© World Scientific Publishing Company

NEURAL NETWORKS APPLIED TO NUCLEAR PHYSICS

Christophe DAVID*
SUBATECH, Université de Nantes, EMN, IN2P3/CNRS
4 Rue Alfred Kastler Nantes, F-44070, France

and

Jörg AICHELIN†
SUBATECH, Université de Nantes, EMN, IN2P3/CNRS
4 Rue Alfred Kastler Nantes, F-44070, France

Received (3 April 1995)

Recently we succeded to solve two problems which are encountered quite frequently in theoretical nuclear physics by employing a neural network. 1) To determine the impact parameter of heavy-ion collisions from the observables is of crucial importance for comparing the experimental results with theory. For central collisions all methods which rely on a single observable have failed. Employing neural network we find that a combination of three observables allows to determine the impact parameter four times more accurate than a single observable. We investigate in detail which combination of observables provide the best results. 2) When simulating heavy-ion reactions one has to get random numbers distributed according to known but not monotonous functions. The standard procedure to invert the function therefore fails. we could show that this type of problem can be quite efficiently solved by means of a neural network.

The current studies of high-energy heavy-ion collisions are of broad scientific interest. This is for two reasons: First, they offer a unique testing ground for newly developed methods to study the behavior of strongly interacting quantum systems with finite particle number far from the ground state. Secondly, one hopes that information on the equation of state (EOS) of dense nuclear matter can be extracted from the experimental data. This knowledge is essential for an understanding of the collapse of supernovae, for neutron-star stability, and for the onset of a possible transition from hadron matter to the quark-gluon plasma. The study of relativistic heavy-ion collisions is plagued by the complexity of the reaction. For the same combination of projectile and target a multitude of different processes take place, depending on the impact parameter. In peripheral collisions we observe only a small energy transfer between projectile and target. Both get excited and behave like compound nuclei or undergo fission. More central reactions cause the multifragmentation of the combined system and many intermediate mass fragments can be observed. In central collisions, finally, we observe the formation of a fireball, which disintegrates mainly by emission of nucleons or very light fragments up to

*DAVID@frcpn11.in2p3.fr
†AICHELIN@nanvs2.in2p3.fr

α particles. There is, however, no clear cut difference between these processes and the fluctuations of the observables are large even for a given impact parameter. To separate these processes is, however, of crucial importance if one would like to investigate the different processes in detail and if one desires to compare it with experiment. This is especially true for the most central collisions. In these reactions the nuclei get compressed and part of the system may reach densities up to three times the normal nuclear matter density ρ_0. If one has no means to select these most central collisions all signals are washed out by the much more numerous more peripheral reactions where such a compression is not present. Up to now in most experiments the observables have been analyzed as a function of the multiplicity of the observed charged particles. As is well known and as we will see later the binning as a function of the multiplicity can give a coarse classification into peripheral, semicentral, and central events but it is by no means sufficient to separate the most central reactions. Recently, efforts have been made for a better selection of the central events by using the stopping and the directivity for the classification.[1,2] Although this transverse momentum analysis yields somewhat better results it is not sufficient for selecting precisely the desired impact parameter range. In a world where theory and experiment agreed completely and where there were no experimental acceptance cuts, the neural network technique would be a perfect tool to determine the impact parameters from the experimental observables. One has only to train the network by theoretical simulations and then to feed the trained network with experimental observables in order to obtain the impact parameter as the output of the network. However, the world is not that perfect and therefore the application is not that straightforward but requires a careful comparison of the filtered simulation with the experimental observables obtained in 4π experiments. In other fields of nuclear and high-energy physics, such as track reconstruction in high energy reactions, neural networks have already been applied quite successfully. First, we will show that for simulations a neural network is well suited to describe the dependence of the impact parameter on the different observables. We will demonstrate that the different observables contain different informations and a simultaneous measurement of several of them allows one indeed to lower the uncertainty of the impact parameter.By comparing the results for different combinations of observables we select that triple of observables which allows the most precise determination of the impact parameter. These results can be used as a guideline for the experimental analysis. Secondly, when simulating heavy-ion reactions we have to get random numbers distributed according to a known but not monotonous function. The standard procedure to invert the function therefore fails. We will show that this type of problem (in case of cross section parametrization) can be quite efficiently solved by means of a neural network. It is not the purpose of this paper to optimize the results with respect to the network design. Rather, we would like to show that with the standard design we can already obtain results which are superior to those yet obtained with other methods. For our impact parameter determination we apply the quantum molecular dynamics (QMD) model which

has been successfully used in recent years to simulate heavy-ion reactions.[3] In this model the nucleons are represented by Gaussian wave functions which move under the influence of mutual interactions given by the Brückner G matrix. The real part of the G matrix acts like a density dependent two body force between the nucleons. The imaginary part can be formulated as a cross section. Initially the nucleons of projectile and target are located in a sphere of radius $r = 1.14A^{1/3}$ and the initial momentum has been randomly selected in between 0 and the local Fermi momentum. For details of this approach we refer the reader to Ref. [3]. The analysis is performed with 1037 events of the reaction Au(600 MeV) + Au with an impact parameter randomly chosen between 0 and $b^{max} = 14fm$. Before we start to discuss the neural network and the results obtained we calculate as a benchmark how well the impact parameter can be determined by use of one observable only. This analysis was done for three observables, the total multiplicity of protons (MULT), the largest fragment observed in each collision (AMAX) and the energy ratio (ERAT) in the center of mass system defined by

$$ERAT = \frac{\sum p_t^2/2m}{\sum p_z^2/2m}. \tag{1}$$

where p_t and p_z are the final momenta of the observed particles transverse and parallel to the beam. For this purpose we define the normalized impact parameter

$$b = \frac{b^{true}}{b^{max}} \tag{2}$$

Thus the range of the reduced impact parameter is [0,1]. In bins of 0.05 we calculated the mean value and the variance of the above mentioned observables and approximated the mean value by a spline fit We see that MULT as well as AMAX is constant for impact parameters smaller than 3.5 fm; thus they cannot be used to select the most central collisions. Only ERAT shows an appreciable dependence on the impact parameter in that impact parameter domain. In peripheral reactions AMAX approaches a constant value and the other observables show a dependence on b. To estimate the precision of the determination of the impact parameter we invert the fit function in order to obtain the reduced impact parameter b^{fit} as a function of the observables. Then we calculate the standard deviation

$$C = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(b_i^{QMD} - b_i^{fit})^2} \tag{3}$$

for central, semicentral and peripheral events. The results are presented in table 1. We see, as already expected from the discussion above, that the results obtained for AMAX and MULT are quite satisfactory for semicentral and central collisions, giving an uncertainty of the impact parameter determination of about $C \cdot b^{max} = 0.7fm$. The methods fails, however, for central events. The seemingly reasonable values for the variable AMAX and MULT are of no use as discussed above. The

Table 1. Uncertainty of the impact parameter determination using only one observable.

$C=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(b_i^{QMD}-b_i^{fit})^2}$	$0 \leq b \leq 0.15$	$0.3 \leq b \leq 0.7$	$0.85 \leq b \leq 1$
AMAX	0.0791	0.0673	0.0449
MULT	0.1046	0.0426	0.0532
ERAT	0.1286	0.0531	0.0813

small standard deviation is a consequence of the fact that in central collisions ($b < 3fm$) a cluster with mass number larger than 5 is rarely produced.
Therefore, the variance of AMAX is small and the same is true - due to particle number conservation - for MULT. The only candidate, ERAT, fails badly giving an uncertainty of $C \cdot b^{max} = 1.8fm$. In fig. 1 we investigate in detail the difference between b^{fit} and b^{QMD} obtained for this variable.

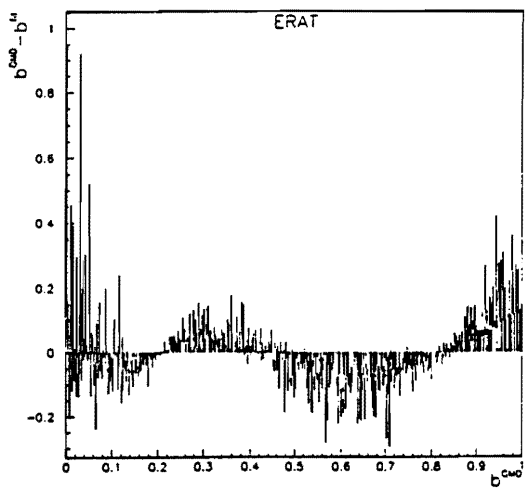

Fig. 1. The difference between b^{QMD} and b^{fit} for the observable ERAT for the impact parameter ordered events.

For this purpose we ordered the events corresponding to their impact parameter and plotted for each simulation the difference between b^{fit} and b^{QMD}. We see that the huge fluctuations in central collisions are the cause for this failure. Although one may think of other observables, there is the general tendency that with one observable it seems to be impossible to select only central collisions. Next we will show that an artificial neural network which makes use of several variables improves the situation. We start with a short description of the neural network we use. In our calculation we have used a feed-forward network with three layers of neurons.

The three layers consist of an input layer of three cells, a hidden layer of five cells and an output layer of one cell. The input layer receives data from outside (i.e., the values for the observables) and the output layer gives the result (i.e., the impact parameter) whose difference from the known impact parameter has to be minimized. For the learning phase we use the method of error backpropagation for the change of the synaptic connections in order to obtain agreement between impact parameter calculated by the network and the known impact parameter of the QMD simulations. The equations governing the state of the network are

$$S = h_1 + \sum_k \omega'_k \sigma_k + \theta, \; h_1 = \sum_j \omega_{1j} s_j \tag{4}$$

$$s_j = f(\bar{h}_j), \; \bar{h}_j = \sum_k \bar{\omega}_{jk} \sigma_k - \bar{\theta}_j \tag{5}$$

with the following notations $\bar{w}_{jk}$ the synaptic connections between input cells and hidden cells, $\bar{\theta}_j$ the threshold of hidden cells, w_{ij} the synaptic connections between hidden cells and output cell (i=1), h_i the output i-cell activation, $\bar{h}_j$ the hidden j-cell activation, w'_k the synaptic connections between input cells and output cells, θ the threshold of output cell (i=1), s_j the output function of the hidden j-neuron, and S_i the output function of the output i=1 neuron. The activation function of the hidden neurons is given by $f(t) = \frac{1}{2}[1 + tanh(t)]$, and the activation function of the output neuron is linear so we have $g(h_1) = h_1$. For training the network we applied the conjugate gradient method as well as a subroutine of the NAGLIB library. The results obtained are identical.The procedure to find the minimum consists of two steps and for details we refer the reader to Ref. [4]. The training proceeds in iterations. The three input observables are chosen from the following set of observables: the mass of the largest fragment (AMAX), the multiplicity of intermediate fragments (IMF), the multiplicity of protons (MULT), the flow of particles $= \frac{1}{N} \sum sign(y_{cm}) \cdot p_x$, x being the direction of the impact parameter (FLOW), the directivity (DIR), $= \frac{|\sum \vec{p}_i|}{\sum |\vec{p}_i|}$, the energy ratio in the center of mass system [Eq. (1)] (ERAT), and FSD defined as

$$\frac{N_c}{N_c - 1} \cdot \frac{(\sum \vec{p}_i)^2 - \sum p_i^2}{(\frac{p_{proj}}{A_{proj}} \cdot \sum A)^2} \tag{6}$$

We tried several combinations of three observables among the seven above and present each event 10000 times at the network. The quality of the network response is quantified by the standard deviation

$$C = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (b_i^{QMD} - b_i^{network})^2} \tag{7}$$

with b_i^{QMD} the QMD impact parameter, $b_i^{network}$ the impact parameter estimated by the network, and N the number of events.The time evolution of the standard

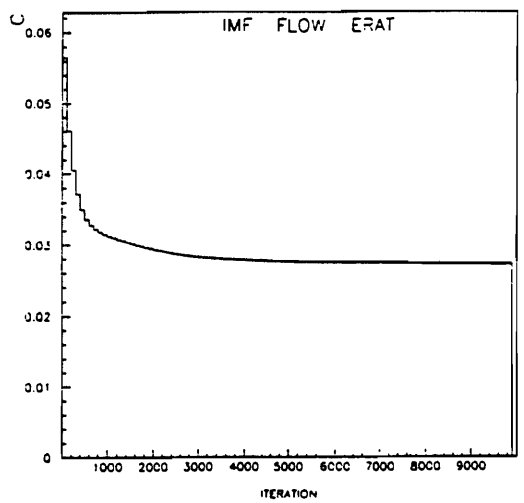

Fig. 2. The standard deviation between b^{QMD} and $b^{network}$ as a function of the number of iterations.

deviation is displayed in fig. 2 for a typical example. We see that after 5000 iterations the network has practically reached its asymptotic value.

To see how well the different impact parameter ranges are reproduced we calculated the standard deviation for three intervals $0 \leq b^{QMD} \leq 0.15$, $0.3 \leq b^{QMD} \leq 0.7$, and $0.85 \leq b^{QMD} \leq 1$ separately. We come now to the results. Table 2 gives a survey of all calculations we performed selected for central, semicentral, and peripheral events. First of all, we observe that especially for central collisions the quality of the impact parameter selection depends strongly on the observables which are used. The combination AMAX, IMF, and FLOW leads to a standard deviation for central collisions of 0.8 fm, whereas the combination IMF, ERAT, and FLOW allows us to determine the impact parameter quite well (standard deviation = 0.31 fm). In any case the values are at least a factor of three better than if one uses ERAT only, as discussed above. Thus a neural network is indeed a proper tool to select the observables which contains the most information. Let us investigate in detail how the network performs. In fig. 3 we display the impact parameter ordered deviation between $b^{network}$ and b^{QMD} for four combinations. In the top row we display the two combinations which give the best result. We see that up to b = 0.8 there are practically no systematic structures. Thus the standard deviation is caused by the fluctuations of the observables in simulations with almost the same impact parameters. Above b = 0.8 (= 11.2 fm) we observe systematic deviations. All the observables MULT, FLOW, ERAT, and IMF tend to zero for large impact parameters. Therefore their value for a determination of the impact parameter of

peripheral reactions is rather limited. Most probably observables like the scattering angle of the large residue are more appropriate.

Table 2. Variance of the impact parameter determination using three (four) observables as the input of the network.

$C=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(b_i^{QMD}-b_i^{net})^2}$	$0 \leq b \leq 0.15$	$0.3 \leq b \leq 0.7$	$0.85 \leq b \leq 1$
AMAX IMF FLOW	0.0569	0.049	0.0481
AMAX IMF ERAT	0.0331	0.0277	0.0441
AMAX IMF FSD	0.0455	0.0433	0.0469
AMAX MULT FLOW	0.0521	0.0489	0.0442
AMAX MULT ERAT	0.0356	0.0281	0.0422
AMAX MULT FSD	0.0475	0.0457	0.0415
AMAX FLOW ERAT	0.04	0.032	0.043
IMF MULT FSD	0.0546	0.0414	0.041
IMF MULT ERAT	0.0335	0.025	0.0406
IMF MULT FLOW	0.0543	0.0455	0.0422
AMAX DIR ERAT	0.0327	0.0357	0.0427
AMAX ERAT FSD	0.0277	0.029	0.0469
IMF DIR ERAT	0.027	0.03	0.0564
IMF ERAT FSD	0.0242	0.027	0.0492
MULT FLOW DIR	0.0354	0.0256	0.0375
MULT DIR ERAT	0.0234	0.0232	0.0387
MULT ERAT FSD	0.0251	0.023	0.0395
MULT FLOW ERAT	0.0234	0.0217	0.0353
IMF FLOW ERAT	0.0224	0.0224	0.0357
IMF MULT FLOW ERAT	0.0251	0.0232	0.0368

The bottom shows the combinations MULT DIR ERAT and AMAX IMF FLOW. the first comes closest to the analysis with which the experimental groups [1,2] tried to select central events. We see that the network produces some systematic deviations close to the most central collisions. AMAX IMF FLOW is the worst case. We observe systematic structures in $b^{network} - b^{QMD}$. Since all of the three values tend towards zero for central collision the network has large problems to perform well in this region. To verify whether an increase of the number of observables improves the situation we also performed a calculation with four input cells using the combination IMF, MULT, FLOW, and ERAT, which represent the observables which give the best prediction using three entry cells. We do not observe an improvement of the prediction. If we change the number of hidden cells for the selection of three observables we do not obtain a better prediction. In summary, we have presented a systematic study of a new method, the application of a neural network, to determine the impact parameter of heavy-ion reactions. The neural network allows one to lower the standard deviation between the known impact parameter of "theoretical data" and the impact parameter derived from the observables by a factor of four as compared to the use of one observable only. Thus the most interesting central events can now be selected with a precision of about 0.3 fm. Using instead of the "theoretical" data those which have been filtered by the acceptance of the experiment, the network - after being trained with this data set - can be directly

used to select the most central experimental reactions by using the measured values of the observables as the input variables.

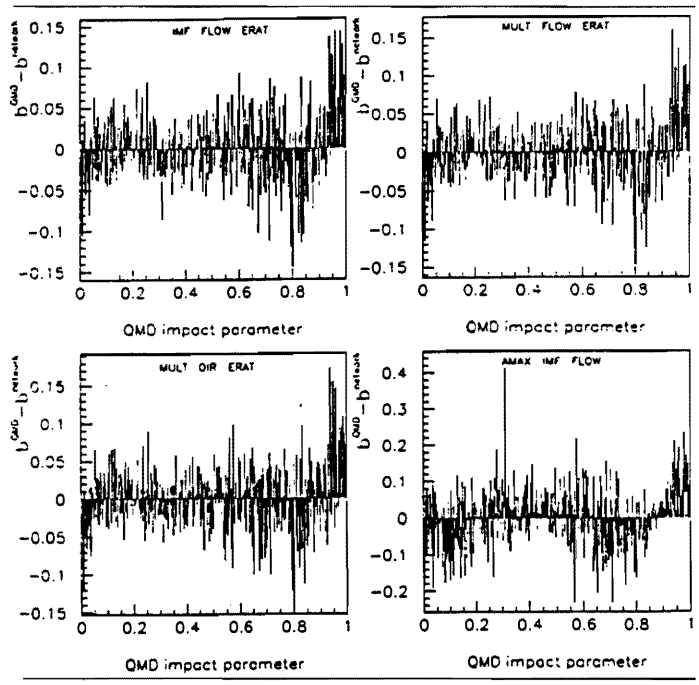

Fig. 3. The difference between b^{QMD} and $b^{network}$ for the impact parameter ordered events for four combinations of observables.

One could even imagine that this method can be used on-line to select already during the experiment the impact parameter range desired for the later analysis. In Monte Carlo simulations of heavy-ion reactions one encounters the problem to find random numbers which have to be distributed according to non invertible functions. For example, if two particles scatter, one has to choose a scattering angle θ with a distribution $\frac{d\sigma}{d\theta}$. We found that this task can be very easily accomplished with a neural network with two input variables. A random number and the energy variables (because $\frac{d\sigma}{d\theta}$ depends on the energy) and one output variable: the scattering angle θ. As an example we show here how to treat[5] K^+p scattering.[6] With the object of parametrization of differential cross sections for elastic K^+p scattering by use of a neural network, we plotted the differential cross section according to cosθ for each value of momentum p, using the data given in Ref. [6]. We used the data from p=0.798 GeV/c to 1.907 GeV/c. We have fitted these data with polynomials and obtain

$$\frac{d\sigma}{d\Omega}(p, cos\theta) = \sum_{i=0}^{n} a_i(p)(cos\theta)^i \tag{8}$$

where a_i(p) are the polynomial coefficients and n the degree of the polynomial. For each momentum we calculate X defined as

$$X(p, cos\theta) = \frac{\int_{-1}^{cos\theta} \frac{d\sigma}{d\Omega}(p, cos\theta')d(cos\theta')}{\int_{-1}^{1} \frac{d\sigma}{d\Omega}(p, cos\theta')d(cos\theta')} \tag{9}$$

X is calculated for several values of cosθ chosen between -0.94 and .94. We see that X is distributed between 0 and 1. We define the normalized momentum $p = \frac{p^{true}}{p^{max}}$ with $p^{max} = 1.907 GeV/c$. We used the same neural network which has been described above. X, and p in input and $cos\theta$ in output have been presented 10000 times at the network. The quality of the network response is quantified by the standard deviation

$$C = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(cos\theta_i^{th} - cos\theta_i^{net})^2} \tag{10}$$

with $cos\theta_i^{th}$ the cosθ presented at the network $cos\theta_i^{net}$ the cosθ estimated by the network, and N the number of events. After 10000 iterations, we obtained $C = 0.033$, an extremely small value which allow us immediately practical application.

References

1. J.P. Alard *et al.*, *Phys. Rev. Lett.* **69**, 889 (1992)
2. T. Wienold, *GSI Report* No. **93-28** (1993)
3. J. Aichelin and H. Stöcker, *Phys. Lett.* **B176**, 14 (1986); J. Aichelin, *Phys. Rep.* **202**, 233 (1991); C. Hartnack *et al.*, *Nucl. Phys.* **A540**, 643 (1994)
4. C. David, M.Freslier and J. Aichelin, *Phys. Rev.* **C51**, 1453 (1995)
5. C. David and J. Aichelin, *in preparation*
6. B.J. Charles *et al.*, *Nucl. Phys.* **B131**, 7-53 (1977)