Fourier transform method to determine the probability density function from a given set of random samples

K. Nanbu
Institute of Fluid Science, Tohoku University, Sendai 980-77, Japan
(Received 10 July 1995)

A method based on the Fourier transform theory is proposed to determine the probability density function from a given set of random samples. First the Fourier transform of the probability density function is obtained from the set. If the transform has a long tail, it is to be smoothed. Then an appropriate window is applied to the transform. The inverse of the resulting transform yields the probability density function. The present method is applied to several examples and is shown to be very effective.

PACS number(s): 02.70.Lq, 02.50.Ng, 51.50.+v, 47.45.−n

I. INTRODUCTION

The Monte Carlo particle simulation has been widely used in the analysis of rarefied flow [1,2] and glow discharge [3,4]. Each particle has velocity (energy) and position. A set of particles in a small cell of computational domain contains information reflecting the structure of the velocity distribution function at the center of the cell. Let $x_1, x_2, \ldots$ be the velocities (energies) of the particles in the cell and $h(x)$ be the velocity (energy) distribution function at the cell center. Then the set $\{x_1, x_2, \ldots\}$ should be regarded as a set of random samples taken out from the distribution function $h(x)$, which is mathematically the probability density function (PDF).

Often we need to determine the PDF from a given set of random samples [5,6]. Simple counting is the most widely used method to determine the PDF: the x space is divided into discrete intervals; the number of samples in each interval is counted. Recently Ventzek and Kitamori [7] proposed an elaborate method to use the distribution of samples in the interval. Here we present a quite different method based on the theory of the Fourier transform [8]. The effectiveness of our method is ascertained by several example calculations.

II. METHOD

Let $\{x_i; i=1, \ldots, N\}$ be a set of random samples for the probability density function $h(x)$, which is assumed to be a continuous function. The number N is called the sample size. It may be the number of particles at a fixed time for unsteady stochastic processes or the whole sum of the number of particles counted at successive sampling times for stationary processes. Here we propose a method to determine $h(x)$ from a given set $\{x_i\}$. If the sample size N is sufficiently large, $h(x)$ can be approximated by [9]

$$\bar{h}(x)=\frac{1}{N}\sum_{i=1}^{N}\delta(x-x_i)\ , \tag{1}$$

where δ is the delta function. Although $\bar{h}(x)$ has a discrete form, integrations of $h(x)$ and $\bar{h}(x)$ over an arbitrary interval yield the same value as $N\to\infty$. The Fourier transform $\bar{H}$ of Eq. (1) is [8]

$$\bar{H}(f)=\frac{1}{N}\sum_{i}\exp(-j2\pi x_i f)\ , \tag{2}$$

where $j=\sqrt{-1}$ and the summation is over $i=1$ to $i=N$ hereafter. Let $H(f)$ be the Fourier transform of $h(x)$. The principle idea of the present work is based on the fact that the functional form of $\bar{H}(f)$ tends to $H(f)$ as $N\to\infty$. The fact makes it possible to obtain an approximation of $h(x)$ from the inverse Fourier transform of $\bar{H}(f)$, not $H(f)$. Let us write $\bar{H}(f)=\bar{R}(f)+j\bar{I}(f)$. We then have

$$\bar{R}(f)=\frac{1}{N}\sum_{i}\cos 2\pi x_i f\ , \tag{3}$$

$$\bar{I}(f)=-\frac{1}{N}\sum_{i}\sin 2\pi x_i f\ . \tag{4}$$

Of course, $\bar{R}$ and $\bar{I}$ are continuous functions of f. Recall that $\bar{h}(x)$ is a discrete function. We now show that $\bar{R}(f)$ and $\bar{I}(f)$ are, respectively, good approximations of $R(f)$ and $I(f)$, where $H(f)=R(f)+jI(f)$. The details of the present method and its effectiveness are best explained by several examples.

A. Gaussian distribution

The velocity distribution function takes the following form in equilibrium:

$$h(x)=\frac{1}{\sqrt{\pi}}\exp(-x^2)\quad(-\infty<x<\infty)\ . \tag{5}$$

The Fourier transform of $h(x)$ is

$$R(f)=\exp(-\pi^2 f^2)\ . \tag{6}$$

The imaginary part vanishes. We make a set of random samples of size N from Eq. (5).

$$x_i=(-\ln U_1)^{1/2}\sin 2\pi U_2\quad(i=1,\ldots,N)\ . \tag{7}$$

Hereafter U_i $(i=1,2,\ldots)$ denotes the independent random number uniformly distributed between 0 and 1.

Substitution of Eq. (7) into Eq. (3) gives $\bar{R}(f)$. Let us determine the PDF from $\bar{R}(f)$. We suppose that the PDF is known in advance to be an even function of x. So we set $\bar{I}(f)=0$. Figure 1 shows $R(f)$ and $\bar{R}(f)$ for $N=10^2$, 10^3, and 10^4. The solid line represents the exact transform $R(f)$. The function $\bar{R}(f)$ depends on the set $\{x_i\}$ employed, especially for small sample size. To find the two $\bar{R}(f)$'s which show the largest positive and negative deviation from the exact $R(f)$, we tried 1000 independent sets of $\{x_i\}$ for $N=10^2$ and 100 sets for $N=10^3$ and $N=10^4$. The dashed and dotted lines in Fig. 1 are these worst two. The two are soon used to determine the error bound for the PDF.

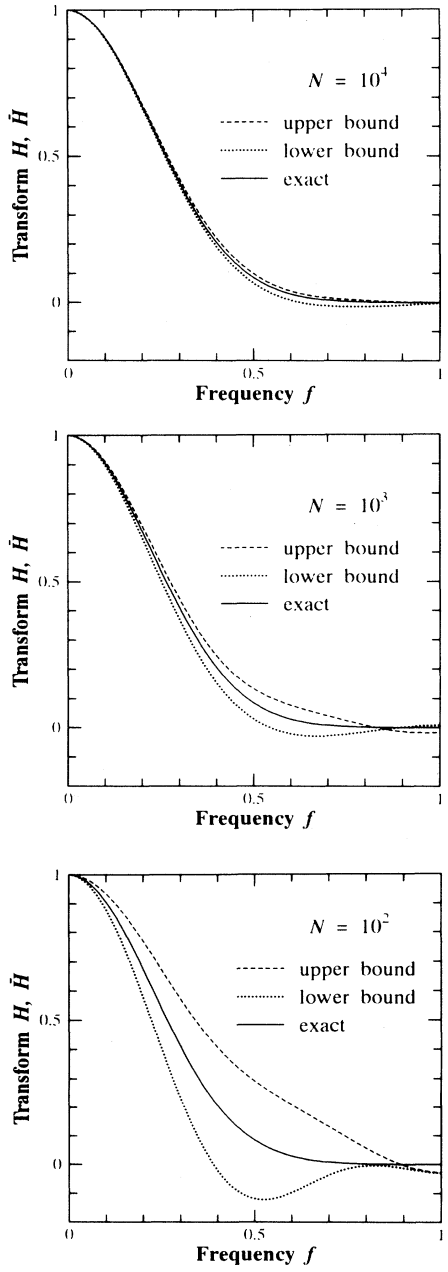

FIG. 1. Fourier transforms for $N=10^2$, 10^3, and 10^4.

Let us consider the inverse Fourier transform of $\bar{R}(f)$. As for the calculation of the inverse by the fast Fourier transform, see Ref. [8]. We denote the inverse by $\tilde{h}(x)$ since $\tilde{h}(x)$ is neither equal to the discrete $\bar{h}(x)$ nor to the exact $h(x)$. Let x_p be the period in space domain. It is divided into n equal intervals, so that the spatial sampling interval Δx is x_p/n. The period f_p in the frequency domain is $1/\Delta x$, the frequency sampling interval Δf being f_p/n $(=1/x_p)$. Both $\Delta x(=1/f_p)$ and Δf $(=1/x_p)$ should be small for the inverse to be accurate, which requires the use of large f_p and x_p. Our choice is $f_p=25$ and $n=2^{10}$, hence $\Delta x=0.04$ and $\Delta f=0.0244$.

We obtained the inverse $\tilde{h}(x)$ of $\bar{R}(f)$ and found that it is poor even for $N=10^4$. This is because the accuracy of $\bar{R}(f)$ is poor for large f. The rectangular window [8] $W(f)$ is applied to $\bar{R}(f)$ to cut off the inaccurate tail of $\bar{R}(f)$; $W(f)=1$ for $|f|\le f_c$ and $W(f)=0$ for $|f|>f_c$, f_c being 0.7 for all N. The dashed and dotted lines in Fig. 2 show the inverse $\tilde{h}(x)$ of $W(f)\bar{R}(f)$. These two lines denote the upper and lower bounds of $\tilde{h}(x)$ for a given N. The solid line represents the exact $h(x)$. The error bound is narrow for $N=10^4$ but wide for $N=10^2$. Some users of the Monte Carlo method say that they found an accurate PDF for such a small sample size as 10^2. However, such agreement should be regarded as accidental. One of the advantages of the present method is to make it possible to examine systematically the error bound of the PDF for a given sample size.

B. Bimodal distribution

The velocity distribution function $h(x)$ of electrons is bimodal in glow discharge [5]. It is

$$h(x)=A\left[\frac{\alpha}{\pi}\right]^{1/2}\exp(-\alpha x^2)+(1-A)\left[\frac{\beta}{\pi}\right]^{1/2}\exp(-\beta x^2)\quad(-\infty<x<\infty)\,, \tag{8}$$

where $0<A<1$, $\alpha>0$, and $\beta>0$. If x denotes a component of velocity, α and β are inversely proportional to the mean energy. The Fourier transform $H(f)$ is

$$H(f)=A\exp\left[-\frac{\pi^2 f^2}{\alpha}\right]+(1-A)\exp\left[-\frac{\pi^2 f^2}{\beta}\right]. \tag{9}$$

We consider the case of $A=\frac{1}{2}$. Then a set of random samples of size N for Eq. (8) is

$$x_i=(-\alpha^{-1}\ln U_1)^{1/2}\sin 2\pi U_2\quad(i=1,\ldots,N/2)\,, \tag{10a}$$

$$x_i=(-\beta^{-1}\ln U_3)^{1/2}\sin 2\pi U_4\quad(i=N/2+1,\ldots,N)\,. \tag{10b}$$

The procedure to obtain $\bar{H}(f)$ and its inverse $\tilde{h}(x)$ is the same as that of case A. The PDF is shown only for $N=10^4$ in Fig. 3. The parameters are $\alpha=1$ and $\beta=0.1$. Since β is small, the second term of Eq. (8) has a longer tail. The cut-off frequency f_c for the rectangular window is set of 0.7. The solid line in Fig. 3 represents the exact $h(x)$ and the dashed line the inverse $\tilde{h}(x)$. The latter is obtained for $f_p=25$ and $n=2^{11}$. We see that $\tilde{h}(x)$ is fairly accurate even for large x, where the PDF is small.

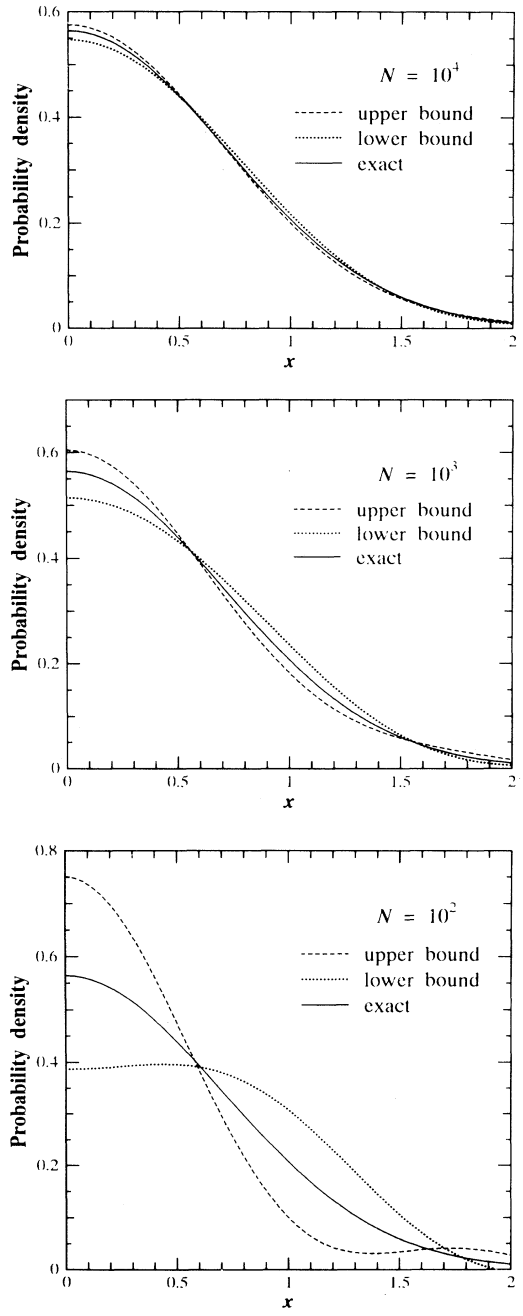

FIG. 2. Probability density functions for $N=10^2$, 10^3, and 10^4.

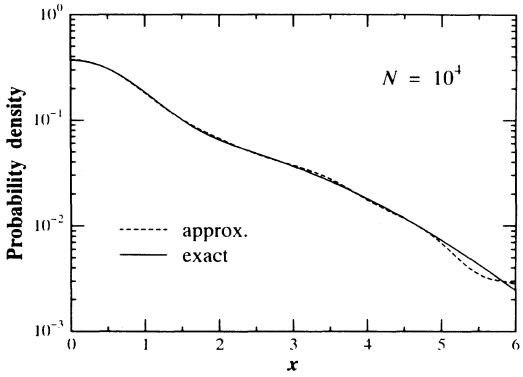

FIG. 3. Probability density function for $N=10^4$.

C. Exponential distribution

The Fourier transforms $H(f)$ for cases A and B rapidly decreases as f increases. A special treatment is necessary if $H(f)$ decays slowly. Such an example is

$$h(x)=e^{-x} \quad (x \geq 0) . \tag{11}$$

The transforms are

$$R(f)=\frac{1}{1+(2\pi f)^2} , \quad I(f)=-\frac{2\pi f}{1+(2\pi f)^2} . \tag{12}$$

We see that $R \sim f^{-2}$ and $I \sim f^{-1}$ as $f \to \infty$. Also we should be careful at the discontinuity of $h(x)$ at $x=0$ [8]. The samples are

$$x_i=-\ln U_1 \quad (i=1, \ldots, N) .$$

The dashed lines in Fig. 4 show the real part $\bar{R}(f)$ for $N=10^2$, 10^3, and 10^4. The imaginary part $\bar{I}(f)$ is shown in Fig. 5. The solid lines represent the exact transform of Eq. (12). We see that $\bar{H}(f)$ is accurate for $N=10^4$ but rather poor for $N=10^2$ and 10^3 except the region of small f. [Note that we here omit the consideration of the error bound for $\bar{H}(f)$; $\bar{H}(f)$ for each N is obtained only for one set of $\{x_i\}$.] In case of $N=10^2$ and 10^3 the inverse of $\bar{H}(f)$ is poor if $\bar{H}(f)$ is used as it stands. Moreover, since the transform has a long tail, we cannot apply the rectangular window to $\bar{H}(f)$ to cut off the inaccurate part of $\bar{H}(f)$. We propose the following method to treat the inaccurate tail of $\bar{H}(f)$. The method is applicable, in general, whenever $\bar{H}(f)$ decays slowly.

Let $\bar{R}(f)$ and $\bar{I}(f)$ be the real and imaginary parts of $\bar{H}(f)$. We use $\bar{R}(f)$ and $\bar{I}(f)$ up to $f=f_0$, at which $\bar{R}(f)$ and $\bar{I}(f)$ are still accurate. The frequency f_0 can be estimated from the behavior of $\bar{R}(f)$ and $\bar{I}(f)$. Since $\bar{R}(f)$ is an even function and $\bar{I}(f)$ is an odd function of f, we write for $f>f_0$ as

$$\bar{R}(f)=\bar{R}(f_0)g^2(1-a+ag^2) , \tag{13a}$$

$$\bar{I}(f)=\bar{I}(f_0)g(1-b+bg^2) , \tag{13b}$$

where $g=f_0/f$. Higher order terms are disregarded. Equation (13) holds at $f=f_0$. The coefficients a and b

are determined by means of the least-squares error-regression analysis of the original data of $\bar{R}(f)$ and $\bar{I}(f)$ for $f > f_0$. They are

$$a = \frac{\sum_k R_k g_k^2(g_k^2-1) - \sum_k g_k^4(g_k^2-1)}{\sum_k [g_k^2(g_k^2-1)]^2} ,$$

$$b = \frac{\sum_k I_k g_k(g_k^2-1) - \sum_k g_k^2(g_k^2-1)}{\sum_k [g_k(g_k^2-1)]^2} ,$$

where $R_k = \bar{R}(f_k)/\bar{R}(f_0)$, $I_k = \bar{I}(f_k)/\bar{I}(f_0)$, $g_k = f_0/f_k$, and f_k's ($> f_0$) are appropriate sampling frequencies such as $f_k = f_0 + 0.01k$, k being 1,2, ...

In the case of $N = 10^2$ and 10^3 the functions $\bar{R}(f)$ and $\bar{I}(f)$ are smoothed for $f > f_0$ by use of Eq. (13). The dotted lines in Figs. 4 and 5 represent Eq. (13). The connecting frequency f_0 is 0.4 for $N = 10^2$ and 0.6 for $N = 10^3$. In the case of $N = 10^4$ there is no need to apply such a smoothing.

Let us consider the window $W(f)$. Since the Fourier transform for case C has a long tail, $W(f)$ is best applied to the whole period in the frequency domain, i.e., to

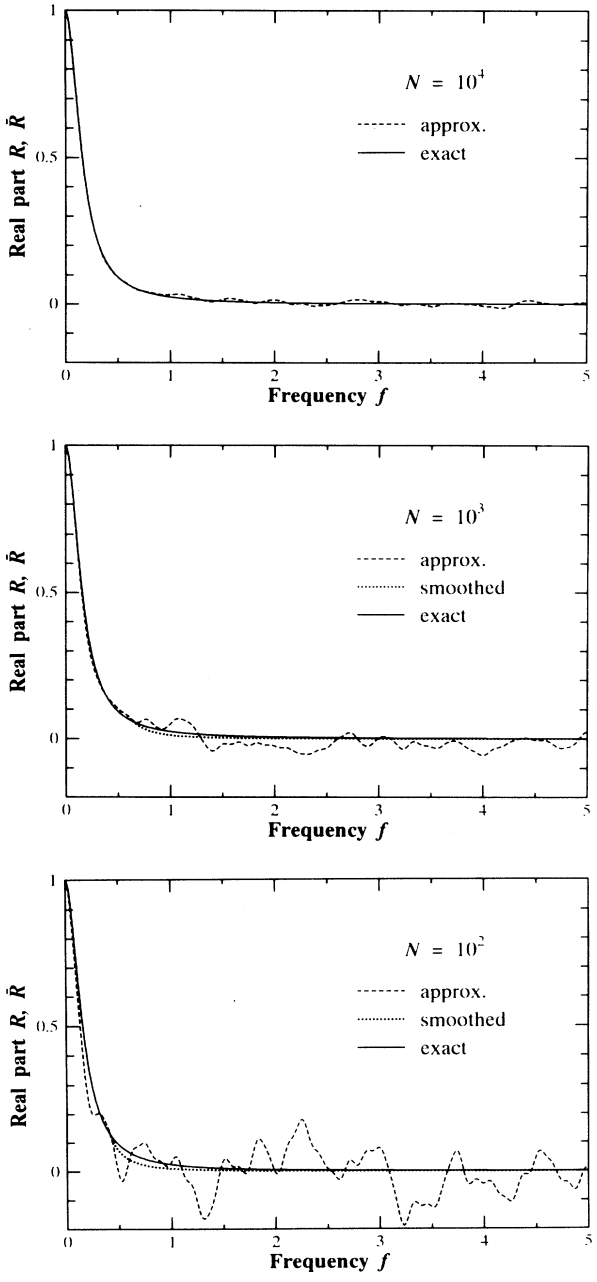

FIG. 4. Real parts of the Fourier transform for $N = 10^2$, 10^3, and 10^4.

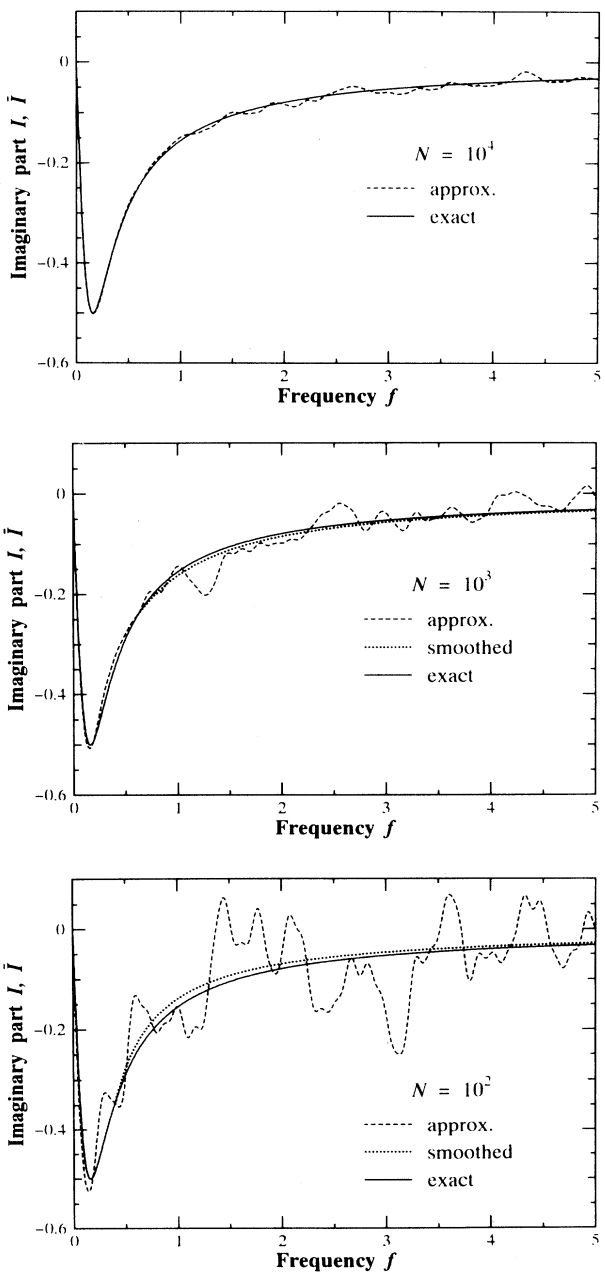

FIG. 5. Imaginary parts of the Fourier transform for $N = 10^2$, 10^3, and 10^4.

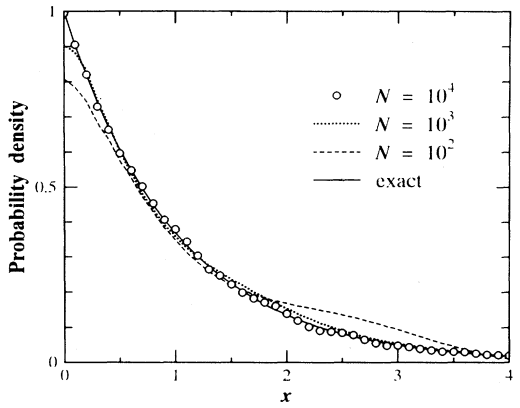

FIG. 6. Probability density functions for $N=10^2$, 10^3, and 10^4.

$|f| \le f_p/2$. When the transform has a long tail, the rectangular window $W(f)$ is not the best one; the inverse of $W(f)\bar{H}(f)$ is oscillatory in a region of small x in the space domain. This is true even if we consider the inverse of $W(f)H(f)$, where $H(f)$ is the exact transform. We found that the oscillation is fairly damped if the Hanning window [8] with the following form is used:

$$W(f)=0.8+0.2\cos(2\pi f/f_p)\ . \tag{14}$$

Now the inverse of $W(f)\bar{H}(f)$ is almost satisfactory but it still shows a slight oscillation. To eliminate this, the inverse is locally smoothed by use of the least-squares analysis. The resulting PDF is shown in Fig. 6 for $N=10^2$ (dashed line), $N=10^3$ (dotted line), and $N=10^4$ (symbols), the solid line being the exact PDF. These are obtained for $f_p=10$ and $n=2^7$. Not only the PDF for $N=10^4$ but also that for $N=10^3$ is satisfactory. Note that $\bar{H}(f)$ for $N=10^4$ is not smoothed in frequency domain. If smoothed, the result would be better.

D. Equilibrium distribution plus smeared beam

In glow discharge electrons in bulk plasma is divided into two groups, thermal electrons and high-energy beam electrons [3]. A model of energy distribution function for such electrons is

$$h(x)=Ah_1(x)+(1-A)h_2(x)\ ,$$

where $0<A<1$, $h_1(x)$ is the equilibrium distribution, and $h_2(x)$ is the Gaussian distribution with narrow width, i.e.,

$$h_1(x)=2\alpha(\alpha x/\pi)^{1/2}e^{-\alpha x}\quad (x\ge 0)\ , \tag{15}$$

$$h_2(x)=(\beta/\pi)^{1/2}\exp[-\beta(x-x_0)^2]\quad (|x-x_0|\le x_0)\ . \tag{16}$$

The function $h_2(x)$ is defined to be zero for $|x-x_0|>x_0$ but Eq. (16) can be used for any x since x_0 is assumed large here. Let $H_1(=R_1+jI_1)$ be the transform of h_1. We have

$$R_1=r\cos\theta\ ,\quad I_1=-r\sin\theta\ ,$$

where $r=\alpha^{3/2}/[\alpha^2+(2\pi f)^2]^{3/4}$ and $\theta=\frac{3}{2}\tan^{-1}(2\pi f/\alpha)$. The transform of h_2 is

$$H_2=\exp\left[-j2\pi f x_0-\frac{\pi^2 f^2}{\beta}\right]\ .$$

We consider the case of $A=\frac{1}{2}$. Then a set of random samples of size N are

$$x_i=-\alpha^{-1}[\ln U_1+(\sin 2\pi U_2)^2\ln U_3]\quad (i=1,\ldots,N/2)\ ,$$

$$x_i=x_0+(-\beta^{-1}\ln U_4)^{1/2}\sin 2\pi U_5\quad (i=N/2+1,\ldots,N)\ .$$

Here we consider the case of $\alpha=1$, $\beta=\pi$, and $x_0=15$; the mean energy of beam electrons is 10 times higher than that of thermal electrons. To save space we limit our attention to the case of $N=10^4$. Let H $(=R+jI)$ be the exact transform and $\bar{H}$ $(=\bar{R}+j\bar{I})$ be the approximate one obtained from Eqs. (3) and (4). Figure 7 shows the real and imaginary parts of the transform. The solid lines represent the exact R and I and the dashed lines the approximate $\bar{R}$ and $\bar{I}$. We see that the damping rates for $\bar{R}$ and $\bar{I}$ are higher than those for R and I. The functions $\bar{R}$ and $\bar{I}$ are smoothed for $f>1$ by use of Eq. (13). We see from Eq. (13) that $\bar{R}\to f^{-2}$ and $\bar{I}\to f^{-1}$ as $f\to\infty$. The

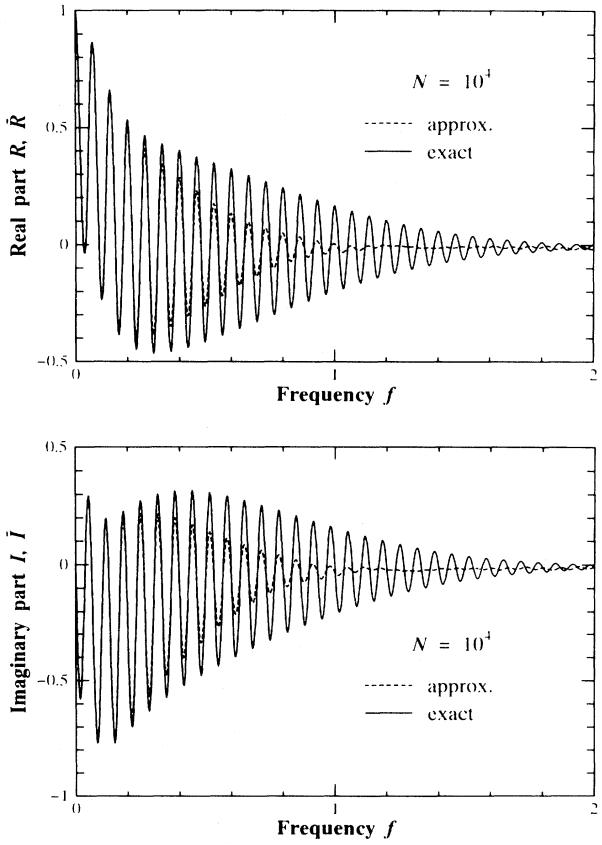

FIG. 7. Real and imaginary parts of the Fourier transform for $N=10^4$.

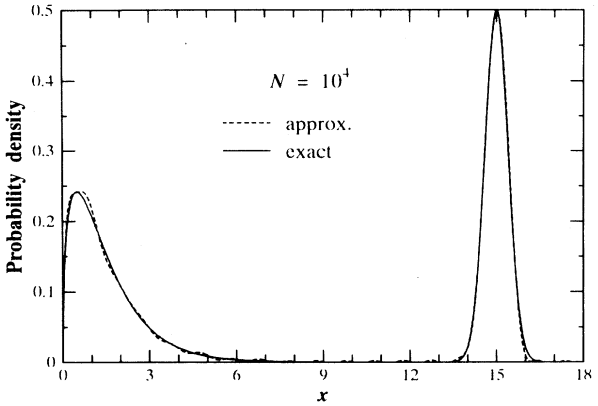

FIG. 8. Probability density function for $N=10^4$.

exact R and I tend to $f^{-3/2}$ as $f\rightarrow\infty$. Nevertheless we use Eq. (13) because we do not know the asymptotic form of $\bar{R}$ and $\bar{I}$ in practical applications. After smoothing, $\bar{H}$ is multiplied by the window, Eq. (14). The inverse of $W\bar{H}$ is accurate enough. However, the local smoothing in space domain slightly improves the PDF. The final PDF is shown in Fig. 8 by the dashed line compared with the exact PDF (solid line). The calculation is done for $f_p=20$ and $n=2^9$. Although $\bar{R}$ and $\bar{I}$ show an appreciable deviation from the exact R and I in frequency domain, the PDF is fairly accurate in the space domain.

E. Application to actual simulation of rf discharge

We are now trying to apply the present method to an actual particle simulation of rf discharge [10]. The energies of ions incident on the grounded electrode are sampled for various discharge conditions. Here we consider only one case: the peak-to-peak rf voltage is 2000 V, the rf frequency is 13.56 MHz, gas (argon) pressure is 42 mTorr, and the distance between parallel electrodes is 64 mm. The number of sampled ions, N, is 12 984. The PDF's of ion energy resulting from the simple counting and the Fourier transform method are compared in Fig. 9. Both are obtained for the same energy interval Δx ($=0.25$ eV). The procedure to determine the transform and its inverse is the same as that in case C. The parameters are now chosen as $f_p=4$, $n=2^{13}$, $f_0=0.3$, and $f_k=f_0+0.0005k$. Since the transform shows a rapid oscillation, a large n (small Δf) is used. Note that Eqs. (13) and (14) are employed without any modification. These equations are applicable in general if the Fourier transform decays slowly. Smoothing the inverse is omitted because it has no effect. We see from Fig. 9 that the noise in the PDF resulting from the present method is much smaller than that of the simple counting. Several spikes in the PDF are typical in the rf discharge. It is most probable that the present method is applicable to the PDF with spikes or discontinuities.

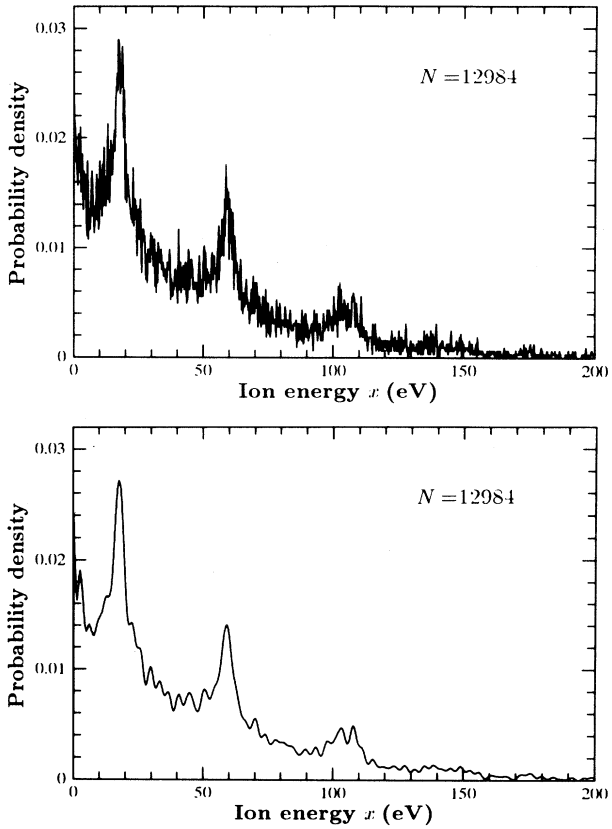

FIG. 9. Probability density functions for $N=12\,984$. The upper is the PDF from the simple counting and the lower is that from the Fourier transform method.

III. CONCLUSIONS

A method to determine the probability density function (PDF) from a given set of random samples is proposed. First, the effect of sample size on the PDF is systematically examined. Next, certain rules for the use of the present method are clarified through various examples. Let $\bar{H}(f)$ be the Fourier transform of the PDF expressed by a set of random samples. General rules are as follows.

(1) When $\bar{H}(f)$ decays rapidly, apply the rectangular window to cut off the inaccurate part of $\bar{H}(f)$ and then obtain the inverse Fourier transform.

(2) When $\bar{H}(f)$ has a long tail, smooth the tail in the frequency domain by using Eq. (13), apply the Hanning window of Eq. (14) over the whole period, and obtain the inverse. If necessary, smooth the PDF in the space domain.

It may be straightforward to extend the present method to the case of two- or three-dimensional probability density function.

[1] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon, Oxford, 1994).

[2] K. Nanbu, in *Proceedings of the 15th International Symposium on Rarefied Gas Dynamics,* edited by V. Boffi and C. Cercignani (Teubner, Stuttgart, 1986), Vol. 1.

[3] G. G. Lister, J. Phys. D **25**, 1649 (1992).

[4] C. K. Birdsall, IEEE Trans. Plasma Science **19**, 65 (1991).

[5] D. Vender and R. W. Boswell, IEEE Trans. Plasma Sci-

ence **18**, 725 (1990).
[6] A. Date, K. Kitamori, Y. Sakai, and H. Tagashira, J. Phys. D **25**, 442 (1992).
[7] P. L. G. Ventzek and K. Kitamori, J. Appl. Phys. **75**, 3785 (1994), and references cited therein.
[8] E. O. Brigham, *The Fast Fourier Transform and its Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[9] K. Nanbu, J. Phys. Soc. Jpn. **49**, 2042 (1980).
[10] K. Nanbu and Y. Kitatani, Vacuum (to be published).