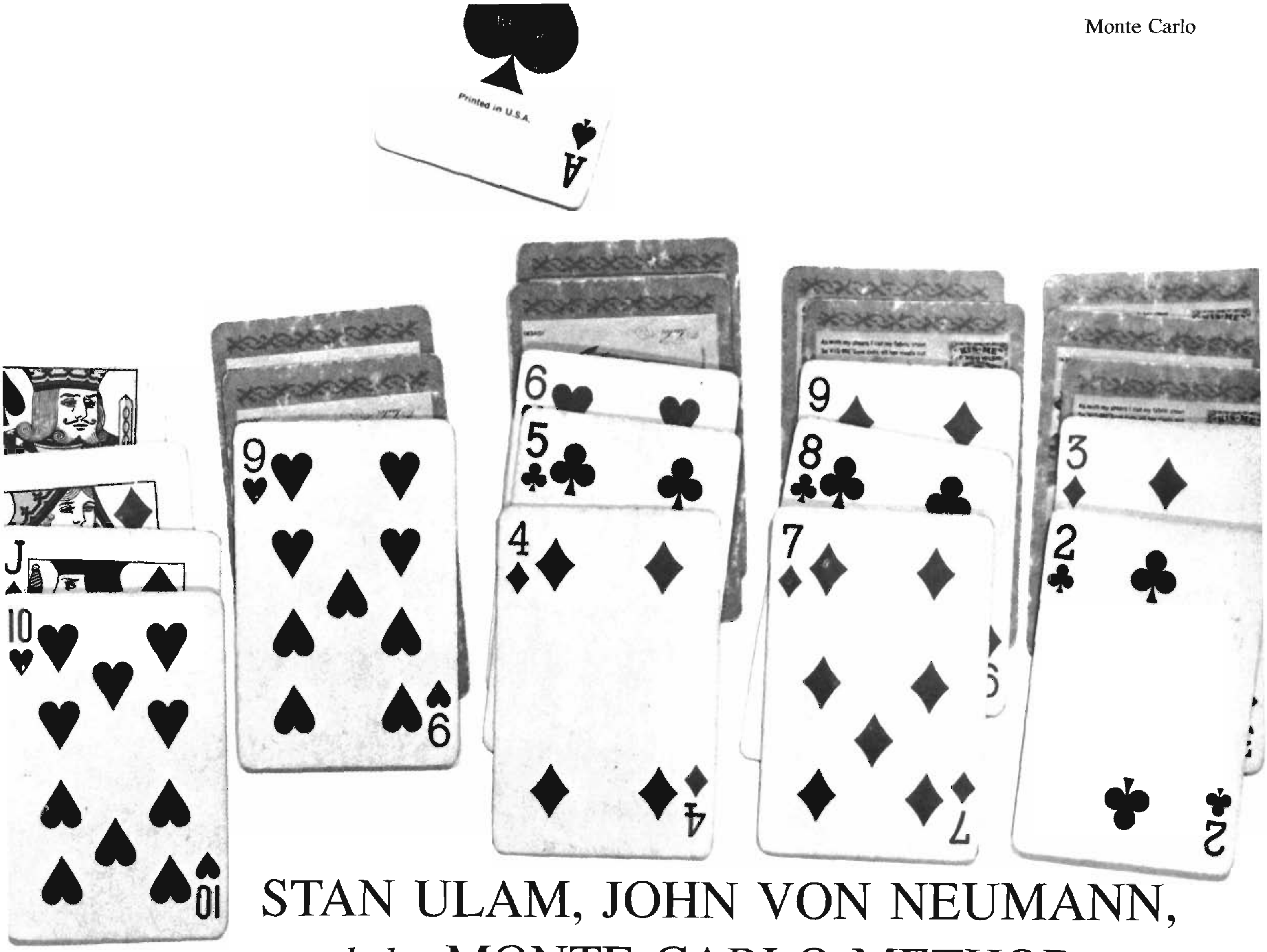


STAN ULAM, JOHN VON NEUMANN, *and the* MONTE CARLO METHOD

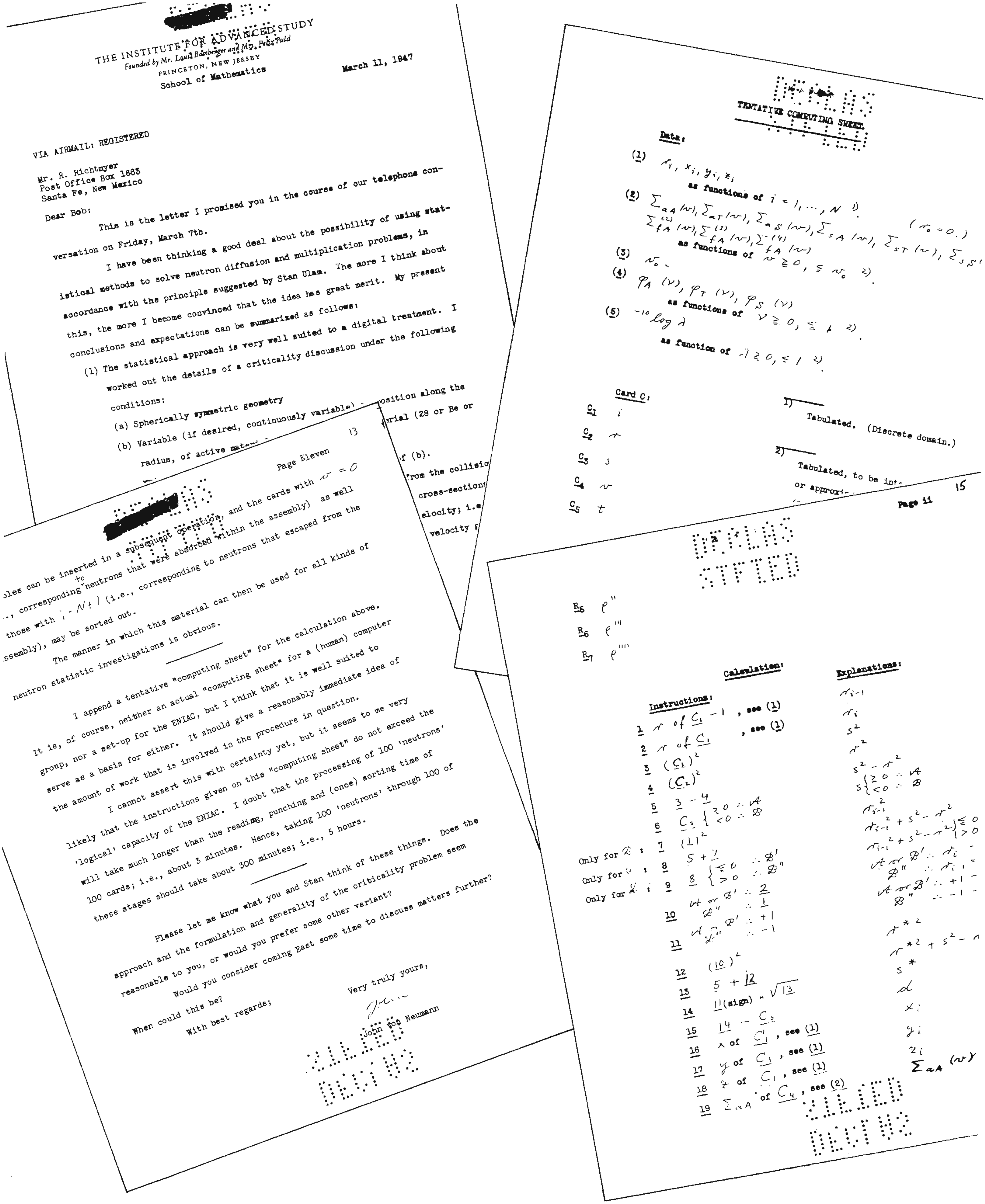
by *Roger Eckhardt*

The Monte Carlo method is a statistical sampling technique that over the years has been applied successfully to a vast number of scientific problems. Although the computer codes that implement Monte Carlo have grown ever more sophisticated, the essence of the method is captured in some unpublished remarks Stan made in 1983 about solitaire.

> "The first thoughts and attempts I made to practice [the Monte Carlo method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later... [in 1946, I] described the idea to John von Neumann and we began to plan actual calculations."

Von Neumann was intrigued. Statistical sampling was already well known in mathematics, but he was taken by the idea of doing such sampling using the newly developed electronic computing techniques. The approach seemed especially suitable for exploring the behavior of neutron chain reactions in fission devices. In particular, neutron multiplication rates could be estimated and used to predict the explosive behavior of the various fission weapons then being designed.

In March of 1947, he wrote to Robert Richtmyer, at that time the Theoretical Division Leader at Los Alamos (Fig. 1). He had concluded that "the statistical approach is very well suited to a digital treatment," and he outlined in some detail how this method could be used to solve neutron diffusion and multiplication problems in fission devices for the case "of 'inert' criticality" (that is, approximated as momentarily static config-



[redacted]

THE INSTITUTE FOR ADVANCED STUDY
Founded by Mr. Louis Bamberger and Mrs. Felix Fuld
PRINCETON, NEW JERSEY
School of Mathematics

March 11, 1947

VIA AIRMAIL: REGISTERED

Mr. R. Richtmyer
Post Office Box 1663
Santa Fe, New Mexico

Dear Bob:

This is the letter I promised you in the course of our telephone conversation on Friday, March 7th.

I have been thinking a good deal about the possibility of using statistical methods to solve neutron diffusion and multiplication problems, in accordance with the principle suggested by Stan Ulam. The more I think about this, the more I become convinced that the idea has great merit. My present conclusions and expectations can be summarized as follows:

(1) The statistical approach is very well suited to a digital treatment. I worked out the details of a criticality discussion under the following conditions:

(a) Spherically symmetric geometry

(b) Variable (if desired, continuously variable) ... position along the radius, of active mate... ...erial (28 or Be or

...f (b).

...rom the collisio

cross-section...

...elocity; i.e

velocity ...

13

Page Eleven

[redacted]

...bles can be inserted in a subsequent operation, and the cards with $v = 0$ ..., corresponding to neutrons that were absorbed within the assembly) as well those with $i = N+1$ (i.e., corresponding to neutrons that escaped from the ...ssembly), may be sorted out.

The manner in which this material can then be used for all kinds of neutron statistic investigations is obvious.

I append a tentative "computing sheet" for the calculation above. It is, of course, neither an actual "computing sheet" for a (human) computer group, nor a set-up for the ENIAC, but I think that it is well suited to serve as a basis for either. It should give a reasonably immediate idea of the amount of work that is involved in the procedure in question.

I cannot assert this with certainty yet, but it seems to me very likely that the instructions given on this "computing sheet" do not exceed the 'logical' capacity of the ENIAC. I doubt that the processing of 100 'neutrons' will take much longer than the reading, punching and (once) sorting time of 100 cards; i.e., about 3 minutes. Hence, taking 100 'neutrons' through 100 of these stages should take about 300 minutes; i.e., 5 hours.

Please let me know what you and Stan think of these things. Does the approach and the formulation and generality of the criticality problem seem reasonable to you, or would you prefer some other variant?

Would you consider coming East some time to discuss matters further? When could this be?

With best regards;

Very truly yours,

John von Neumann

TENTATIVE COMPUTING SHEET

Data:

(1) r_i, x_i, y_i, z_i

as functions of $i = 1, \cdots, N$ 1).

(2) $\Sigma_{aA}(v), \Sigma_{aT}(v), \Sigma_{aS}(v), \Sigma_{sA}(v), \Sigma_{sT}(v), \Sigma_{sS'}$ $(r_0 = 0.)$
$\Sigma^{(2)}_{fA}(v), \Sigma^{(3)}_{fA}(v), \Sigma^{(4)}_{fA}(v)$

as functions of $v \geq 0, \leq v_0$ 2).

(3) v_0.

(4) $\varphi_A(\nu), \varphi_T(\nu), \varphi_S(\nu)$

as functions of $\nu \geq 0, \leq 1$ 2).

(5) $-{}^{10}\log \lambda$

as function of $\lambda \geq 0, \leq 1$ 2).

Card C:

$\underline{C_1}$ i

$\underline{C_2}$ r

$\underline{C_3}$ s

$\underline{C_4}$ v

$\underline{C_5}$ t

1) Tabulated. (Discrete domain.)

2) Tabulated, to be int... or approxi...

15

Page 11

$\underline{R_5}$ ρ''

$\underline{R_6}$ ρ'''

$\underline{R_7}$ ρ''''

Calculation:

Instructions:

1 r of $\underline{C_1} - 1$, see (1)

2 r of $\underline{C_1}$, see (1)

3 $(\underline{C_3})^2$

4 $(\underline{C_2})^2$

5 $\underline{3} - \underline{4}$

6 $\underline{C_3} \begin{cases} \geq 0 \therefore \mathcal{A} \\ < 0 \therefore \mathcal{B} \end{cases}$

Only for $\mathcal{B}$: 7 $(\underline{1})^2$

Only for $\mathcal{B}$: 8 $5 + \underline{7}$ $\begin{cases} \leq 0 \therefore \mathcal{B}' \\ > 0 \therefore \mathcal{B}'' \end{cases}$

Only for $\mathcal{B}$: 9 $\underline{8}$

10 $\mathcal{A}$ or $\mathcal{B}' \therefore \underline{2}$, $\mathcal{B}'' \therefore \underline{1}$

11 $\mathcal{A}$, $\mathcal{B}''$, $\mathcal{B}' \therefore +1$, $\therefore -1$

12 $(\underline{10})^2$

13 $\underline{5} + \underline{12}$

14 $\underline{11}$(sign) $\times \sqrt{\underline{13}}$

15 $\underline{14} - \underline{C_3}$

16 x of $\underline{C_1}$, see (1)

17 y of $\underline{C_1}$, see (1)

18 z of $\underline{C_1}$, see (1)

19 Σ_{aA} of $\underline{C_4}$, see (2)

Explanations:

r_{i-1}

r_i

s^2

r^2

$s^2 - r^2$

$s \begin{cases} \geq 0 \therefore \mathcal{A} \\ < 0 \therefore \mathcal{B} \end{cases}$

r_{i-1}^2

$r_{i-1}^2 + s^2 - r^2 \begin{cases} \leq 0 \\ > 0 \end{cases}$

$r_{i-1}^2 + s^2 - r^2$

$\mathcal{A}$ or $\mathcal{B}' \therefore r_i$ -

$\mathcal{B}'' \therefore r_{i-1}$ =

$\mathcal{A}$ or $\mathcal{B}' \therefore +1$ -

$\mathcal{B}'' \therefore -1$ -

r^{*2}

$r^{*2} + s^2 - r$

s^*

d

x_i

y_i

z_i

$\Sigma_{aA}(v)$

Fig. 1. The first and last pages of von Neumann's remarkable letter to Robert Richtmyer are shown above, as well as a portion of his tentative computing sheet. The last illustrates how extensivly von Neumann had applied himself to the details of a neutron-diffusion calculation.

urations). This outline was the first formulation of a Monte Carlo computation for an electronic computing machine.

In his formulation von Neumann used a spherically symmetric geometry in which the various materials of interest varied only with the radius. He assumed that the neutrons were generated isotropically and had a known velocity spectrum and that the absorption, scattering, and fission cross-sections in the fissionable material and any surrounding materials (such as neutron moderators or reflectors) could be described as a function of neutron velocity. Finally, he assumed an appropriate accounting of the statistical character of the number of fission neutrons with probabilities specified for the generation of 2, 3, or 4 neutrons in each fission process.

The idea then was to trace out the history of a given neutron, using random digits to select the outcomes of the various interactions along the way. For example, von Neumann suggested that in the compution "each neutron is represented by [an 80-entry punched computer] card . . . which carries its characteristics," that is, such things as the zone of material the neutron was in, its radial position, whether it was moving inward or outward, its velocity, and the time. The card also carried "the necessary random values" that were used to determine at the next step in the history such things as path length and direction, type of collision, velocity after scattering—up to seven variables in all. A "new" neutron was started (by assigning values to a new card) whenever the neutron under consideration was scattered or whenever it passed into another shell; cards were started for several neutrons if the original neutron initiated a fission. One of the main quantities of interest, of course, was the neutron multiplication rate—for each of the 100 neutrons started, how many would be present after, say, 10^{-8} second?

At the end of the letter, von Neumann attached a tentative "computing sheet" that he felt would serve as a basis for setting up this calculation on the ENIAC. He went on to say that "it seems to me very likely that the instructions given on this 'computing sheet' do not exceed the 'logical' capacity of the ENIAC." He estimated that if a problem of the type he had just outlined required "following 100 primary neutrons through 100 collisions [each]. . . of the primary neutron or its descendants," then the calculations would "take about 5 hours." He further stated, somewhat optimistically, that "in changing over from one problem of this category to another one, only a few numerical constants will have to be set anew on one of the 'function table' organs of the ENIAC."

His treatment did not allow "for the displacements, and hence changes of material distribution, caused by hydrodynamics," which, of course, would have to be taken into account for an explosive device. But he stated that "I think that I know how to set up this problem, too: One has to follow, say 100 neutrons through a short time interval Δt; get their momentum and energy transfer and generation in the ambient matter; calculate from this the displacement of matter; recalculate the history of the 100 neutrons by assuming that matter is in the middle position between its original (unperturbed) state and the above displaced (perturbed) state;. . . iterating in this manner until a "self-consistent" system of neutron history and displacement of matter is reached. This is the treatment of the first time interval Δt. When it is completed, it will serve as a basis for a similar treatment of the second time interval. . . etc., etc."

Von Neumann also discussed the treatment of the radiation that is generated during fission. "The photons, too, may have to be treated 'individually' and statistically, on the same footing as the neutrons. This is, of course, a non-trivial complication, but it can hardly consume much more time and instructions than the corresponding neutronic part. It seems to me, therefore, that this approach will gradually lead to a completely satisfactory theory of efficiency, and ultimately permit prediction of the behavior of all possible arrangements, the simple ones as well as the sophisticated ones."

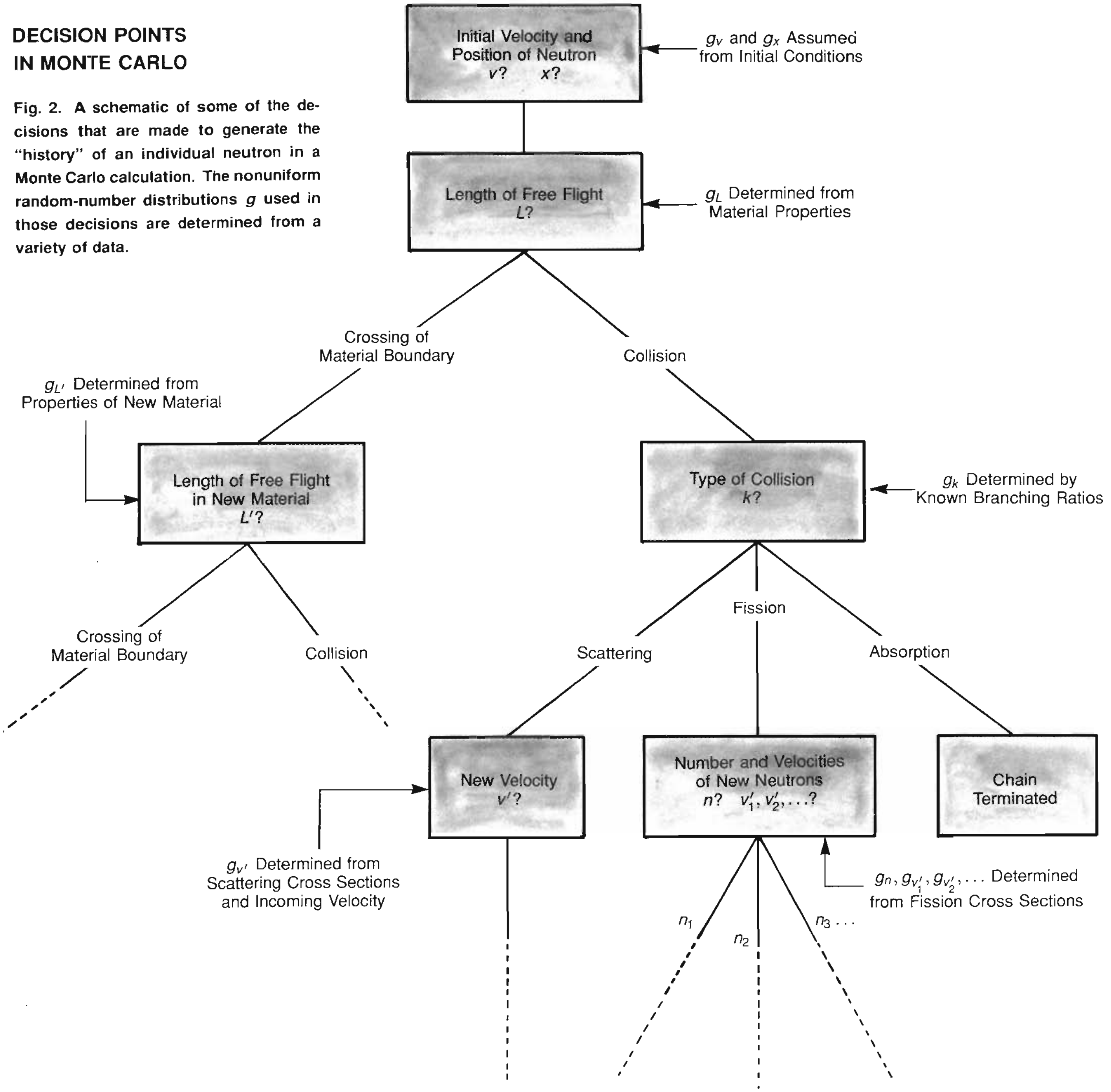
And so it has. At Los Alamos in 1947, the method was quickly brought to bear on problems pertaining to thermonuclear as well as fission devices, and, in 1948, Stan was able to report to the Atomic Energy Commission about the applicability of the method for such things as cosmic ray showers and the study of the Hamilton Jacobi partial differential equation. Essentially all the ensuing work on Monte Carlo neutron-transport codes for weapons development and other applications has been directed at implementing the details of what von Neumann outlined so presciently in his 1947 letter (see "Monte Carlo at Work").

In von Neumann's formulation of the neutron diffusion problem, each neutron history is analogous to a single game of solitare, and the use of random numbers to make the choices along the way is analogous to the random turn of the card. Thus, to carry out a Monte Carlo calculation, one needs a source of random numbers, and many techniques have been developed that pick random numbers that are *uniformly* distributed on the unit interval (see "Random-Number Generators"). What is really needed, however, are *nonuniform* distributions that simulate probability distribution functions specific to each particular type of decision. In other words, how does one ensure that in random flights of a neutron, on the average, a fraction $e^{-x/\lambda}$ travel a distance x/λ mean free paths or farther without colliding? (For a more mathematical discussion of random variables, probability distribution functions, and Monte Carlo, see pages 68–73 of "A Tutorial on Probability, Measure, and the Laws of Large Numbers.")

The history of each neutron is gener-

DECISION POINTS IN MONTE CARLO

Fig. 2. A schematic of some of the decisions that are made to generate the "history" of an individual neutron in a Monte Carlo calculation. The nonuniform random-number distributions g used in those decisions are determined from a variety of data.

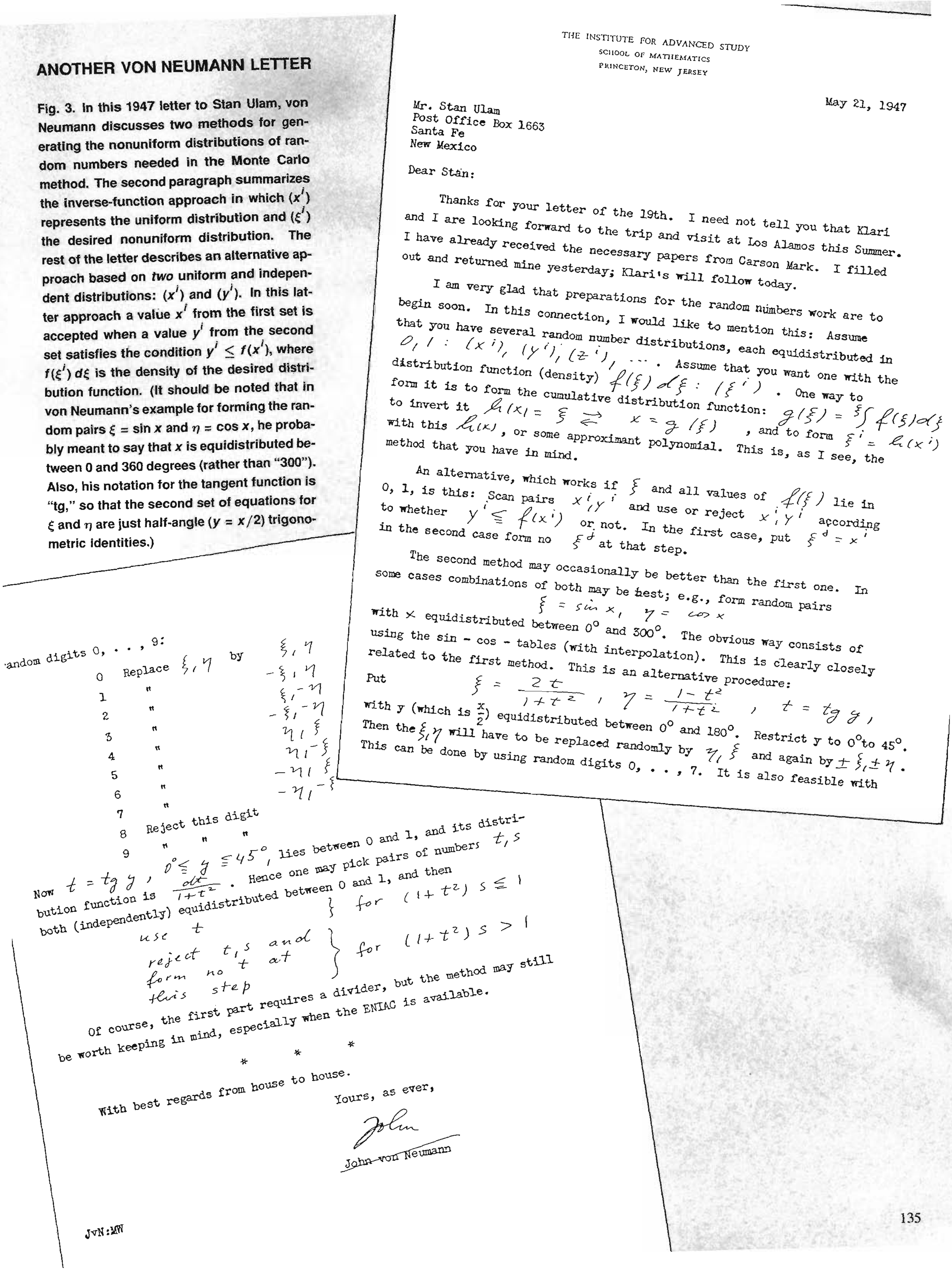

Initial Velocity and Position of Neutron v? x?
g_v and g_x Assumed from Initial Conditions
Length of Free Flight L?
g_L Determined from Material Properties
Crossing of Material Boundary
Collision
$g_{L'}$ Determined from Properties of New Material
Length of Free Flight in New Material L'?
Type of Collision k?
g_k Determined by Known Branching Ratios
Crossing of Material Boundary
Collision
Scattering
Fission
Absorption
New Velocity v'?
Number and Velocities of New Neutrons n? $v'_1, v'_2, \ldots$?
Chain Terminated
$g_{v'}$ Determined from Scattering Cross Sections and Incoming Velocity
$g_n, g_{v'_1}, g_{v'_2}, \ldots$ Determined from Fission Cross Sections
n_1
n_2
$n_3 \ldots$

ated by making various decisions about the physical events that occur as the neutron goes along (Fig. 2). Associated with each of these decision points is a known, and usually nonuniform, distribution of random numbers g that mirrors the probabilities for the outcomes possible for the event in question. For instance, returning to the example above, the distribution of random numbers g_L used to determine the distance that a neutron travels before interacting with a nucleus is exponentially decreasing, making the selection of shorter distances more probable than longer distances. Such a distribution simulates the observed exponential falloff of neutron path lengths. Similarly, the distribution of random numbers g_k used to select between a scattering, a fission, and an absorption must reflect the known probabilities for these different outcomes. The idea is to divide the unit interval $(0, 1)$ into three subintervals in such a way that the probability of a uniform random number being in a given subinterval equals the probability of the outcome assigned to that set.

In another 1947 letter, this time to Stan Ulam, von Neumann discussed two techniques for using uniform distributions of random numbers to generate the desired nonuniform distributions g (Fig. 3). The first technique, which had already been

ANOTHER VON NEUMANN LETTER

Fig. 3. In this 1947 letter to Stan Ulam, von Neumann discusses two methods for generating the nonuniform distributions of random numbers needed in the Monte Carlo method. The second paragraph summarizes the inverse-function approach in which (x^i) represents the uniform distribution and (ξ^i) the desired nonuniform distribution. The rest of the letter describes an alternative approach based on *two* uniform and independent distributions: (x^i) and (y^i). In this latter approach a value x^i from the first set is accepted when a value y^i from the second set satisfies the condition $y^i \leq f(x^i)$, where $f(\xi^i)\,d\xi$ is the density of the desired distribution function. (It should be noted that in von Neumann's example for forming the random pairs $\xi = \sin x$ and $\eta = \cos x$, he probably meant to say that x is equidistributed between 0 and 360 degrees (rather than "300"). Also, his notation for the tangent function is "tg," so that the second set of equations for ξ and η are just half-angle ($y = x/2$) trigonometric identities.)



THE INSTITUTE FOR ADVANCED STUDY
SCHOOL OF MATHEMATICS
PRINCETON, NEW JERSEY

May 21, 1947

Mr. Stan Ulam
Post Office Box 1663
Santa Fe
New Mexico

Dear Stan:

Thanks for your letter of the 19th. I need not tell you that Klari and I are looking forward to the trip and visit at Los Alamos this Summer. I have already received the necessary papers from Carson Mark. I filled out and returned mine yesterday; Klari's will follow today.

I am very glad that preparations for the random numbers work are to begin soon. In this connection, I would like to mention this: Assume that you have several random number distributions, each equidistributed in $0, 1$: $(x^i), (y^i), (z^i), \ldots$. Assume that you want one with the distribution function (density) $f(\xi)\,d\xi : (\xi^i)$. One way to form it is to form the cumulative distribution function: $g(\xi) = \int^{\xi} f(\xi)\,d\xi$ to invert it $h(x) = \xi \rightleftarrows x = g(\xi)$, and to form $\xi^i = h(x^i)$ with this $h(x)$, or some approximant polynomial. This is, as I see, the method that you have in mind.

An alternative, which works if ξ and all values of $f(\xi)$ lie in 0, 1, is this: Scan pairs x^i, y^i and use or reject x^i, y^i according to whether $y^i \leqq f(x^i)$ or not. In the first case, put $\xi^j = x^i$ in the second case form no ξ^j at that step.

The second method may occasionally be better than the first one. In some cases combinations of both may be best; e.g., form random pairs
$$\xi = \sin x, \quad \eta = \cos x$$
with x equidistributed between 0° and 300°. The obvious way consists of using the sin - cos - tables (with interpolation). This is clearly closely related to the first method. This is an alternative procedure:

Put
$$\xi = \frac{2t}{1+t^2}, \quad \eta = \frac{1-t^2}{1+t^2}, \quad t = tg\, y,$$
with y (which is $\frac{x}{2}$) equidistributed between 0° and 180°. Restrict y to 0° to 45°. Then the ξ, η will have to be replaced randomly by η, ξ and again by $\pm\xi, \pm\eta$. This can be done by using random digits 0, . . . , 7. It is also feasible with random digits 0, . . . , 9:

0	Replace ξ, η by	ξ, η
1	"	$-\xi, \eta$
2	"	$\xi, -\eta$
3	"	$-\xi, -\eta$
4	"	η, ξ
5	"	$\eta, -\xi$
6	"	$-\eta, \xi$
7	"	$-\eta, -\xi$
8	Reject this digit	
9	" " "	

Now $t = tg\, y$, $0° \leqq y \leqq 45°$, lies between 0 and 1, and its distribution function is $\frac{dt}{1+t^2}$. Hence one may pick pairs of numbers t, s both (independently) equidistributed between 0 and 1, and then

$$\left.\text{use } t\right\} \text{ for } (1+t^2)\,s \leqq 1$$
$$\left.\begin{array}{l}\text{reject } t, s \text{ and} \\ \text{form no } t \text{ at} \\ \text{this step}\end{array}\right\} \text{ for } (1+t^2)\,s > 1$$

Of course, the first part requires a divider, but the method may still be worth keeping in mind, especially when the ENIAC is available.

* * *

With best regards from house to house.

Yours, as ever,

John

~~John von Neumann~~

JvN:MW

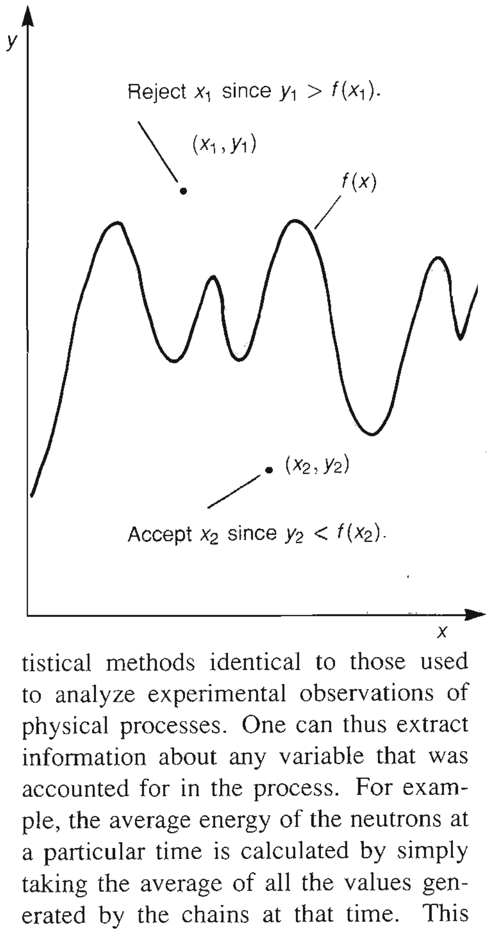
proposed by Stan, uses the inverse of the desired function $f = g^{-1}$. For example, to get the exponentially decreasing distribution of random numbers on the interval $(0, \infty)$ needed for path lengths, one applies the inverse function $f(x) = -\ln x$ to a uniform distribution of random numbers on the open interval $(0, 1)$.

What if it is difficult or computationally expensive to form the inverse function, which is frequently true when the desired function is empirical? The rest of von Neumann's letter describes an alternative technique that will work for such cases. In this approach *two* uniform and independent distributions (x^i) and (y^i) are used. A value x^i from the first set is accepted when a value y^i from the second set satisfies the condition $y^i \leq f(x^i)$, where $f(\xi^i)\,d\xi$ is the density of the desired distribution function (that is, $g(x) = \int f(x)dx$).

This acceptance-rejection technique of von Neumann's can best be illustrated graphically (Fig. 4). If the two numbers x^i and y^i are selected randomly from the domain and range, respectively, of the function f, then each pair of numbers represents a point in the function's coordinate plane (x^i, y^i). When $y^i > f(x^i)$ the point lies above the curve for $f(x)$, and x^i is rejected; when $y^i \leq f(x^i)$ the point lies on or below the curve, and x^i is accepted. Thus, the fraction of accepted points is equal to the fraction of the area below the curve. In fact, the proportion of points selected that fall in a small interval along the x-axis will be proportional to the average height of the curve in that interval, ensuring generation of random numbers that mirror the desired distribution.

THE ACCEPTANCE-REJECTION METHOD

Fig. 4. If two independent sets of random numbers are used, one of which (x^i) extends uniformly over the range of the distribution function f and the other (y^i) extends over the domain of f, then an acceptance-rejection technique based on whether or not $y^i \leq f(x^i)$ will generate a distribution for (x^i) whose density is $f(x^i)\,dx^i$.


y
Reject x_1 since $y_1 > f(x_1)$.
(x_1, y_1)
$f(x)$
(x_2, y_2)
Accept x_2 since $y_2 < f(x_2)$.
x

After a series of "games" have been played, how does one extract meaningful information? For each of thousands of neutrons, the variables describing the chain of events are stored, and this collection constitutes a numerical model of the process being studied. The collection of variables is analyzed using statistical methods identical to those used to analyze experimental observations of physical processes. One can thus extract information about any variable that was accounted for in the process. For example, the average energy of the neutrons at a particular time is calculated by simply taking the average of all the values generated by the chains at that time. This value has an uncertainty proportional to $\sqrt{V/(N-1)}$, where V is the variance of, in this case, the energy and N is the number of trials, or chains, followed.

It is, of course, desirable to reduce statistical uncertainty. *Any* modification to the stochastic calculational process that generates the same expectation values but smaller variances is called a variance-reduction technique. Such techniques frequently reflect the addition of known physics to the problem, and they reduce the variance by effectively increasing the number of data points pertinent to the variable of interest.

An example is dealing with neutron absorption by weighted sampling. In this technique, each neutron is assigned a unit "weight" at the start of its path. The weight is then decreased, bit by bit at each collision, in proportion to the absorption cross section divided by the total collision cross section. After each collision an outcome *other* than absorption is selected by random sampling and the path is continued. This technique reduces the variance by replacing the sudden, one-time process of neutron absorption by a gradual elimination of the neutron.

Another example of variance reduction is a technique that deals with outcomes that terminate a chain. Say that at each collision *one* of the alternative outcomes terminates the chain and associated with this outcome is a particular value x_t for the variable of interest (an example is x_t being a path length long enough for the neutron to escape). Instead of collecting these values *only* when the chain terminates, one can generate considerably more data about this particular outcome by making an extra calculation at *each* decision point. In this calculation the know value x_t for termination is multiplied by the probability that that outcome will occur. Then *random* values are selected to continue the chain in the usual manner. By the end of the calculation, the "weighted values" for the terminating outcome have been summed over all decision points. This variance-reduction technique is especially useful if the probablity of the alternative in question is low. For example, shielding calculations typically predict that only one in many thousands of neutrons actually get through the shielding. Instead of accumulating those rare paths, the small probabilities that a neutron will get through the shield on its