SUPPLEMENTO AL VOLUME IX, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
3° Trimestre

A Monte-Carlo Method to Calculate Multiple Phase Space Integrals - II.

F. CERULUS and R. HAGEDORN

CERN - Geneva

(ricevuto il 18 Agosto 1958)

1. – Error estimates.

In a foregoing paper (1), which will be quoted henceforward as (I), we presented a Monte-Carlo (MC) approach to the problem of calculating multiple phase space integrals. We shall use the same notation as in (I).

The proposed calculation can be done only on automatic computers and of course only with a limited accuracy. As a MC calculation is a statistical process, one must know how accurate its result actually—or better: probably—is. Automatic computers allow one to include a check of accuracy within the programme, so that the machine stops or goes over to the next case as soon as the prescribed accuracy is reached.

We therefore begin with an estimate of the error. This error estimate applies to a single calculation of a particular $\varrho_n^*(E, 0)$. Actually, however, one will calculate this function for $n = 3, 4, 5 \ldots$ and also perhaps for different energies. Since we know that it is a smooth function with one single maximum if we vary either n or E (keeping the other fixed), we may get a better

(1) *A Monte-Carlo method to calculate multiple phase space integrals* (I). Equations of that paper are quoted as *e.g.* (I, 15), whereas equations of this paper are referred to as *e.g.* (12), etc.

statistical accuracy for a whole table, than we would have for a single case. If, *e.g.* for fixed E, one calculates a table for $n = 3, 4, \ldots, n'$, then the accuracy of a number from such a table will be better by a factor of roughly $1/\sqrt{n'-3}$ provided the values in the table are smoothed out by using any method of balancing the fluctuations.

For the estimation of the individual error we proceed as follows:

First a certain number N_0 is fixed ($N_0 \gg 1$) which is the basic number of samples or the basic number of contributions to the sum (I, 19). This number should be chosen so small—say 10 or 50 or 100—that one does not expect that N_0 terms make already a good approximation to ϱ^*. The process has then to be repeated several times and its convergence watched. Each repetition will add again N_0 contributions. Thus, if the desired accuracy is reached with λ sets of N_0 contributions, the number of samples is $N = \lambda N_0$. We introduce the following abbreviations:

(1) $\quad \varrho_\lambda \equiv \varrho_n^{*(\lambda)}(E, 0)$ — *i.e.* the MC approximation to $\varrho_n^*(E, 0)$ [2] with $N = \lambda N_0$ contributing samples (see (I, 19)). We want this approximation to fulfil some condition of accuracy to be specified later.

(2) $\quad S_\mu = \sum_{i=(\mu-1)N_0+1}^{\mu N_0} \psi(\varepsilon_1^{(i)} \ldots \varepsilon_n^{(i)})$ — *i.e.* the partial sum of the ψ's from the μ-th set of N_0 samples

(3) $\quad \Sigma_\lambda = \sum_{\mu=1}^{\lambda} S_\mu = \sum_{i=1}^{\lambda N_0} \psi(\varepsilon_1^{(i)} \ldots \varepsilon_n^{(i)})$ — *i.e.* the total sum of ψ's from the $N = \lambda N_0$ samples.

With these definitions is follows from (I, 19) that,

(4) $\quad \varrho_\lambda = \frac{A}{\lambda} \Sigma_\lambda ; \quad A = \frac{(E-M)^{n-1}(2\pi)^{n-1}}{(n-1)!\,(n-3)!\,N_0} ; \quad \varrho_n^*(E, 0) = \lim_{\lambda \to \infty} \varrho_\lambda \equiv \varrho .$

Consider first λ as a fixed number and suppose that ϱ_λ has been calculated. To get an estimate of the accuracy we can use only the accumulated data from the λN_0 samples and we therefore deal with *a posteriori* estimates only.

We can say what would be the mean error (more accurately its r.m.s. value [3]) if we would calculate ϱ_λ many times independently. In fact, we calculate it only once and may have found a value which differs much more from the exact value than just by one mean error. There is no way out of this si-

---

[2] We omit throughout the star and write simply ϱ_λ and ϱ for convenience. The quantities, however, correspond to ϱ^* as defined in (I).

[3] r.m.s. means: root mean square.

tuation, because it is a logical consequence of the MC method, and common to all sampling procedures. Even if we would really calculate many times and take the mean value, then this mean value would simply be another single (though more accurate) calculation of $\varrho_{\lambda'}$ where now λ' is many times larger than λ was. We can then carry through the same reasoning for λ', which we gave above for λ, and we come to precisely the same result. Only the mean error will be smaller.

Let us consider a typical situation. If ϱ_λ has been calculated, then this is the mean value of λ contributions S_μ (all independent of each other), and each one of these comes from N_0 samples. We may here forget about the latter fact, because N_0 influences only the magnitude of the fluctuations of the S_μ. We shall consider therefore the S_μ as random variables. We draw a histogram, which shows this situation:

$$\frac{\varrho_\lambda}{A} = \overline{S}^{(\lambda)} = \frac{1}{\lambda} \sum_{\mu=1}^{\lambda} S_\mu$$

is the mean value and it is supposed that

$$\lim_{\lambda \to \infty} \frac{\varrho_\lambda}{A} = \frac{\varrho}{A} .$$

Actually we do not know this limit and the only quantities which are at our disposal are those which we can extract from the above picture.

For the following we need two different mean values:

i) The mean value over the λ contributions which have so far been calculated. This mean value can be inferred numerically from the actual data; it will be denoted by a bar.

ii) We may imagine the above actual case to be one of a large ensemble, *i.e.* we may imagine ϱ_λ to be calculated L times in the same way (statistically independent) and all the L histograms drawn. They will yield L values $\varrho_\lambda^{(l)}$ ($l = 1 \ldots L$). The second kind of mean value is then over the L-ensemble. Here we take the limit $L \to \infty$ and write $\langle\ \rangle$.

The contributions S_μ have from now on a second label l such that $S_\lambda^{(l)}$ is the μ-th contribution to the l-th histogram. Let F be any function of S, then i) and ii) mean

$$(5) \qquad \begin{cases} \text{i)} \quad \overline{F^{(l)}}^{(\lambda)} = \dfrac{1}{\lambda} \sum_{\mu=1}^{\lambda} F(S_\mu^{(l)}) ; & \lambda \text{ always finite} \\ \text{ii)} \quad \langle F \rangle = \lim\limits_{L \to \infty} \dfrac{1}{L} \sum\limits_{l=1}^{L} F(S_\mu^{(l)}) . \end{cases}$$

Whereas $\overline{F^{(l)}}^{(\lambda)}$ still depends on l, $\langle F\rangle$ is clearly independent of μ. Of course

$$\lim_{\lambda\to\infty} \overline{F^{(l)}}^{(\lambda)} = \langle \overline{F}^{(\lambda)}\rangle = \langle F\rangle ,$$

but we are just interested in the deviations for finite λ.

We proceed now to estimate the error of ϱ_λ as a function of λ. With $\lim_{\lambda\to\infty} \varrho_\lambda = \langle\varrho_\lambda\rangle = \varrho = A\langle S\rangle$ we have

$$(\varrho_\lambda - \varrho)^2 = \frac{A^2}{\lambda^2}\Big(\sum_{\mu=1}^{\lambda} S_\mu - \varrho/A\Big)^2 = \frac{A^2}{\lambda^2}\sum_{\nu,\mu}^{\lambda} \Delta S_\nu\, \Delta S_\mu ; \qquad \Delta S_\mu = S_\mu - \langle S\rangle .$$

Since all contributions are statistically independent, we find with (5)

$$\langle(\varrho_\lambda - \varrho)^2\rangle = \frac{A^2}{\lambda^2}\sum_{\mu=1}^{\lambda}\langle(\Delta S_\mu)^2\rangle = \frac{A^2}{\lambda^2}\sum_{\mu=1}^{\lambda}\langle\Delta S^2\rangle = \frac{A^2}{\lambda}\langle\Delta S^2\rangle .$$

We now define $(\delta\varrho)_\lambda \equiv \sqrt{\langle(\varrho_\lambda - \varrho)^2\rangle}$ and have

$$\frac{(\delta\varrho)_\lambda}{\varrho} \equiv \frac{\sqrt{\langle(\varrho_\lambda - \varrho)^2\rangle}}{\varrho} = \frac{1}{\sqrt{\lambda}}\frac{\sqrt{\langle\Delta S^2\rangle}}{\langle S\rangle} . \tag{6}$$

In the following discussion we shall assume that the distribution of S_μ around S is more or less a Gaussian distribution. We use this assumption only for estimates of orders of magnitude. If N_0 is large enough, the fluctuations of the S_μ will be small and in the neighbourhood of $\langle S\rangle$ the Gauss distribution will be a good approximation. This has the advantage that we can calculate some quantities explicitly. Though the results do not hold strictly, they represent at least good estimates.

If $\langle\Delta S^2\rangle$ (and hence $\delta\varrho = \sqrt{\langle\Delta\varrho^2\rangle}$) were known, we could state confidence limits for ϱ, starting from the calculated ϱ_λ. Assuming a nearly Gaussian distribution for S_μ it is well-known that

the probability that	$\lvert\varrho - \varrho_\lambda\rvert \leqslant \delta\varrho$,	is	$\approx 68\,\%$
»	»	»	$\lvert\varrho - \varrho_\lambda\rvert \leqslant 2\cdot\delta\varrho$,	is	$\approx 95\,\%$
»	»	»	$\lvert\varrho - \varrho_\lambda\rvert \leqslant 3\cdot\delta\varrho$,	is	$\approx 99.95\,\%$.

Then, in order to obtain a given probability that the relative error is smaller than a given (small) number, we only have to choose the corresponding λ large enough. Unfortunately neither $\langle\Delta S^2\rangle$ nor $\langle S\rangle$ are actually known. We may, however, without great error replace in (6) $\langle S\rangle$ by $\bar{S}^{(\lambda)}$, because $\delta\varrho/\varrho$

will be changed only by an amount of the second order. So we redefine

$$\frac{(\delta\varrho)_\lambda}{\varrho} = \frac{1}{\sqrt{\lambda}} \frac{\sqrt{\langle \Delta S^2 \rangle}}{\overline{S}^{(\lambda)}} . \tag{7}$$

There are now many ways to estimate $\langle \Delta S^2 \rangle$ from the sample of λ elements S_μ, *e.g.* using the largest occuring deviation or the r.m.s. deviation from the mean value. These methods, of which Student's test is the most adequate, have several practical drawbacks, the main one being that after each addition of a new set of samples the mean value changes and all deviations have to be calculated anew.

These disadvantages may be avoided in the following way: We use the quantity $\overline{\delta S}^{(\lambda)}$ which we define as follows:

$$\overline{\delta S}^{(\lambda)} = \frac{1}{\lambda} \sum_{\mu=2}^{\lambda} \left| \left(\frac{1}{\mu} \sum_{\tau=1}^{\mu} S_\tau \right) - S_\mu \right| = \frac{1}{\lambda} \sum_{\mu=2}^{\lambda} | \overline{S}^{(\mu)} - S_\mu | . \tag{8}$$

That means: After each step of the calculation we take the mean value $\overline{S}^{(\mu)}$ and find the absolute value of the deviation of the last contribution S_μ from this mean value. The average of all these absolute values is our $\overline{\delta S}^{(\lambda)}$. The The advantage is that the mean values $\overline{S}^{(\mu)}$ are calculated anyway, because they yield the ϱ_μ, and S_μ is just calculated as the last contribution. The absolute value of only this one difference is taken, added to the corresponding quantities of the foregoing steps and the mean is taken; neither are new differences calculated, nor are square roots involved. Since $\overline{S}^{(\mu)}$ tends to $\langle S \rangle$ as $\mu \to \infty$, it is easy to show that

$$\lim_{\lambda \to \infty} \overline{\delta S}^{(\lambda)} = \langle |\Delta S| \rangle . \tag{9}$$

But, since we do not intend to use large λ, we must relate the new quantity to $\langle |\Delta S| \rangle$ or $\langle \Delta S^2 \rangle$ also for finite λ.

For this purpose we use the assumption of a Gauss distribution, since this allows explicit calculations. We shall calculate the expectation value of $\overline{\delta S}^{(\lambda)}$ and express it by $\langle \Delta S^2 \rangle$ under the assumption that the distribution of S_μ about $\langle S \rangle$ follows a Gauss law

$$\Delta S = S - \langle S \rangle \tag{10}$$

and

$$w(S) = \frac{1}{\sqrt{2\pi \langle \Delta S^2 \rangle}} \exp\left[- \frac{\Delta S^2}{2 \langle \Delta S^2 \rangle} \right] ; \quad \int_{-\infty}^{+\infty} w(S)\, \mathrm{d}S = 1 . \tag{11}$$

Then the expectation value (5 ii) is defined as

$$\langle F\rangle = \int_{-\infty}^{+\infty} F(S)\, w(S)\, \mathrm{d}S\,. \tag{12}$$

We need for the following a simple consequence of (11): As is well known, the mean value of a sum of random variables follows again a Gauss distribution, if the random variables themselves follow such a distribution. The r.m.s. deviation of the mean value of $\bar{S}^{(n)}$ from $\langle S\rangle$ is then, as shown in the derivation of (6)

$$\langle (S^{(n)} - \langle S\rangle)^2\rangle = \langle \Delta_n S^2\rangle = \frac{1}{n}\langle \Delta S^2\rangle\,. \tag{13}$$

The probability that $\bar{S}^{(n)}$ lies between x and $x+\mathrm{d}x$, is therefore according to (11) and (13) given by

$$W_n(x)\,\mathrm{d}x = \frac{\mathrm{d}x}{\sqrt{2\pi\cdot(1/n)\langle \Delta S^2\rangle}} \exp\left[-\frac{(x-\langle S\rangle)^2}{2(1/n)\langle \Delta S^2\rangle}\right]. \tag{14}$$

We determine now the probability distribution of the difference

$$\sigma_\mu \equiv \bar{S}^{(\mu)} - S_\mu \tag{15}$$

in (8). We must take care of the fact, that $\bar{S}^{(\mu)}$ is not independent of S_μ and write therefore

$$\sigma_\mu = \bar{S}^{(\mu)} - S_\mu = \frac{1}{\mu}\Big(\sum_{\tau=1}^{\mu-1} S_\tau + S_\mu\Big) - S_\mu = \frac{\mu-1}{\mu}\,(\bar{S}^{(\mu-1)} - S_\mu)\,,$$

where now $\bar{S}^{(\mu-1)}$ is independent of S_μ. We put $\bar{S}^{(\mu-1)} - S_\mu = y_\mu$. The probability that $y \leqslant y_\mu \leqslant y+\mathrm{d}y$, is then obviously

$$v_\mu(y)\,\mathrm{d}y = \mathrm{d}y \int_{-\infty}^{+\infty} W_{\mu-1}(y+S_\mu)\, w(S_\mu)\,\mathrm{d}S_\mu$$

and with (14) and (11):

$$v_\mu(y) = \frac{1}{\sqrt{(\mu/(\mu-1))\,2\pi\langle \Delta S^2\rangle}} \exp\left[-\frac{y^2}{(\mu/(\mu-1))\,2\langle \Delta S^2\rangle}\right]. \tag{16}$$

The distribution of σ_μ is then according to the general formula

$$V\{f(x)\}\,\mathrm{d}f(x) = r(x)\,\mathrm{d}x$$

given by

$$V_\mu(\sigma) = \frac{1}{\sqrt{((\mu-1)/\mu)\, 2\pi\langle \Delta S^2\rangle}} \exp\left[-\frac{\sigma^2}{((\mu-1)/\mu)\, 2\langle \Delta S^2\rangle}\right]. \tag{17}$$

This is again a Gauss distribution with, as a glance on (11) shows:

$$\langle \sigma_\mu^2\rangle = \frac{\mu-1}{\mu}\langle \Delta S^2\rangle . \tag{18}$$

A simple calculation shows now that for a Gauss distribution about zero

$$\langle |x| \rangle = 2\int_0^\infty x\, w(x)\, \mathrm{d}x = \sqrt{\frac{2}{\pi}}\sqrt{\langle x^2\rangle} . \tag{19}$$

This gives

$$\langle |\overline{S}^{(\mu)} - S_\mu| \rangle = \sqrt{\frac{2}{\pi}}\sqrt{\frac{\mu-1}{\mu}}\sqrt{\langle \Delta S^2\rangle} ,$$

which we insert now in the expectation value of (8), namely

$$\langle \overline{\delta S}^{(\lambda)}\rangle = \frac{1}{\lambda}\sum_{\mu=2}^{\lambda}\langle |\overline{S}^{(\mu)} - S_\mu| \rangle ,$$

with the result

$$\langle \overline{\delta S}^{(\lambda)}\rangle = \sqrt{\frac{2}{\pi}}\sqrt{\langle \Delta S^2\rangle}\,\frac{1}{\lambda}\sum_{\mu=2}^{\lambda}\sqrt{\frac{\mu-1}{\mu}} \tag{20}$$

or

$$\sqrt{\langle \Delta S^2\rangle} = \sqrt{\frac{\pi}{2}}\,\frac{\langle \overline{\mathrm{d}S}^{(\lambda)}\rangle}{\frac{1}{\lambda}\sum_{\mu=2}^{\lambda}\sqrt{\frac{\mu-1}{\mu}}} = \sqrt{\frac{\pi}{2}}\,\frac{\langle \sum_{\mu=2}^{\lambda} |\overline{S}^{(\mu)} - S_\mu| \rangle}{\sum_{\mu=2}^{\lambda}\sqrt{\frac{\mu-1}{\mu}}} . \tag{21}$$

This can now be introduced into (7). Of course, we find ourselves here before the same difficulty as with (7). What we actually know is not $\langle \overline{\delta S}^{(\lambda)}\rangle$, but only that particular $\overline{\delta S}^{(\lambda)}$ which we can calculate using the actually computed S_μ. As pointed out at the beginning, this is a logical consequence of the fact that we do not know the distribution, but only a finite number of drawn samples. We could now continue the argument and proceed to calculate the distribution of $\overline{\delta S}^{(\lambda)}$ around its expectation value $\langle \overline{\delta S}^{(\lambda)}\rangle$. We would then end up with essentially the same situation.

In fact the situation is not so bad; if we now replace

$$\langle \overline{\delta S}^{(\lambda)} \rangle$$

by the calculated value

$$\overline{\delta S}^{(\lambda)}$$

then for not too small λ the actual $\overline{\delta S}^{(\lambda)}$ lies near its expectation value. The error in the error estimate, which we thereby introduce, is therefore only of higher order. Even if we would estimate the error wrongly by a factor 2 or so (which is very improbable), this would not cause any serious trouble. We shall therefore adopt the following convention for the estimate of the error:

$$\frac{(\delta\varrho)_\lambda}{\varrho} = \sqrt{\frac{\pi}{2\lambda}} \cdot \frac{1}{\overline{S}^{(\lambda)}} \cdot \frac{\sum_{\mu=2}^{\lambda} |\overline{S}^{(\mu)} - S_\mu|}{\sum_{\mu=2}^{\lambda} \sqrt{\frac{\mu-1}{\mu}}} = z_\lambda \leqslant z_0 ; \qquad \lambda \geqslant \lambda_0 \gg 1 . \tag{22}$$

Here z_0 is a given small number fixing the accuracy. $\lambda_0 \gg 1$ means λ_0 greater that at least 5. This is necessary, since otherwise the mean value over the actual deviations does not mean very much. Note that for $\lambda = 1$ the expressions would not be defined since numerator and denominator become zero. The sum in the denominator,

$$\sum_{\mu=2}^{\lambda} \sqrt{\frac{\mu-1}{\mu}} ,$$

can be calculated for $\lambda = 5, 6, \ldots, 20$, or so, and be stored in the computer. Then no square roots are necessary during the programme and the check of the accuracy by means of (22) becomes extremely simple in an automatic calculation.

If, depending on the construction of the computer, it is simpler to take a square than to take the absolute value of a given number, then one can use instead of (22) the following formula, which is derived in essentially the same way:

$$\left(\frac{\delta\varrho}{\varrho}\right)_\lambda^2 = \frac{1}{\lambda}\left(\frac{1}{\overline{S}^{(\lambda)}}\right)^2 \frac{\sum_{\mu=2}^{\lambda} (\overline{S}^{(\mu)} - S_\mu)^2}{\sum_{\mu=2}^{\lambda} \frac{\mu-1}{\mu}} = z_\lambda^2 \leqslant z_0^2 ; \qquad \lambda \geqslant \lambda_0 \gg 1 , \tag{23}$$

with the same z_0 as above. The factor $\sqrt{\pi/2}$ and the roots in the denominator have disappeared. The sum in the denominator can here be calculated for

each check by the recurrence formula, since the sum for $\lambda - 1$ is already there from the foregoing check.

It should be noted again that if

$$\frac{(\delta\varrho)_\lambda}{\varrho} = z_\lambda ,$$

this means that the probability that the actual error is $\leqslant z_\lambda$ is $\approx 68\,\%$ and so on (see discussion following eq. (6)). Even that is only true for large λ, since we replaced $\langle\overline{\delta S}^{(\lambda)}\rangle$ by the calculated $\overline{\delta S}^{(\lambda)}$. To be sure, one can always check a few results by Student's test and make z_0 smaller if necessary.

In the following flow diagram (22) and (23) are considered separately.

2. – Flow diagram.

The following flow diagram has been developed in collaboration with members of the « Institut für praktische Mathematik » Technische Hochschule, Darmstadt, Germany.

It may therefore show some features which are due to the particular properties of the IBM 650 computer that was used for the calculation. Still, it is felt that this flow diagram will remain essentially the same for any type of computer. Therefore it was considered useful to reproduce it here, since it involves som little tricks.

A few comments:

a) Input data

T, N_0, n, A, $\Delta\varepsilon$ are defined in (I).

λ_0 and z_0 are defined here by eqs. (22) and (23).

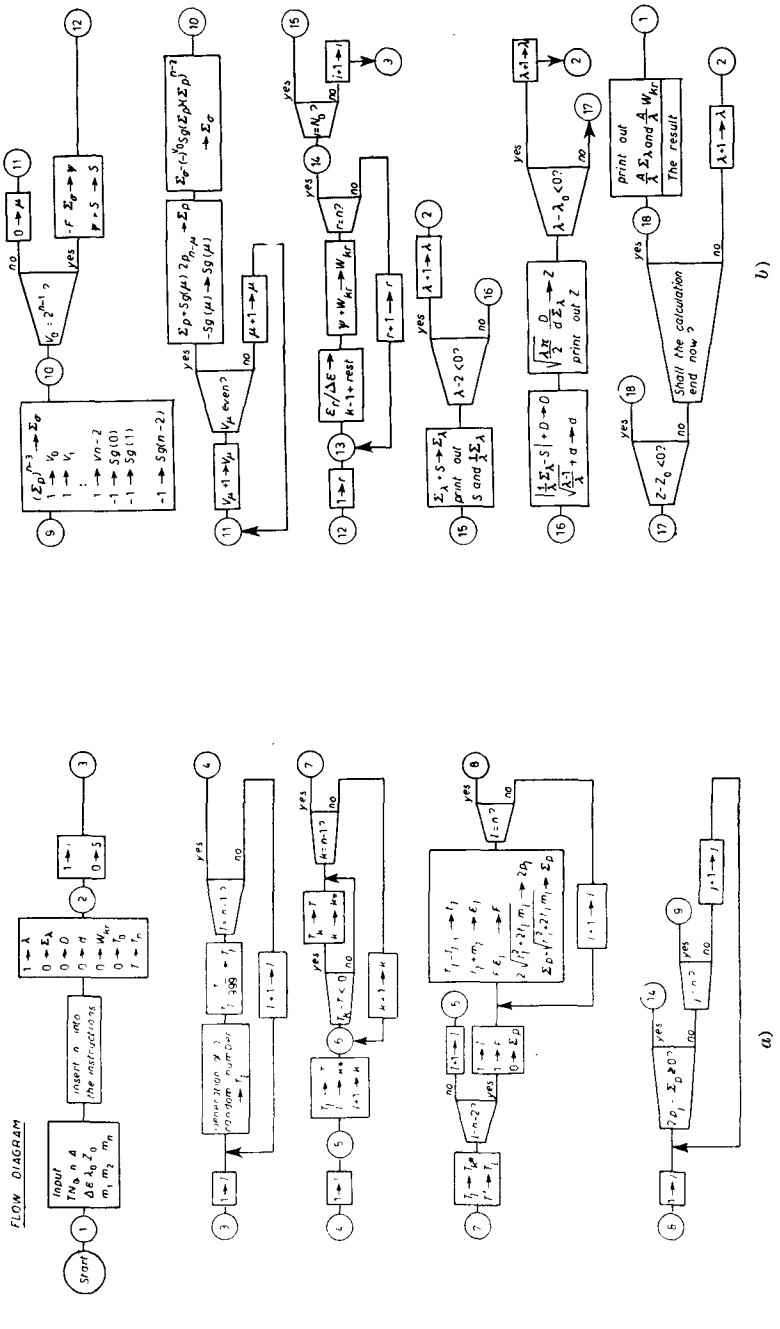
$m_1 \ldots m_n$ are the masses of the particles.

The most convenient values for λ_0 and N_0 and reasonable values for z_0 depend on the total energy and on the number and masses of the particles. λ_0 should at least be of the order of 5. The smaller we choose N_0, the larger λ becomes and the better one can observe the convergence: see discussion under *h*).

b) Instructions before (2)

$W_{\lambda r}$ gives the spectrum.

$(A/\lambda)W_{kr}$ will be the $\varrho_n^{(r)}(E, 0, \varepsilon_r)\Delta\varepsilon$ (of (I), Fig. 2) in the k-th cell. In this programme each particle gets its own spectrum, whether there are identical particles or not. If there are equal particles, one takes in the end the mean value of their spectra thus gaining statistical accuracy. It is convenient to normalize the spectra of the different kinds of masses and



Flow diagram.

to print out directly the normalized spectra together with the mean energies. Since the kinetic energies are as readily available in the computer as the total energies are, one may express the spectra as functions of either variable. (See discussion under *g*)).
λ, Σ_λ, D, d are quantities used during the calculation and the error estimate. (See below).

c) Between (3) and (4)
Here the random numbers are produced. In the present flow diagram it has been assumed that there is a subroutine available which produces (pseudo-) random numbers of three digits between 0 and 999. They are called T' and are mapped on the interval $0 \leqslant T_l \leqslant T$.

d) Between (4) and (8)
Here the $n-1$ random numbers $0 \leqslant T_l \leqslant T$ are reordered according to magnitude and the variables which determine ψ (I, 18)) are prepared.
$F = \varepsilon_1 \cdot \varepsilon_2 \cdot \varepsilon_3 \ldots \varepsilon_n$ is an abbreviation only.

e) Between (8) and (9) (or (14) respectively):
It is checked if perhaps one p_j is larger than the sum of the other p's. If that is the case all the calculations up to (14) are unnecessary, since according to (I, A.10) they yield zero. It is doubtful whether this check gives a gain in computing time, as the probability for « yes » is $(\frac{1}{2})^{n-1}$ (n particles). Thus it saves only in a very few cases a long, useless calculation, but on the other hand it takes itself a time proportional to n.

f) Between (9) and (12)
the function ψ is calculated.

$$\Sigma_\sigma \text{ is the square bracket of (I, 18).}$$

$$\Sigma_p \equiv \sum_i \sigma_i \sqrt{\varepsilon_i^2 - m_i^2}, \qquad \text{(I, 18)}.$$

That part, which begins at (11) and feeds itself back into (10) calculates the sum Σ_σ following exactly the example given in (I, Appendix). It is meant rather symbolically, since most computers do not have an operation which checks whether a number is even. There are many ways of replacing this by an operation which is much faster and implies essentially the same.

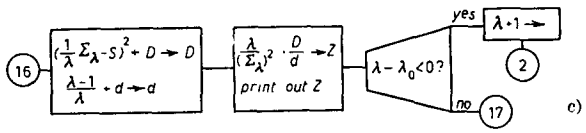
g) Between (12) and (15):
The spectra are calculated for each particle separately according to (I). W_{kr} is a storage matrix, which contains (for r fixed) the spectrum of particle with number r. See discussion under *b*).

h) Between (15) and the end.

The accuracy is checked. S contains now the sum of the ψ's coming from N_0 samples and Σ_λ is the sum of λ such S. It seems desirable that the machine prints out S_λ and $(1/\lambda)\,\Sigma_\lambda$ which allows one to draw histograms such as shown in Fig. 1. This visual control often gives a much better insight than a r.m.s. value or similar quantities. Since the convergence of the MC approach seems to depend on the energy, number of particles, etc., it is advisable to have this extra check here. In order to have a summary of the actual situation, also z_λ (or z_λ^2 respectively) (see here (22) and (23)) are printed out. D and d are numerator and denominator respectively in (22) and (23) of this paper.

Finally, one may have chosen the desired accuracy so high, *i.e.* z_0 so small, that it would never be reached in a reasonable computing time. Therefore the possibility is provided for finishing the calculation and printing out the present results, even if z_λ is not $\leqslant z_0$.

If the square of the relative error is estimated by using the squared differences to the mean value, the flow diagram is changed only between (16) — ... — (17) as follows:



Flow diagram.

Since here the quadratic error is estimated, z_0 has here to be the square of the z_0 in the other method.

In the other method $\sqrt{(\lambda-1)/\lambda}$ and $\sqrt{\pi\lambda/2}$ can be calculated for the interesting λ-values and stored as a small table. Then the lengthy square roots are eliminated from the program.

3. – Practical experiences.

At present a programme is running for nucleon nucleon collisions at 25 GeV primary (laboratory) energy. The cases $2N+\pi$ and $2N+2\pi$ could be checked with the formulae given by BLOCK [4] and showed good agreement. As the phase space integral can be solved rigorously for massless particles, we used

---

[4] M. M. BLOCK: *Phys. Rev.*, **101**, 796 (1956).

a few such cases for various other checks. All cases in which a check was possible showed that the actual error was of the order of the predicted one and in fact in most cases it was even smaller (compare the predictions in the text between Eq. (6) and (7)).

We shall now give one typical example in some detail and summarize the others in a table.

Example: We took the following data:

$$n = 3\ ; \quad m_1 = m_2 = m_3 = 0\ ; \quad E = T = 0.9916\ (^5)\ ; \quad N_0 = 10\ ; \quad \lambda = 37\ .$$

The total number of samples was therefore 370. The prescribed accuracy z_0 was taken to be zero, so that the machine continued to calculate, until it was stopped. It seems most convenient to represent everything in the form of histograms as in Fig. 1 and 2.

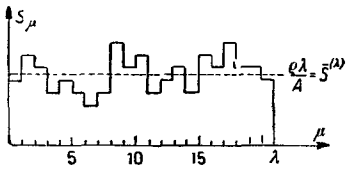


Fig. 1.

Fig. 2 shows the contributions S_μ, each coming from $N_0 = 10$ samples. We have drawn in the mean value $\bar{S}^{(\lambda)} = (1/\lambda)\Sigma_\lambda$ ($\lambda = 37$) and the estimated error limits as well as the exact value, which in this case lies inside the estimated limits.

Fig. 3 shows $(1/\mu)\Sigma_\mu$ for $\mu = 2, 3, \ldots, 37$ again together with the exact value. Here the MC value approaches the exact one from above. In other cases it came from below and in still others it oscillated.

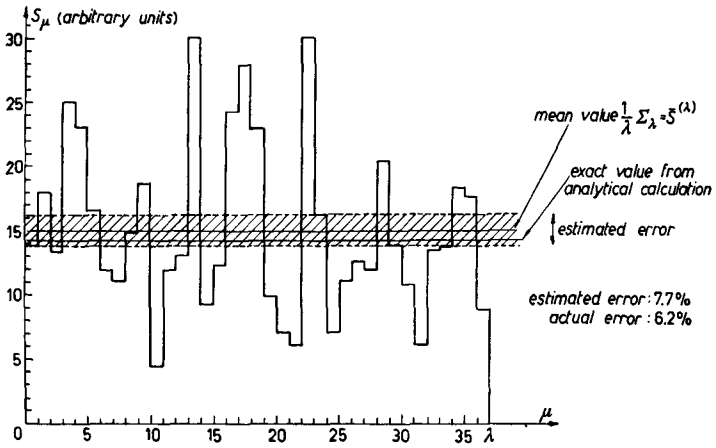


Fig. 2.

(5) Units are $\hbar = c = 1$; the energy unit is the nucleonic mass: $M = 1$.

Fig. 4 shows the estimated error in percent. One sees here that it takes some time before this *a posteriori* error estimate begins to have a meaning. In the beginning it fluctuates very much and only for λ greater than about

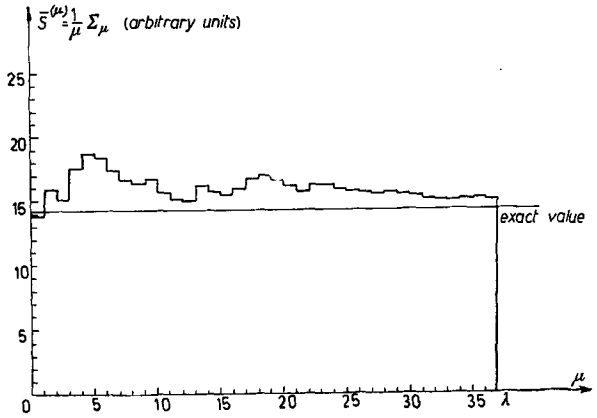


Fig. 3.

15 does it become reasonably smooth. This shows that in this case $\lambda_0 = 15$ would have been reasonable (λ_0 is that number of sets of N_0 samples, beyond which the machine would take the error estimate seriously, compare it with

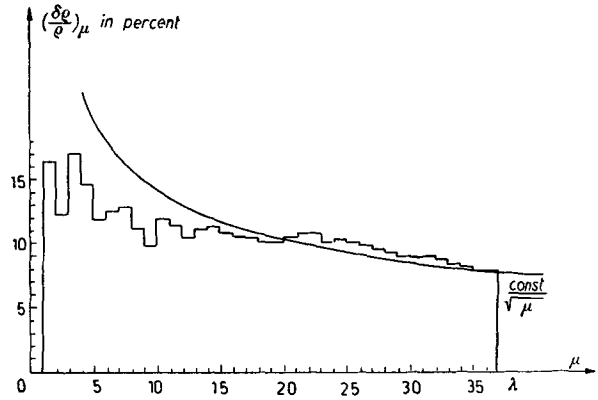


Fig. 4.

the prescribed error z_0 and stop as soon as the estimate is less than that). Of course, λ_0 is roughly inversely proportional to N_0 and here we took N_0 small on purpose, so as to obtain an illustrative example. For comparison a curve const/$\sqrt{\mu}$ has been drawn in. The constant was adjusted such that this curve passed through $(\delta\varrho/\varrho)_{(37)}$.

Fig. 5 shows the MC spectrum as a histogram together with the exact analytical expression (full curve), which was normalized to the same area. It is interesting to compare this with not only the histogram, but with the full information, which one could obtain from it: the dotted curve shows a fit by a polynomial, the coefficients of which were calculated by a least squares method. In this calculation the normalization was accounted for by a supplementary condition. It is this dotted curve which one would consider as the calculated spectrum and it seems to come out quite well. It should be noted that it would come out even better if the spectrum were smooth at the upper end. As one sees from the analytical curve, it has there the character of a step function, which makes it a bit difficult for the interpolation polynomial to reproduce it very well. For particles with mass and for higher particle numbers nothing like that will happen.

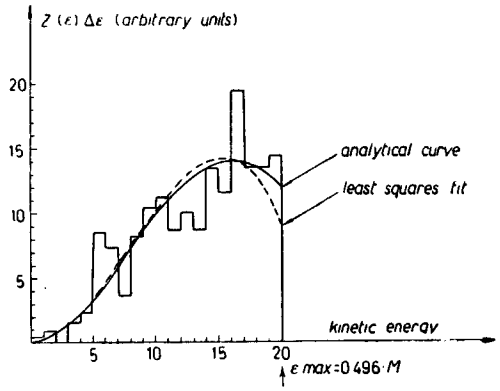


Fig. 5.

Further results for checking the method: Before application, the method was checked in a few further cases for which exact values were known. The calculations were done with only a few samples in the first two examples. The results are given in Table I.

TABLE I.

Particles	CM kinetic energy	ϱ^*_{MC}	ϱ^*_{exact}	Error	
				predicted	actual
6 particles $m = 0$	0.9916	$1.29 \cdot 10^{-6}$	$1.65 \cdot 10^{-6}$	± 23 %	+ 27.9%
6 particles $m = 0$	5.57	$5.1 \cdot 10^{4}$	$5.4 \cdot 10^{4}$	± 23.5%	+ 5.9%
3 particles $m_1 = m_2 = 1$ $m_3 = 0.15$	0.9916	15.72	14.67	± 7.6%	− 7.2%

4. – Estimates of computing time.

As already mentioned, the calculations were performed on a IBM 650 computer. Our computing times could be translated into those for other computers by a suitable factor which may be inferred from well known data on such machines.

Fig. 6 shows the computing time for 50 samples as a function of the total number of particles, n. The dotted part of the curve is an extrapolation based on estimates. It is seen that this function increases rapidly. It should be noted that:

i) No attempt has been made in our case to achieve « optimal programming »,

ii) there are computers available, which are much faster than the IBM 650.

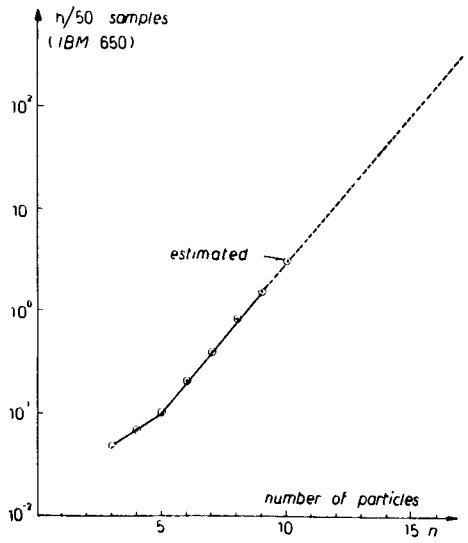


Fig. 6.

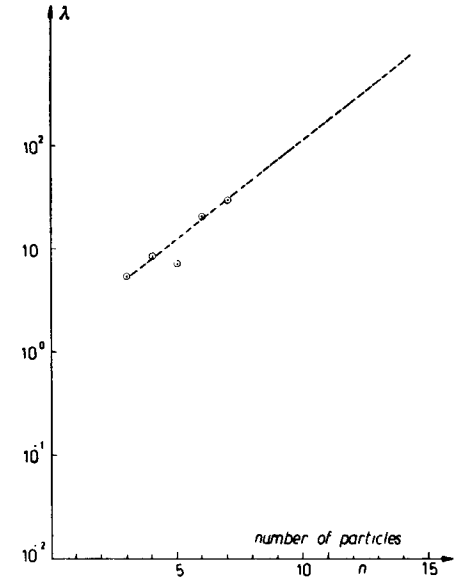


Fig. 7.

Both facts together lead to the guess that on the fastest computers this time can be reduced by a factor of about 100.

We have relatively less experience of the number of samples needed for a certain accuracy, since only a few cases were carried through until a high accuracy was achieved. Fig. 4, however, shows that the accuracy goes roughly with $1/\sqrt{\lambda}$.

In the following Fig. 7 we show the number of samples versus the number n of particles which gives an accuracy below $\sim 8\%$. Some of the points were

found by extrapolating the trend of the error by means of const/$\sqrt{\lambda}$ as shown in Fig. 4. The points partly represent mean values over several different calculations. The straight line is drawn tentatively. The present information does not allow a better determination. (For $n = 5$ the error has obviously been underestimated).

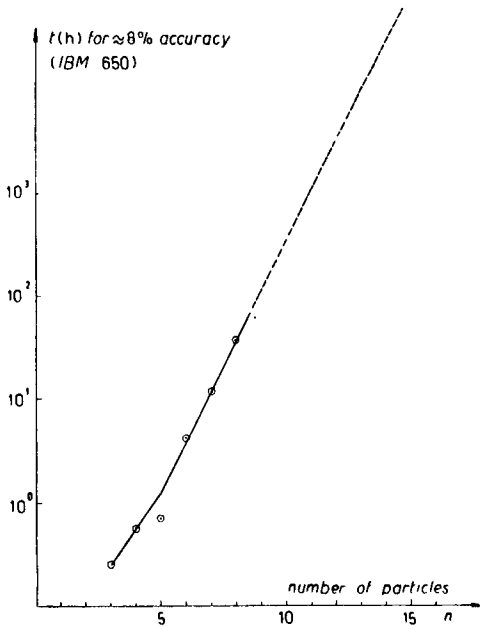


Fig. 8.

It seems that in cases where the kinetic energy T is small compared to the total mass, the convergence does not depend very much on the number of particles. In contrast if T is larger than M, the number of necessary samples increases rapidly with the number of particles. This can be made plausible by a discussion of the formulae for ψ (I, 18). Taking two different samples and looking at what effect the differences of the samples have on ψ, it is found that ψ is more sensitive for differences of samples when $T \gg M$ than when the mass is large, since in the latter case the masses determine the value of ψ more than the kinetic energies do. Sensitivity with respect to differences between samples means slow convergence. It seems that Fig. 6 shows a pessimistic estimate, since it is drawn using mostly results with cases where T was larger than M.

The total computing time necessary for reaching an accuracy of about 8 % is then given by the product of the curves of Figs. 6 and 7. The result is shown in Fig. 8.

This figure should not be taken too seriously because:

i) it is based partly on very rough estimates,

ii) in most applications for higher numbers of created particles less accuracy is necessary, since the total phase space volume is already so small that it contributes very little to everything which one normally calculates, as for instance the mean number of created particles, total spectra, mean kinetic energies, etc.

Assuming that optimal programming on a faster computer leads to perhaps a reduction by a factor 100, then the method is expected to break down for 13 particles if 8 % accuracy is demanded throughout. If, however, the higher particle numbers are not very important, so that (20 ÷ 30) % accuracy is enough,

then one may use the method up to 14 or perhaps 15 particles. At energies where *e.g.* in a nucleon-nucleon collision 13 mesons are produced with non-vanishing probability, the whole Fermi theory becomes rather doubtful on one hand and pure thermodynamical approaches begin to yield reasonable results on the other hand. Thus the MC method seems to cover essentially the most interesting region.

5. – Comparison with the usual approximation.

The usual procedure for calculating a phase space integral for, say, N nucleons and n mesons has been to consider the nucleons as very heavy and the mesons as particles without mass. Then one puts

$$E = \sqrt{p^2 + m^2} \approx \begin{cases} p & \text{for mesons} \\ m + \dfrac{p^2}{2m} & \text{for nucleons} \end{cases} \tag{24}$$

and one can now calculate exactly on either assumption the integral

$$\varrho_k^n(E, P) = \int \delta(E - \sum_i E_i)\, \delta(\boldsymbol{P} - \sum_i \boldsymbol{p}_i)\, \mathrm{d}\boldsymbol{p}_1 \ldots \mathrm{d}\boldsymbol{p} \tag{25}$$

by introducing the Fourier representations of the two δ-functions. The total ϱ^* is then given by the convolution

$$\varrho_{n,N}^*(E, P) = \int \varrho_n^*(E - \varepsilon,\ |\boldsymbol{P} - \boldsymbol{p}|)\, \varrho_N^*(\varepsilon, p)\, \mathrm{d}\varepsilon\, \mathrm{d}\boldsymbol{p}\,, \tag{26}$$

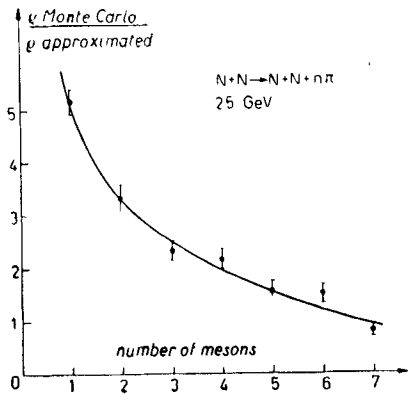


Fig. 9.

which in turn is quite involved but can be calculated by various approximations, *e.g.* the method of steepest descent or the use of tabulated functions. As far as we know, only BELEN'KIJ *et al.* [6] have considered the convolution (26), whereas other authors tacitly assume that each system of mesons and nucleons, separately has total momentum zero. This would be a fair approximation only for the case that any of these systems contains many particles. Since this is not true, the result of such calculations will be even worse than with

[6] S. BELEN'KIJ *et al.*: *Usp. Fiz. Nauk*, **62**, 1 (1957).

the use of (26). As an illustration, Fig. 9 shows the factor between ϱ^* calculated using (24)–(26) and ϱ^* calculated by our method. We took 25 GeV primary energy.

* * *

This paper owes its existence to the very pleasant and fruitful collaboration with the « Institut für praktische Mathematik » of the Technische Hochschule Darmstadt, Germany. Its Director, Professor A. WALTHER, showed much interest and offered us every help and to one of us hospitality at the Institute. Our special thanks are due to dipl. math. G. HUND and dipl. math. D. STEPHAN for many discussions, programming and running the calculations.

We also thank Professor FERRETTI (CERN) for criticism and stimulating discussions.