to reproduce the core with Gaussian PDFs for a small number (four or five) of excellent hits where that core was quite similar to a Gaussian. At the same time, the systematic comparison of the real PDFs with Gaussian PDFs gives an indication of the deviations from optimality. Actually, the Gaussian PDFs are the optimum for the schematic model. Evidently, this selection produces effective standard deviations very near to those of the super-lucky model almost perfect for excellent hits. Such hits turned out to be far more numerous [2] than our best expectations.

In any case, a different selection of the cuts can change this convergence and perhaps the fit quality for the loss of optimality of those excellent hits. At the moment of these selections, we were in a very early stage of the work, with no idea of the results.

Other interesting comparisons are possible: the $\sigma_\eta(x_{g2})$ of the lucky model from our simulations and the data. The calculation of these expressions are very fast (contrary to the schematic model that requires a huge number of numerical integrations). Figure 11 shows these comparisons for the two types of detectors. The blue dots are now the $\sigma_\eta(x_{g2})$ of the super-lucky model given by the data from a test-beam [20], and the red ones are those of the low side of Figure 6. The overlaps are excellent, showing the high quality of our method of simulation [2]. Although, the histograms of the simulated signal distributions on the strips show excellent overlaps with those from data; this type of comparison had never been explored before for evident reasons. The $\sigma_\eta(x_{g2})$ -values have a non-trivial relation with the data, and these strict correlations enforce our system of extracting/defining [2] the average signal distribution on the strips.

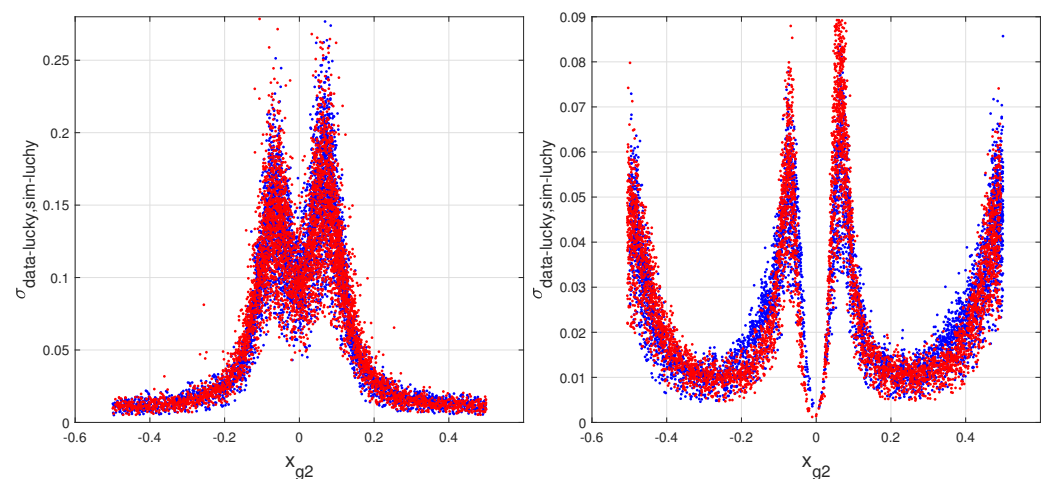


Figure 11. Blue dots: the $\sigma_\eta(x_{g2})$ of Equation (29) from data of the test beam [20]. Red dots: the $\sigma_\eta(x_{g2})$ of Equation (29) from simulations. **Left:** Normal strip detectors. **Right:** Floating strip detectors.

It is reasonable to expect that the use of the $\sigma_\eta(x_{g2})$ -values, in real track fitting, provide the results comparable to those of Figures 7 and 8 for the detector types or combinations (Figures 9 and 10).

6. Conclusions

This is a part of a study for COG PDFs, essential tools to go beyond the methods based on variance minimizations. The long analytical equations, reported here, are indispensable components to operate with the maximum likelihood method. Although complex and slow, the maximum likelihood could be able to obtain results impossible with other methods.

For example, the elimination of the effects introduced by the outliers. The increase of the track parameter resolution could also reduce the complexity of the tracker hardware, requiring fewer detection layers to obtain the resolution of the standard least squares method (or of its equivalent the Kalman filter). These equations were on our desk for a long time, but the huge length of the standard demonstrations forbade their publications. The method, illustrated here, allowed manageable demonstrations. The produced expressions can be handled with the essential help of MATHEMATICA. The numerical simulations with MATLAB complete the verification of the full process. The analytical expressions also allow an explanation of the lucky model and to find an advanced form (super-lucky model) able to extend its use in trackers composed of detectors with very different properties. The reported simulations show a substantial increase of the parameter resolution of the super-lucky model, well beyond the results of the standard least squares and very near to those of the schematic model. This method, for the easy availability of its composing elements, adds negligible complications to the fits with substantial increases of resolution.

Author Contributions: Conceptualization, G.L. and G.E.L.; methodology, G.L.; and software, G.L. and G.E.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used were simulated by the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADC	analog-to-digital converter
PDF	probability density function
MIP	minimum ionizing particle
COG	center of gravity
COG _n	center of gravity with n strips

Appendix A. The Partial COG₂ PDF and Its Cumulative Probability Distribution

We report a synthetic calculation of the PDF for $\xi/(\xi + \eta)$ along the lines of [13] for the ratio of two random variables. The PDF is obtained differentiating the cumulative probability distribution for the random variable x . The cumulative distribution is defined as the probability to have $\xi/(\xi + \eta) \leq x$. Thus, the product of $P_1(\xi)P_2(\eta)$ must be integrated on regions of the plane (η, ξ) compatible with the defined condition. Figure A1 illustrates the geometry of this calculation.

We have to select two different procedures, one for $x \leq 0$ and one for $x > 0$. The two lines of equation, $\xi + \eta = 0$ and $\xi(1 - x)/x = \eta$, are the boundaries of the integration regions. The first line is fixed and separates the two regions with different signs of the denominator of $\xi/(\xi + \eta)$. The other line rotates around the origin as x increases, and it is the second boundary of the integration regions. It overlaps the line $\xi + \eta = 0$ when $x \rightarrow \pm\infty$. The η -axis separates the two regions with $x \neq 0$.

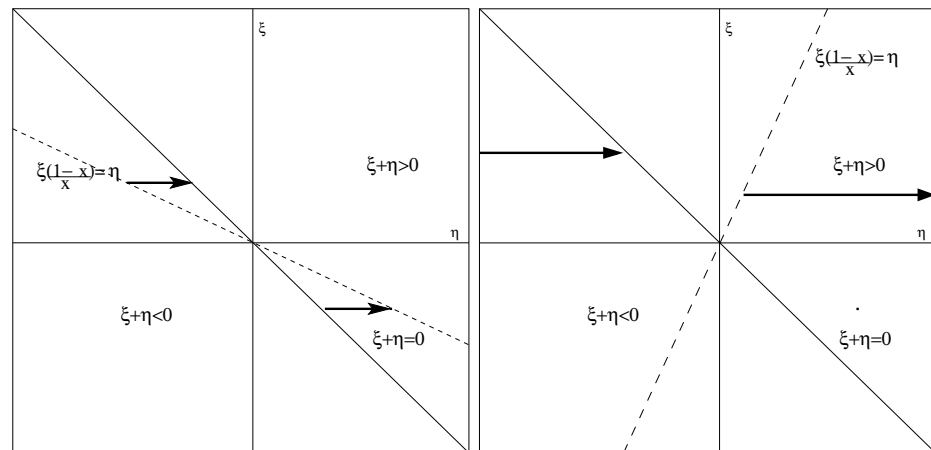


Figure A1. The integration regions of the plane (η, ξ) for $x \leq 0$ (left) and for $x > 0$ (right). The integration regions are indicated by thick arrows along the η integrations. The ξ integrations are not indicated; they are orthogonal to the thick arrows to cover the sector of the plane with the arrows.

For $x \leq 0$, we obtain:

$$F_2^-(x) = \int_{-\infty}^0 d\xi P_1(\xi) \int_{-\xi}^{\xi(1-x)/x} P_2(\eta) d\eta + \int_0^{\infty} d\xi P_1(\xi) \int_{\xi(1-x)/x}^{-\xi} P_2(\eta) d\eta \quad (\text{A1})$$

and, for $x > 0$, $F_2^+(x)$ is

$$F_2^+(x) = \int_{-\infty}^0 d\xi P_1(\xi) \int_{-\xi}^{+\infty} P_2(\eta) d\eta + \int_0^{+\infty} d\xi P_1(\xi) \int_{\xi(1-x)/x}^{+\infty} P_2(\eta) d\eta \\ + \int_{-\infty}^0 d\xi P_1(\xi) \int_{-\infty}^{\xi(1-x)/x} P_2(\eta) d\eta + \int_0^{+\infty} d\xi P_1(\xi) \int_{-\infty}^{-\xi} P_2(\eta) d\eta. \quad (\text{A2})$$

It is easy to prove that $F_2^-(x) = 0$ for $x \rightarrow -\infty$ and $F_2^+(x) = 1$ for $x \rightarrow +\infty$.

The PDF $P_{xg2R}(x)$ is given by a differentiation of $F_2^-(x)$ and $F_2^+(x)$ with respect to x , obtaining

$$P_{xg2R}(x) = \frac{dF_2^+(x)}{dx} = \frac{1}{x^2} \int_{-\infty}^{+\infty} d\xi |\xi| P_1(\xi) P_2\left(\xi \frac{1-x}{x}\right), \quad (\text{A3})$$

an identical result is obtained differentiating $F_2^-(x)$. The cumulative distribution for the random variable, $-\beta/(\beta + \eta)$ could be obtained similarly.

The construction of the cumulative distribution for the complete COG₂ algorithm of Equation (14) implies the insertion of another random variable β . The integration regions are defined in the space (ξ, η, β) . The cumulative distribution is expressed by a large number of integrals on sectors of the (ξ, η, β) -space. The differentiation and the collection of the various terms reproduce Equation (16). Further details and other PDFs are reported in Ref. [19] (and references therein.)

Appendix B. An Advanced Approximation of $P_{xg2}(x)$ Beyond Equation (20)

For very inclined tracks, the MIP signal is spread among various strips, and the histograms of COG₂ algorithm show very large gaps around zero. In this case, the approximations, described above, show perceptible deviations from the simulated data and the numerical integrations of Equation (17). In this case, a better approximation is useful.

The following approximation shows negligible differences from the numerical integrations. For its construction, the Fubini theorem is applied to invert the order of the double integrals of Equation (16), and variable transformations are selected to have a zero as the lowest limit of an internal integration region. In this way, the two integrations become independent and can be executed in any order. The neglecting of the change of

sign introduced by the absolute values of Equation (17) allows one to obtain the following analytical results:

$$\begin{aligned}
 P_{xg2}(x) = & \frac{1}{2\sqrt{2\pi}} \frac{a_2(1-x)\sigma_1^2 + a_1x\sigma_2^2}{[(1-x)^2\sigma_1^2 + x^2\sigma_2^2]^{3/2}} \exp \left[- \left(x - \frac{a_1}{a_1 + a_2} \right)^2 \frac{(a_1 + a_2)^2}{2(\sigma_1^2(1-x)^2 + x^2\sigma_2^2)} \right] \\
 & \left\{ 1 - \operatorname{erf} \left[\frac{(1-x)[a_3(1-x) - a_2x]\sigma_1^2 - (a_1 - a_3)x^2\sigma_2^2}{\sqrt{2((1-x)^2\sigma_1^2 + x^2\sigma_2^2)(x^2\sigma_1^2\sigma_2^2 + (1-x)^2\sigma_1^2\sigma_3^2 + x^2\sigma_2^2\sigma_3^2)}} \right] \right\} + \\
 & + x \rightarrow -x, \quad a_1 \leftrightarrow a_3 \quad \sigma_1 \leftrightarrow \sigma_3
 \end{aligned} \tag{A4}$$

The approximation does not reproduce the absolute value (actually, $A \operatorname{erf}(A)$) of the previous Equation (20). However, for realistic values of the parameters $\{a_j\}$, it is irrelevant. In any case, it is a trivial completion, as in (20). Here, the argument of the erf-function is more complete than that given for the small $|x|$ of Equation (20) (it is easy to reduce this erf-function to that for small $|x|$). The difference of Equation (A4) with a numerical integration is negligible in many significant cases.

To complete the approximation, we report the Cauchy-like terms (even if they are of scarce relevance). These terms are exactly given by MATHEMATICA being the first terms of an integration by part of Equation (16). The Cauchy tails are evident, and the factor x in the numerator compensates the $\sqrt{x^2}$ in the denominator. The expression of $P_{xg2}^{\text{Cauchy}}(x)$ is:

$$\begin{aligned}
 P_{xg2}^{\text{Cauchy}}(x) = & \exp \left[- \frac{(a_3(1-x) - a_2x)^2\sigma_1^2 + (a_1(1-x) - a_2x)^2\sigma_3^2 + (a_1 - a_3)^2x^2\sigma_2^2}{2(x^2\sigma_2^2\sigma_3^2 + (1-x)^2\sigma_1^2\sigma_3^2 + x^2\sigma_2^2\sigma_1^2)} \right] \\
 & \times \frac{x \sigma_1^2 \sigma_2^2}{2\pi((1-x)^2\sigma_1^2 + x^2\sigma_2^2) \sqrt{(1-x)^2\sigma_3^2\sigma_1^2 + x^2\sigma_2^2(\sigma_1^2 + \sigma_3^2)}} \\
 & \times \operatorname{erf} \left[\frac{(1-x)a_2\sigma_1^2\sigma_3^2 + \sigma_2^2(\sigma_1^2a_3 + \sigma_3^2a_1)x}{\sqrt{2}\sigma_1\sigma_2\sigma_3 \sqrt{\sigma_2^2\sigma_3^2x^2 + \sigma_1^2(\sigma_3^2(1-x)^2 + \sigma_2^2x^2)}} \right] \\
 & + \exp \left[- \frac{a_1^2}{2\sigma_1^2} - \frac{a_2^2}{2\sigma_2^2} \right] \frac{[1 - \operatorname{erf}(a_3/\sqrt{2}\sigma_3)] \sigma_1\sigma_2}{2\pi[x^2\sigma_2^2 + (1-x)^2\sigma_1^2]} \\
 & + (x \rightarrow -x, \quad a_1 \leftrightarrow a_3 \quad \sigma_1 \leftrightarrow \sigma_3).
 \end{aligned} \tag{A5}$$

Actually, these terms are a very small fraction of the main terms (around 10^{-5}) but become of the order of 10^{-1} for very inclined tracks. The exponential term has maxima around $x = a_3/(a_3 + a_2)$ (due to the term with $(a_1(1-x) + xa_2)$) and $x = a_1/(a_1 + a_3)$ (due to the term with $(a_3(1-x) + xa_2)$); these two maxima are very close one to another, with large overlaps.

References

1. Landi, G.; Landi, G.E. Beyond the $\sqrt{N}$ -limit of the least squares resolution and the lucky model. *Instruments* **2022**, *6*, 10. [\[CrossRef\]](#)
2. Landi, G.; Landi, G.E. Improvement of track reconstruction with well tuned probability distributions. *J. Instrum.* **2014**, *9*, P10006. [\[CrossRef\]](#)
3. Landi, G.; Landi, G.E. Optimizing momentum resolution with a new fitting method for silicon-strip detectors. *Instruments* **2018**, *2*, 22. [\[CrossRef\]](#)
4. Devore, J.L.; Berk, K.N. *Modern Mathematical Statistics with Applications*; Springer: New York, NY, USA, 2018. [\[CrossRef\]](#)
5. Landi, G.; Landi, G.E. Generalized inequalities to optimize the fitting method for track reconstruction. *Physics* **2020**, *2*, 608–623. [\[CrossRef\]](#)

6. Landi, G. Problems of position reconstruction in silicon microstrip detectors. *Nucl. Instr. Meth. Phys. Res. A* **2005**, *554*, 226–246. [[CrossRef](#)]
7. Landi, G.; Landi, G.E. The Cramer-Rao inequality to improve the resolution of the standard least-squares method in track fitting. *Instruments* **2020**, *4*, 2. [[CrossRef](#)]
8. Landi, G. The center of gravity as an algorithm for position measurements. *Nucl. Instr. Meth. Phys. Res. A* **2002**, *485*, 698–719. [[CrossRef](#)]
9. MATHEMATICA 6; Wolfram Inc.: Champaign, IL, USA, 2010. Available online: <https://www.wolfram.com/> (accessed on 1 July 2022).
10. MATLAB; The MathWork Inc.: Natick, MA, USA, 2020. Available online: <https://www.mathworks.com/> (accessed on 1 July 2022).
11. Hartmann, F. Silicon tracking detectors in high-energy physics. *Nucl. Instrum. Meth. Phys. Res. A* **2012**, *666*, 25–46. [[CrossRef](#)]
12. The CMS Collaboration. The performance of the muon detector in proton-proton collision at $\sqrt{s} = 7$ TeV at LHC. *J. Instrum.* **2013**, *8*, P11002. [[CrossRef](#)]
13. Gnedenko, B.V. *The Theory of Probability and Elements of Statistics*; AMS Chelsea Publishing: New York, NY, USA, 1989.
14. Gauss, C.F. *Méthode des Moindres Carrés. Mémoires sur la Combination des Observations*; Mallet-Bachelier: Paris, France, 1855. Available online: https://books.google.it/books?id=_qzpB3QqQkQC (accessed on 1 September 2018).
15. The CMS Collaboration. Description and Performance of track and primary vertex reconstruction with the CMS tracker. *J. Instrum.* **2014**, *9*, P10009. [[CrossRef](#)]
16. Samedov, V.V. Inaccuracy of coordinate determined by several detectors' signals. *J. Instrum.* **2012**, *7*, C06002. [[CrossRef](#)]
17. Landi, G.; Landi, G.E. Silicon microstrip detectors. *Encyclopedia* **2021**, *1*, 1076–1083. [[CrossRef](#)]
18. Belau, E.; Klanner, R.; Lutz, G.; Neugebauer, E.; Seebrunner, H.J.; Wylie, A. Charge collection in silicon strip detector. *Nucl. Instrum. Methods Phys. Res. A* **1983**, *214*, 253–260. [[CrossRef](#)]
19. Landi, G.; Landi, G.E. Positioning error probabilities for some forms of center-of-gravity algorithm calculated with the cumulative distributions. Part II. *arXiv* **2021**, arXiv:2103.03464. [[CrossRef](#)]
20. Adriani, O.; Bongi, M.; Bonechi, L.; Bottai, S.; Castellini, G.; Fedele, D.; Grandi, M.; Landi, G.; Papini, P.; Ricciarini, S.; et al. In-flight performance of the PAMELA magnetic spectrometer. *PoS* **2008**, *Vertex 2008*, 048. [[CrossRef](#)]
21. The CMS Collaboration. Performance of the reconstruction and identification of high-momentum muons in proton proton collision at $\sqrt{s} = 13$ TeV. *J. Instrum.* **2020**, *15*, P02027. [[CrossRef](#)]