INTERMITTENCY ANALYSIS OF CORRELATED DATA

By Barbara Wosiek

Institute of Nuclear Physics, Cracow*

(Received September 20, 1990)

In this report we describe the method of the analysis of the dependence of the factorial moments on the bin size in which the correlations between the moments computed for different bin sizes are taken into account. For large multiplicity nucleus-nucleus data inclusion of the correlations does not change the values of the slope parameter, but gives errors significantly reduced as compared to the case of fits with no correlations.

PACS numbers: 13.85.Hd

Recently there has been a great interest in studying fluctuations of particle density in small domains of phase space. It was found in a variety of processes that multiparticle production may show the feature of intermittency i.e. large fluctuations increasing as the size of the domain is decreased [1, 2]. Searches for intermittency have been performed by applying the method proposed by Białas and Peschanski [1]. It is based on measuring the scaled factorial moments of the multiplicity distribution in rapidity bins, defined as:

$$\overline{F^q(\delta)} = \frac{\overline{\langle n(n-1)\ldots(n-q+1)\rangle}}{\langle \bar{n}\rangle^q}, \tag{1}$$

where n is the number of particles in bin of length δ, the brackets denote averaging over the $\Delta Y/\delta$ bins into which the investigated rapidity interval of length ΔY has been divided and the bar denotes averaging over events in the sample. It was suggested [3] that one should correct for the rapidity dependence of the single-particle density distribution by changing the normalization in (1) as follows

$$\overline{F^q(\delta)} = \frac{\overline{\langle n(n-1)\ldots(n-q+1)\rangle}}{\langle \bar{n}^q\rangle}. \tag{2}$$

One defines intermittency as the power-law increase of the factorial moments with decreasing δ, $F \sim \delta^{-b}$, where the intermittency exponent, b, measures the strength of the effect.

* Address: Instytut Fizyki Jądrowej, Kawiory 26A, 30-055 Kraków, Poland.

The procedure to obtain F's at different δ is as follows. For a given event the total pseudorapidity interval ($\Delta\eta = 5$ for the analysis presented in this report) was first divided into bins of the largest size equal to one pseudorapidity unit. Next, the original $\Delta\eta$ interval was histogrammed into M ($5 < M \leqslant M_{\max}$) smaller bins of size $\delta = \Delta\eta/M$. In the present analysis the bin size was varied down to $\delta = 0.1$ corresponding to $M_{\max} = 50$. Thus, the same data were used to obtain the moments for different subdivisions and smaller bins are contained in larger ones. In effect the moments computed for various δ are strongly correlated. Indeed, our analysis confirms this fact as it is shown below. As a consequence the errors of average factorial moments on the $\bar{F}(\delta)$ vs δ plots do not have any significance. For example, one sees [4, 5] that the scattering of experimental points around the smooth curve is much smaller than expected from the quoted error bars.

In the previous analysis [4, 5] the correlations between the factorial moments computed at different δ were ignored and the best fits were obtained by minimizing the χ^2-function:

$$\chi^2 = \sum_{i=1}^{N} (\bar{F}^q_{i,\mathrm{exp}} - \bar{F}^q_{i,\mathrm{th}})^2/\sigma_i^2, \tag{3}$$

where $\bar{F}^q_{i,\mathrm{exp}}$ is the average factorial moment calculated according to (2), $\bar{F}^q_{i,\mathrm{th}} = c \cdot (\delta_i)^{-b}$ and N denotes the number of points included in the fit. As the result we obtained fit parameters with very large errors, especially for the slope parameter, b. In some cases the error exceeded 100% of the value of the slope. At the same time the fits were too "good" (χ^2 per degree of freedom was of the order 10^{-2}–10^{-1}). These problems were due to the fact that the procedure applied was not correct since the moments measured for different bin sizes are not independent. We made some attempts [5] in order to reduce the errors of the fitted parameters, but up to now the correlations between the factorial moments were not properly taken into account.

Obvious way to include the correlations is to compute the whole covariance matrix of various measurements:

$$C_{ij} = \frac{\overline{F_i^q F_j^q} - \bar{F}_i^q \cdot \bar{F}_j^q}{N-1}, \tag{4}$$

where indices i, and j label different bin sizes. To find the values of the parameters of a fitted function $\bar{F}^q_{\mathrm{th}}(c, b)$ we minimize [6]:

$$\chi^2 = \sum_i \sum_j (\bar{F}^q_{i,\mathrm{exp}} - \bar{F}^q_{i,\mathrm{th}})\,(C_{ij}^{-1})\,(\bar{F}^q_{j,\mathrm{exp}} - \bar{F}^q_{j,\mathrm{th}}) \tag{5}$$

with respect to c and b. It is obvious that, if measured F_i^q's are independent, the covariance matrix is diagonal and formula (5) reduces to (3). There are several advantages of including the correlations in the fits. First, we get sensible estimates of the errors of the parameters and what is more important we get a reliable estimate of the goodness of the fit. Moreover positive correlations lead to reduced errors as compared to the case with no correlations. On the other hand the parameters that minimize the correct χ^2-function (5) may differ from those that minimize the simplified one (3). A curious feature of the fits with strong correlations is that the best fit may lie below (or above) all the measured points rather than

passing above some and below others [7]. Such a behaviour would clearly be wrong for uncorrelated data. However, the reader may easily convince himself that this is correct in case of correlated data. Indeed, if one point deviates statistically from its exact mean value, the rest of them "must do the same" because of strong positive correlations. Hence, all data points lie below or above theoretical prediction. A more detailed discussion of this problem is given in the Appendix.

Let us check what are the effects of correlations for the dependence of the scaled factorial moments on δ in experimental data. We analysed two samples of events consisting of semi-central interactions of oxygen and sulfur primaries with Ag/Br nuclei at 200 GeV/n. The data used in the analysis are provided by the KLM Collaboration and the details about the experimental material can be found in [8]. For both samples of interactions the average multiplicities of produced particles are very large. In average about 270 (150) particles are produced within 5 units of pseudorapidity for sulfur (oxygen) projectile. For comparison a sample with smaller number of produced particles (in average 20 particles per event within $\Delta\eta = 5$) consisting of proton interactions with Ag/Br targets at 200 GeV was also analysed. The events statistics are correspondingly 124, 146 and 833 events for sulfur, oxygen and proton primaries.

Figures 1a and 1b illustrate the correlations between the moments measured for $^{16}O + AgBr$ data for two combinations of bin sizes. As expected, the correlations are strong and positive especially between moments computed for adjacent values of δ (Fig. 1a). They remain strong even for maximally different δ (Fig. 1b). The same conclusion can be

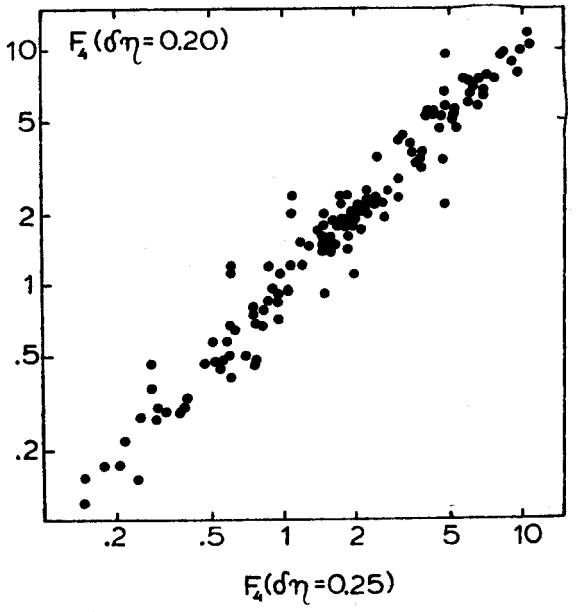

Fig. 1a. Scatter-plot of the 4-th factorial moments measured in individual events for two neighbouring bin sizes, $\delta\eta = 0.25$ and $\delta\eta = 0.20$

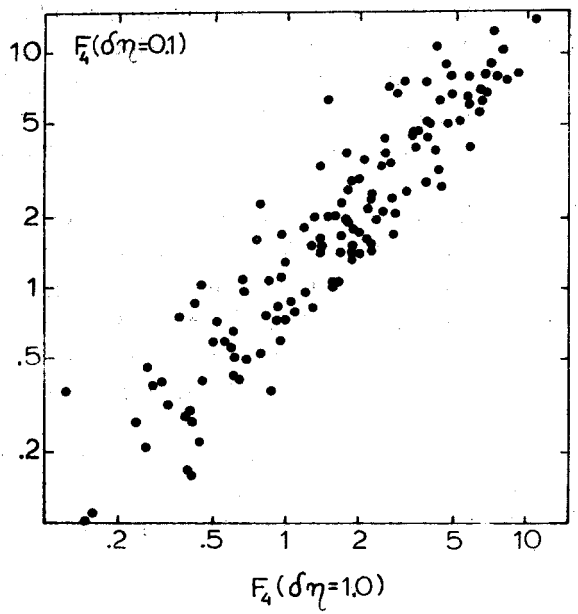

Fig. 1b. The same plot as in Fig. 1a, but for two extremal values of bin sizes, $\delta\eta = 1.0$ and $\delta\eta = 0.1$

TABLE I

Correlation coefficients, $\varrho_{ij} = C_{ij}/\sqrt{C_{ii} \cdot C_{jj}}$, calculated for the 4-th factorial moment for $^{16}O+AgBr$ collisions. Indices $i, j = 1, \ldots 11$ correspond to $\delta\eta$ varying from 1.0 to 0.1

j \ i	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.96	0.93	0.92
2	0.99	1.00	0.98	0.99	0.98	0.98	0.97	0.97	0.95	0.94	0.92
3	0.98	0.98	1.00	0.98	0.98	0.97	0.97	0.97	0.95	0.95	0.93
4	0.99	0.99	0.98	1.00	0.98	0.98	0.97	0.98	0.96	0.95	0.93
5	0.98	0.98	0.98	0.98	1.00	0.99	0.99	0.99	0.97	0.95	0.94
6	0.98	0.98	0.97	0.98	0.99	1.00	0.98	0.98	0.97	0.95	0.94
7	0.97	0.97	0.97	0.97	0.99	0.98	1.00	0.98	0.98	0.96	0.95
8	0,97	0.97	0.97	0.98	0.99	0.98	0.98	1.00	0.97	0.96	0.95
9	0.96	0.95	0.95	0.96	0.97	0.97	0.98	0.97	1.00	0.97	0.96
10	0.93	0.94	0.95	0.95	0.95	0.95	0.96	0.96	0.97	1.00	0.95
11	0.92	0.92	0.93	0.93	0.94	0.94	0.95	0.95	0.96	0.95	1.00

drawn from Table I where the correlation coefficients are listed. One can see that all values are positive and even for the most different δ's they are bigger than 90%. For the moments of smaller order, the distant correlations are still stronger exceeding 99% for the second factorial moment. Similar analysis performed for ^{32}S and p collisions with Ag/Br targets reveals following regularity. More complicated are the events (i.e. larger multiplicity per bin) the stronger are correlations. For the 4-th factorial moment the most distant correla-

tions exceed 98% for $^{32}S+AgBr$ collisions, while for $p+AgBr$ data we measured about 55%. We think that for low-multiplicities geometrical inclusion of progressively smaller bins does not necessarily cause strong correlations. It can also be seen from Figs 1 that widths of the moment distributions (projections of target diagrams on each of the axis) are very large.

In Fig. 2 we show examples of the best fits to the data obtained by minimizing correct χ^2-function, Eq. (5), for the three data samples. Note, that the individual errors are not very meaningful in case of strongly correlated measurements. In Table II the numerical values of the fit parameters and χ^2 per degree of freedom obtained from two fitting procedures are given for the dependence of the 4-th factorial moment on $\delta\eta$ for each data sample.

In Figs. 3, 4 and 5 the values of the parameters of the fit are shown as a function of the order of the factorial moment for both correct fitting procedure and simplified one. In

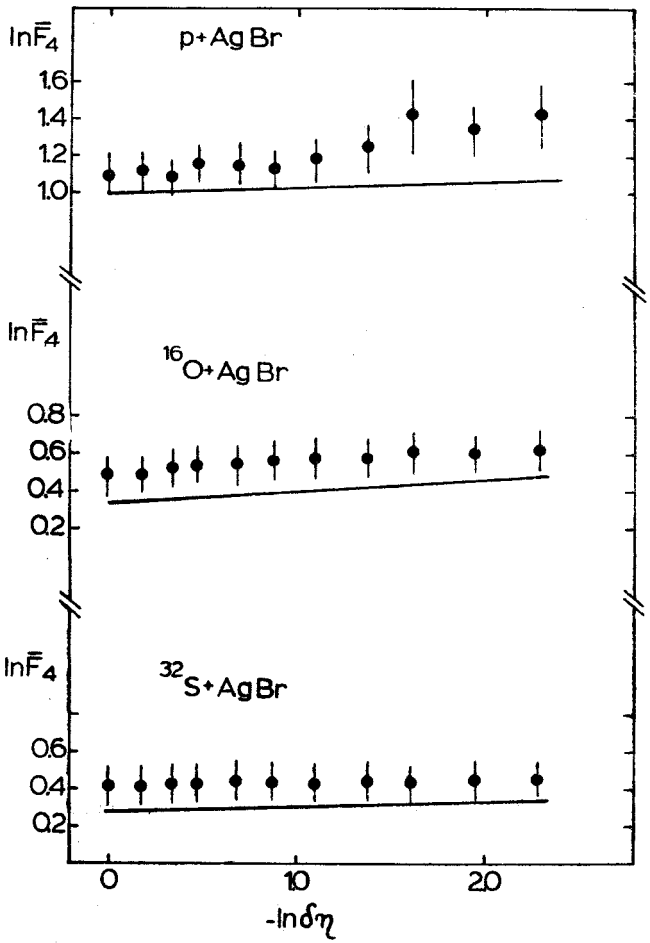

Fig. 2. Best fits to the $\bar{F}^4$ dependence on $\delta\eta$ resulting from minimizing formula (5) for ^{32}S, ^{16}O and p collisions with Ag/Br targets at 200 GeV/n

TABLE II

Parameters resulting from the fit $\bar{F}^{\alpha} = c \cdot (\delta\eta)^{-b}$ for $\delta\eta$ from 1.0 to 0.1

		Correct χ^2	Simplified χ^2
$^{32}S+AgBr$	$\ln c$	0.282 ± 0.092	0.405 ± 0.047
	b	0.025 ± 0.006	0.023 ± 0.038
	χ^2/NDF	10.1/9	0.1/9
$^{16}O+AgBr$	$\ln c$	0.331 ± 0.078	0.495 ± 0.047
	b	0.064 ± 0.011	0.060 ± 0.039
	χ^2/NDF	8.8/9	0.3/9
$p+AgBr$	$\ln c$	0.996 ± 0.078	1.068 ± 0.047
	b	0.037 ± 0.045	0.139 ± 0.051
	χ^2/NDF	10.4/9	1.6/9

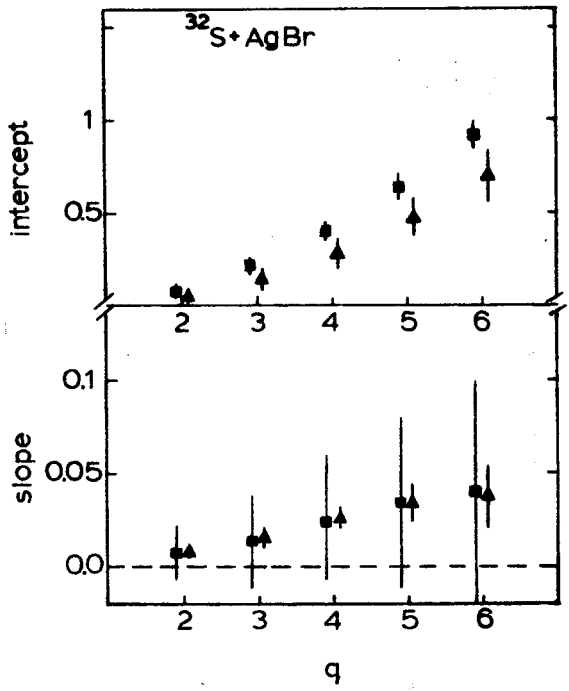

Fig. 3. Comparison of the values of fit parameters: intercept, $a = \ln c$ and slope, b obtained from the simplified fit (■) and fit with correlations (▲) for $^{32}S+AgBr$ collisions

case of nucleus-nucleus interactions with large multiplicities the effects of including correlations are not very large. Systematic shift toward lower values of the intercept parameter is observed. The slope values do not differ significantly, but the errors are smaller by about a factor of 4 as compared to the case with no correlations. A reasonable estimates of the

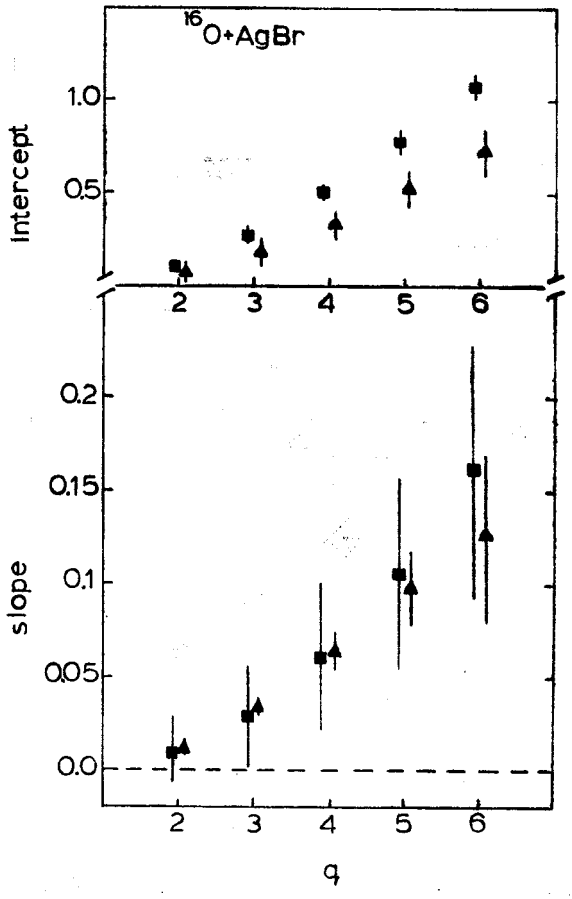

Fig. 4. The same as Fig. 3, but for ^{16}O+AgBr collisions

goodness of the fits were obtained (χ^2 per degree of freedom between 0.7 and 1.4). For p+AgBr data the effect of including correlations is much stronger. From the fits with correlations we obtain the values of the slope for the moments of the order greater than 3 consistent with zero. Thus, the results for low-multiplicity data are uncertain and unstable. It is possible that more reliable results can be obtained for much higher statistics of events. But still, one should realize that in case of low multiplicities only single events give a non--zero contributions to higher order moments computed for the smallest bin sizes.

Concluding, we have shown that the correlations can be easily taken into account in fitting the dependence of the factorial moments on bin size. For large multiplicity nucleus--nucleus data inclusion of the correlations does not change the values of the slope parameter, but gives errors significantly reduced as compared to the case of fits with no correlations. For low-multiplicity data including correlations may change the values of the slope parameter. It would be very interesting to see what are the effects of including the correlations in the fits for other experimental data, especially e^+e^- and hh collisions (low multiplicities

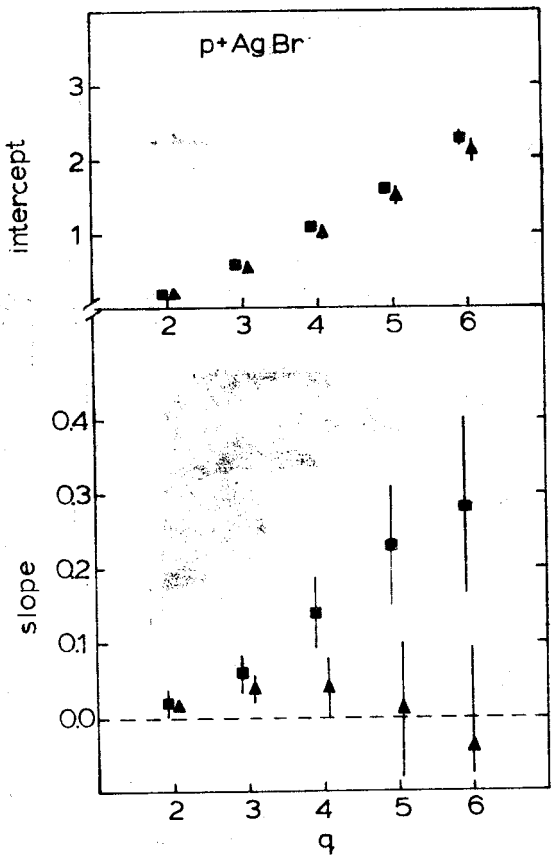

Fig. 5. The same as Figs. 3 and 4, but for p+AgBr collisions

of produced particles) in which the intermittency signal was observed to be stronger than in nuclear interactions [2].

The author would like to thank the KLM Collaboration for providing the data used in the analysis. I would also like to thank W. Ochs, K. Zalewski and K. Fiałkowski for the instructive discussions.

APPENDIX

Since our results seem at first glance to counter common intuition (the best fit passing on one side of all data points), we have decided to present in this Appendix a simple analytically solvable example. This example was born during the discussions with W. Ochs, J. Wosiek, K. Zalewski and K. Fiałkowski. It proves that the intuition developed for the uncorrelated data is misleading in the case of correlated measurements.

In one case however expectations for the uncorrelated and correlated data agree. When all data points happen to lie exactly on the curve which belongs to the family we are

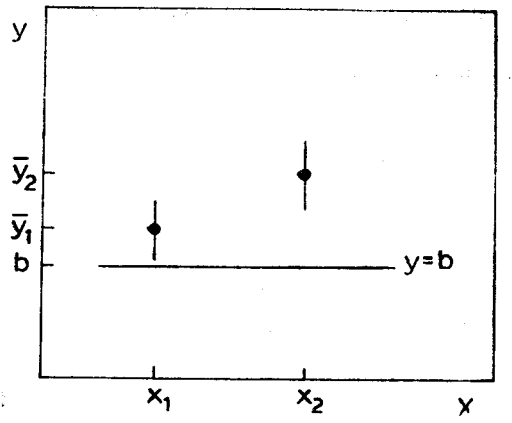

Fig. 6. Hypothetical flat line fit to two data points $\bar{y}_1$ and $\bar{y}_2$

fitting, then of course the best solution, correlated or not, gives $\chi^2 = 0$ and the fit passes through all the points. This observation is trivial for the uncorrelated data. In general, it follows from the semi-positive definitness of the covariance matrix, Eq. (4). This is just an academic possibility since in practice the data never lie exactly on the fitted curve. However, it was important to distinguish if for the sake of the following discussion.

Consider two data points $(\bar{y}_1, x_1)$, $(\bar{y}_2, x_2)$, and $\bar{y}_2 > \bar{y}_1$ (Fig. 6) with the covariance matrix:

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad C_{12} = C_{21}. \tag{A1}$$

Imagine fitting this data set by the straight line, $y = ax+b$. To avoid above mentioned trivial solution, let us look for the best flat fit, $y = b$. In this case our family (labeled by one free parameter b) does not contain the exact solution, and, as we shall see, the interesting "counter-intuitive" result is possible and easily understandable. According to Eq. (5) we minimize:

$$\chi^2(b) = \sum_{i=1}^{2} \sum_{j=1}^{2} (\bar{y}_i - b)\chi_{ij}(\bar{y}_j - b) \tag{A2}$$

with

$$(\chi_{ij}) = \frac{1}{C_{11}C_{22} - C_{12}^2} \begin{pmatrix} C_{22} & -C_{12} \\ -C_{12} & C_{11} \end{pmatrix}. \tag{A3}$$

The minimum lies at:

$$b_0 = \frac{(\chi_{11}+\chi_{12})\bar{y}_1 + (\chi_{22}+\chi_{12})\bar{y}_2}{\chi_{11}+2\chi_{12}+\chi_{22}}. \tag{A4}$$

Hence, there exists a range of correlations for which the best flat line fit would lie below or above both data points. Namely, when $\chi_{22} < -\chi_{12}(\chi_{11} < -\chi_{12})$, the fitted line passes below (above) two data points. In addition the positivity constraint: $\chi_{11}+2\chi_{12}+\chi_{22} > 0$ implies that the other diagonal element of the χ matrix (χ_{11} or χ_{22}) is always bigger than

$-\chi_{12}$. Note that the negative sign of χ_{12} means, according to Eqs. (A3) and (4), positive correlations. Thus, it follows from the conditions ($\chi_{22} < -\chi_{12}$ or $\chi_{11} < -\chi_{12}$) that only for strong enough positive correlations the best fit passes on one side of both measured data points.

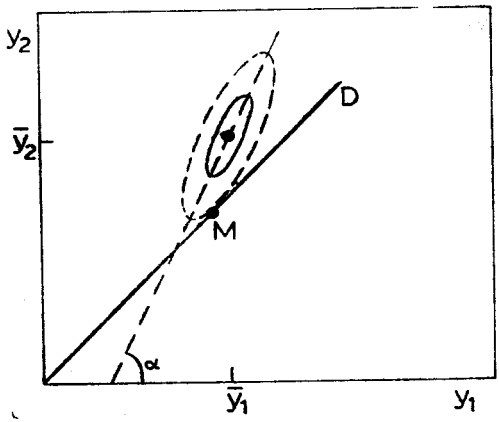

Fig. 7. Schematic illustration of the phase space of all possible measurements of y_1 and y_2

All these conclusions can be immediately seen from Fig. 7 where the phase space of all possible measurements of y_1 and y_2 is shown. The ellipses represent the contours of constant χ^2 and the inner one corresponds to one standard deviation (1 st. dev.) confidence area. The probability that the exact result for y_1 and y_2 lies within the inner ellipse is 68.33 %. The tilt of the ellipse and its major axes are determined by the covariance matrix via Eqs. (A2) and (A3). Minimizing Eq. (A2) for the flat fit, corresponds to searching for the minimum of χ^2 (y_1, y_2) along the straight line $y_1 = y_2$ (labeled as D in Fig. 7), with b parametrizing position on this line. Obviously, the minimum lies at point M, where the diagonal D is tangent to the contour of $\chi^2_{\min}$. It is also clear that, for strong enough positive correlations (positive tilt angle α), the best fit M lies above ($\alpha < 45°$) or below ($\alpha > 45°$) both $\bar{y}_1$ and $\bar{y}_2$. Note also that, in some cases, the best fit can deviate from both data points by more than 1 st. dev. in the same direction, and the best solution M can still be acceptable.

REFERENCES

[1] A. Białas, R. Peschanski, *Nucl. Phys.* **B273**, 70 (1986); *Nucl. Phys.* **B308**, 857 (1988).
[2] W. Kittel, R. Peschanski, Nijmegen Preprint HEN-325, Saclay Preprint SPhT/89-143.
[3] K. Fiałkowski, B. Wosiek, J. Wosiek, *Acta Phys. Pol.* **B20**, 639 (1989).
[4] R. Hołyński et al. (The KLM Collab.), *Phys. Rev. Lett.* **62**, 733 (1989).
[5] R. Hołyński et al. (The KLM Collab.), *Phys. Rev.* **C40**, R2449 (1989). The errors of fit parameters given in this paper are the values obtained from the fit multiplied by the square root of chi-square divided by the number of degrees of freedom.
[6] W. T. Eadie et al., *Statistical Methods in Experimental Physics*, North Holland Publ. Co., Amsterdam 1971.
[7] S. Gottlieb et al., *Phys. Rev.* **D38**, 2245 (1988).
[8] L. M. Barbier et al. (The KLM Collab.), *Phys. Rev. Lett.* **60**, 405 (1988).