PHYSICAL REVIEW D VOLUME 48, NUMBER 5 1 SEPTEMBER 1993

Integral correlation measures for multiparticle physics

H. C. Eggers*
Institut für Theoretische Physik, Universität Regensburg, D-93040 Regensburg, Germany

P. Lipa† and P. Carruthers
Department of Physics, University of Arizona, Tucson, Arizona 85721

B. Buschbeck
Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfergasse 18, A-1050 Vienna, Austria
(Received 29 March 1993)

We report on a considerable improvement in the technique of measuring multiparticle correlations via integrals over correlation functions. A modification of measures used in the characterization of chaotic dynamical systems permits fast and flexible calculation of factorial moments and cumulants as well as their differential versions. Higher-order correlation integral measurements even of large multiplicity events such as encountered in heavy ion collisons are now feasible. The change from "ordinary" to "factorial" powers may have important consequences in other fields such as the study of galaxy correlations and Bose-Einstein interferometry.

PACS number(s): 24.60.Ky, 13.85.Hd, 25.75.+r, 98.62.Py

I. INTRODUCTION

Wherever statistical analyses are done, whether in physics, biology, or psychology, the measurement of the correlation function is a basic element of understanding. While each discipline has its own set of questions for which it seeks answers, the underlying statistical mechanisms are very similar: given a set of variables, one first finds the distribution of how often each of these variables takes on a certain value, and much that is useful can be learned from these one-variable distributions. Following immediately is the question how two variables behave simultaneously, whether they are independent or in some way correlated: the two-variable correlation function provides the answers. Higher orders provide additional information, but with escalating difficulty of measurement and diminishing returns.

Conversely, a knowledge of correlation functions to all orders provides complete information on any statistical system.

Of special interest to us here are point distributions. Typical examples of point distributions include galaxies in the sky, cows in a field, and particles in phase space (the exact size of the object under study is irrelevant as long as it is small compared to the embedding space). The aim of this paper is to develop and extend methods of measuring correlation functions of point distributions. While we shall be considering particle correlations in high-energy collisions, the formalism developed here should be suitable, with appropriate modifications, for problems in a number of other situations.

Beyond traditional methods, recent advances in the theory of fractals and scaling have spawned a new way of approaching correlations: scaling behavior manifests itself in power-law behavior of the correlation function, which in particle physics can be measured experimentally through the factorial moments revived by Białas and Peschanski [1]. On the other hand, a measure termed the correlation integral [2–4] has been in use in the characterization of strange attractors and other contexts for some time. As an improvement on the factorial moments and following the suggestion of Dremin [5], we previously introduced into multiparticle physics two forms of the correlation integral which we termed the "snake" and the symmetrized "Grassberger-Hentschel-Procaccia" (GHP) integrals [6]. In this paper, we advocate a slightly different form, which for obvious reasons we name "star" integral. Yet another form useful for pion interferometry takes the invariant mass of the q-tuple as a measure of its size [7,8].

To illustrate the concept of the correlation integral and the difference between the various forms, we consider the phase-space plot (e.g., in rapidity and azimuthal angle) of the pions produced in a particular collision and ask the question how clusters consisting of q particles may be characterized, i.e., what "size" should be assigned to each cluster. Taking $q = 5$, for example, we show in Fig. 1 one particular selection of five particles from this event and assign the following sizes to it: in (a), the five particles are joined into a "snake," and the 5-tuple is assigned a size ϵ corresponding to the longest of the four joining lines. In (b), the same 5-tuple is assigned a size given by the maximum of all ten pairwise distances; this defines the GHP correlation integral. The "star" integral in Fig. 1(c), finally, is assigned a size corresponding to the longest of the four lines to the chosen center particle. A line represents a particular interparticle distance being *tested* and found to be smaller than ϵ; those particles *not*

*Present address: Department of Physics, McGill University, Montréal, Québec, Canada H3A 3T8.

†Present address: Institut für Hochenergiephysik, Nikolsdorfergasse 18, A-1050 Vienna, Austria.

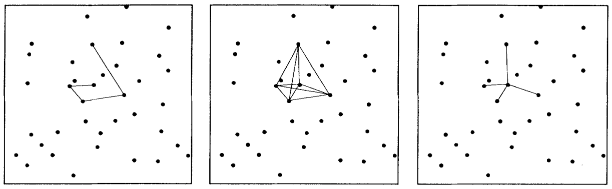

FIG. 1. Different version of the correlation integral: (a) snake, (b) GHP, and (c) star. For a given set of q particles in phase space (here $q = 5$) taken from the N particles of a particular event, pairwise distances are tested according to the topology of joining lines shown. The longest of the lines characterizes the size of the given q-tuple in every prescription. The star count is much more efficient than the other prescriptions, see Sec. II.

connected by a line may or may not be within a distance ϵ of each other.

Every correlation integral of order q thus assigns a size to every possible q-tuple of particles in an event. The way this assignment is done distinguishes the different versions of correlation integrals. Once such an assignment is made, they all count the number of q-tuples that are smaller than a given size. For a large data sample this corresponds to an integration over the qth-order correlation function, as shown explicitly in Eqs. (4)–(7) below.

We shall show that the star shares all the advantages of the snake and GHP forms but is more amenable to intuitive understanding and is computationally more efficient by orders of magnitude. Furthermore, our formalism leads naturally to a conceptual cleanup of the heuristic fractal measure used in the study of strange attractors, in galaxy distributions, and in multiparticle correlations [3–6,9]. Being derived directly and rigorously from the underlying correlation function, it necessitates the use of *factorial* powers rather than the ordinary powers used previously. The difference becomes significant when small particle numbers are involved, a situation which occurs inevitably when the integration domain becomes sufficiently small.

In addition to this cleanup, we provide here a technique to measure integrals of cumulant correlation functions which are the genuine higher-order correlations and thus very useful observables. Previously, the only way to extract information on higher-order cumulants had been via combinations of factorial moments [10] and in third order for some very special configurations in rapidity space [11]. By clarifying and extending the technique of event mixing, we show how cumulant functions can be integrated directly and thus share all the advantages of correlation integrals over bin-averaged factorial moments.

The paper is organized as follows. In Sec. II, the star correlation integral is introduced, while in Sec. III we lay a solid theoretical foundation for the practice of event mixing. In Sec. IV, we remind the reader of the basic relations defining true correlations and show how these may be measured directly in the correlation integral scheme. Differential versions and their considerable advantages are discussed in Sec. V. We conclude with some remarks and recommendations.

II. THE CORRELATION INTEGRAL

A. Basic concepts

The starting point for all correlation analyses is the qth-order correlation function or "inclusive density" $\rho_q(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_q)$. The choice of the q variables $\boldsymbol{x}_k$ living in d dimensions is determined by our particular problem; in high-energy physics, we can have, for example, vector three-momenta $\boldsymbol{x} \equiv (p_x, p_y, p_z)$, some combination of boost-invariant variables such as rapidity, azimuthal angle, and transverse momentum $(y, \phi, p_\perp)$, or just one of these alone.[1]

The set of correlation functions ρ_q, $q = 1, 2, \ldots$ can be defined for a fixed total number of particles and/or for specific particle types such as positively and negatively charged pions; for the purposes of this paper, we shall mostly consider only one type of particle within an inclusive distribution where the total number of particles per event is not held constant. For this case, ρ_q is defined operationally as

$$\rho_q(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q) = \frac{1}{\sigma_I} \frac{d^q \sigma_{\text{incl}}}{d\boldsymbol{x}_1\, d\boldsymbol{x}_2 \cdots d\boldsymbol{x}_q} \,, \tag{1}$$

where σ_I is the total interaction cross section and σ_{incl} the inclusive cross section. Integrating in d dimensions over the total window $\Delta\boldsymbol{x} = (\Delta x)^d$, we get

$$\int_{\Delta\boldsymbol{x}} d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_q\, \rho_q(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q) = \left\langle N^{[q]} \right\rangle \,, \tag{2}$$

where N is the total number of particles in $\Delta\boldsymbol{x}$ and $\left\langle N^{[q]} \right\rangle = \langle N(N-1) \cdots (N-q+1) \rangle = \xi_q(\Delta x)$ is the unnormalized qth-order factorial moment over the same region.

Integrating ρ_q over various domains of integration, one can obtain any number of possible moments. For example, the vertical normalized factorial moment revived by Białas and Peschanski [1],

$$\begin{aligned} F_q^v &\equiv \frac{1}{M^d} \sum_{m_1,\ldots,m_d=1}^{M} \frac{\left\langle n^{[q]}_{m_1,\ldots,m_d} \right\rangle}{\langle n_{m_1,\ldots,m_d} \rangle^q} \\ &= \frac{1}{M^d} \sum_{m_1,\ldots,m_d=1}^{M} \frac{\int_{\Omega(m)} \prod_k d\boldsymbol{x}_k\, \rho_q(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)}{\int_{\Omega(m)} \prod_k d\boldsymbol{x}_k\, \rho_1(\boldsymbol{x}_1) \cdots \rho_1(\boldsymbol{x}_q)} \,, \end{aligned} \tag{3}$$

integrates ρ_q in a Cartesian lattice of M^d cubes, each of size $\Omega(m) = \Omega(m_1, \ldots, m_d) = (\delta x)^d$, normalizing each bin separately. For the purposes of searching for power-law behavior of the correlation function, F_q is measured as a function of M, with the bin edge length decreasing correspondingly, $\delta x = \Delta x / M$. The "horizontal" facto-

[1]In addition to a suitable choice of variables, some "preprocessing" of the data may be required to eliminate unwanted effects, for example the transformation to jet coordinates in e^+e^- collisions [12], and, if desired, the creation of data subsamples of fixed total multiplicity.

rial moments [1], the differential versions of Sec. V, and indeed the traditional way of presenting correlation functions in constant bin sizes are all integrations over different domains of the same correlation function.

In the same way, the three correlation integrals are simply integrals over specific domains. The qth-order snake correlation integral is defined in terms of ρ_q and the "cluster size" ϵ as

$$F_q^{\text{snake}}(\epsilon) \equiv \frac{\int \rho_q(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q)\,\Theta_{12}\Theta_{23}\cdots\Theta_{q-1,q}\,d\boldsymbol{x}_1\cdots d\boldsymbol{x}_q}{\int \rho_1(\boldsymbol{x}_1)\cdots\rho_1(\boldsymbol{x}_q)\,\Theta_{12}\Theta_{23}\cdots\Theta_{q-1,q}\,d\boldsymbol{x}_1\cdots d\boldsymbol{x}_q}, \tag{4}$$

where $\Theta_{ij} \equiv \Theta(\epsilon - |\boldsymbol{x}_i - \boldsymbol{x}_j|)$. Similarly, the GHP moment is defined with all interparticle distances restricted:[2]

$$F_q^{\text{GHP}}(\epsilon) \equiv \frac{\int \rho_q(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q)\,\prod_{i<j}\Theta_{ij}\,d\boldsymbol{x}_1\cdots d\boldsymbol{x}_q}{\int \rho_1(\boldsymbol{x}_1)\cdots\rho_1(\boldsymbol{x}_q)\,\prod_{i<j}\Theta_{ij}\,d\boldsymbol{x}_1\cdots d\boldsymbol{x}_q}, \tag{5}$$

where

$$\prod_{i<j=1}^{q} \Theta_{ij} = \Theta_{12}\Theta_{13}\cdots\Theta_{1q}\Theta_{23}\cdots\Theta_{q-1,q} \tag{6}$$

restricts all possible pairs of coordinates. The star integral is defined as

$$F_q^{\text{star}}(\epsilon) \equiv \frac{\int \rho_q(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q)\,\Theta_{12}\Theta_{13}\cdots\Theta_{1q}\,d\boldsymbol{x}_1\cdots d\boldsymbol{x}_q}{\int \rho_1(\boldsymbol{x}_1)\cdots\rho_1(\boldsymbol{x}_q)\,\Theta_{12}\Theta_{13}\cdots\Theta_{1q}\,d\boldsymbol{x}_1\cdots d\boldsymbol{x}_q}, \tag{7}$$

involving all coordinates paired with $\boldsymbol{x}_1$. The "topology" of the different correlation integrals shown in Fig. 1 is thus visible already in the selection of Θ functions.

B. The star integral

While the definition (1) of ρ_q is exact and unambiguous, we shall need for the derivation of the correlation integral an alternative but equivalent formulation written down by, e.g., Klimontovich [13]. Let the N particles of a specific event be situated at the points $\boldsymbol{X}_i$ in phase space, $i = 1,\dots,N$. Then the "event correlation function" $\hat{\rho}_q$ is defined as adding 1 at every point $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q)$ if there is simultaneously at each $\boldsymbol{x}_k$ a particle $\boldsymbol{X}$, independently of the positions of the other $N-q$ particles. This is done for all $N^{[q]} = N!/(N-q)!$ selections of q-tuples out of the total N particles. For a specific event a this defines a function

$$\hat{\rho}_q^a(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q) = \sum_{\substack{i_1\neq i_2\neq\dots\neq i_q\\=1}}^{N} \delta(\boldsymbol{x}_1 - \boldsymbol{X}_{i_1}^a)\,\delta(\boldsymbol{x}_2 - \boldsymbol{X}_{i_2}^a)\cdots\delta(\boldsymbol{x}_q - \boldsymbol{X}_{i_q}^a)\,, \tag{8}$$

where $\delta(\boldsymbol{x})$ is the product of d one-dimensional δ functions. This function, when averaged over all events, yields the q-particle distribution function of Eq. (1),

$$\rho_q(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q) = \left\langle \hat{\rho}_q^a(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q)\right\rangle = N_{\text{ev}}^{-1}\sum_{a=1}^{N_{\text{ev}}} \hat{\rho}_q^a(\boldsymbol{x}_1,\dots,\boldsymbol{x}_q)\,, \tag{9}$$

where N_{ev} is the number of events in the experimental sample. For finite resolution, Eq. (9) corresponds to building up a qd-dimensional histogram.[3]

Inserting Eqs. (8) and (9) into the numerator of Eq. (7), we find immediately the (unnormalized) star integral factorial moment to be

[2] In Ref. [6], we erroneously denied that the GHP integral could be written down as an analytical integral of the correlation function.

[3] Since both definitions (1) and (9) are implemented for a finite number of events, they are, strictly speaking, estimators of the true correlations.

$$\xi_q^{\rm star}(\epsilon) = \left\langle \sum_{i_1 \neq i_2 \neq \ldots \neq i_q}^{N} \Theta(\epsilon - X_{i_1 i_2})\Theta(\epsilon - X_{i_1 i_3}) \cdots \Theta(\epsilon - X_{i_1 i_q}) \right\rangle , \tag{10}$$

where $X_{i_1 i_k} = |\boldsymbol{X}_{i_1} - \boldsymbol{X}_{i_k}|$. Pulling out the first sum, we can factorize the remaining sums:

$$\xi_q^{\rm star}(\epsilon) = \left\langle \sum_{i_1} \left(\sum_{i_2 \neq i_1} \Theta(\epsilon - X_{i_1 i_2}) \right)^{[q-1]} \right\rangle , \tag{11}$$

where the factorial power $[q-1]$ in the exponent came about because the sum indices are restricted to $i_\alpha \neq i_\beta \neq i_1$ for all $\alpha \neq \beta$. The quantity inside the parentheses is so important that we give it a special name, the *sphere count*:

$$\hat{n}(\boldsymbol{X}_{i_1}, \epsilon) \equiv \sum_{i_2=1}^{N} \Theta(\epsilon - X_{i_1 i_2}) , \qquad i_2 \neq i_1 , \tag{12}$$

which represents the number of particles within a sphere of radius ϵ centered on the particle $\boldsymbol{X}_{i_1}$, excluding the center particle itself. [For the given ϵ and center particle shown in Fig. 2(a), we would have $\hat{n}(\boldsymbol{X}_{i_1}, \epsilon) = 9$.] In a similar form, it is used extensively in the description of galaxy distributions. Introducing the shorthand notation $\hat{n}(\boldsymbol{X}_{i_1}^a, \epsilon) \equiv a$ the unnormalized factorial moment can be written compactly as (we henceforth drop the "star" superscript)

$$\xi_q(\epsilon) = N_{\rm ev}^{-1} \sum_a \sum_{i_1} \hat{n}(\boldsymbol{X}_{i_1}^a, \epsilon)^{[q-1]} \equiv \left\langle \sum_{i_1} a^{[q-1]} \right\rangle . \tag{13}$$

An alternative derivation of $\xi_q(\epsilon)$ proceeds by coordinate transformation [6,14]. One first defines a "particle-centered" correlation function around the particle at $\boldsymbol{X}_{i_1}$, fixing to it the coordinate $\boldsymbol{x}_1$ by a δ function:

$$\hat{\rho}_q(\boldsymbol{X}_{i_1}; \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_q) = \delta(\boldsymbol{x}_1 - \boldsymbol{X}_{i_1}) \sum_{i_2 \neq i_3 \neq \ldots \neq i_q}^{N} \delta(\boldsymbol{x}_2 - \boldsymbol{X}_{i_2})\delta(\boldsymbol{x}_3 - \boldsymbol{X}_{i_3}) \cdots \delta(\boldsymbol{x}_q - \boldsymbol{X}_{i_q}) , \tag{14}$$

where the sum indices are all restricted additionally by $i_\alpha \neq i_1$, and then transforms to relative coordinates. These are the distinctive hallmark of correlation integrals: for the snake integral, we used the coordinate transformation $R = \sum_{k=1}^{q} x_k/q$ and $r_k = x_{k+1} - x_k$ [6]. For the star integral, on the other hand, all coordinates are defined relative to $\boldsymbol{x}_1$:

$$\boldsymbol{r}_1 = \boldsymbol{R} = \boldsymbol{x}_1 , \quad \boldsymbol{r}_k = \boldsymbol{x}_k - \boldsymbol{x}_1 , \quad k = 2, \ldots, q . \tag{15}$$

Inserting these into the δ functions of Eq. (14), we find

$$\hat{\rho}_q(\boldsymbol{X}_{i_1}; \boldsymbol{R}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_q) = \delta[\boldsymbol{R} - \boldsymbol{X}_{i_1}] \sum_{i_2 \neq i_3 \neq \ldots \neq i_q} \delta[\boldsymbol{r}_2 - (\boldsymbol{X}_{i_2} - \boldsymbol{X}_{i_1})] \cdots \delta[\boldsymbol{r}_q - (\boldsymbol{X}_{i_q} - \boldsymbol{X}_{i_1})] . \tag{16}$$

Since at this point we are only interested in correlations as a function of relative distances, we integrate out $\boldsymbol{R}$ over the entire interval $\Delta \boldsymbol{R} = \Delta \boldsymbol{x}$. The relative coordinates $\boldsymbol{r}_k$ we want to restrict to a maximum length $r_k = |\boldsymbol{r}_k| \leq \epsilon$. For higher dimensions $d > 1$, we must first integrate out the angular parts $d\Omega_k$ of $d\boldsymbol{r}_k$. Since we shall eventually normalize our correlations using exactly the same domain of integration, however, the constants resulting from the angular integrations will cancel and we henceforth ignore them. The remaining integral over the lengths r_k is given by $\int_0^\epsilon \prod_k dr_k\, r_k^{d-1}$. At the same time, on integrating out the angular coordinates, the remaining δ functions of Eq. (16) become

$$\int d\Omega_k \, \delta[\boldsymbol{r}_k - \boldsymbol{X}_{i_1 i_k}] = \frac{\delta[r_k - X_{i_1 i_k}]}{X_{i_1 i_k}^{d-1}} , \tag{17}$$

and the factors r_k^{d-1} will on integration cancel exactly with $X_{i_1 i_k}^{d-1}$. To express this entire process of simplification, we shall write, in shorthand,

$$\int_0^\epsilon \prod_k d\boldsymbol{r}_k , \tag{18}$$

which is just an integral of the lengths r_k. For $d = 1$, $\int_0^\epsilon \prod_k d\boldsymbol{r}_k$ is shorthand for $\int_{-\epsilon}^\epsilon \prod_k dr_k$.

Integrating in this way over all relative coordinates, we again get a factorial product of single sums:

$$\int_{\Delta \boldsymbol{x}} d\boldsymbol{R} \int_0^\epsilon \prod_{k=2}^{q} d\boldsymbol{r}_k \, \hat\rho_q(\boldsymbol{X}_{i_1}; \boldsymbol{R}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_q) = \sum_{i_2 \neq \ldots \neq i_q} \Theta(\epsilon - X_{i_1 i_2})\Theta(\epsilon - X_{i_1 i_3}) \cdots \Theta(\epsilon - X_{i_1 i_q})$$

$$= \left(\sum_{i_2} \Theta(\epsilon - X_{i_1 i_2}) \right)^{[q-1]} , \tag{19}$$

which, on summing over i_1 and averaging over all events, again yields Eq. (11).

As derived above, the $\xi_q(\epsilon)$ of Eqs. (7) and (11)–(13) is the unnormalized factorial moment over the domain shown in Fig. 2(a):

$$\Omega_S(\epsilon) \equiv \{\boldsymbol{R}, \boldsymbol{r}_k | \boldsymbol{R} \in \Delta\boldsymbol{x}, r_k \in [0,\epsilon], k = 2, \ldots, q\} ; \tag{20}$$

this "floating sphere" form is best whenever the nature of our variables $\boldsymbol{x}$ in d dimensions is such that a length can be sensibly defined, the most obvious being the Euclidean distance

$$r_k \equiv \sqrt{r_{k,x}^2 + r_{k,y}^2 + \cdots} , \tag{21}$$

but often this is not so good. The variables $(y, \phi, p_\perp)$, for example, have very different behavior, and it may be better to treat each one separately. For such cases, one may use the "floating box" form [15], where each of the d components is treated as a one-dimensional correlation integral and the total domain as shown in Fig. 2(b) is

$$\Omega_B(\epsilon) \equiv \{R_f, r_{k,f} | R_f \in \Delta x, r_{k,f} \in [-\epsilon, \epsilon] \; \forall k = 2, \ldots, q; f = y, \phi, \ldots\} . \tag{22}$$

This corresponds to inserting $d(q-1)$ Θ functions into Eq. (7), one for each component.

The results of this section have the following important consequences.

(1) As shown in Eqs. (19)–(11), the $q-1$ sums factorize nicely into a single sum of sphere counts $\hat n(\boldsymbol{X}_{i_1}, \epsilon)$. This means that $\xi_q(\epsilon)$ can be calculated in an algorithm of order N^2, independently of the order q. As emphasized previously [3], this represents a tremendous savings in CPU time over other correlation measures, including the snake and GHP integrals advocated by us previously [6], which run under N^q and $N^q/q!$ algorithms respectively.

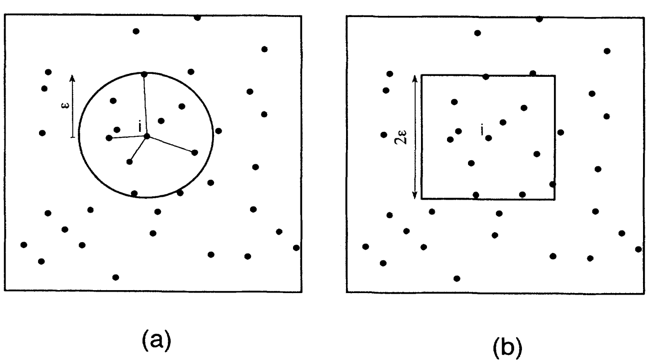

FIG. 2. The conceptual advantage of the star integral over other versions stems from the fact that counting q-stars can be reduced to computing factorials of *sphere counts* $\hat n(\boldsymbol{X}_i, \epsilon)$. (a) "Floating sphere": $\hat n(\boldsymbol{X}_i, \epsilon)$ = number of neighboring particles within a sphere of radius ϵ centered at particle i with coordinates $\boldsymbol{X}_i$. The center particle itself is not included in the sphere count. (b) When the coordinates $\boldsymbol{x}$ have very different physical properties in their components (such as y, ϕ and $p_\perp$), the "floating box" may be a better choice as it treats the distances along different coordinate axes independently.

(2) This means that the correlation integral can now be used also for correlation analysis in heavy ion collisions, something hitherto impossible due to the large event multiplicities. The big improvements in statistics over the conventional vertical factorial moments will allow for much more accurate measurements.

(3) Unlike the star integral found so far in the literature [3–5,16], we get a *factorial* power of the sphere count $\hat n^{[q-1]}$ rather than the ordinary power $\hat n^{q-1}$. This result we obtained rigorously from first principles merely by restricting the sum indices to be unequal because the same particle may not be counted more than once. Even for the large multiplicities encountered in astronomy, the change is not inconsequential, as the difference between $\hat n^{[q-1]}$ and $\hat n^{q-1}$ is important when ϵ becomes sufficiently small.

(4) We obtained this factorial power without drawing on distinctions between "dynamical" and "statistical" fluctuations [1].

(5) To first order, we have ignored the variation of $\xi_q(\epsilon)$ with the center coordinate $\boldsymbol{R}$; this is equivalent to assuming that the physics is the same for all parts of the defined window. When statistics permit, it may be very useful to measure ξ_q also as a function of $\boldsymbol{R}$. For example, one expects the correlation structure at small transverse momentum to be very different from that at large $p_\perp$, so that a separate measurement of ξ_q for small and large $R \equiv p_{\perp 1}$ may be most revealing.

(6) The advantages of correlation integrals over the traditional factorial moments arise because the former use interparticle distances directly while the latter rely on fixed bins and grids rather than the particle positions

themselves.

(7) We must emphasize that the star integral as derived here is different from the snake and GHP versions we used in earlier papers. All three are correlation integrals, but they differ in the topology of the interparticle distances measured. For $q = 2$, all three are the same, while the snake and star versions are the same even at the level $q = 3$. Only in fourth order do the differences between the latter two appear. While the snake and GHP versions are not wrong and an improvement over previous work, the present star integral represents a big step forward.

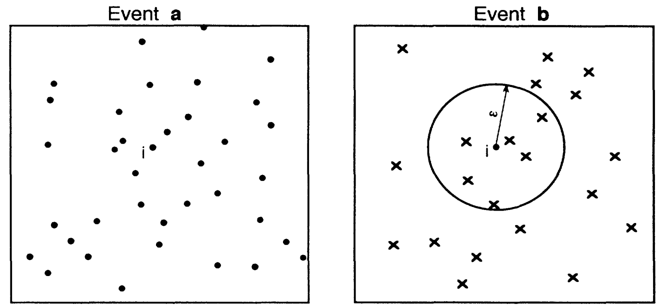

FIG. 3. The basic building blocks for computing the normalization of correlation integrals [Eqs. (23)–(26)] as well as cumulants [Eqs. (43)ff] is the sphere count $\hat{n}_b(\boldsymbol{X}_i^a, \epsilon)$. While similar to the count of Fig. 2(a), $\hat{n}_b(\boldsymbol{X}_i^a, \epsilon)$ performs a sphere count around particle $\boldsymbol{X}_i^a$ taken from event a (shown as a dot) by placing it in the event b and counting the b-particles (shown as crosses) within the sphere. In the example shown, $\hat{n}_b(\boldsymbol{X}_i^a, \epsilon) = 6$. Averages over many events b are taken while $\boldsymbol{X}_i^a$ is kept fixed.

III. NORMALIZATION BY EVENT MIXING

We now proceed to consider the denominator of Eq. (7). Since the topic of normalization is complex and full of pitfalls, we do not address the full range of issues here and defer such discussion to future work. Instead, we concentrate on the normalization scheme we consider most suitable for the measurement of correlations in high-energy physics, a version resembling the so-called vertical normalization used in Eq. (3).

The normalization of the correlation integral is done by means of *event mixing*, a seemingly heuristic technique commonly used in Bose-Einstein correlations [12]. It is well founded, however, both for our purposes here and in the Bose-Einstein context [8]. We recall that if all q coordinates $\boldsymbol{x}_k$ are statistically independent of each other, the correlation function factorizes: $\rho_q = \rho_1^q$. We hence normalize the numerator (11) of the star integral by integrating $\prod_{k=1}^q \rho_1(\boldsymbol{x}_k = \boldsymbol{R}+\boldsymbol{r}_k)$ over the same domain:

$$\xi_q^{\rm norm}(\epsilon) \equiv \int_{\Delta \boldsymbol{x}} d\boldsymbol{R}\,\rho_1(\boldsymbol{R}) \int_0^\epsilon \prod_{k=2}^{q} d\boldsymbol{r}_k\,\rho_1(\boldsymbol{R}+\boldsymbol{r}_2)\cdots\rho_1(\boldsymbol{R}+\boldsymbol{r}_q) = \int_{\Delta \boldsymbol{x}} d\boldsymbol{R}\,\rho_1(\boldsymbol{R})\left[\int_0^\epsilon d\boldsymbol{r}_2\,\rho_1(\boldsymbol{R}+\boldsymbol{r}_2)\right]^{q-1}, \qquad (23)$$

(here and below it is understood that one transforms from $\boldsymbol{x}$ to $\boldsymbol{r}$ coordinates before integration). Inserting $\rho_1(\boldsymbol{x}_k) = N_{\rm ev}^{-1}\sum_{e_k}\sum_{i_k}\delta(\boldsymbol{x}_k - \boldsymbol{X}_{i_k}^{e_k})$ for each factor, we find, after integration,

$$\xi_q^{\rm norm}(\epsilon) = N_{\rm ev}^{-1}\sum_{e_1}\sum_{i_1}\left[N_{\rm ev}^{1-q}\sum_{e_2,\ldots,e_q}\ \sum_{i_2,\ldots,i_q}\Theta(\epsilon - X_{i_1 i_2}^{e_1 e_2})\cdots\Theta(\epsilon - X_{i_1 i_q}^{e_1 e_q})\right], \qquad (24)$$

where now $X_{i_1 i_k}^{e_1 e_k} \equiv |\boldsymbol{X}_{i_1}^{e_1} - \boldsymbol{X}_{i_k}^{e_k}|$ measures the distance between two particles *taken from different events* e_1 and e_k. This much resembles the numerator expressions of Eq. (11), and, indeed, the sums also factorize here, so that

$$\xi_q^{\rm norm}(\epsilon) = \left\langle \sum_{i_1}\left\langle\sum_{i_2}\Theta(\epsilon - X_{i_1 i_2}^{e_1 e_2})\right\rangle^{q-1}\right\rangle \equiv \left\langle\sum_i \langle\hat{n}_b(\boldsymbol{X}_i^a, \epsilon)\rangle^{q-1}\right\rangle, \qquad (25)$$

or, in shorthand,

$$\xi_q^{\rm norm}(\epsilon) = \left\langle \sum_i \langle b\rangle^{q-1}\right\rangle. \qquad (26)$$

Comparing the numerator (11) to the denominator (26), we note that the exponent of the latter is an ordinary power $q-1$ instead of the factorial power $[q-1]$ of the former; this follows from the fact that in Eq. (24) there are no restrictions on the sum indices.

Figure 3 shows how the denominator sphere count $\hat{n}_b(\boldsymbol{X}_i^a, \epsilon) \equiv b$ is found by inserting the particle $\boldsymbol{X}_i^a$ of event a into another event b of the sample and doing the count $\hat{n}_b(\boldsymbol{X}_i^a, \epsilon) = \sum_j \Theta(\epsilon - X_{ij}^{ab})$ around it. This is done for many events b to obtain the inner event average of Eq. (25). Note that one has to distinguish carefully between a- and b-event averages: the computation of $\langle b\rangle = \langle\hat{n}_b(\boldsymbol{X}_i^a, \epsilon)\rangle$ involves an average over different events b *while the position of the sphere center is kept fixed* at $\boldsymbol{X}_i^a$. Expressions of the form $\langle\sum_i\langle b\rangle\rangle$ then denote sums of contributions when the center of the sphere "floats" over the $i = 1, \ldots, N$ particle positions of event a and finally over all events $a = 1, \ldots, N_{\rm ev}$.

While the derivation of Eq. (25) from (23) is exact, this expression for the normalization is correct only for $N_{\rm ev} \to \infty$: it contains a hidden bias due to correlations induced by use of the same events in every factor ρ_1 in (23). For the case where the experimental sample size is not infinite, statistical theory provides *estimators*, which from the limited-size sample estimate quantities for the "true" (i.e., infinitely large) sample. Applying the theory of estimators to our problem, we find that the product of distributions ρ_1 in Eq. (23) must be modified precisely in such a way that the event indices $e_1, \ldots, e_q$ are all mutually unequal. This is in agreement with the heuristic procedure of creating "fake events" where each track is taken from a different (real) event.

The corrections to obtain such an *unbiased* form of the normalization can be written as a series in powers of $1/N_{\rm ev}$, with the leading term given by $\langle b \rangle^{q-1}$. For the relatively small number of events and great sensitivity to bias found in heavy ion samples, the correction can be quite important, while the situation for hadronic data is less acute. We defer the details of this very technical discussion and the exact expressions to a future publication [17].

Subject to the above corrections, the normalized qth-order moment in its star integral form is

$$F_q(\epsilon) \equiv \frac{\xi_q(\epsilon)}{\xi_q^{\rm norm}(\epsilon)} \simeq \frac{\left\langle \sum_i a^{[q-1]} \right\rangle}{\left\langle \sum_i \langle b \rangle^{q-1} \right\rangle} . \tag{27}$$

As in the case of the traditional moments (3), one would measure $F_q(\epsilon)$ as a function of decreasing ϵ; a straight line in a plot of $\ln F_q$ versus $\ln \epsilon$ would, as before, be interpreted as scaling behavior of ρ_q. We caution, however, that there are important issues which must be addressed before such conclusions can be drawn, among them normalization effects.

The following points should be noted.

(1) Theoretical models are easily compared to data obtained with the star integral: when they predict the single-particle distribution ρ_1 and correlation functions ρ_q, their corresponding star integral moment is just given by the analytic integral expression (7). Monte Carlo simulations should, of course, take into account the proper removal of bias induced by finite sample size.

(2) The measurement of the star integral is very economical. Just as the numerator ξ_q can be measured in an algorithm of order N^2 (the number of particles per event), so the denominator requires only an algorithm of the order of the square of the sample size, $N_{\rm ev}^2$, a large savings over the previous GHP and snake integral algorithms [6]. (This savings is not destroyed by the above-mentioned corrections to obtain an unbiased normalization.) When the order of the events in the sample is random, this can be reduced even further by taking for the "inner" event average $\langle b \rangle$ only a fraction of the full sample, e.g., $b \in \{a-1, a-2, \ldots, a-A\}$. The smaller A, the faster the calculation but the larger the statistical error. Since a small value of A introduces a considerable bias, which disappears as it is increased, great care must be taken that A is of a size where the normalization becomes independent of its exact value. An optimal value for A can be found for a given sample and length ϵ by experimentation. Whenever doubt arises, the full $A = N_{\rm ev}-1$ sum should be taken as this is exact.

For smaller values of A, the error in the denominator will be non-negligible and should be combined with the numerator error, including covariances.

(3) For $\epsilon = \Delta x$, all Θ functions become unity by default and so $F_q(\Delta x) = \langle N^{[q]} \rangle / \langle N \rangle^q$, which is unity only when the event multiplicities in $\Delta \boldsymbol{x}$ are Poisson distributed. Furthermore, when the distribution is purely random for a given ϵ, then $F_q(\epsilon)$ becomes unity; see the discussion around Eq. (50) below.

(4) Annoying boundary effects due to the finiteness of $\Delta \boldsymbol{x}$ are largely canceled out, because the vertical normalization used means that sphere counts for centers $\boldsymbol{X}_i$ close to the boundary are reduced for both numerator and denominator. In the business of galaxy distributions, this remains a much-discussed topic [9]; since unfortunately only a single event exists in this case, the horizontal normalization, which is vulnerable to such boundary effects, has to be used there.

(5) Nevertheless, there may be instances where the horizontal normalization may be preferred. The definition of fractal dimensions, for example, is often couched in terms of the horizontally normalized correlation integral.

(6) The vertical normalization has the additional advantage that the effects of the single-particle distribution $\rho_1(\boldsymbol{x})$ are canceled out to some extent. While in many cases ρ_1 is constant ("stationary") or varying only weakly, it may in high-energy collisions vary quite strongly; especially the transverse momentum distribution $\rho_1(p_\perp)$ is strongly peaked and then falls off exponentially. Since sphere counts in both numerator and denominator vary about equally as a function of ρ_1, the trivial effects on F_q of a strongly varying single particle distribution are compensated. However, this compensation is only partial: when ϵ is small, each small domain is approximately flat and the cancellation is fairly reliable. When ϵ is large, though, the variation of a nonstationary ρ_1 within the domain is averaged out before the ratio is taken instead of the better reverse order. Ideally, one would divide up such large domains into many small ones and sum up the contributions only after normalization, but this is usually made impossible by bad statistics. Also, even the so-called vertical factorial moments (3) used traditionally suffer for large bins from the same effect, so that this problem can be ignored at the present level of sophistication.

(7) Of greater concern is the possibility that the integration of the center coordinate $\boldsymbol{R}$ over the entire domain $\Delta \boldsymbol{x}$ in Eqs. (19) and (23) suffers from the same problem for any value of ϵ. For ρ_1's that are weakly varying this does not matter very much, but when they change drastically [such as $\rho_1(p_\perp)$] a more vertical form should be used. At some cost in CPU time, this can be implemented straightforwardly. Instead of letting $\boldsymbol{R}$ range over the entire space, we introduce discrete binning; for the floating box,

$$\int_{\Delta \boldsymbol{x}} d\boldsymbol{R} \longrightarrow \prod_{f=1}^{d} \left(\frac{1}{L} \sum_{\ell_f=1}^{L} \int_{(\ell_f-1)\delta R}^{\ell_f \delta R} dR_f \right) , \tag{28}$$

with $\delta R = \Delta x / L$, while the floating sphere of Eq. (20)

yields, after integrating out the angles,

$$\int_{\Delta \boldsymbol{x}} d\boldsymbol{R} \longrightarrow \frac{1}{L} \sum_{\ell=1}^{L} \int_{(\ell-1)\delta R}^{\ell \delta R} R^{d-1}\, dR\,. \tag{29}$$

By splitting the $\boldsymbol{R}$ integration in this way, the correlation integral can be made largely independent of the shape of ρ_1. As one is usually interested in F_q as a function of ϵ only, the number of subdivisions L and bin sizes δR can be kept fixed throughout a computation; the only requirement is that the number of subdivisions should be large enough to ensure that the final correlation integral depends on L only weakly.

For systems where ρ_1 varies strongly only with one of its variables while varying weakly with the others, it may be possible to define a "hybrid" correlation integral with mixed normalization, implementing an R sum as in Eq. (28) for this component only and not for the weakly varying components. This would be useful to boost the number of counts.

As discussed in Sec. II B, splitting up $\boldsymbol{R}$ into different regions may in itself be useful in isolating different physical effects. No sums over ℓ would be taken for such cases.

(8) There remains the question how the traditional factorial moments (3) and star correlation integrals (27) are to be compared. The simplest answer is that they should not be compared at all, since by their different definitions they cannot be expected to yield exactly the same results even for the same data sample. Under ideal circumstances ($N_{\rm ev}\to\infty$, weakly varying ρ_1, etc.), the two should yield a similar slope asymptotically. It may be helpful to compare the two at equal ϵ, but deviations should not be interpreted as revealing anything fundamental. If one finds large differences between traditional factorial moments (3) and correlation integrals, then they are mostly due to the different normalization procedures used.

One can, however, check for consistency between the BP factorial moments and the star integrals. Both should be the same for $\epsilon = \Delta x$. Also, for integer M, one can do the star integral *separately* for every bin, with ϵ set to $\Delta x/M$, and then add up the contributions. This sum over star integrals should then be identical to the vertical factorial moment $F_q^v(M)$.

IV. CUMULANTS

A. Definition and use

The measurement of factorial moments and/or correlation integrals may be useful in itself in searching for a power law in the correlation function. Moments do not, however, reveal the true correlations, because the correlation function contains uncorrelated parts which have to be subtracted. This becomes clear on considering the effect that statistical independence has on the correlation function. Statistical independence of two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ means that the correlation function $\rho_2(\boldsymbol{x}_1,\boldsymbol{x}_2)$ factorizes into a product $\rho_1(\boldsymbol{x}_1)\rho_1(\boldsymbol{x}_2)$. Similarly, when $\boldsymbol{x}_1$ becomes statistically independent of all other points $\boldsymbol{x}_{k\neq 1}$, the higher-order correlation functions factorize accordingly: $\rho_q(\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_q) \to \rho_1(\boldsymbol{x}_1)\rho_{q-1}(\boldsymbol{x}_2,\ldots,\boldsymbol{x}_q)$. All possible combinations of such factorizations have to be subtracted from the original correlation function before one can speak of the "true" correlations.

These reduced quantities, known as cumulants, are basic to statistical analysis of any sort [18]. They are constructed precisely in such a way as to become zero whenever any one or more of the points x_k becomes statistically independent of the others. (The often-used factorization $\rho_q \to \rho_1^q$ is only the most drastic form of statistical independence, assuming that every point becomes independent of every other.) Cumulants of different distributions are also additive under convolution of the distributions [19] as well as being invariant under change of origin [18].

The first few cumulants C_q are, in terms of the correlation functions,

$$C_2(\boldsymbol{x}_1,\boldsymbol{x}_2) = \rho_2(\boldsymbol{x}_1,\boldsymbol{x}_2) - \rho_1(\boldsymbol{x}_1)\rho_1(\boldsymbol{x}_2)\,, \tag{30}$$

$$\begin{aligned} C_3(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3) = {}& \rho_3(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3) - \rho_1(\boldsymbol{x}_1)\rho_2(\boldsymbol{x}_2,\boldsymbol{x}_3) - \rho_1(\boldsymbol{x}_2)\rho_2(\boldsymbol{x}_3,\boldsymbol{x}_1) - \rho_1(\boldsymbol{x}_3)\rho_2(\boldsymbol{x}_1,\boldsymbol{x}_2) \\ & + 2\,\rho_1(\boldsymbol{x}_1)\rho_1(\boldsymbol{x}_2)\rho_1(\boldsymbol{x}_3)\,, \end{aligned} \tag{31}$$

$$\begin{aligned} C_4(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4) = {}& \rho_4(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4) - \sum_{(4)} \rho_1(\boldsymbol{x}_1)\rho_3(\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4) - \sum_{(3)} \rho_2(\boldsymbol{x}_1,\boldsymbol{x}_2)\rho_2(\boldsymbol{x}_3,\boldsymbol{x}_4) \\ & + 2\sum_{(6)} \rho_1(\boldsymbol{x}_1)\rho_1(\boldsymbol{x}_2)\rho_2(\boldsymbol{x}_3,\boldsymbol{x}_4) - 6\,\rho_1(\boldsymbol{x}_1)\rho_1(\boldsymbol{x}_2)\rho_1(\boldsymbol{x}_3)\rho_1(\boldsymbol{x}_4)\,. \end{aligned} \tag{32}$$

The bracketed numbers under the sums indicate the number of permutations of the arguments $\boldsymbol{x}_k$ which have to be included. Further, omitting the arguments,

$$C_5 = \rho_5 - \sum_{(5)} \rho_1\rho_4 - \sum_{(10)} \rho_2\rho_3 + 2\sum_{(10)} \rho_1\rho_1\rho_3 + 2\sum_{(15)} \rho_1\rho_2\rho_2 - 6\sum_{(10)} \rho_1\rho_1\rho_1\rho_2 + 24\rho_1^5\,, \tag{33}$$

and so on for higher orders.

These equations have been utilized to find simple relations between the vertical factorial moments F_q of Eq. (3) and the integrated normalized cumulants [10]. With $m = (m_1, \ldots, m_d)$ as usual, the integrated normalized cumulant is defined as

$$K_q^v(M) \equiv \frac{1}{M^d} \sum_m \frac{\int_{\Omega_m} \prod_k d\boldsymbol{x}_k\, C_q(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)}{\left[\int_{\Omega_m} d\boldsymbol{x} \rho_1(\boldsymbol{x})\right]^q}, \quad (34)$$

which yields relations such as $K_2^v = F_2^v - 1$, $K_3^v = F_3^v - 3F_2^v + 2$, etc. which can thus be utilized directly by experimentalists. These relations hold exactly for $q \leq 3$ and approximately for $q > 3$. They are not true for horizontally normalized moments.

B. Correlation integral cumulants

In contrast with Eq. (34), the correlation integral cumulant is defined as the integral of C_q over the domains Ω_S of Eq. (20) or Ω_B of Eq. (22) after appropriate transformation to relative coordinates:

$$K_q(\epsilon) \equiv \frac{f_q(\epsilon)}{\xi_q^{\mathrm{norm}}(\epsilon)}, \quad (35)$$

with

$$\begin{aligned} f_q(\epsilon) &\equiv \int_{\Delta \boldsymbol{x}} d\boldsymbol{R} \int_0^\epsilon \prod_k d\boldsymbol{r}_k\, C_q(\boldsymbol{R}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_q) \\ &= \int C_q(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)\, \Theta_{12}\Theta_{13}\cdots\Theta_{1q}\, d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_q \quad (36) \end{aligned}$$

the un-normalized factorial cumulant. The latter can be written entirely in terms of the sphere counts introduced previously:

$$\begin{aligned} a &= \sum_j \Theta(\epsilon - X_{ij}^{aa}) = \hat{n}(\boldsymbol{X}_i^a, \epsilon), \qquad j \neq i, \\ b &= \sum_j \Theta(\epsilon - X_{ij}^{ab}) = \hat{n}_b(\boldsymbol{X}_i^a, \epsilon) \quad (37) \end{aligned}$$

(see Figs. 2 and 3 for terms contributing to a and b respectively). To demonstrate this, we start with $q = 2$. Here $f_2 = \int(\rho_2 - \rho_1\rho_1)$, which by Eqs. (11)–(13) and (23)–(26) is seen to yield (henceforth we suppress the dependence on ϵ)

$$f_2 = \left\langle \sum_i (a - \langle b \rangle) \right\rangle. \quad (38)$$

For $q = 3$, the first term in the expansion (31) of C_3 is just $\xi_3 = \langle \sum_i a^{[2]} \rangle$, while the last term $\rho_1\rho_1\rho_1$ yields $\xi_3^{\mathrm{norm}} = \langle \sum_i \langle b \rangle^2 \rangle$. The three "mixed terms" involving both ρ_2 and ρ_1 must be worked out explicitly. On the one hand,

$$\begin{aligned} \int d\boldsymbol{R} \int_0^\epsilon d\boldsymbol{r}_2\, d\boldsymbol{r}_3 \rho_1(\boldsymbol{x}_1)\rho_2(\boldsymbol{x}_2, \boldsymbol{x}_3) &= N_{\mathrm{ev}}^{-2} \sum_{a,b} \sum_i \sum_{j \neq k} \Theta(\epsilon - X_{ij}^{ab})\Theta(\epsilon - X_{ik}^{ab}), \\ &= \left\langle \sum_i \langle b^{[2]} \rangle \right\rangle, \end{aligned} \quad (39)$$

while on the other hand, if $\boldsymbol{x}_1$ is contained in ρ_2,

$$\int d\boldsymbol{R} \int_0^\epsilon d\boldsymbol{r}_2\, d\boldsymbol{r}_3 \rho_1(\boldsymbol{x}_2)\rho_2(\boldsymbol{x}_1, \boldsymbol{x}_3) = \left\langle \sum_i a\langle b \rangle \right\rangle, \quad (40)$$

so that, putting all the pieces together,

$$f_3 = \left\langle \sum_i \left(a^{[2]} - \langle b^{[2]} \rangle - 2a\langle b \rangle + 2\langle b \rangle^2 \right) \right\rangle. \quad (41)$$

The constant recurrence of the outer event average and i sum suggests that we define an "i-particle cumulant" by

$$\left\langle \sum_i \hat{f}_q(i) \right\rangle \equiv f_q, \quad (42)$$

in terms of which we find

$$\begin{aligned} \hat{f}_2(i) &= a - \langle b \rangle, & (43) \\ \hat{f}_3(i) &= a^{[2]} - \langle b^{[2]} \rangle - 2a\langle b \rangle + 2\langle b \rangle^2, & (44) \\ \hat{f}_4(i) &= a^{[3]} - \langle b^{[3]} \rangle - 3a^{[2]}\langle b \rangle - 3a\langle b^{[2]} \rangle + 6\langle b \rangle\langle b^{[2]} \rangle + 6a\langle b \rangle^2 - 6\langle b \rangle^3, & (45) \\ \hat{f}_5(i) &= a^{[4]} - \langle b^{[4]} \rangle - 4a^{[3]}\langle b \rangle - 4a\langle b^{[3]} \rangle - 6a^{[2]}\langle b^{[2]} \rangle + 8\langle b \rangle\langle b^{[3]} \rangle + 12a^{[2]}\langle b \rangle^2 + 6\langle b^{[2]} \rangle\langle b^{[2]} \rangle & \\ &\quad + 24a\langle b \rangle\langle b^{[2]} \rangle - 36\langle b \rangle^2\langle b^{[2]} \rangle - 24a\langle b \rangle^3 + 24\langle b \rangle^4, & (46) \end{aligned}$$

which are then summed over all particles i and averaged over all events to yield f_q.

How these sums can be obtained graphically is illustrated for $q = 3$ in Fig. 4(a) and $q = 4$ in Fig. 4(b). The black squares represent individual particles; those enclosed by a circle belong to the same event. The center particle at $\boldsymbol{X}_i^a$ is connected to the $q-1$ other particles by the lines representing the Θ functions. One now draws all possible event topologies with $q-1$ lines connected to one center particle. For p joining lines within the $\boldsymbol{X}_i^a$ event, one writes down a factor $a^{[p]}$; while lines connecting $\boldsymbol{X}_i^a$ to p particles in the same (other) event yields a factor $\langle b^{[p]}\rangle$. p lines going to different events $b, c, \ldots$ results in a factor $\langle b\rangle^p$. Putting all such factors together and assigning to each the appropriate sign and prefactor from Eqs. (30)ff, one obtains the cumulant expansions (43)–(46).

Writing higher orders recursively in terms of lower-order cumulants,

$$\begin{aligned}
\hat{f}_3(i) &= a^{[2]} - \langle b^{[2]}\rangle - 2\langle b\rangle \hat{f}_2(i)\,, \\
\hat{f}_4(i) &= a^{[3]} - \langle b^{[3]}\rangle - 3\langle b\rangle \hat{f}_3(i) - 3\langle b^{[2]}\rangle \hat{f}_2(i)\,, \\
\hat{f}_5(i) &= a^{[4]} - \langle b^{[4]}\rangle - 4\langle b\rangle \hat{f}_4(i) - 6\langle b^{[2]}\rangle \hat{f}_3(i) - 4\langle b^{[3]}\rangle \hat{f}_2(i)\,,
\end{aligned} \tag{47}$$

we are led to the conjecture that

$$\hat{f}_q(i) = a^{[q-1]} - \langle b^{[q-1]}\rangle - \sum_{p=2}^{q-1} \binom{q-1}{p-1} \langle b^{[q-p]}\rangle \hat{f}_p(i)\,, \tag{48}$$

which, if proven for arbitrary $q \geq 5$, could open the way for an easy calculation of cumulants to arbitrary order

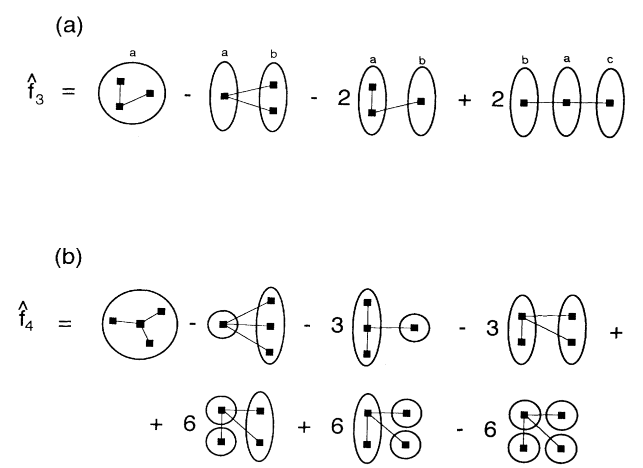

FIG. 4. Schematic representation of event mixing terms entering the (a) third- and (b) fourth-order cumulant [Eqs. (44)–(45)]. For a given center particle i in event a, the other particles in the q-tuple can be either within the same event a or in different events $b, c, \ldots$ (see text). p particles in event a lead to a factor $a^{[p-1]}$; p particles in event b give a factor $\langle b^{[p]}\rangle$ and a particle in p different events b, c, ..., gives a factor $\langle b\rangle^p$. With the appropriate combinatorial prefactors, the sum of these terms yields the third- and fourth-order cumulant integrals.

without doing the messy algebra involved.

Just as the normalization ξ_q^{norm} must be corrected for bias, we must for third and higher order cumulants also correct for limited sample size. Here, too, using estimators to correct for the effect of the limited-size sample requires that all event sums go over unequal events. This again results in corrections of the order $1/N_{\mathrm{ev}}$, $1/N_{\mathrm{ev}}^2, \ldots$ so that, for example, $\hat{f}_3(i)$ above would acquire the additional correction term $2(\langle b\rangle^2 - \langle b^2\rangle)/(N_{\mathrm{ev}} - 2)$. While these corrections can be quite important, we defer a discussion of their origins and exact expressions to future work [17].

When an event sample has no correlations, the count around a particle in event a would on average be the same as when it was inserted into other events,

$$a^{[q-1]} \to \langle b^{[q-1]}\rangle, \tag{49}$$

which results in $\hat{f}_q = 0$. This would of course be true on average only. Less obvious but also true is that if any one of the q variables of C_q becomes independent, the integrated cumulants f_q become zero also; this can be shown graphically too. A stronger condition of randomness, comparable to the Poisson distribution in fixed bins, would be reflected by the behavior

$$\begin{aligned}
\left\langle \sum_i a^{[q-1]} \right\rangle &\longrightarrow \left\langle \sum_i \langle b\rangle^{q-1} \right\rangle, \\
\left\langle \sum_i \langle b^{[q-1]}\rangle \right\rangle &\longrightarrow \left\langle \sum_i \langle b\rangle^{q-1} \right\rangle,
\end{aligned} \tag{50}$$

so that the moments would go to unity, $F_q \to 1$, for this case.

It must be stressed that the qth-order cumulant contains no correlations of order lower than q. Thus even if $f_2 > 0$, f_3 can still be zero when there are no true third-order correlations; Equation (47) is merely a convenient grouping of the terms.

We further see that, since only the basic quantities a and b are needed to construct cumulants, they can also be calculated very economically with order N^2 algorithms.

The estimation of errors always provides a headache since mostly one has to deal with the intricacies of error propagation and covariances. In the star integral formulation, however, this process is much simplified; for the statistical error on K_q, one merely has to calculate

$$\sigma^2(f_q) = \frac{\left\langle [\sum_i \hat{f}_q(i)]^2 \right\rangle - f_q^2}{N_{\mathrm{ev}}} \tag{51}$$

and combine this with the corresponding statistical error for the event mixing denominator. Errors for the factorial moments F_q and differentials of Sec. V are obtained with the same ease.

C. Presenting cumulants

While there may be many useful ways to plot correlations, depending on what one is looking for, we recommend the following format for cumulants. The second-order cumulant

$$K_2(\epsilon) = \frac{\left\langle \sum_{i\neq j} \Theta(\epsilon - X_{ij}^{aa}) \right\rangle}{\left\langle \sum_{i,j} \Theta(\epsilon - X_{ij}^{ab}) \right\rangle} - 1 \tag{52}$$

cannot be smaller than -1. Its theoretical maximum is harder to find; but an estimate can be made by using the extreme case where the events are extremely "spiky" but the spikes are uniformly distributed in phase space (from event to event). The event-averaged count will be approximately $\bar{n} = \langle N \rangle (\epsilon/\Delta x)^d$, while the Θ functions of the numerator are all unity for our narrow spikes; this extreme case thus gives an approximate upper limit

$$-1 < K_2(\epsilon) < \frac{\langle N(N-1)\rangle}{\langle \bar{n}(\bar{n}-1)\rangle} - 1 \simeq \left(\frac{\Delta x}{\epsilon}\right)^{2d}, \tag{53}$$

to which a given sample K_2 can be compared. For higher orders, one may then plot the K_q directly and/or as the ratios K_q/K_2, which would express qth-order correlations as a fraction of second-order correlations. Testing for linking, on the other hand [10], one would plot the ratios K_q/K_2^{q-1}.

Apart from calculating cumulants which are averaged over the entire event sample, it may in specific cases be interesting to look at "single-event cumulants" $\hat{f}_q$ for rare events, for example if a certain event or group of events exhibits unusual patterns in a phase-space plot. To see whether such cumulants differ significantly from mere statistical noise in the fluctuations, they should be plotted on top of the corresponding event-averaged cumulant plus/minus twice (or three times) the error $f_q \pm 2\left(\left\langle [\sum_i \hat{f}_q(i)]^2 \right\rangle - f_q^2\right)^{1/2}$. Whether such single-event cumulants are a good idea will have to be established in practice. We remind the reader that the distribution of factorial moments is not necessarily Gaussian [20] and that a large-deviation analysis may also be appropriate in such cases [21].

V. DIFFERENTIAL VERSIONS

Correlation integrals and their cumulants described so far are defined always in terms of a maximum distance ϵ; the ubiquitous Θ functions ensure that all interparticle distances X_{ij} involved must be smaller than ϵ. The simplicity of this definition allows one to test clusters of many particles at once, i.e., probe correlations of order 3 or higher, something not possible using the conventional methods of measuring correlation functions. It makes good sense, however, to ask not only whether some interparticle distance is smaller than some value, $X_{ij} < \epsilon$, but also whether it falls within a certain *distance interval* $[\epsilon_{t-1}, \epsilon_t]$.

To this purpose, we define a sequence of distances $\epsilon_1, \epsilon_2, \ldots$ up to some maximum distance determined by the total domain of integration. This sequence can be either linear,

$$\epsilon_t = t\epsilon_1, \tag{54}$$

or exponential,

$$\epsilon_t = \epsilon_1 c^{t-1}, \quad c > 1, \tag{55}$$

the second definition being useful because data presented as function of $\ln \epsilon_t$ will be equally spaced. These two sequences divide up the whole phase space into adjacent and disjoint domains; for $d = 1$ and $q = 2$, these domains are shown in Fig. 5 as the sequence of strips filling the entire domain. We also introduce the *indicator function*

$$I_t(X) \equiv \Theta(\epsilon_t - X) - \Theta(\epsilon_{t-1} - X), \tag{56}$$

which is unity when $\epsilon_{t-1} < X < \epsilon_t$ and zero otherwise.

The differential forms are defined as follows (see also Fig. 6). Given a center particle $\boldsymbol{X}_{i_1}^a$ in event a, the number of particles situated a distance $X_{i_1 i_k} \in [\epsilon_{t-1}, \epsilon_t]$ away from $\boldsymbol{X}_{i_1}$ is

$$\Delta\hat{\xi}_2(i,t) = \hat{n}(\boldsymbol{X}_{i_1}^a, \epsilon_t) - \hat{n}(\boldsymbol{X}_{i_1}^a, \epsilon_{t-1}) \equiv a_t - a_{t-1}, \tag{57}$$

the latter defining the shortened notation we shall be using. We next ask how many clusters of $q-1$ particles exist for which the *maximum distance* to $\boldsymbol{X}_{i_1}$ is in this interval, $\max(X_{i_1 i_2}, \ldots, X_{i_1 i_q}) \in [\epsilon_{t-1}, \epsilon_t]$. The answer is surprisingly simple: the number of such clusters is

$$\begin{aligned}\Delta\hat{\xi}_q(i,t) &= \hat{n}(\boldsymbol{X}_{i_1}^a, \epsilon_t)^{[q-1]} - \hat{n}(\boldsymbol{X}_{i_1}^a, \epsilon_{t-1})^{[q-1]} \\ &\equiv a_t^{[q-1]} - a_{t-1}^{[q-1]},\end{aligned} \tag{58}$$

since through use of

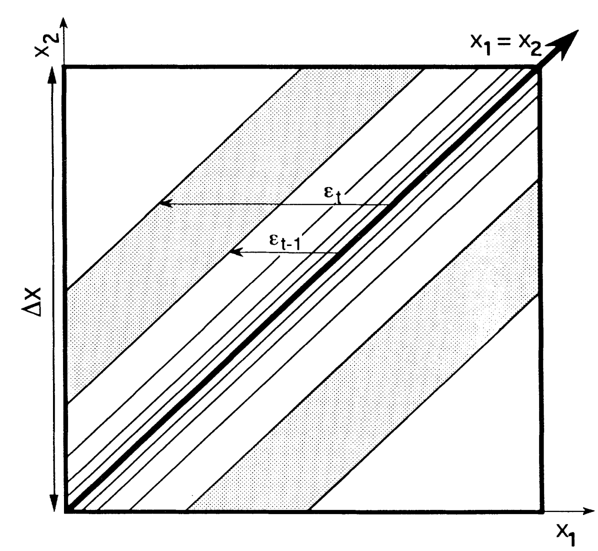

FIG. 5. The exponential sequence of distances ϵ_t ($t = 1, 2, \ldots$) of Eq. (55), used to define the differential forms of correlation integrals of Sec. V. The shaded regions represent the integration area of the differential integral $\Delta F_2(t)$ over the two-particle density $\rho(x_1, x_2)$. Note that the coordinates x in this figure are one-dimensional ($d = 1$) and the labels 1 and 2 refer to two different particles within the interval Δx.

$$\Theta(\epsilon_T - X) = \sum_{t=1}^{T} I_t(X) \tag{59}$$

and

$$\Theta(\epsilon - \max(X_1, \ldots, X_q)) = \prod_{p=1}^{q} \Theta(\epsilon - X_p) \tag{60}$$

we can show that

$$\Delta\hat{\xi}_q(i_1, t) = \left(\sum_{i_2} \Theta(\epsilon_t - X_{i_1 i_2})\right)^{[q-1]} - \left(\sum_{i_2} \Theta(\epsilon_{t-1} - X_{i_1 i_2})\right)^{[q-1]} = \sum_{i_2 \neq \ldots \neq i_q} I_t(\max(X_{i_1 i_2}, \ldots, X_{i_1 i_q}))\,. \tag{61}$$

In Fig. 6, one such cluster is shown, where at least one particle other than the central one is in the shaded regions for the floating sphere domain of Eq. (20).

The normalization proceeds along the same lines as in Sec. III, so that not surprisingly we find that the normalized differential moment is

$$\Delta F_q(t) = \frac{\left\langle \sum_i a_t^{[q-1]} - a_{t-1}^{[q-1]} \right\rangle}{\langle \sum_i \langle b_t \rangle^{q-1} - \langle b_{t-1} \rangle^{q-1} \rangle}\,, \tag{62}$$

while the cumulants are similarly

$$\Delta K_q(t) = \frac{\left\langle \sum_i \hat{f}_q(i,t) - \hat{f}_q(i,t-1) \right\rangle}{\langle \sum_i \langle b_t \rangle^{q-1} - \langle b_{t-1} \rangle^{q-1} \rangle}\,. \tag{63}$$

Both ΔF_q and ΔK_q are thus accessible to measurement with minimal additional effort. Usually, they will be plotted as a function of t, i.e., the distance interval within which the maximum interparticle distance would fall. ΔF_q can be thought of as the analogue to the "factorial correlator" defined in Ref. [1] and ΔK_q to their cumulants [22].

We conclude this section by listing the advantages of using the differential moments and cumulants as a measure of cluster size.

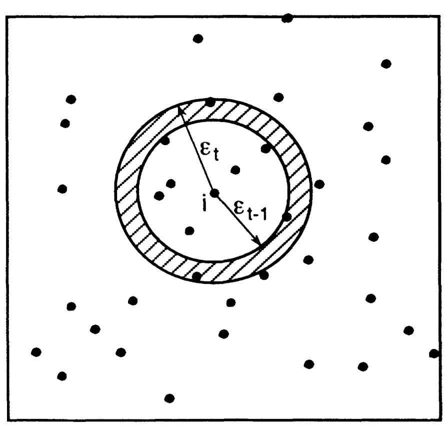

FIG. 6. An example of the differential sphere count within the shaded area in two dimensions ($d = 2$) gives $\Delta\hat{\xi}_2(i,t) = a_t - a_{t-1} = 9 - 6 = 3$ [Eq. (57)]. For higher orders, the number of all q-stars with size within the interval $[\epsilon_{t-1}, \epsilon_t]$, i.e., with at least one of the $q-1$ neighboring particles within the shaded area, is given by $\Delta\hat{\xi}_q(i,t) = a_t^{[q-1]} - a_{t-1}^{[q-1]}$ of Eq. (58). In the present figure, for example, $\Delta\hat{\xi}_3(i,t) = 42$.

First of all, they have all the advantages of the correlation integrals in that their statistics will be higher and the data points calculated more stable than the corresponding correlator. Especially for higher dimensions, the large gain in statistics will permit measurements that would be impossible otherwise.

Differential moments and cumulants are almost immune to the problems of converting from biased to unbiased estimators in the normalization and in $\hat{f}_q$. Because these corrections for bias take the form of an additive series, taking the difference $\hat{f}_q(i,t) - \hat{f}_q(i,t-1)$ causes them to cancel to a large degree.

When a unique distance can be defined as in Eq. (21), the differential count is unambiguous even in higher dimensions, unlike the correlator, where distance definitions are ambiguous for multidimensional analyses [23].

Very importantly, the domains of integration of the normalized differential moments and cumulants are disjoint (Fig. 5), meaning that the data points will not be correlated amongst themselves, a constant bug in ordinary moments and cumulants.

A special status must be accorded to $\Delta F_2(t)$: it behaves as a "roaming distance," looking for all particle pairs that are a certain distance apart. This is suggestive of interpreting ΔF_2 as a kind of Fourier transform of the distances [24].

VI. CONCLUSION

We have developed a general formalism for measuring correlations of point distributions. The language used has been that of high-energy physics, but we believe that the method may be of use in other fields also. The use of a δ function notation has enabled us to derive the star integral from first principles and through its greater clarity pointed to a number of important extensions. Most important of these is that the correlation integral used in the astronomy literature [3,9] and suggested for high-energy physics [5] appears to be in need of modification from the form $\hat{n}^{q-1}$ to the "factorial power" form $\hat{n}^{[q-1]}$. To assess whether such a modification is possible or practical in galaxy distributions is beyond the scope of this paper; for the limited number of particles encountered in high-energy physics, however, it seems an unavoidable and clearly superior formulation.

The felicitous definition of relative coordinates leading to the star integral makes the latter very economical to calculate, including all cumulants and differential quantities. Since, in addition, the domains of integration are the largest possible, we believe that the star integral extracts the maximum available correlation information from the data for the minimum price in CPU time. In this respect, it has proven superior to the traditional factorial moments and older correlation measurements, especially when correlations are measured in higher-dimensional spaces. As computers continue their evolution to previously unimaginable speeds and capacity, the present method should become routine even for the higher orders.

Three issues have been dealt with only cursorily or not at all: the question of eliminating the influence of the total multiplicity distribution, the measurement of correlations between different species of particles (e.g., distinguished by their charge), and the problem of properly defining fractal dimensions for nonstationary ensembles of events rather than the usual single-event time series. We hope to return to the latter in future.

The present paper has emphasized the *how to* rather than the *what* of measurement. At first sight, it may seem unnecessary to devote so much effort to the mere process of measurement. However, high-energy hadronic collisions, and to an even greater extent nucleus-nucleus collisions, are presently in a state where very few exact calculations on correlations can be done and most theoretical work relies on assumptions which often are hard to support or believe. In this context, we believe that it is of cardinal importance that there should be a clear and unambiguous method for analyzing correlations, and that, if possible, a standard should be established by which different experiments can be compared.

The confusion underlying the dynamics of correlations is reflected in the cursory way in which we have treated the choice of variables. While there are some theoretical preferences [25], we are fairly ignorant of the dynamics of the soft component and hence the best choice of variables in this case. The original proposal regarding self-similarity in particle production [1] did not have theoretical grounding in currently acceptable physical theories but was based on a toy model to illustrate the idea, while exact calculations of correlations, e.g., in a QCD framework [26] can be applied only to high-$p_\perp$ processes.

The occurrence of different forms of correlation integrals might cause unpleasant confusion among experimentalists who would prefer a unique recipe to extract information of higher-order correlation functions. Unfortunately, there is *a priori* no best choice. The different forms merely reflect the freedom of choice of the particular *shape* of the integration domain; but all commonly probe the correlation functions by decreasing the *size* of the integration domain. While the numerical values of the various integrals may differ, the functional dependence is supposed to be similar (this was shown numerically for the snake and GHP integrals [6]). Moreover, by suitable normalization most of the numerical differences between the various forms can be divided out, so that the choice of a particular form can be guided by practical efficiency arguments.

On very general grounds, the choice of relative coordinates seems a wise one, in whatever variables one may prefer. This is true especially for cases where there is some degree of stationarity in the distribution (i.e., invariance under translation), which is generally assumed to be true for galaxy distributions and perhaps for pion distributions at higher collision energies. The ubiquitous use of such stationarity assumptions testifies to their popularity. Not least, measurements in Bose-Einstein interferometry rely on relative coordinates, whether in their three-vector form $\boldsymbol{q} = \boldsymbol{p}_1 - \boldsymbol{p}_2$ or as a function of $Q^2 = -(p_1 - p_2)^2$.

In this context, we have aimed to provide a framework that is adaptable to any future choice of variables and dynamical theories. This will hopefully allow for clean measurements to guide theoretical thinking, while remaining flexible in its implementation.

ACKNOWLEDGMENTS

H.C.E. and P.C. gratefully acknowledge support by the Alexander von Humboldt Foundation. H.C.E. and P.L. thank the Institut für Hochenergiephysik der ÖAW, and the Theory Groups at the University of Arizona and the University of Regensburg, respectively, for support. P.C. acknowledges the hospitality of the GSI and the University of Frankfurt. Special thanks to the EMC and NA35 groups in Munich for insisting on clear thinking, and to F. Botterweck, M. Charlet, and D. Weselka for useful discussions. This work was supported in part by the U.S. Department of Energy, High Energy and Nuclear Physics Divisions, Grant No. DE-FG02-88ER40456.

---

[1] A. Białas and R. Peschanski, Nucl. Phys. **B273**, 703 (1986); **B308**, 857 (1988).
[2] P. Grassberger, Phys. Lett. **97A**, 227 (1983); H.G.E. Hentschel and I. Procaccia, Physica **8D**, 435 (1983).
[3] G. Paladin and A. Vulpiani, Lett. Nuovo Cimento **41**, 82 (1984).
[4] G. Paladin and A. Vulpiani, Phys. Rep. **156**, 147 (1987).
[5] I. M. Dremin, Mod. Phys. Lett. A **3**, 1333 (1988); **4**, 1027 (1989).
[6] P. Lipa, P. Carruthers, H. C. Eggers, and B. Buschbeck, Phys. Lett. B **285**, 300 (1992).
[7] I. Juricic *et al.*, Phys. Rev. D **39**, 1 (1989).
[8] H. C. Eggers, P. Lipa, P. Carruthers, and B. Buschbeck, Phys. Lett. B **301**, 298 (1993).
[9] P. H. Coleman and L. Pietronero, Phys. Rep. **213**, 311 (1992).
[10] P. Carruthers, H. C. Eggers, and I. Sarcevic, Phys. Lett. B **254**, 258 (1991).
[11] ABCDHW Collaboration, A. Breakstone *et al.*, in *Proceedings of the 24th International Conference on High Energy Physics*, Munich, West Germany, 1988, edited by R. Kotthaus and J. Kuhn (Springer, Berlin, 1988).
[12] H. Aihara *et al.*, Phys. Rev. D **31**, 996 (1985).
[13] Yu. L. Klimontovich, *The Statistical Theory of Non-Equilibrium Processes in a Plasma* (MIT Press, Cambridge, MA, 1967).

[14] P. Carruthers, Astrophys. J. **380**, 24 (1991).
[15] G. Jancso and M. Charlet (private communication).
[16] H. Atmanspacher, H. Scheingraber, and G. Wiedenmann, Phys. Rev. A **40**, 3954 (1989).
[17] P. Lipa and H. C. Eggers (unpublished).
[18] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, 5th ed. (Oxford University Press, New York, 1987), Vol. 1.
[19] P. Carruthers and C. C. Shih, Int. J. Mod. Phys. A **2**, 1447 (1987).
[20] P. Lipa, H.C. Eggers, F. Botterweck, and M. Charlet, Z. Phys. C **54**, 115 (1991).
[21] J. Seixas and R. Vilela Mendes, Nucl. Phys. **B383**, 622 (1992).
[22] H. C. Eggers, P. Carruthers, P. Lipa, and I. Sarcevic, Phys. Rev. D **44**, 1975 (1991).
[23] NA22 Collaboration, M. Charlet, in *Proceedings of the Ringberg Workshop on Multiparticle Production*, 1991, edited by R. C. Hwa, W. Ochs, and N. Schmitz (World Scientific, Singapore, 1992).
[24] P. Carruthers and T. Hakioglu, Phys. Rev. D **45**, 4046 (1992).
[25] B. Andersson, P. Dahlqvist, and G. Gustafsson, Phys. Lett. B **314**, 604 (1988); J. D. Bjorken, Phys. Rev. D **45**, 4077 (1992).
[26] W. Ochs and J. Wosiek, Phys. Lett. B **289**, 159 (1992); P. Brax and R. Peschanski, Saclay Report No. SPhT–92–005 (unpublished).