PHYSICAL REVIEW D VOLUME 43, NUMBER 3 1 FEBRUARY 1991

Information theory and phenomenology of multiple hadronic production

G. Wilk
Soltan Institute for Nuclear Studies, Warsaw, Poland

Z. Włodarczyk
Institute of Physics, Pedagogical University, Kielce, Poland
(Received 31 May 1990)

Multiple hadronic production is usually visualized as a two-step process: the formation of some well-defined intermediate objects such as strings or fireballs and their subsequent hadronization (decays). It is shown how information theory provides us with a model-independent tool in dealing with the hadronization step for which the most plausible distributions of hadrons are formed.

I. INTRODUCTION

Multiple hadronic production processes are usually visualized as proceeding in two main steps. At first a number of more or less well-defined intermediate objects such as strings[1-3] or fireballs[4-6] is formed. Next follows their hadronization: usually one says that strings "fragment" and fireballs "decay" into finally observed hadrons. That terminology reflects the essential dynamical difference between them: strings are supposed to be essentially one-dimensional whereas fireballs are three-dimensional objects. However, confrontation with experimental data has washed out this difference substantially: strings are now allowed to "bend" in phase space before fragmentation (so they can produce more hadrons than before)[7] whereas fireballs are usually forced to decay anisotropically (therefore reducing their hadronic multiplicities).[4] In both cases the agreement with data is claimed as satisfactory, their apparently very different dynamic origins notwithstanding.

For somebody not interested in the hadronization part of the process it causes the problem of how to tell the difference between "fat" strings and "lean" fireballs; i.e., which description more accurately depicts the hadronization stage to follow? Dynamical origins aside, in both cases we are faced with apparently similar objects of well-defined invariant energies (masses) M which decay into $N=N(M)$ secondaries with obviously limited transverse momenta P_{Ti}. A possible way out is to look for experiments in which the first step of the process can be argued to be trivial, i.e., in which only one object is supposed to be produced ($e^+e^-\rightarrow$hadrons and $l+N\rightarrow$hadrons are the usual choices here), and just use the corresponding data, either directly or via their "best" parametrization, as an input for the hadronization stage.[1,3]

In this work we would like to quantify such an approach by providing *least-biased* (or *most plausible*) model-independent distributions of hadrons from objects, which we shall for definiteness call fireballs and of which we only know their masses M, the number of secondaries they produce, N, and that suitably defined transverse momenta of these secondaries are strongly damped. Implicit here is the assumption that the distribution of the number and masses of those fireballs should be provided by some other dynamical considerations (cf. for example, Ref. 5). Also the limitation of the transverse momenta of secondaries has to be regarded as a dynamical input which we shall only make use of but which we are not going to explain in what follows.

We shall argue that the only way to achieve this goal is to make use of the information-theory approach (already explored along similar lines in other branches of science[8]). That is because only this mathematical theory tells us how to produce model-independent distributions of quantities of interest. By this we understand that for given values of some external dynamical parameters (such as M, N, etc.) there is only one possible outcome which does not depend further on the dynamics of the hadronization process. Therefore such a distribution should have a status comparable to that of the phase space of reaction and should facilitate the subsequent investigations of different dynamical models of hadronic collisions.

II. INFORMATION THEORY IN HADRONIC PHYSICS

Let us first briefly summarize the essence of the information-theory approach from the perspective of multihadron production. The reasoning goes as follows: the proliferation of phenomenological models, which more or less satisfactorily describe experimental data in quite different (if not contradictory) dynamical terms, indicates that these data contain in fact only a rather limited amount of *information*. It is then quite possible that precisely this information, reflecting some important features of strong interactions or just conservation laws, is always used in formulating the basic assumptions of otherwise apparently different models. This would explain the common feature of these models—their phenomenological success.

To check this hypothesis one has to be able to answer, in a model-independent way, the following question: given some physical assumptions of a model (plus phase space and conservation law), what are then their *most trivial* consequences regardless of other details of the model?

In order to answer this question one has to define what "triviality" means and how to treat it quantitatively (i.e., how it can be measured). The most natural thing is to accept that *triviality* corresponds to *the lack of information* (i.e., the less informative, the more trivial). Thus, again, the *information* appears in a natural way as a useful concept in characterizing the hadronic production processes. One is then led to the information theory for definitions and measures of information and for the methods of interference of distributions of interest, which would correspond to imposed constraints and which were both least biased and unique. In what follows we shall provide for completeness a brief review of the necessary concepts, definitions, and formulas from the information theory.[9–11]

Suppose that from some experiment we have obtained a set of data which can be represented by the probability assignment $P^0=\{p_i^0\ (i=1,2,\ldots,n)\}=\{$*probabilities of occurrence of different events in this set*$\}$. This assignment represents our original state of knowledge about the outcome of this experiment. Let there be now a new set of data from the same experiment, which changes our state of knowledge to $P=\{p_i\ (i=1,2,\ldots,n)\}$. Then the quantity

$$I(P;P^0)=\sum_{i=1}^{n} p_i \ln(p_i/p_i^0) \tag{1}$$

represents a unique measure for the information provided by the new data. It is called *relative information* or just *information*.[9] It can be readily generalized to the continuous probability spaces [where probability density $\rho(x)$ replaces p_i]:

$$I(\rho;\rho^0)=\int_a^b \rho(x)\ln\left[\frac{\rho(x)}{\rho^0(x)}\right]dx \ . \tag{2}$$

Notice that $I(\rho;\rho^0)$ is invariant under a monotone one-to-one transformation $x\to y=g(x)$. This ensures that the information in the probability assignment ρ, relative to ρ^0, does not depend on the variable(s) chosen to describe the sample space (i.e., the space of events of our experiment, which in our case will aways be a phase space of reaction). Notice also that we obtain information only when we learn something new, i.e., whenever the final state of knowledge P differs from the initial one P^0.

The prior probabilities P^0 settle then, in a sense, the level from which we measure our information. We say that they represent the state of zero information, i.e., the state corresponding to the knowledge of only the space of events, in our case the phase space. In Eq. (1) it would mean $P_i^0=1/n$, in Eq. (2) $\rho^0(x)=\text{const}=1/(b-a)$, in general $\rho^0(x)=\text{const}=1/\Gamma$ where $\Gamma=\{$*phase space*$\}$.

In what follows the quantity important to us will be the so-called *missing information* (or *uncertainty*) in the probability assignment P defined as

$$U(P)=U(P;P^m;P^0)=I(P^m;P^0)-I(P;P^0) \ . \tag{3}$$

Here P and P^0 are defined as before and the probability assignment $P^m=\{p_i^m\ (i=1,2,\ldots,n)\}$ corresponds to the maximum knowledge which can be obtained from the outcome of our experiment.

In most cases of interest one deals with

$$P^0=\{p_i^0=\frac{1}{n}(i=1,2,\ldots,n)\} \ ,$$

$$P^m=\{\delta_{ik}\ (i=1,2,\ldots,n;\ k=\text{fixed})\} \ ,$$

which leads to

$$U(P)=U(\{p_i\})=-\sum_{i=1}^{n} p_i \ln p_i \ . \tag{4}$$

Correspondingly, in the continuous case,

$$\rho^0(x)=\text{const}=\frac{1}{\Gamma},\quad \Gamma=\{\text{phase space}\} \ ,$$

$$\rho^m(x)=\begin{cases} L^{-1} & \text{for } x\subset L,\ L \text{ being a small cell in a phase space } \Gamma \ , \\ 0 & \text{otherwise} \end{cases}$$

which leads to

$$U(P)=U[\rho(x)]=-\int \rho(x)\ln[L\rho(x)]dx \ . \tag{5}$$

The $U(P)$ of Eq. (4) is identical to the so-called Shannon information measure (or Shannon's *information entropy*).[10] In the continuous case $U(P)$ is similar (but not identical) to Shannon's information entropy, which is defined simply as

$$S_I(p)\equiv-\int \rho(x)\ln[\rho(x)]dx \ . \tag{6}$$

This formula is the one most used and the choice of P^0 leading to it is sometimes called[11] "the first principle of statistical mechanics": equal "*a priori* probability" is assigned to equal volumes of phase space. (Its quantum version assigns it instead to all quantum states of a complete orthogonal set.)

The last element we need is to specify the way to choose distributions $\{p_i\}$ or $\rho(x)$ once the limited amount of information about our experiment is given. The fact that $\{p_i\}$ or $\rho(x)$ is a probability assignment implies that they are non-negative and normalized, i.e.,

$$\sum_{i=1}^{n} p_i=1 \quad\text{or}\quad \int d\Gamma\,\rho(\Gamma)=1 \ . \tag{7}$$

Let the known information be of the form of the expectation values of k dynamical quantities $\langle F_k \rangle = f_k$, $k=1,\ldots,m<n$:

$$\langle F_k \rangle = \sum_{i=1}^{n} F_k(x_i)p_i \quad \text{or} \quad \langle F_k \rangle = \int d\Gamma\, F_k(\Gamma)\rho(\Gamma) \ . \tag{8}$$

We are looking for $\{p_i\}$ or $\rho(x)$ which tells us "the truth, the whole truth" about our experiment, i.e., reproduces known information. It still leaves, however, a great deal of freedom so we demand also that our distributions tell us "nothing but the truth," which means that they should convey the least information, i.e., to be those with the maximum missing information. This leads to our "second (and final) principle": we are going to assign to our system such a distribution $\{p_i\}$ [or $\rho(x)$] which reproduces all available information and has the maximum missing information.[11]

We have to determine then the maximum of $U(P)$ (or S_I) as given by Eqs. (4) and (6) subject to conditions (7) and (8). We can take them into account by introducing Lagrange multipliers $-(1+\Omega)$ and λ_i, $i=1,\ldots,k$, and by varying U with respect to p_i or ρ we find as a solution

$$R = \left[\begin{matrix}\{P_i\}\\ \rho(x)\end{matrix}\right] = \exp\left[\Omega - \sum_{k=1}^{m} \lambda_k F_k\right] \ . \tag{9}$$

The R is thus expressed in terms of Lagrange multipliers, which in turn are to be determined from the constraints. Thus, in the discrete case we have

$$\Omega = -\ln \sum_{i=1}^{n} \exp\left[-\sum_{k=1}^{m} \lambda_k F_k(x_i)\right] \equiv -\ln Z \ , \tag{10}$$

$$\begin{aligned} f_k = \langle F_k \rangle &= \frac{1}{Z}\sum_{i=1}^{n} F_k(x_i)\exp\left[-\sum_{k=1}^{m}\lambda_k F_k(x_i)\right] \\ &= \frac{\partial\Omega(\lambda_1,\ldots,\lambda_m)}{\partial\lambda_k} \ , \end{aligned} \tag{11}$$

and similar expressions (with obvious replacement $\sum_{i=1}^{n} \rightarrow \int d\Gamma$) for the continuous case.

This is the answer to our question. The result represented by Eq. (9) provides the most trivial and model-independent estimate of a distribution function of interest which is consistent with imposed constraints (i.e., which uses only the available information and nothing more). It now should be confronted with experimental data. A good agreement means that there is no more information left in the data. Otherwise there is still some unexplored information left. It thus can be used to determine new additional constraints (i.e., assumptions of the model) and, by repeating the whole procedure again, to obtain a new, more refined distribution (to be again confronted with experimental data).

Notice that although Eq. (9) resembles closely formulas used extensively in all thermodynamical-statistical models,[4] the resulting picture (when applied to multiple hadronic production phenomenology) differs substantially in that information (or information entropy), being a general mathematical quantity, can be used without referring to any kind of thermodynamical equilibrium. It applies thus to essentially all experimental situations without limitations on energies or multiplicities as we shall see later on. This is one of the examples of the model independence of our results mentioned before in Sec. I.

The information-theory (or specifically information entropy) approach has been used in multiple hadronic production only sporadically (although from the very beginning[12]) and without much reverberation. Nevertheless, it was this method which showed[13] that the minimal dynamical input sufficient for a highly successful description of virtually all relevant inclusive data (in the CERN ISR energy range of this period) consists of (i) limitation of the transverse momenta of secondaries, and (ii) limitation of the energy W available for production of secondaries (i.e., existence of leading particles or inelasticity of the reaction $K=W/\sqrt{s}$ being smaller than 1, in fact, of the order of 0.5; $\sqrt{s}$ is c.m. energy of reaction). As all successful phenomenological models at that time had both conditions built in (explicitly or implicitly), no wonder that they were simply bound to provide essentially the same results (in terms of different parameters, of course).

Similarly, knowledge of only the mean multiplicity $\langle n \rangle$ results in the geometrical (or Bose-Einstein) multiparticle distribution $P(n)$ as the most probable one, but additional allowance for k independent, equally strongly emitting sources leads immediately to the so-called negative-binomial form of $P(n)$ (Ref. 14) (which has recently emerged as a successful and fashionable description of experimental data).[15] It is interesting to note here that should we add the *indistinguishability* of the particles as additional information we would obtain the Poisson rather then geometrical distribution $P(n)$ [roughly speaking, because of $L \rightarrow n!L$ in Eq. (5) now]. Therefore this *additional information* converts the resultant multiparticle distribution from the most broad to the most narrow one [in the sense that the behavior of the dispersion $D=(\langle n^2\rangle - \langle n \rangle^2)^{1/2}$ changes from $D \approx \langle n \rangle$ for the Bose-Einstein distribution to $D \approx \sqrt{\langle n \rangle}$ for the Poisson one].

The information-entropy approach has also been used in investigations of asymptotic behavior of multiparticle production processes[16] and of the information content of multiparticle production models.[17] It served as a guiding principle in formulating different hypotheses of scaling behavior for single-particle distributions.[18] The so-called mutual information[10] has been proposed as a particularly suitable tool for investigations of forward-backward correlations in multiparticle processes.[19] The information-theory approach was also recently shown to be especially suitable for the final processing of raw experimental data.[20]

III. THE LEAST BIASED SINGLE-PARTICLE INCLUSIVE DISTRIBUTIONS

In what follows we shall concentrate on the discussion of the least biased (or most plausible) single-particle distributions resulting from the information-theory approach.[21]

Let us start with semi-inclusive normalized rapidity distribution

$$f_N(y)=\frac{1}{N}\frac{dN}{dy}=\frac{1}{\sigma_N N}\int d^2p_T\left[\frac{E\,d\sigma}{d^3p}\right]_{y\ \text{fixed}}$$

for an event in which a fireball of mass M decays (in its rest frame) into exactly N secondaries with strongly damped transverse momenta P_T [which is represented by fixed average transverse mass $\mu_T=(\mu^2+\langle p_T\rangle^2)^{1/2}$]. It is tacitly assumed that we know the distribution of M and $P(N)$ from elsewhere. We shall, in what follows, consider always only one kind of produced hadron (namely, pions of mass μ) and in this example $\langle p_t\rangle$ represents their mean transverse momentum. Effectively we are then dealing with a one-dimensional decay represented fully by the distribution $f_N(y)=f_N(y;M,\mu_T)$.

Notice that $f_N(y)$ defined in such a way is nothing else but the probability density to find a single particle (the only one traced down in full detail by the experimental apparatus) at rapidity y and as such it falls into the category of continuous distributions discussed in Sec. II [cf. Eqs. (5) and (6)]. Taking now as the prior probability assignment $f_N^0(y)=\text{const}$ in the allowed rapidity space $y\in[-Y_M,Y_M]$ one gets, for the relative information,

$$I_R(y)=\int_{-y_M}^{Y_M}dy\,f_N(y)\ln[f_N(y)]\ . \tag{12}$$

According to the procedure outlined in the previous section, we are now looking for the distribution $f_N(y)$ which maximizes I_R subject to constraints of normalization and energy conservation:

$$\begin{gathered}\int_{-Y_M}^{Y_M}dy\,f_N(y)=1\ ,\\ N\int_{-Y_M}^{Y_M}dy\,Ef_N(y)=N\mu_T\int_{-Y_M}^{Y_M}dy\,\cosh y\,f_N(y)=M\ .\end{gathered} \tag{13}$$

In the rest frame of the fireball $f_N(y)=f_N(-y)$, therefore the momentum-conservation constraint is satisfied automatically (i.e., it leads to the vanishing of corresponding Lagrange multiplier). Notice that our constraints represent only conservation laws, not dynamics. The latter is supposed to govern the production of our fireballs [i.e., their mass spectrum $\tau(M)$] and its decay pattern [i.e., the multiplicity distribution $P(N)$ for a given mass M]; in other words, it enters into the interrelations between our external parameters $\{M,N,\mu_T\}$ and it settles then the initial conditions for the constraints given by Eqs. (13).

Following the methods outlined in Sec. II we get as most probable the following distribution:

$$f_N(y)=\frac{1}{Z}\exp(-\beta\mu_T\cosh y)\ , \tag{14}$$

where

$$Z=Z(M,N,\mu_T)=\int_{-Y_M}^{Y_M}dy\,\exp(-\beta\mu_T\cosh y) \tag{15}$$

is the normalization and $\beta=\beta(M;N,\mu_T)$ is the Lagrange multiplier to be obtained from conditions (13):

$$\int_0^{Y_M}dy\left[\cosh y-\frac{M}{\mu_T N}\right]\exp(-\beta\mu_T\cosh y)=0\ , \tag{16}$$

here $Y_M=Y_M(M;N,\mu_T)$ is the maximal rapidity available:

$$Y_M=\ln\left\{\frac{M'}{2\mu_T}\left[1+\left[1-\frac{4\mu_T^2}{M'^2}\right]^{1/2}\right]\right\}\ ,\qquad M'=M-(N-2)\mu_T\ . \tag{17}$$

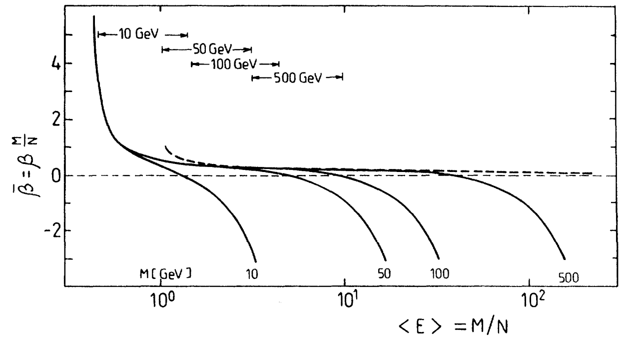

FIG. 1. The parameter $\bar\beta=\beta(M/N)$ in one-dimensional case [Eqs. (14), (16), and (19)] as a function of the mean available energy $\langle E\rangle=M/N$ for different energies M; $\mu_T=0.4$ GeV. The dashed curve corresponds to $\bar\beta$ being a solution of the approximate Eq. (22). The regions of $\langle E\rangle$ corresponding to the $N_\pm=\bar N(M)\pm\sigma$ as given in Table I are also indicated.

Notice that for $M'\gg 2\mu_T$, i.e., for $N\ll N_{\max}=M/\mu_T$, which is usually the case, $Y_M\simeq\ln(M/\mu_T)=\ln N_{\max}$ and does not depend on N. This case was already investigated in Ref. 21 to some extent. In general Eq. (16) has to be solved numerically for β once M, N, and μ_T are given.

The main features of distribution $f_N(y)\simeq f_N(y;\mu_T,N_{\max}/N)$ can be summarized as follows (cf. Figs. 1 and 2).

(i) Contrary to all statistical models exploiting Eq. (14) with $1/\beta$ interpreted as a kind of "temperature" with tacit understanding that $\beta>0$ always, here β can assume any value. In fact $\beta>0$ only if

$$\begin{aligned}N>N_0&=N_0(M,N,\mu_T)\\ &=\frac{MY_M}{\mu_T\sinh Y_M}\underset{\text{large}N_{\max}}{\longrightarrow}2\ln\frac{M}{\mu_T}\\ &=2\ln(N_{\max})\ .\end{aligned} \tag{18}$$

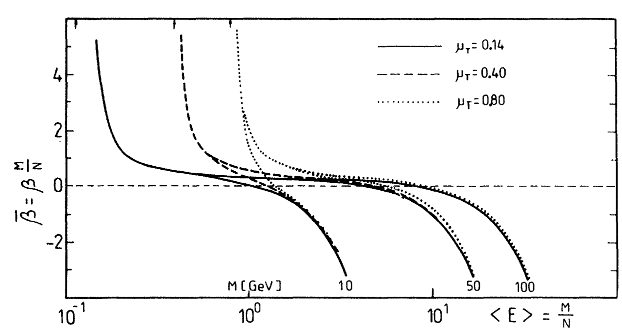

FIG. 2. Example of the sensitivity of $\bar\beta$ as defined in Fig. 1 on different choices of μ_T for different energies. Arrows indicate values of $\langle E\rangle$ $(=\mu_T)$ for which $\bar\beta\to+\infty$.

(ii) It means that for $N = N_0$ we have $\beta \equiv 0$, i.e., $f_N(y)$ is then *strictly flat.* Therefore *exact* Feynman scaling (or plateau in the rapidity distribution) is realized *only* if the mean multiplicity of the reaction $N = \bar{N}(M)$ follows rule (18) [i.e., $N = N_0(M)$] as a function of the energy M. In fact this seems to be the case at CERN I.S.R. energies[13] (at which the concept of Feynman scaling was conceived). However, data at present energy range seem to favor much faster growth of the multiplicity with energy $\bar{N} \simeq \ln^2 M$ or even $\bar{N} \simeq M^{0.5}$.[22] If applied to a single cluster or fireball, it would then mean that Feynman scaling was only a transient phenomenon seen in an energy range for which it just happened that the $\bar{N} \simeq \ln M$ law was a locally valid numerical representation of data.

(iii) One can also write $f_N(y)$ in the form

$$f_N(y) = \frac{1}{Z} \exp\left[-\bar{\beta} \frac{\mu_T \cosh y}{\langle E \rangle} \right], \tag{19}$$

where $\langle E \rangle = M/N$ is the mean energy available to the particle in a given event and $\bar{\beta} = \beta(M/N)$. We have found such a parametrization particularly useful because, as can be seen in Fig. 1, for large enough masses M (in fact, for large enough $N_{\max} = M/\mu_T$, cf. Fig. 2) in the usually explored range of multiplicities $N = N_{\pm} = \bar{N}(M) \pm \sigma$, the resulting $\bar{\beta}$ is (almost) constant as a function of $\langle E \rangle$ (i.e., $\beta \approx N/M$ here) which suggests a kind of (approximate) scaling in the variable $z = E/\langle E \rangle$, $E = \mu_T \cosh y$, in this region.

(iv) For $N < N_0$, β is negative, $\beta < 0$. In fact $\beta \underset{N \to 2}{\longrightarrow} -\infty$ and in this limit $f_N(y)$ takes a double δ-function shape:

$$f_{N=2}(y) = \tfrac{1}{2}[\delta(y - y_M) + \delta(y + y_M)] . \tag{20}$$

(v) For $N_0 < N \ll N_{\max}$ and for large energies M, one can replace Y_M in Eq. (16) by infinity and perform the integration over rapidity to obtain

$$\int_0^\infty dy \left[\cosh y - \frac{M}{\mu_T N} \right] \exp(-\beta \mu_T \cosh y) = K_1(\beta \mu_T) - \frac{M}{\mu_T N} K_0(\beta \mu_T) = 0 . \tag{21}$$

In the limit of $z = \beta \mu_T \ll 1$ (i.e., $\beta \ll 1/\mu_T$) where $K_0(z) \simeq -\ln z$ and $K_1(z) \simeq 1/z$, we have then the following simple transcendental equation for β:

$$\beta \ln(\beta \mu_T) = \frac{N}{M} \quad \text{or} \quad \bar{\beta} \ln \left[\bar{\beta} \frac{N}{N_{\max}} \right] = -1 . \tag{22}$$

As can be seen in Fig. 1 its solution follows quite nicely the exact one, especially for larger masses M. However, its fails completely for $M < 50$ GeV and for large values of $\langle E \rangle$ where the exact $\bar{\beta}$ becomes negative. If instead $z = \beta \mu_T > 1$, which is realized for large multiplicities N (actually for $N > 0.2 N_{\max}$), where

$$K_\nu(z) = \left[\frac{\pi}{2z} \right]^{1/2} e^{-z} \left[\sum_{l=0}^{L-1} \frac{\Gamma(\frac{1}{2}+\nu+1)}{l!\,\Gamma(\frac{1}{2}+\nu-1)} \frac{1}{(2z)^l} + O(|z|^{-L}) \right] \tag{23}$$

we get (to leading order in $1/z$)

$$\beta(N, N_{\max}) = \frac{0.375}{\mu_T} \frac{5(N/N_{\max}) - 1}{1 - N/N_{\max}} \underset{N \to N_{\max}}{\longrightarrow} \frac{3}{2\mu_T} \frac{1}{1 - N/N_{\max}} . \tag{24}$$

Notice that $\beta(N = N_{\max}) = +\infty$ which means that in this limit (corresponding to $\langle E \rangle = \mu_T$) $f_N(y) \to \delta(y)$, i.e., all particles tend to be maximally concentrated in the middle of the rapidity plot (cf. Fig. 2).

Some of these features were already noticed and used in phenomenological analyses of data[23–25] but always with an implicit assumption that a "partition temperature" (as $1/\beta$ is usually called) is necessarily positive. That is justified, as seen in Fig. 1 (cf. also Table I), only when one works in a rather narrow band of multiplicities N centered on the mean multiplicity $N = \bar{N}(M)$ or when the average available energy per particle $\langle E \rangle = M/N$ is small enough. In many situations, however, especially where the large energy M has to be distributed among a very small number of secondaries ($N < N_0$), the resulting

TABLE I. Values of $\beta_{\pm}$, $\bar{\beta}_{\pm}$, and $\alpha_{\pm}$ in one- (1D) and three-dimensional (3D) cases for different energies M and for $N_{\pm} = \bar{N}(M) \pm \sigma$ with $N(M) = 4.5 M^{0.5}$ and $\sigma = \frac{1}{2}\bar{N}(M)$ (which corresponds to the maximal possible width of the multiplicity distribution); $\mu_T = 0.4$ GeV.

M (GeV)	10		50		100		500	
$N_{\pm} = \bar{N} \pm \sigma$	14.2±7.1		31.8±15.9		45.0±22.5		100.6±50.3	
$\langle E_- \rangle = \frac{M}{N_-}$	1.40		3.14		4.44		9.94	
$\langle E_+ \rangle = \frac{M}{N_+}$	0.47		1.05		1.48		3.31	
	1D	3D	1D	3D	1D	3D	1D	3D
β_-	−0.11	0.03	0.06	0.06	0.05	0.06	0.02	0.02
β_+	5.74	6.60	0.52	0.48	0.27	0.27	0.09	0.09
$\bar{\beta}_-$	−0.15	0.04	0.20	0.20	0.20	0.25	0.20	0.20
$\bar{\beta}_+$	2.70	3.10	0.55	0.50	0.40	0.40	0.30	0.30
α_-		5.20		5.30		5.40		5.50
α_+		−1.5		4.70		5.05		5.35

β is completely different from the naive expectations;[23–25] in fact, in the above-mentioned situation β becomes negative. And precisely this is the region where the standard models[23–25] always had trouble in adequately describing the experimental data.

Let us proceed now to a more general semi-inclusive distribution, namely, to

$$\rho_N(y,p_T)=\frac{1}{\sigma N}\frac{d\sigma}{(dy\,d^2p_T)} .$$

Assuming azimuthal symmetry one needs in this case an additional constraint to fix the p_T dependence. For the presentation that will follow we shall use as a constraint the value of a mean transverse mass $\mu_T=(\mu^2+p_T^2)^{1/2}$:

$$\langle\mu_T\rangle=\langle\mu_T(M,N)\rangle=\int d\Gamma\,\mu_T\rho_N(y,p_T) . \tag{25}$$

Here $(p_Tdp_T=\mu_Td\mu_T)$

$$\int d\Gamma\equiv 2\pi\int_{-Y_M}^{Y_M}dy\int_{\mu}^{\mu_{TM}(y)}d\mu_T\mu_T ,$$
$$\mu_{TM}(y)=\frac{M'}{2\cosh y},\quad M'=M-(N-2)\mu . \tag{26}$$

Such a choice allows us to perform some of the integrations analytically.[26] Taking again the prior probability assignment $\rho^0(y,\mu_T)$ constant in the phase space Γ one gets, for a relative information formula in this case,

$$I_R(\rho)=\int d\Gamma\,\rho_N(y,\mu_T)\ln[\rho_N(y,\mu_T)] . \tag{27}$$

The distribution ρ_N which maximizes I_R subject to the constraint (25) as well as to normalization and energy-conservation constraints:

$$\int d\Gamma\,\rho_N(\Gamma)=1 ,$$
$$N\int d\Gamma\,E\rho_N(\Gamma)=N\int d\Gamma\,\mu_T\cosh y\,\rho_N(y,\mu_T)=M \tag{28}$$

is given by

$$\rho_N(y,\mu_T)=\frac{1}{Z}\exp(-\alpha\mu_T-\beta\mu_T\cosh y) \tag{29}$$

with Z, α, and β being, respectively, the normalization and the Lagrange multipliers corresponding to $\langle\mu_T\rangle$ and M. They are given as solutions of the equations

$$2\pi\int_{-Y_M}^{Y_M}dy\begin{Bmatrix}\cosh y\\ 1\end{Bmatrix}F_N(y;\alpha,\beta)=Z\begin{Bmatrix}M\\ \langle\mu_T\rangle\end{Bmatrix} , \tag{30}$$

where

$$F_N(y;\alpha,\beta)=\int_{\mu}^{\mu_{TM}(y)}d\mu_T\mu_T^2\rho_N(y,\mu_T)$$
$$=-\frac{e^{-a\mu_T}}{a^3}[1+(1+a\mu_T)^2]\Big|_{\mu}^{\mu_{TM}(y)} ,$$
$$a=a(y;\alpha,\beta)=\alpha+\beta\cosh y . \tag{31}$$

The normalization is given now by

$$Z=\int d\Gamma\exp(-\alpha\mu_T-\beta\mu_T\cosh y) \tag{32}$$

and

$$H_N(y;\alpha,\beta)=2\pi\int_{\mu}^{\mu_{TM}}d\mu_T\mu_T\rho_N(y,\mu_T)$$
$$=-2\pi\frac{e^{-a\mu_T}}{Za^2}(1+a\mu_T)\Big|_{\mu}^{\mu_{TM}(y)} \tag{33}$$

represents the single-particle rapidity distribution in the three-dimensional case (integrated over transverse momenta p_T).

The results are summarized in Figs. 3–5. In Fig. 3 the solutions for $(\alpha,\bar\beta)$ are given for different energies and $\langle\mu_T\rangle=0.4$ GeV (to compare with $\mu_T=0.4$ GeV in one-dimensional case presented in Figs. 1 and 2). Notice that essentially, except for unavoidable differences in details, the general features are the same now as before for $f_N(y)$. The main difference is that now the values of N_0 below which β becomes negative depend also on α; with smaller (fixed) α the corresponding N_0 gets also smaller and for $\alpha=0$ parameter β is always positive, cf. Fig. 4. That is because of the opening of the p_T part of phase space. We can now conserve energy even for $N=2$ and $y=0$ just by sending off particles to the maximal p_T regions at this rapidity. Also the (integrated) single-particle distribution $H_N(y)$, Eq. (33), behaves for $N<N_0$ slightly differently than the corresponding strictly one-

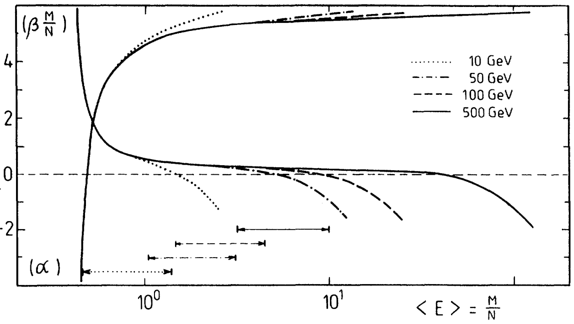

FIG. 3. The parameters $\bar\beta=\beta(M/N)$ and α in three-dimensional case [Eqs. (29) and (30)] as a function of the $\langle E\rangle=M/N$ for different M; $\langle\mu_T\rangle=0.4$ GeV. Indicated are regions of $\langle E\rangle$ corresponding to $N_\pm=\bar N(M)\pm\sigma$ as given in Table I.

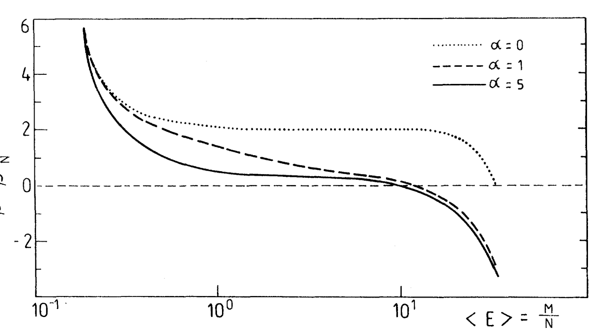

FIG. 4. The parameter $\bar\beta$ in three-dimensional case for different choices of α (kept now constant) as a function of $\langle E\rangle$; $\mu=0.14$ GeV.

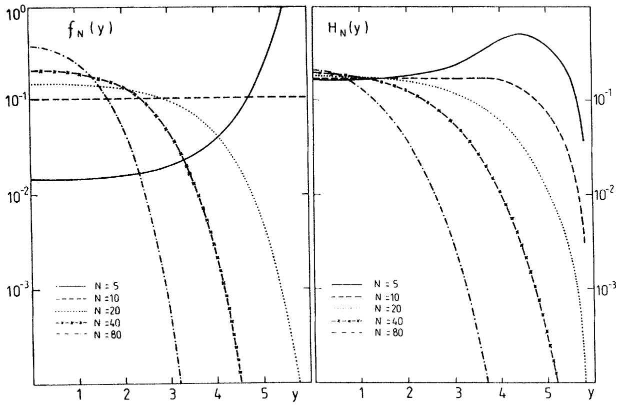

FIG. 5. The examples of $f_N(y)$ [Eq. (14)] and $H_N(y)$ [Eq. (33)] for $M=50$ GeV and different N; $\mu_T=\langle\mu_T\rangle=0.4$ GeV.

dimensional distribution $f_N(y)$, cf. Fig. 5. Because of the closing of the p_T part of the phase space, i.e., because $\mu_{TM}(y=\pm y_M)=\mu$, $H(y=\pm Y_m)=0$ now whereas f_N is maximal there. Notice, that in our approach the growth of the $dN/dy|_{y=0}$ with energy is essentially given by $dN/dy|_{y=0}=\bar{N}(M)H(y=0;N=\bar{N}(M))\simeq\bar{N}(M)/\ln M$.

A characteristic feature of $\rho_N(y,\mu_T)$ for small mean available energies $\langle E\rangle=M/N$ is the fact that in this region (in which parameter $\bar{\beta}$ becomes large), the $\langle\mu_T\rangle$ constraint forces α to decrease and ultimately to become negative, cf. Fig. 3. However, in the region of $N=\bar{N}\pm\sigma$ both α and $\bar{\beta}$ are essentially positive and only slowly varying as functions of $\langle E\rangle$. This can be regarded as *a posteriori* justification of all more or less *ad hoc* applications of formula (29) (Refs. 23–25) except for both very small and very large values of $\langle E\rangle$. In these regions physical justifications provided usually for Eq. (29) are already invalid whereas the formula itself, i.e., as derived here, is still correct and can be used. But one has to realize that it should be regarded only as the best possible model-independent *parametrization* of data based on conservation laws rather than dynamics. Dynamics enters only, as was already stressed, via the actual value and possible energy dependence of the multiplicity N and via the possible energy and multiplicity dependence of $\langle\mu_T\rangle$. But once this information is provided the distribution $\rho_N(y,p_T)$ *has to be* of the form of Eq. (29).

IV. SUMMARY AND CONCLUSIONS

Using the tools of the information theory we have derived the most probable (or least biased) single-particle distributions for one- and three-dimensional cases [Eqs. (19) and (29), respectively] for a decay of an isolated fireball of a given mass M producing exactly N secondaries with restricted transverse momenta or masses $\langle\mu_T\rangle$. In our approach M, N, and $\langle\mu_T\rangle$ [with the understanding that essentially $N=N(M)$ and $\langle\mu_T\rangle=\langle\mu_T(M,N)\rangle$] are the only external, dynamical parameters. Together with the energy conservation constraint they *fix uniquely* the corresponding Lagrange multipliers $\alpha=\alpha(M,N,\langle\mu_T\rangle)$ and $\beta=\beta(M,N,\langle\mu_T\rangle)$. We would like to stress particularly this feature because in all phenomenological models,[23–25] which use essentially the same distributions, α and β were always regarded as free parameters chosen to provide the best fit to data. That was because although formally the same, the corresponding distributions were obtained from completely different premises. In our case the only way to change the final distributions would be through changes in the mutual relations between parameters M, N, and $\langle\mu_T\rangle$, i.e., through changes in our dynamical input, not in the parametrization itself.

As presented in Figs. 1–5 our parametrization applies to the whole range of multiplicities and $\langle\mu_T\rangle$ for any M, including both very small and very large values of the mean energy available to the produced secondary, $\langle E\rangle$, i.e., including the region where all above-mentioned models are either not applicable or have serious problems. It is thus particularly useful in all applications demanding a maximal objectivity and neutrality of presentation. In the Appendix we have thus enclosed a possible scheme for the Monte Carlo event generator based on our parametrization.

The method can be rather easily extended to cover also the multiparticle distributions $P(N)$. We shall cover this subject here repeating only what was already said in Sec. II, that one can rather easily identify the minimal information necessary to obtain $P(N)$ ranging from a geometrical to Poissonian distributions. The extension to the multifold distributions, although not so straightforward, is also possible and very promising.[19]

ACKNOWLEDGMENTS

We are grateful to Professor R. Weiner for the discussions and hospitality extended to us at Marburg University where part of this work was done. This work was supported in part by the Heraeus-Stiftung, Hanau.

APPENDIX: ALGORITHM FOR MONTE CARLO EVENT GENERATOR

We shall provide here the scheme of the Monte Carlo event generator algorithm which can be used to produce single-particle distributions $\{y_i,\mu_{Ti}\}$ according to formulas (29)–(30). For a given mass M and multiplicity N, once α and β are calculated from Eq. (30), we do the following:

(1) Choose $y\in[-Y_M,Y_M]$ according to $H_N(y;\alpha,\beta)$, Eq. (33) (the sign is chosen randomly from a uniform distribution).

(2) Choose $\mu_T\in[\mu,\mu_{TM}]$ according to

$$G_N(\mu_T)=\frac{\rho_N(y,\mu_T)}{H_N(y;\alpha,\beta)}\ . \tag{A1}$$

Repeating (1) and (2) N times one gets a first raw distribution $\{y_i,\mu_{Ti}\}$, $i=1,\ldots,N$. To correct for possible energy-momentum nonconservations resulted from the selection procedure:

(3) Form

$$\sum_{i=1}^{N} \mu_{Ti}\cosh y_i = E, \quad \sum_{i=1}^{N} \mu_{Ti}\sinh y_i = P \ . \tag{A2}$$

If $E \neq M$ and $P \neq 0$:

(4) Look for rapidity shift Δ and scale factor δ such that

$$\delta \sum_{i=1}^{N} \mu_{Ti}\cosh(y_i+\Delta) = M \ , \quad \sum_{i=1}^{N} \mu_{Ti}\sinh(y_i+\Delta) = 0 \ . \tag{A3}$$

From (A3) and (A2) one gets

$$\Delta = -\frac{1}{2}\ln\frac{E+P}{E-P}, \quad \delta = \frac{M}{\sqrt{(E^2-P^2)}} \ . \tag{A4}$$

Therefore, the final distribution is given by $\{y_i+\Delta; \delta\mu_{Ti}\}$, $i=1,\ldots,N$.

[1] A. Capella, U. Sukhatme, C. -I. Tan, and J. Trân Thanh Vân, Phys. Lett. **81B**, 68 (1979); A. Capella and J. Trân Thanh Vân, *ibid.* **114B**, 450 (1982), Z. Phys. C **10**, 249 (1981); **18**, 85 (1983); **38**, 177 (1988); A. B. Kaidalov and K. A. Martirosyan, Yad. Fiz. **39**, 1545 (1984) [Sov. J. Nucl. Phys. **39**, 979 (1984)]; **40**, 211 (1984) [**40**, 135 (1984)]; F. W. Bopp, P. Aurenche, and J. Ranft, Phys. Rev. D **33**, 1867 (1986).

[2] B. Anderson, G. Gustafson, and B. Nilson-Almqvist, Nucl. Phys. **B281**, 289 (1987) [cf. also *Hadronic Multiparticle Production,* edited by P. Carruthers and J. Rafelski, Advanced Series on Directions in High Energy Physics, Vol. 2 (World Scientific, Singapore, 1988), pp. 376 and 428, and references therein].

[3] K. Werner, Phys. Rev. D **39**, 780 (1989).

[4] Compare, for example, E. L. Feinberg, Usp. Fiz. Nauk. **139**, 3 (1983) [Sov. Phys. Usp. **26**, 1 (1983)], and in *Hadronic Matter under Extreme Conditions,* edited by G. M. Zinovjev and W. P. Shelest (Naukova Dumka, Kiev, 1986), and references therein (cf. also Refs. 23–25 below).

[5] G. N. Fowler, F. S. Navarra, M. Plümer, A. Vourdas, R. M. Weiner, and G. Wilk, Phys. Rev. C **40**, 1219 (1989).

[6] R. A. M. Nazareth, N. Prado, and T. Kodama, Phys. Rev. D **40**, 2861 (1989); W. Ochs, Z. Phys. C **15**, 227 (1982); **23**, 131 (1984).

[7] B. Anderson, P. Dahlqvist, and G. Gustafson, Phys. Lett. B **214**, 604 (1988); Nucl. Phys. **B328**, 76 (1989).

[8] H. Haken, *Information and Self-Organization* (Springer-Verlag, Berlin, 1988); J. Skilling, Nature (London) **309**, 748 (1984).

[9] R. Balian, M. Veneroni, and N. Balazs, Europhys. Lett. **1**, 1 (1986); A. Hobson, *Concepts in Statistical Mechanics* (Gordon and Breach, New York, 1971).

[10] C. E. Shannon, Bell System Tech. J. **27**, 379 (1948); **27**, 623 (1948); cf. also S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959); R. J. McEliece, *The Theory of Information and Coding* (Addison-Wesley, Reading, MA, 1977).

[11] The classical introduction to the use of the information theory as the basis of all statistical mechanics is provided by E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957); cf. also A. Katz, *Principles of Statistical Mechanics—The Information Theory Approach* (Freeman, San Francisco, 1967). More recent reviews can be found in W. T. Grandy, Jr., Phys. Rep. **62**, 175 (1980); F. Schlögel, *ibid.* **62**, 267 (1980); R. Balian and N. L. Balazs, Ann. Phys. (N.Y.) **179**, 97 (1987), and references therein.

[12] Actually the first work of this type we could trace back did not even mention the notion of information but still made use of its formalism, cf. Y. Takahashi, Nuovo Cimento **4**, 531 (1956).

[13] Y.-A. Chao, Nucl. Phys. **B40**, 475 (1972).

[14] M. I. Gorenstein, Yad. Fiz. **31**, 1630 (1980) [Sov. J. Nucl. Phys. 31, 845 (1980)]; B. Carazza and Gandolfi, Lett. Nuovo Cimento **15**, 553 (1976).

[15] A. Giovannini and L. Van Hove, Z. Phys. C **30**, 391 (1986); Acta Phys. Polon. **B19**, 917 (1988); **B19**, 931 (1988). The other two parameters form of $P(n)$ is truncated Gaussian which results from the assumed knowledge of both the mean multiplicity $\langle n \rangle$ and the dispersion $D = (\langle n^2 \rangle - \langle n \rangle^2)^{1/2}$. It was even checked (with success) against experimental data a long time ago by G. W. Parry and P. Rotelli, J. Phys. A **6**, 1565 (1973).

[16] P. Rotelli, Phys. Rev. **182**, 1622 (1969).

[17] G. Camisassa, P. Frè, and L. Sertorio, Nuovo Cimento **B30**, 364 (1975).

[18] I. Kita, R. Nakamura, R. Nakamija, and J. Yotsuyanagi, Prog. Theor. Phys. **63**, 919 (1980); W. Ernst and I. Schmitt, Nuovo Cimento **A41**, 217 (1977); **46**, 203 (1978).

[19] P. Carruthers and C. C. Shih, Phys. Rev. Lett. **62**, 2073 (1989); P. Carruthers, Int. J. Mod. Phys. A **4**, 5587 (1989).

[20] UA5 Collaboration, R. E. Ansorge *et al.*, Z. Phys. C **43**, 357 (1989); cf. also C. Fuglesang, Nucl. Instrum. Methods **A278**, 765 (1989).

[21] First approximate analytical formula for the one-dimensional case of rapidity distribution was presented by F. Takagi and T. Tsukamoto, Phys. Rev. D **38**, 2288 (1988) [cf. also G. Wilk and Z. Wlodarczyk, in *Frontiers in Particle Physics,* proceedings of the XII Warsaw Symposium on Elementary Particle Physics, Kazimierz, Poland, 1989, edited by Z. Adjuk (World Scientific, Singapore, 1990), p. 555]. Notice also that our semi-inclusive distributions differ slightly from those in Refs. 13 and 18 where only averaged events were considered (i.e., summed over all possible multiplicities N). In a sense we consider semi-inclusive single-particle distributions in exclusive events, i.e., events with fixed total multiplicity N.

[22] C. Geich-Gimbel, Int. J. Mod. Phys. A **4**, 1527 (1989).

[23] V. Girija and J. Hüfner, Phys. Lett. B **212**, 501 (1988); Y. Kurkara, J. Hüfner, and J. Aichelin, *ibid.* **205**, 549 (1988); Z. Phys. C **42**, 485 (1989).

[24] Y. T. Chou and C. N. Yang, Phys. Rev. D **32**, 1692 (1985); K. Watanabe and S. Tone, Phys. Lett. B **213**, 370 (1988); Phys. Rev. D **39**, 195 (1989), and references therein.

[25] T. F. Hoang, Phys. Rev. D **38**, 2739 (1988); S. Y. Lo and A. Schreiber, Int. J. Mod. Phys. A **3**, 1603 (1988), and references therein.

[26] It can also be argued that such a choice leads to a better representation of experimental data, cf. V. Blobel *et al.*, Nucl. Phys. **B69**, 454 (1974); M. J. Tannenbaum, Report No. BNL-43818 (unpublished).