RADIATION RESEARCH **113,** 437–446 (1988)

Frequency Distributions and Density Functions of Distances with Simulated Linear Track Structures

J. B. SCHÄFFER[1]

Neuherberg, Federal Republic of Germany

AND

H. SCHERB AND G. WELZL

Medis-Institut, Gesellschaft für Strahlen- und Umweltforschung, D-8042 Neuherberg, Federal Republic of Germany

SCHÄFFER, J. B., SCHERB, H. AND WELZL, G. Frequency Distributions and Density Functions of Distances with Simulated Linear Track Structures. *Radiat. Res.* **113,** 437–446 (1988).

Track structures of high-LET particles can be simulated by various linear approaches. The distribution of distances seems to be an important parameter in understanding the type of interactions which occur and the biological effects which these excitations and ionizations will create; therefore, the distance distributions of these simulated track structures were calculated. Three presentations show that their exact appearance depends on the scaling parameter: the number of classes. In one approach the theoretical density of the distances was calculated by the techniques of convolution and by forming mixed distributions which confirm the findings of the simulation.

1. INTRODUCTION

If radiation interacts with biological matter, excitations and/or ionizations are caused in molecules depending on the nature and energy of the radiation field and the specificity of the target.

Tracing the sites of these events in space yields a characteristic pattern—the so-called "track structure"—of the interactions. An important variable of this complex pattern is the distribution of distances of these events, which should allow one to draw conclusions about the types of reaction as well as the energy depositions which have occurred. As a first approximation, the track structures of α particles or high-LET particles can be dealt with as rectilinear tracks in which the only variance is from the spatial sequence of depositions (ionizations) along the axis (energy strag-

[1] Present address: Medis-Institut, Gesellschaft für Strahlen- und Umweltforschung, D-8042 Neuherberg, Federal Republic of Germany.

0033-7587/88 $3.00

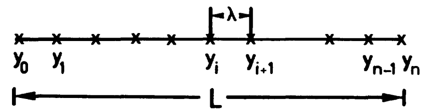

FIG. 1. $n + 1$ equidistant points along the line L.

gling). Therefore, we restricted our Monte Carlo simulations and calculations to a one-dimensional space (*1, 2*).

2. LINEAR SIMULATION OF FAST ION TRACK STRUCTURES AND DETERMINATION OF THE DISTRIBUTION OF THE DISTANCES

2.1. Frequency Distribution of Equidistant Points

Suppose that $n + 1$ points equally spaced λ apart make up a distance L so that $y_{i+1} - y_i = \lambda$ = constant for all $i = 0, 1, \ldots, n - 1$ (Fig. 1). The total number of distances, d_{ij}, represented by these points is $n(n + 1)/2$, where

$$d_{ij} = y_i - y_j; \qquad i, j = 0, 1, \ldots, n \qquad (i > j).$$

Minimum distance: $y_{i+1} - y_i = \lambda$.
Maximum distance: $y_n - y_0 = n\lambda = L$.

Taking the minimum distance as the bin width we can classify the distances d_{ij} into n intervals $J_\nu =](\nu - 1)\lambda, \nu\lambda]$ for $\nu = 1, \ldots, n$ as shown in Fig. 2. Figure 2 shows a linear decline of the frequency with increasing distance. The behavior of randomly spaced distances is treated in the following sections.

2.2. Frequency Distribution Produced by Randomly Spaced Points

2.2.1. Frequency distribution with a logarithmic transformation (3). Along an axis, $n + 1$ points are distributed at increments Δy_i with expected value λ (Fig. 3),

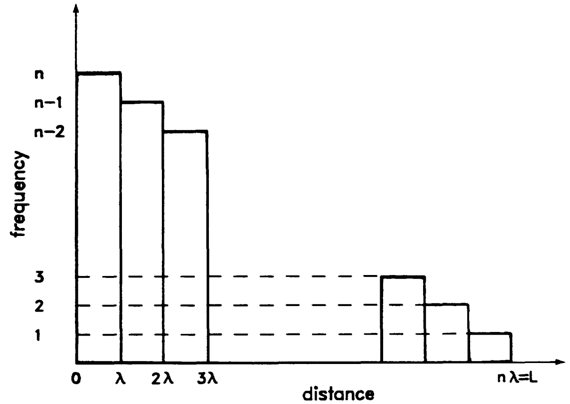

FIG. 2. Histogram of the distance distribution from a group of equidistant points.

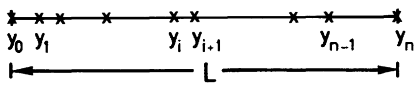

FIG. 3. $n + 1$ nonequidistant points along an axis of length L.

$$\Delta y_i = y_{i+1} - y_i, \qquad \text{where} \qquad i = 0, 1, \ldots, n-1,$$

with the coordinates of the $n + 1$ points generated by the following algorithm which we used for our first Monte Carlo simulations:

$$y_0 = 0;$$
$$y_{i+1} = y_i + (-\lambda)\cdot\ln(x_i);$$
$$\lambda = \text{constant} \qquad (\text{expected value of the increments } \Delta y_i);$$
$$0 < x_i < 1 \qquad (\text{uniformly distributed random numbers}) \text{ with } \qquad i = 0, 1, \ldots, n-1.$$

From this geometry, one can calculate $n(n + 1)/2$ different distances:

$$d_{ij} = y_i - y_j; \qquad i, j = 0, 1, \ldots, n \qquad (i > j).$$

Maximum value for $d_{ij\max}$: $d_{n0} = y_n - y_0 = y_n$.
The expected value for the maximum value: $E(d_{n0}) = n\cdot\lambda = L$.

Figure 4 shows distance distributions for two realizations differing in the number of points. It shows the relationship between the distribution of the distances and the percentile frequencies for two different λ with logarithmic transformation. In its histograms the number of classes is arbitrarily set to 100.

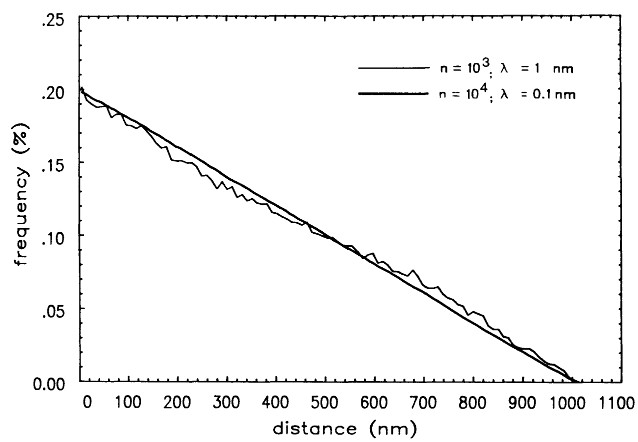

FIG. 4. Distance distributions for $n = 1000$ and 10,000 points with corresponding λ values of 1 and 0.1 nm, respectively.

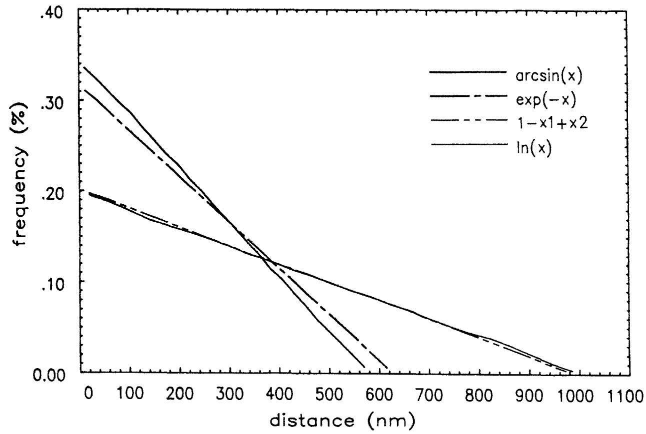

FIG. 5. Frequency distribution of the distances for four different transformations with n = 2000 points and 25 classes (bin width = 40 nm, λ = 0.5 nm).

With a fixed number of classes it is evident that the distribution of the distances is triangular in form, and increasing the number of points results in a decreasing of the relative statistical variations.

2.2.2. Frequency distributions with other transformations. To investigate the effect of the specificity of the transformation on the increments Δy_i we also used for our simulations three other functions: arcsin(x), exp($-x$), and $f(x1, x2) = 1 - x1 + x2$, where x, $x1$, and $x2$ are uniformly distributed random numbers in the open interval]0, 1[. For every transformation, one realization was performed and plotted for n = 2000 points and 25, 100, and 5000 classes. Apart from statistical variations the relationship between frequency and distance can be called linear for all four graphs. To get a sufficient enlargement of the graphs in Fig. 5 it is necessary to take a smaller bin width (more classes).

Inspection of Fig. 6 reveals that apart from statistical variations only the logarithmic transformation shows a linear behavior of the distribution for distances in the range $0 < d < 2\lambda$. The other three distributions show an increase in the same region.

Theoretical investigations will have to be made to determine the exact relationship between frequency and distance with the logarithmic transformation.

2.3. Theoretical Derivation of the Distribution of Distances for the Logarithmical Transformation

The distribution gained by Monte Carlo simulation is now derived for an arbitrary finite n. Referring again to Fig. 3, where y_0 represents the origin and y_i the position of the ith point from it, the increments Δy_i of neighboring points are distributed according to:

$$\Delta y_i = -\lambda \cdot \ln(x_i;) \qquad i = 1, 2, \ldots, n, \tag{1}$$

with x_i being uniformly distributed random numbers in the range

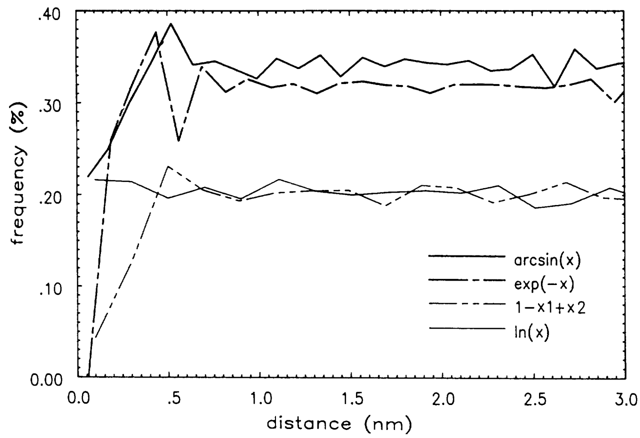

FIG. 6. Enlargement of the four distance distributions with n = 2000 points and 5000 classes (bin width = 0.2 nm; λ = 0.5 nm).

$$0 < x_i < 1.$$

We are looking for the distribution of the distances between the $n + 1$ points.

To derive the density function we set (a) $n = 1$. We want to determine the distribution of the distances between two points. For the sake of applying the rule of transformation for densities unmodified (*4*) we transform

$$(1) \rightarrow (1') \qquad Y = \lambda \cdot \ln(X).$$

X and Y are random variables. Therefore

$$Y = h(X).$$

Let density of the random variable X be $f(x)$ with (Fig. 7):

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{else.} \end{cases}$$

The rule of transformation for densities reads

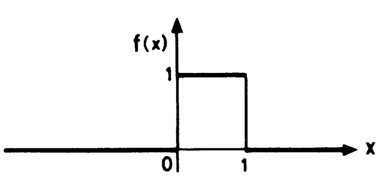

FIG. 7. Density of the random variable X.

$$g(y) = \begin{cases} f(h^{-1}(y))\dfrac{d}{dy}(h^{-1}(y)) & \text{for } a < y < b \\ 0 & \text{else,} \end{cases} \tag{2}$$

where $g(y)$ is the density of the random variable Y and the original set of h is confined to the interval $]0, 1[$.

$$h:]0, 1[\rightarrow]-\infty, 0[$$
$$x \rightarrow \lambda \cdot \ln(x)$$

Hence

$$a = \lim_{x\rightarrow 0} h(x) = -\infty$$

and

$$b = \lim_{x\rightarrow 1} h(x) = 0.$$

Applying the rule of transformation with
(i)

$$y = h(x) = \lambda \cdot \ln(x),$$

(ii) the inverse

$$h^{-1}(y) = \exp\left(\frac{y}{\lambda}\right),$$

and (iii) its derivative

$$\frac{dh^{-1}}{dy}(y) = \frac{1}{\lambda} \cdot \exp\left(\frac{y}{\lambda}\right)$$

yields as density of the distance distribution for one point

$$g_1(y) = f\left(\exp\left(\frac{y}{\lambda}\right)\right) \cdot \left(\frac{1}{\lambda}\right) \cdot \exp\left(\frac{y}{\lambda}\right) \qquad \text{for } -\infty < y < 0$$

$$g_1(y) = \begin{cases} \dfrac{1}{\lambda} \cdot \exp\left(-\dfrac{y}{\lambda}\right) & \text{for } y > 0, \\ 0 & \text{else.} \end{cases} \tag{3}$$

This is the density function for the distances of next neighbor points.

(b) $n = 2$. We want to determine the function $g_2(z)$, the distribution of all distances between the given points. These comprise the two distances of next neighbor points $Y_1 = (y_1 - y_0)$ and $Y_2 = (y_2 - y_1)$ and one distance of the next but one points $(y_2 - y_0)$. With Z we denote the sum of the single distances

$$Z = Y_1 + Y_2. \tag{4}$$

Y_1 and Y_2 are identically distributed random variables with the densities

$$g_1(y_1) = g_1(y_2) = g_1(y) = \begin{cases} \dfrac{1}{\lambda} \cdot \exp\left(-\dfrac{y}{\lambda}\right) & y > 0 \\ 0 & \text{else.} \end{cases}$$

The density $g_2(z)$ of the random variable Z can be calculated by a convolution (*4*):

$$g_2(z) = \int_{-\infty}^{\infty} g_1(y)\cdot g_1(z-y)dy$$

$$= \int_0^z \frac{1}{\lambda}\cdot \exp\left(-\frac{y}{\lambda}\right)\cdot \frac{1}{\lambda}\cdot \exp\left(\frac{z-y}{-\lambda}\right)dy$$

$$g_2(z) = \frac{1}{\lambda}\cdot\left(\frac{z}{\lambda}\right)\cdot \exp\left(-\frac{z}{\lambda}\right). \tag{5}$$

The density of the distances results from the weighted sum of the single densities (density of a mixed distribution)

$$f(x) = a_1\cdot f_1(x) + a_2\cdot f_2(x), \tag{6}$$

where $f_1(x), f_2(x)$ are the densities and a_1, a_2 the respective weights

$$a_1 + a_2 = 1; \qquad a_1 = \frac{2}{3}, \qquad a_2 = \frac{1}{3};$$

$$f(y) = \frac{2}{3}\cdot g_1(y) + \frac{1}{3}\cdot g_2(y)$$

$$f(y) = \frac{1}{3}\cdot\frac{1}{\lambda}\cdot\left(2+\frac{y}{\lambda}\right)\cdot \exp\left(-\frac{y}{\lambda}\right).$$

(c) $n = 3$. We obtain the density $g_3(z)$ for the distribution of the sum of the distances from three points analogously to (b) where

$$g_3(z) = \int_{-\infty}^{\infty} g_1(y)\cdot g_2(z-y)dy$$

$$= \frac{1}{\lambda^3}\cdot\int_0^z (z-y)\cdot \exp\left(-\frac{z}{\lambda}\right)dy$$

$$g_3(z) = \frac{1}{2}\cdot\frac{1}{\lambda}\cdot\left(\frac{z}{\lambda}\right)^2\cdot \exp\left(-\frac{z}{\lambda}\right).$$

(d) Generalization for an arbitrary natural n. Now we can generalize to the density of the distribution of the distances of the nth point from the origin.

The proposition

$$g_n(z) = \frac{1}{(n-1)!}\cdot\frac{1}{\lambda}\cdot\left(\frac{z}{\lambda}\right)^{n-1}\cdot \exp\left(-\frac{z}{\lambda}\right) \tag{7}$$

is proved by complete induction. According to the proposition, it holds that

$$g_n(z) = \int_0^z g_{n-1}(y)\cdot g_1(z-y)dy$$

$$g_n(z) = \int_0^z \frac{1}{(n-2)!}\cdot\frac{1}{\lambda}\cdot\left(\frac{y}{\lambda}\right)^{n-2}\cdot \exp\left(-\frac{y}{\lambda}\right)\cdot\left(\frac{1}{\lambda}\right)\cdot \exp\left(\frac{z-y}{-\lambda}\right)dy$$

$$g_n(z) = \frac{1}{(n-1)!} \cdot \frac{1}{\lambda} \cdot \left(\frac{z}{\lambda}\right)^{n-1} \cdot \exp\left(-\frac{z}{\lambda}\right).$$

The general forms for the mixed distribution of n points are, $n = 2$,

$$f_2(z) = \frac{2}{3} \cdot g_1(z) + \frac{1}{3} \cdot g_2(z),$$

$n = 3$,

$$f_3(z) = \frac{3}{6} \cdot g_1(z) + \frac{2}{6} \cdot g_2(z) + \frac{1}{6} \cdot g_3(z)$$

and n arbitrary,

$$f_n(z) = \sum_{i=0}^{n-1} \frac{(n-i)}{n(n+1)/2} \cdot g_{i+1}(z)$$

$$f_n(z) = \frac{2}{n(n+1)} \cdot \sum_{i=0}^{n-1} \frac{(n-i)}{i!} \cdot \frac{1}{\lambda} \cdot \left(\frac{z}{\lambda}\right)^i \cdot \exp\left(-\frac{z}{\lambda}\right). \tag{8}$$

This formula describes the exact distribution of the density for an arbitrary finite n.

2.4. *Estimating the Limit of the Density Function for Infinite n*

Considering the case $n \to \infty$ and $\lambda \to 0$ with $n\lambda = L$ = constant, we put $\lambda = L/n$ in $f_n(z)$ which yields

$$f_n(z) = \frac{\frac{2}{L} \cdot \sum_{i=0}^{n-1} \frac{(n-i)}{(n+1)} \cdot \frac{1}{i!} \cdot \frac{z^i \cdot n^i}{L^i}}{\exp\left(z \cdot \frac{n}{L}\right)}$$

$$f_n(z) = \frac{2}{L} \cdot \frac{\frac{n}{n+1} \cdot \sum_{i=0}^{n-1} \frac{z^i \cdot n^i}{i! L^i} - \frac{n}{n+1} \cdot \frac{z}{L} \cdot \sum_{i=0}^{n-1} \frac{i \cdot z^{i-1} \cdot n^{i-1}}{i! L^{i-1}}}{\exp\left(z \cdot \frac{n}{L}\right)}.$$

Describing the exponential function by a power series we get

$$f_n(z) = \frac{n}{n+1} \cdot \frac{2}{L} \cdot \frac{\sum_{i=0}^{n-1} \frac{n^i}{i!} \cdot \left(\frac{z}{L}\right)^i - \frac{z}{L} \cdot \sum_{i=0}^{n-2} \frac{n^i}{i!} \cdot \left(\frac{z}{L}\right)^i}{\sum_{i=0}^{\infty} \frac{n^i}{i!} \cdot \left(\frac{z}{L}\right)^i}.$$

Thus it remains to be discussed the limiting behavior of the expression

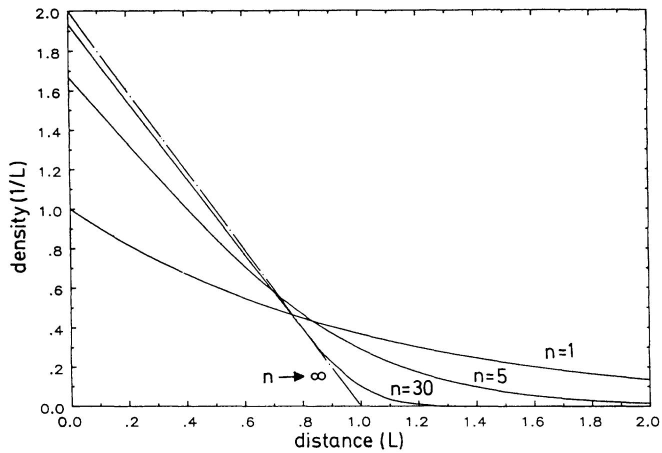

FIG. 8. Density function for $\lambda = 1$ and $n = 1, 5, 30$ as well as the limiting case $n \rightarrow \infty$. The distance is measured in units of L.

$$R_k\left(\frac{z}{L}\right) = \frac{\sum_{i=0}^{k} \frac{n^i}{i!} \cdot \left(\frac{z}{L}\right)^i}{\sum_{i=0}^{\infty} \frac{n^i}{i!} \cdot \left(\frac{z}{L}\right)^i} = \frac{\exp\left(\frac{nz}{L}\right) - \sum_{i=k+1}^{\infty} \frac{n^i}{i!} \cdot \left(\frac{z}{L}\right)^i}{\exp\left(\frac{nz}{L}\right)} \tag{9}$$

for $n \rightarrow \infty$ when $k = n - 1$ and $k = n - 2$.

It can be shown (e.g., by Laplace transformation) that $\lim_{n\rightarrow\infty} R_k(z/L) = 1$ for $z < L$ and $\lim_{n\rightarrow\infty} R_k(z/L) = 0$ for $z > L$. From this, we obtain the following function for the limiting distribution (Fig. 8):

$$f(z) = \frac{2}{L} \cdot \left(1 - \frac{z}{L}\right). \tag{10}$$

This finding is in accordance with what can be expected from Crofton's theorem [see (5), pp. 24–26]. Our model differs slightly from Crofton's insofar as we do not restrict the randomly spaced points on a fixed interval. In our case the maximum distance is in fact unrestricted; only its expectation has the fixed value L.

3. RESULTS AND CONCLUSIONS

Track structures of α particles and high-LET particles have been simulated linearly using various transformations of a uniformly distributed random variable. For large but finite n, all the resulting frequency distributions of distances were triangular in shape when represented in a sufficiently ('coarse') way, i.e., the bin width of the histogram is larger than the expected value of the distances of next neighbor points. But when the frequency distribution from the same track structure is pictured in a rather ('fine') way, i.e., when the expected value of the distances of next neighbor points is greater than the bin width, differences show up in the region of small distances.

Deducing mathematicallly the distance distribution of logarithmically transformed distances by calculating the density function using the techniques of convoluting densities and forming mixed distributions leads to a formula that allows one to determine the distance distribution for arbitrary finite n.

By forming the limit of the density function ($n \rightarrow \infty$, $\lambda \rightarrow 0$, $n\lambda = L$ = constant), we get an exact triangular distribution which confirms the results gained by simulation.

RECEIVED: April 29, 1986; REVISED: March 26, 1987

REFERENCES

1. A. M. KELLERER and D. CHMELEVSKY, Criteria for the applicability of LET. *Radiat. Res.* **63,** 226–234 (1975).
2. J. B. SCHÄFFER, *Eine Studie zur Spurstrukturanalyse durch Simulationen und Berechnung von Anregungszentren und Abstandsverteilungen,* GSF-Bericht 6/86 der Gesellschaft für Strahlen- und Umweltforschung, 8042 Neuherberg, FRG, 1986.
3. L. L. CARTER and E. D. CASHWELL, *Particle Transport Simulation with the Monte Carlo Method,* Chapter 2. USERDA Technical Information Center, Oak Ridge, TN, 1975.
4. J. HEINHOLD and K.-W. GAEDE, *Ingenieur-Statistik.* E. Oldenbourg, München und Wien, 1968.
5. M. G. KENDALL and P. A. P. MORAN, *Geometrical Probability.* Griffin, London, 1963.