Analysis of random laser scattering pulse signals with lognormal distribution

Yan Zhen-Gang(闫振纲)[†], Bian Bao-Min(卞保民)[‡], Wang Shou-Yu(王绶玙), Lin Ying-Lu(林颖璐), Wang Chun-Yong(王春勇), and Li Zhen-Hua(李振华)

Department of Information Physics and Engineering, Nanjing University of Science and Technology, Nanjing 210094, *China*

(Received 10 January 2013; revised manuscript received 13 March 2013)

The statistical distribution of natural phenomena is of great significance in studying the laws of nature. In order to study the statistical characteristics of a random pulse signal, a random process model is proposed theoretically for better studying of the random law of measured results. Moreover, a simple random pulse signal generation and testing system is designed for studying the counting distributions of three typical objects including particles suspended in the air, standard particles, and background noise. Both normal and lognormal distribution fittings are used for analyzing the experimental results and testified by chi-square distribution fit test and correlation coefficient for comparison. In addition, the statistical laws of three typical objects and the relations between them are discussed in detail. The relation is also the non-integral dimension fractal relation of statistical distributions of different random laser scattering pulse signal groups.

Keywords: random pulse signal, statistical self-similarity, lognormal distribution, laser scattering

PACS: 05.40.Ca, 02.50.–r, 05.45.Df **DOI:** 10.1088/1674-1056/22/6/060505

1. Introduction

Studying the statistical distribution of natural phenomena is very important, for it could help us understand the laws of nature. Gaddum[1] has shown the importance of lognormal distribution and measured many natural phenomena that fit well to lognormal distribution. In the area of geochemistry, Ahrens[2] analyzed the frequency distribution of elements in granite at various locations. The final results indicated that the frequency distribution is a lognormal distribution. Aitchison and Brown[3] have applied lognormal distribution to economic statistics and proposed a better model for economic analysis. Besides, Spratt and John[4] showed that growth rates, mass, and survival of human cancer in natural history approximately tend to be of lognormal distribution. Becker *et al.*[5] obtained that the distributions of city size fit to the lognormal distribution from both the census data and Nighttime Lights of the World data in the field of sociology. Peter[6] obtained droplet and particle imaging results in the lognormal distribution of pixel intensities. Limpert *et al.*[7] have particularized many measured results in different areas which fit well to lognormal distribution and also analyzed the similarities and differences between the normal distribution and lognormal distribution. Chen *et al.*[8] have studied sharp peak landscapes of the Eigen model from a new perspective about how the quasispecies are distributed in the sequence space and found that the distribution of clustering coefficient $C(k)$ follows lognormal distribution.

The distribution of random tiny particles is an important natural phenomenon which attracts lots of research interest. Till now, many researchers have done lots of work on particle size measurement. Granqvist and John[9] have observed ultra-fine metal powder using a microscope and found that particle size distribution satisfies the lognormal distribution. Moreover, combining with the central limit theorem, a mathematical model of lognormal distribution in small deviation was established.[9] Hwang *et al.*[10] have developed a simple and fast method of measuring the sizes of sub-micron particles. Real-time measurement of sub-micron aerosol particles presented a lognormal distribution of particle size.[10] Baker *et al.*[11] have measured the particle size distribution of a small dispersion particle group which obeys the lognormal distribution. In our previous work[12–14] we studied the statistical distribution of the millions of particles, and the results showed that the random signal characteristics fit well to the lognormal distribution.[12] We[13] have also calculated the normalized coefficient of the fitting statistical distribution and measured distributions under different threshold voltages experimentally. Besides, we[14] have studied the signal amplitude distributions of aerosol scattering pulses and signal width distributions of background noise. They match well with the lognormal distribution. Though two physical parameters are discussed, the amplitude and width distributions belong to different facts which could not show the relation between measuring objects and instruments in the same measuring system but in different measuring processes.

In the above work the statistical law of the random particle size distributions has been studied by using different methods and measuring instruments, many results have concluded

[†]Corresponding author. E-mail: yzg.njust@gmail.com
[‡]Corresponding author. E-mail: bianbaomin_56@yahoo.com.cn

© 2013 Chinese Physical Society and IOP Publishing Ltd

that the particle size distributions obey lognormal distribution. However, they only studied the statistical distribution law on a single specific physical parameter, which only reflects partial properties of the natural phenomenon. However, a natural phenomenon often contains more than one characteristic at the same time. Besides, the fitting functions are chosen based on the curve shape without a fitting test. As far as we know, the lognormal distribution function has not previously been compared with experimental size distributions for several typical processes in the same device, in spite of the numerous articles published on this subject.

In order to have a better understanding of a natural phenomenon, it becomes important to study statistical distribution laws on all characteristics and the relations of these statistical distribution laws. In this paper, we firstly propose a random process model theoretically which clearly provides the relation between measured results and the different random factors in the measuring system. Then, we design a simple random pulse signal generation and testing system. This system has characteristics of numerous data sampling, high measurement accuracy, short measurement time, simple operation, easy control of the parameters, and good repeatability. In addition, many different random objects can be measured in the same system, such as the scattering pulse signal of a particle suspended in the air, the pulse signal of a standard particle, and the background noise pulse signal of a photoelectric sensor. Finally, we analyze statistical laws of three typical objects and the relations between them, including the relation between amplitude and width distributions, the relation between global and local distributions of a particle suspended in the air and the background noise, and the relation between subsets of amplitude and width distributions of the same object. This work is significant for analyzing random signals and particle measurements.

2. Statistical model of random pulse signal

Generally, random signals are composed of a set of $f(t)$, which are time series recorded values. Figure 1 shows the partial data of the random set $\{V_{\text{noise}}(t)\}$.

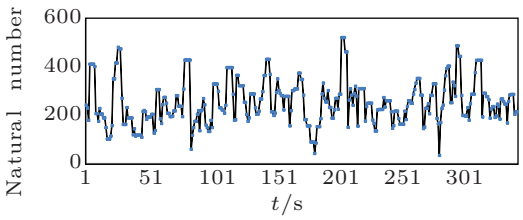


Fig. 1. (color online) Partial data of the background noise signals $\{V_{\text{noise}}(t)\}$. The x axis represents the time series and the y axis refers to the natural number representing the amplitude of signal.

The number of the data capacity of output signal $\{V_{\text{noise}}(t)\}$ is N. The time interval of contiguous sampling data is $\Delta\tau$. The difference between two contiguous sampling data is ΔV. The whole measuring time consumption is $t = \sum_{i=1}^{N} \Delta\tau$. The maximum and the minimum values of $\{V_{\text{noise}}(t)\}$ are $V_{\max}$ and $V_{\min}$ ($V_{\max} \geq V_i \geq V_{\min}$), respectively, which are both limited. In the experiment, $V_{\text{noise}}(t)$ could be regarded as $\{V_{\text{noise}}(t) \equiv V_{\text{c}} + V_{\Delta}\}$, where V_{c} is the constant part while V_{Δ} is the random fluctuating part. $V_{\text{noise}}(t)$ is influenced by various random factors $(A_1, A_2, \ldots, A_J)$, and these random factors could also be divided into constant part $\{A_{1c}, A_{2c}, \ldots, A_{Jc}\}$ and random fluctuating part $\{\Delta_1, \Delta_2, \ldots, \Delta_J\}$.

$$\{A_1, A_2, \ldots, A_J\} \equiv \{A_{1c}, A_{2c}, \ldots, A_{Jc}\} + \{\Delta_1, \Delta_2, \ldots, \Delta_J\}, \quad (\Delta_j \geq \delta_j). \tag{1}$$

$\{\delta_1, \delta_2, \ldots, \delta_J\}$ represents the minimum measuring unit of actions, $\{A_{1c}, A_{2c}, \ldots, A_{Jc}\}$ keeps constant in the measuring environment of the random process, and $\{\Delta_1, \Delta_2, \ldots, \Delta_J\}$ is uncertainty. Therefore, V_{c} corresponds to $\{A_{1c}, A_{2c}, \ldots, A_{Jc}\}$ and V_{Δ} corresponds to $\{\Delta_1, \Delta_2, \ldots, \Delta_J\}$ and $\{V_{\Delta}\}$ is generated by the interaction among random factors $(A_1, A_2, \ldots, A_J)$.

Moreover, we set δ_V as the minimum measuring unit of the output signal, so the maximum and minimum values of the output signal can be written as $V_{\max} \equiv l_{\max}\delta_V$ and $V_{\min} \equiv l_{\min}\delta_V$, where l_i $(l_i \geq 1)$ is natural number. If the capacity of the sampling data is large enough, the minimum value of the output signal can be treated as the measuring accuracy of the system ($V_{\min} = \delta_V$ and $l_{\min} = 1$). Generally, the minimum value could not attain the measuring accuracy, but the minimum value should be l_i times the measuring accuracy. Since the mathematical expectation of the constant part $E(V_{\text{c}}) = \text{const}$ and deviation $D(V_{\text{c}}) = 0$, $V_{\text{noise}}(t)$ without the constant part could be mapped to the mathematical model

$$\frac{V_{\text{noise}} - V_{\text{c}} + \delta_V}{\delta_V} \equiv l_i. \tag{2}$$

In Eq. (2), l_i is the natural number and its minimal value is 1. The number of total counting channels in the measuring system is $l_{\max} \equiv (V_{\max} - V_{\min})/\delta_V + 1$. In the range $(l_{\min} \leq l \equiv V/\delta_V \leq l_{\max})$, counting statistics of $l(i)$ is done according to the size order of l_i, the frequency of l_i in $\{l(i)\}$ is N_{l_i}, the number of the whole data is $N \equiv \sum_{l_{\min}}^{l_{\max}} N_{l_i}$. The distribution density function of $\{V_{\text{noise}}(t)\}$ is

$$\tilde{p}_V(l_i) \equiv \lim_{N > N_{l_i} \to \infty} \frac{N_{l_i}}{N} > 0, \quad (l_{\min} \leq l_i \leq l_{\max}). \tag{3}$$

Discrete distribution $\tilde{p}_V(l_i)$ becomes stable with the increase of N. As in the limited measuring range $[V_{\min}, V_{\max}]$, when the number of total counting channels $\lim_{\substack{\delta_V \to 0 \\ N \to \infty}} l_{\max} \to \infty$, $\tilde{p}_V(l_i)$ tends to become the continuous form of density function $p(l_i)$. Since $\{\delta_1, \delta_2, \ldots, \delta_J\}$ is the minimum measuring unit of $\{A_1, A_2, \ldots, A_J\}$, $\{A_1, A_2, \ldots, A_J\}$ could be mapped by

a natural number set $\{l_{A_1}, l_{A_2}, \ldots, l_{A_J}\}$. $\{p_1, p_2, \ldots, p_J\}$ is a set of statistical distribution density functions corresponding to $\{A_1, A_2, \ldots, A_J\}$.

Generally, the output signal $V(t)$ of a laser particle counter is induced by a series of random independent factors, usually, including the shape randomness of particle P, the equivalent light intensity I when the particle passes through photosensitive area, the equivalent cross section of photosensitive area S, the homogeneity of photosensitive area ε, scattered light collection efficiency by sensor ρ, the photoelectric transformation efficiency of detector η, the gain of the circuit H, etc.

We find that these random factors have positive correlation with the output signals while with different weights to measurement results:

$$V(t) \propto P^{\alpha_P} I^{\alpha_I} S^{\alpha_S} \varepsilon^{\alpha_\varepsilon} \rho^{\alpha_\rho} \eta^{\alpha_\eta} \cdots H^{\alpha_H}, \quad \alpha > 0. \tag{4}$$

Dimensionless constants α represent the relation with positive correlation between different random factors in a particle signal generating environment and output voltages. Let C indicate the stable measurement environment, then at time t the random process model of generating a single particle pulse signal with different random factors could be expressed as

$$C \equiv \frac{V(t)}{P(t)^{\alpha_P} I(t)^{\alpha_I} S(t)^{\alpha_S} \varepsilon(t)^{\alpha_\varepsilon} \rho(t)^{\alpha_\rho} \eta(t)^{\alpha_\eta} \cdots H(t)^{\alpha_H}}, \quad \alpha > 0. \tag{5}$$

Equation (5) indicates the influence of the changes of all random factors on measured result $V(t)$ in the whole measuring process of a single particle pulse signal. For a certain measuring process, C is determined. α represents the weight on measured results induced by different factors, and $\alpha > 1$ means the positive correlation relation between different factors and measured results. The randomness of measured results corresponds to the randomness of these independent factors. According to Eq. (2), Eq. (5) could be rewritten as

$$C = \frac{l_{V_i}}{(l_{P_i})^{\alpha_P} (l_{I_i})^{\alpha_I} (l_{S_i})^{\alpha_S} (l_{\varepsilon_i})^{\alpha_\varepsilon} (l_{\rho_i})^{\alpha_\rho} (l_{\eta_i})^{\alpha_\eta} \cdots (l_{H_i})^{\alpha_H}}, \quad \alpha > 0. \tag{6}$$

Equation (6) shows the random process of single particle pulse signal, while for N measuring signals in a sampling process, Eq. (6) should be

$$C \equiv \frac{\sqrt[N]{\prod_{i=1}^{N} l_{V_i}}}{\sqrt[N]{\prod_{i=1}^{N} (l_{P_i})^{\alpha_P}} \sqrt[N]{\prod_{i=1}^{N} (l_{I_i})^{\alpha_I}} \sqrt[N]{\prod_{i=1}^{N} (l_{S_i})^{\alpha_S}} \sqrt[N]{\prod_{i=1}^{N} (l_{\varepsilon_i})^{\alpha_\varepsilon}} \sqrt[N]{\prod_{i=1}^{N} (l_{\rho_i})^{\alpha_\rho}} \sqrt[N]{\prod_{i=1}^{N} (l_{\eta_i})^{\alpha_\eta}} \ldots \sqrt[N]{\prod_{i=1}^{N} (l_{H_i})^{\alpha_H}}}. \tag{7}$$

Equation (7) could be simplified as

$$\begin{aligned} C &\equiv \sqrt[N]{\frac{\prod_{i=1}^{N} l_{V_i}}{\prod_{i=1}^{N} \left[\prod_{j=1}^{J} l_{j_i}^{\alpha_j}\right]}} = \frac{\sqrt[N]{\prod_{i=1}^{N} l_{V_i}}}{\sqrt[N]{\prod_{j=1}^{J} \left[\prod_{i=1}^{N} l_{j_i}^{\alpha_j}\right]}} \\ &= \frac{\bar{l}_V}{\prod_{j=1}^{J} \sqrt[N]{\prod_{i=1}^{N} l_{j_i}^{\alpha_j}}} = \frac{\bar{l}_V}{\prod_{j=1}^{J} \overline{l^{\alpha_j}}}, \end{aligned} \tag{8}$$

where the statistical parameter $\bar{l}$ is the geometric mean and satisfies

$$\overline{l^{\alpha}} \equiv \sqrt[N]{\prod_{i=1}^{N} l_i^{\alpha}} = \left(\sqrt[N]{\prod_{i=1}^{N} l_i}\right)^{\alpha} = (\bar{l})^{\alpha}.$$

Applying a natural logarithm operation on both sides, we can obtain

$$\begin{aligned} \ln \overline{l^{\alpha}} &= \ln \left[\sqrt[N]{\prod_{i=1}^{N} (l_i)^{\alpha}}\right] = \ln \left[\prod_{i=1}^{N} (l_i)^{\alpha/N}\right] = \ln \left[\prod_{l=1}^{l_{\max}} l^{\alpha \tilde{p}(l)}\right] \\ &= \alpha \left[\sum_{l=1}^{l_{\max}} \ln l \cdot \tilde{p}(l)\right] = \alpha \ln \bar{l}. \end{aligned} \tag{9}$$

Equation (9) shows a nonlinear constraint relation with respect to $\tilde{p}(l)$. The mathematical expectation of $\ln l$ is obtained if substituting continuous $p(l)$ into equation

$$E(\ln l) = \lim_{\substack{\delta \to 0 \\ l_{\max} \to \infty}} \sum_{l=1}^{l_{\max}} \ln l \cdot \tilde{p}(l) = \int_1^{\infty} \ln l \cdot p(l)\, \mathrm{d}l = \ln \bar{l}. \tag{10}$$

In the measuring system influenced by a lot of independent random factors, we propose a stable measurement environment C in Eq. (5) which could be used for establishing a multi-factor random process model. The separation of the constant and random fluctuating parts suggests an easy way to study the random law of the measured results. The complicated measured results are simplified into a natural number sequence based on measuring accuracy. In order to study the random fluctuating part by statistical methods, the statistical distribution function should satisfy the characteristics of Eq. (9). Moreover, we have also introduced the geometric mean which is important in random study as a statistical parameter and should be positive: the positive geometric mean corresponds to the positive property of measured results.

3. Experimental system

In order to testify the proposed model and further study the random law, we use the laser particle counter as the main experimental system, which is shown in Fig. 2 with sampling rate 28.3 L/min. The laser scattering signals of the irregular particles are collected by the photomultiplier tube, and the output signals of the photoelectric sensor are acquired by the high-speed data acquisition card PCI-9812. Data are imported into a computer for multi-channel counting statistics.

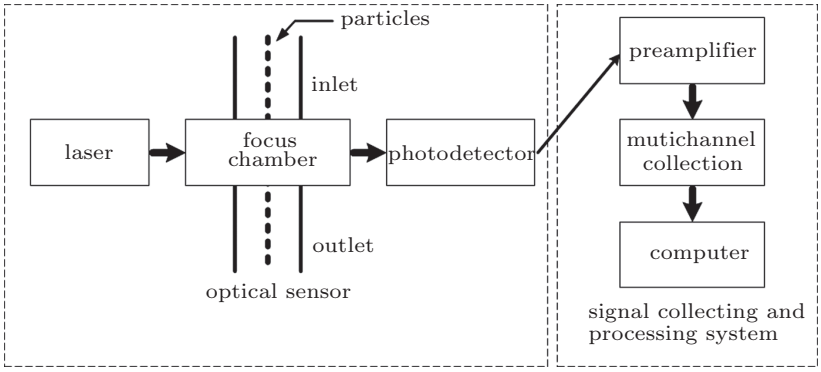


Fig. 2. Measuring system of the laser particle counter.

The experimental apparatus and measuring environment are easy to realize, and the measuring accuracy could be guaranteed. Besides, the system can obtain enough data with good repeatability. The higher the accuracy of the measuring system in recording signals, the larger sample size is measured in unit time, so the statistical characteristics of random signals will be more obvious.

The main technical parameters in experiments are listed as follows: a collimated continuous laser beam (808 nm, 50 mW) is used as the light source; the photosensitive area is about 0.5 mm×1 mm×8 mm; reversed biased voltage is set to be 12 V; dark current is no more than 10^{-9} A; sampling rate is 28.3 L/min; maximum range of voltage is 5 V, which is divided into 2048 counting channels and $\Delta V = 5$ V/2048 ≈ 2.44 mV; maximum signal sampling rate is 20 MHz (corresponding to the time-accuracy $\Delta\tau \sim 0.05$ μs); continuously sampled data of signal is up to 1×10^7; a 1000 level ultra-clean room is the measuring environment only for measuring the background noise.

The basic characteristics of random signals are usually studied by the Fourier spectrum analysis method based on the time dimension. Besides, our group has proposed another method called the two-parameter model.[13] In our method, a pulse signal could be discerned from a series of signals by choosing a suited threshold V_0 as shown in Fig. 3.

The obtained pulse unit has only two basic parameters: amplitude V_j and width τ_j. Once the threshold parameter V_0 is given, the measured signals $\{\tilde{f}(t,V_0)\}$ generally composed of a discrete sequence of pulses, could be mapped to $\{(V_j,\tau_j)\}$ uniquely. The element (V_j,τ_j) represents the basic geometric parameter of the pulse signal. The natural number j corresponds to the time sequence. According to Eq. (2), in the range $([V_{\min},V_{\max}],[\tau_{\min}=2\delta_\tau,\tau_{\max}])$, combining with measuring accuracy (δ_V,δ_τ), the signal set $\{(V_j,\tau_j)\}$ can be mapped to natural number set (l_{V_j},l_{τ_j}). As a result, a mathematical model of two-parameter set $\{(l_{V_j},l_{\tau_j})\}$ can be established as

$$\begin{cases} l_{V_j} = \dfrac{V_j - V_c + \delta_V}{\delta_V}, & l_{V_j} \in \left(1,2,\ldots,\dfrac{V_{\max}-V}{\delta_V}+1\right), \\ l_{\tau_j} = \dfrac{\tau_j - \tau_c}{\delta_\tau}, & l_{\tau_j} \in \left(1,2,\ldots,\dfrac{\tau_{\max}-\tau_c}{\delta_\tau}\right). \end{cases} \tag{11}$$

The values of (l_{V_j},l_{τ_j}) are finite natural numbers. It is obvious that $\{(l_{V_j},l_{\tau_j})\}$ is the mapping form from $\{\tilde{f}(t,V_0)\}$. In fact, Eq. (11) is known as a two-parameter model for a random signal pulse.

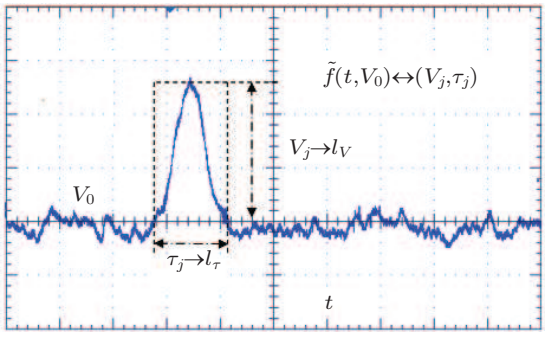


Fig. 3. (color online) Diagram of the two-parameter pulse model, where the x axis denotes the time series and the y axis represents the amplitude of the signal. The solid curve represents $\{\tilde{f}(t,V_0)\}$. The chosen pulse (V_j,τ_j) is in the box of the dotted line according to threshold V_0.

Supposing that N_{l_V} and N_{l_τ} are the numbers of signals which satisfy $V \in l_V\delta_V$ and $\tau \in l_\tau\delta_\tau$, respectively. According to Eq. (3), the probability distribution functions of the random

pulse signals can be expressed as

$$\tilde{p}(l_V) \equiv \frac{N_{l_V}}{N}, \quad \tilde{p}(l_\tau) \equiv \frac{N_{l_\tau}}{N}. \tag{12}$$

The values of $\tilde{p}(l_V)$ and $\tilde{p}(l_\tau)$ become stable with the increase of the number of N and at least one signal should be obtained in each measuring channel. In essence, this tendency of stability represents the stability of the statistical characteristics of the random pulse signals.

The widths of signals are different when the amplitudes from signals $\{N_{l_V}\}$ satisfy $V \in l_V \delta_V$. Therefore, the measuring channels could be further divided according to l_τ, and the distribution function of l_τ in subset N_{l_τ} is defined as

$$\tilde{p}(l_\tau, l_V = c) \equiv \frac{N_{l_V l_\tau}}{N_{l_V}} = \frac{1}{\tilde{p}(l_V)} \frac{N_{l_V l_\tau}}{N}. \tag{13}$$

Similarly, the distribution function of l_V in subset N_{l_V} should be

$$\tilde{p}(l_V, l_\tau = c) \equiv \frac{N_{l_\tau l_V}}{N_{l_\tau}} = \frac{1}{\tilde{p}(l_\tau)} \frac{N_{l_\tau l_V}}{N}, \tag{14}$$

where $N_{l_\tau l_V} = N_{l_V l_\tau}$, $N_{l_\tau l_V}$ and $N_{l_V l_\tau}$ are the numbers of pulse signals with width l_τ and amplitude l_V, respectively.

However, with the experimental system and two-parameter model, in a certain measuring environment, the interaction results between the measured samples and measuring system in which results are measured could be obtained only experimentally.

4. Experimental results

4.1. Statistical distributions of particles suspended in the air

The randomness of voltage pulse signals generated by the particles suspended in the air not only reflects the particle characteristics (mass, material, morphology, etc.), but also shows the random factors of the measuring process. The amplitude and width distributions of the random particle scattering pulse signal group are measured, and the experimental results are shown in Fig. 4.

As shown in Fig. 4, the discrete distribution $\tilde{p}(l)$ looks skewed and has obvious non-negative characteristics. So the normalized lognormal distribution is selected to fit the experimental results into continuous distributions. The well fitting results are shown by the solid lines. The fitting equation in Fig. 4(a) is

$$p(l_V) \equiv \frac{1}{\Omega_{L_V} \sigma_{\ln V} l_V} \exp\left[-\frac{\left(\ln l_V - \ln \bar{l}_V\right)^2}{2\sigma^2_{\ln V}}\right], \quad l_V \in (1, 2, \ldots L_V), \tag{15}$$

where Ω is the normalization coefficient whose value is related to both the range of the signal group characteristic parameter and the total number of channels (the accuracy of channel). The statistical parameters are defined by

$$\ln \overline{l_V} = \sum_{l_V=1}^{l_{V\max}} \ln l_V \cdot \frac{N_{l_V}}{N}, \tag{16}$$

and

$$\sigma_{\ln V} == \left[\sum_{l_V=1}^{l_{V\max}} (\ln l_V - \ln \overline{l_V}) \cdot \frac{N_{l_V}}{N}\right]^{1/2}, \tag{17}$$

where $N = 1825470$, $\ln \overline{l_V} = 5.77$, and $\sigma_{\ln V} = 0.87$. It should be noted that we use the geometric mean and geometric standard deviation for describing the distribution function instead of the arithmetic mean and standard deviation, in which way we could obtain a better fitting result. The geometric standard deviation is dimensionless.

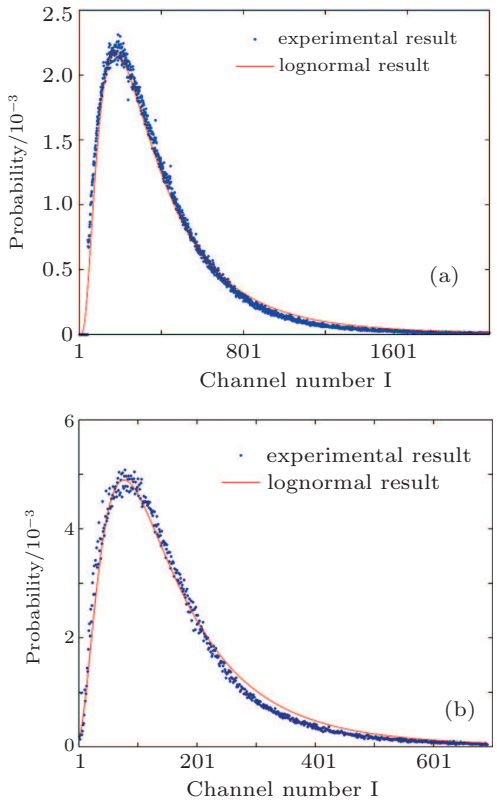


Fig. 4. (color online) Experimental results of particles suspended in the air, with the x axis denoting the channel number and the y axis representing the probability of distribution: (a) amplitude and (b) width.

The equation of width corresponding to the solid line in Fig. 4(b) is

$$p(l_\tau) \equiv \frac{1}{\Omega_{L_\tau} \sigma_{\ln \tau} l_\tau} \exp\left(-\frac{\left(\ln l_\tau - \ln \bar{l}_\tau\right)^2}{2\sigma^2_{\ln \tau}}\right), \quad l_\tau \in (1, 2, \ldots L_\tau), \tag{18}$$

where the statistical parameters $\ln \bar{l}_\tau$ and $\sigma_{\ln \tau}$ are 4.995 and 0.71, respectively. The measured results indicate that the distributions of amplitude $p(l_V)$ and width $p(l_\tau)$ of the two-parameter model match well with the form of the lognormal distribution when the natural number is used as the independent variable.

Table 1. Chi-square distribution fit test and correlation coefficients of particles suspended in the air.

Sample	Chi-square distribution fit test of lognormal distribution						C.C
	N	$\sum f_i^2/N\hat{p}_i$	χ^2	$k-s-1$	$\chi^2_{0.05}(k-s-1)$	accept?	r
Amplitude	1825470	1826535.701	1065.701	2005	2110.221	accept	0.956
Width	1825470	1825795.255	325.255	697	759.529	accept	0.982

According to the chi-square distribution fit test as shown in Table 1, when significance level $\alpha = 0.05$, the presumptions of distribution fitting function as Eqs. (15) and (18) are accepted.

Since the particles suspended in the air could be treated as the sum of standard particles with various diameters and background noise and the statistical distributions of particles suspended in the air satisfy the lognormal distribution, in order to study the reason why statistical distributions of particles suspended in the air fit well to lognormal distribution, we further study the statistical distributions of standard particles with various diameters and background noise with the same experimental system and environment. Besides, we also want to know the random distribution law of these three typical objects.

4.2. Statistical distributions of standard particles

Then, we study the pulse signal of a standard particle. The standard particles are Duke Scientific Corporation products with particle diameter coefficient variation within 2%. The polystyrene standard particles with certified mean diameters 0.38 μm, 0.499 μm, 0.6 μm, 0.8 μm, and 0.99 μm are atomized, dried, and pumped into the measuring device by using a PG-100 particle generator of the U.S. PMS company. The measured results of standard particle amplitude distribution are shown in Fig. 5

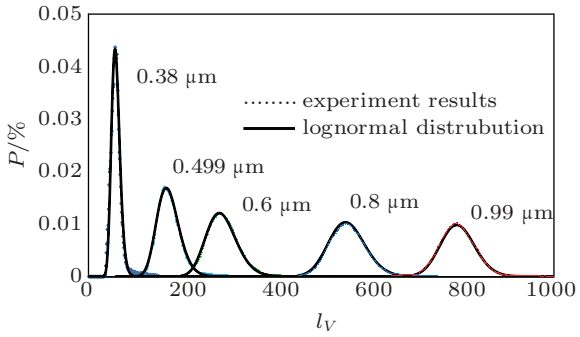


Fig. 5. (color online) Measured results and the fitting distributions of standard particles, with the x axis representing the channel number and the y axis denoting the probability of distribution. The dotted line represents the experiment results and the solid line denotes the lognormal fitting distributions.

The lognormal distribution curves are obtained from the statistical parameters $\ln\bar{l}_V$ and $\sigma^2_{\ln V} = E(\ln l_V - \ln\bar{l}_V)^2$ in Eq. (15). It is obvious that the five standard particle pulse amplitude distributions all satisfy the lognormal distribution, and the fitting values of different standard particles are shown in Table 2.

Table 2. Statistic parameter values of the standard particle.

Sample		Counting number	Geometric mean	Geometric standard deviation	Peak position
Certified mean diameter/μm	Standard deviation	N	$\ln\bar{l}_V$	$\sigma_{\ln V}$	l_V
0.38	13	14802	4.54	0.15	57
0.499	12.6	21010	5.14	0.14	154
0.6	12.2	17886	5.60	0.12	280
0.8	4.8	34062	6.32	0.08	553
0.99	0.01	30223	7.20	0.05	790

The standard particles are round particles with the same size and the same optical characteristics, for they are made of the same material. In the actual measurement, we only focus on its single specific physical parameter of particle diameter. The statistical distribution law on a single specific physical parameter which only reflects partial properties of a natural phenomenon is often of normal distribution. Compared with the particles in the air, the standard particles avoid different effects of randomness such as in shape, material, size, etc. The amplitude distribution of particles in the air could only be described by skewed lognormal distribution. In the condition of fewer random factors and low dispersion, both normal and lognormal distributions could be used. The normal distribution fitting function is widely applied for its simplicity. It is often applied to the statistical distribution of a standard particle such as the measured results of Fig. 5. As shown in Table 3, the result of the chi-square distribution fit test verifies this conclusion.

For large sample measurements, we estimate two parameters: geometric mean and geometric standard deviation in channel k. Even in the condition of $\alpha = 0.995$, the lognormal distribution fitting function could satisfy the test.

Table 3. Results of the chi-square distribution fit test and the correlation coefficient test of lognormal distributions of standard particles.

Sample	Chi-square distribution fit test of lognormal distribution						Acceptable significance level		Correlation coefficient
	N	$\sum f_i^2/N\hat{p}_i$	χ^2	$k-s-1$	$\chi^2_{0.05}(k-s-1)$	accept?	α	$\chi^2_\alpha(k-s-1)$	r
0.38 μm	14802	14818.481	16.481	50	67.505	accept	0.995	27.99	0.993
0.499 μm	21010	21058.83	48.83	128	155.405	accept	0.995	90.54	0.998
0.6 μm	17886	17960.391	74.391	195	228.58	accept	0.995	147.89	0.999
0.8 μm	34062	34150.867	88.867	276	315.749	accept	0.995	219.24	0.999
0.99 μm	30223	30295.288	72.288	142	170.810	accept	0.995	102.35	0.997

The fitting quality of normal distribution is examined by correlation coefficient and chi-square distribution fit test. Table 4 shows the results of the chi-square distribution fit test of normal distribution.

Table 4. Results of the chi-square distribution fit test and the correlation coefficient test of normal distributions of standard particles.

Sample	Chi-square distribution fit test of normal distribution						Acceptable significance level		Correlation coefficient
	N	$\sum f_i^2/N\hat{p}_i$	χ^2	$k-s-1$	$\chi^2_{0.05}(k-s-1)$	accept?	α	$\chi^2_\alpha(k-s-1)$	r
0.38 μm	14802	14855.426	53.426	50	67.505	accept	0.25	56.33	0.962
0.499 μm	21010	21124.279	114.279	128	155.405	accept	0.7	119.1	0.993
0.6 μm	17886	18069.772	183.772	195	228.58	accept	0.7	184.2	0.859
0.8 μm	34062	34303.997	241.997	276	315.749	accept	0.9	246.4	0.977
0.99 μm	30223	30342.995	119.995	142	170.810	accept	0.9	120.9	0.991

Comparing Table 3 with Table 4, the correlation coefficients of lognormal distribution fitting for different diameter particles are all higher than 0.993, while the correlation coefficient of normal distribution fitting is lower than that of lognormal distribution fitting. In Table 4, when significance level $\alpha = 0.05$, the presumptions of normal distribution fitting function are accepted. However, the presumptions of normal distribution fitting function could not be accepted with the increase of significance level in high accuracy condition. Moreover, for the same standard particles, the maximum acceptable significance level of lognormal distribution fitting function is larger than that of normal distribution fitting function.

Based on the above analysis when $\sigma_{\ln V} \leq 0.1$, the normal distribution and the lognormal distribution are accepted for describing the distribution of standard particles.

4.3. Statistical distributions of background noise

The background noise is measured in a 1000 level ultra-clean environment. The gain of the amplifier circuit should be set at a high level, for the noise amplitude is low. The result of the random noise $p_{\text{noise}}(l_V)$ is shown as the discrete points in Fig. 6(a) with the whole sampling time of 5 s and about 10^7 sampling data which look like being of central symmetry. However, if the peak of $p_{\text{noise}}(l_V)$ is taken as the symmetric axis, the solid line is the mirror reverse of $p_{\text{noise}}(l_V)$. It is obvious that the distribution does not have central symmetry. Besides, the fitting result from Eq. (15) matches quite well with the measured amplitude result, and Fig. 6(b) shows the measured width result and its fitting result from Eq. (18).

In the fitting result, $\ln \bar{l}_V = 6.32$, $\sigma_{\ln V} = 0.13$, $\ln \bar{l}_\tau = 3.624$, and $\sigma_{\ln \tau} = 0.252$. They fit well to the lognormal distribution. Since the measured distribution is similar to the normal distribution, in order to evaluate the quality of the fitting, the chi-square distribution fit tests of normal/lognormal distributions are shown in Table 5.

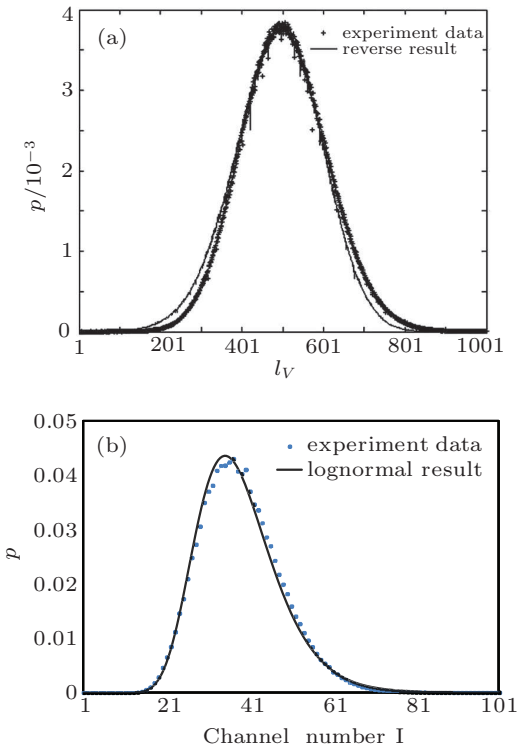


Fig. 6. Experimental results of background noise, with the x axis representing the channel number and the y axis denoting the distribution probability. In panel (a), the dotted line represents the measured amplitude distribution, the solid line denotes the mirror reversing, and the dashed line refers to the fitting result, with the unit of x axis being 2.44 mV. In panel (b), the dotted line represents the measured width distribution and the solid line denotes the fitting curve, with the unit of x axis being 0.1 μs.

Table 5. Results of the chi-square distribution fit test of background noise.

Sample	Chi-square distribution fit test of lognormal distribution						Correlation coefficient
	N	$\sum f_i^2/N\hat{p}_i$	χ^2	$k-s-1$	$\chi^2_{0.05}(k-s-1)$	accept?	r
Amplitude	1047558	1048156.745	598.745	582	639.232	accept	0.995
Width	1047558	1047634.905	76.905	70	90.531	accept	0.998
	Chi-square distribution fit test of normal distribution						
Amplitude	1047558	1048181.668	623.668	582	639.232	accept	0.994
Width	1047558	1047651.962	93.962	70	90.531	reject	0.892

From Table 5, the presumptions of lognormal distribution fitting function for both amplitude and width distributions can be accepted. However, the presumption of normal distribution fitting function for width distribution cannot be accepted. Moreover, the correlation coefficient of normal distribution fitting is lower than that of lognormal distribution fitting.

5. Discussion

5.1. Randomness of three typical measuring objects

These results of three typical random processes measured by the same light scattering particle counting system indicate that all the pulse signal statistical distributions can be expressed by the same distribution $p(l)$, though the values of statistical parameter are different. We believe that this statistical regularity has a connection with the pulse signal measuring method and the randomness in the measuring process.

The random factors in the measuring process of three kinds of measuring objects represent different levels of randomness in the measurement. From the measurement results we can draw the following conclusions. 1) Since the shape of particles suspended in the air is much more complicated than that of standard particles, the distribution of particles in the air has a larger distribution width than the standard particle. While the background noise from the effect of randomness of standard particles is avoided, its distribution has the narrowest distribution width. 2) The diameter of standard particles corresponds to the peak of the signal amplitude distribution curve, and the sensor resolving power of particle size can be obtained intuitively according to the distribution curves. 3) Since the standard particles are not ideal particles, geometric standard deviations of their distributions not only reflect the change of photosensitive area and the randomness of photodetector noise, but also show the randomness of standard particles. 4) The pulse signal of photoelectric sensor noise only reflects the nature of the sensor itself, which means the geometric standard deviation should be minimal, generally, $\sigma_{\text{air}} > \sigma_{\text{standard}} > \sigma_{\text{noise}}$. 5) Normal distribution could be treated as the approximation of lognormal distribution at $\sigma_{\ln V} \leq 0.1$. Considering the non-negative characteristics of lognormal distribution, it owns obvious advantage in random measurements, which is exactly illustrated by the experimental data in Tables 1–5.

5.2. Statistical distributions of subsets

With the same experimental device, the statistical distributions of the light scattering signal of particles suspended in the air and background noise at different amplitude and width intervals are measured. In the width interval of $\tau_{l_\tau} \sim \tau_{l_\tau} + \Delta\tau$, the measured counting distribution function $\tilde{p}(l_V, l_\tau = C)$ of random pulse signal amplitude can be shown by discrete points in Fig. 7(a). For the same sample, in the amplitude interval of $V_{l_V} \sim V_{l_V} + \Delta V$, the measured counting distribution function $\tilde{p}(l_\tau, l_V = C)$ of random pulse signal width can be shown by dots in Fig. 7(b).

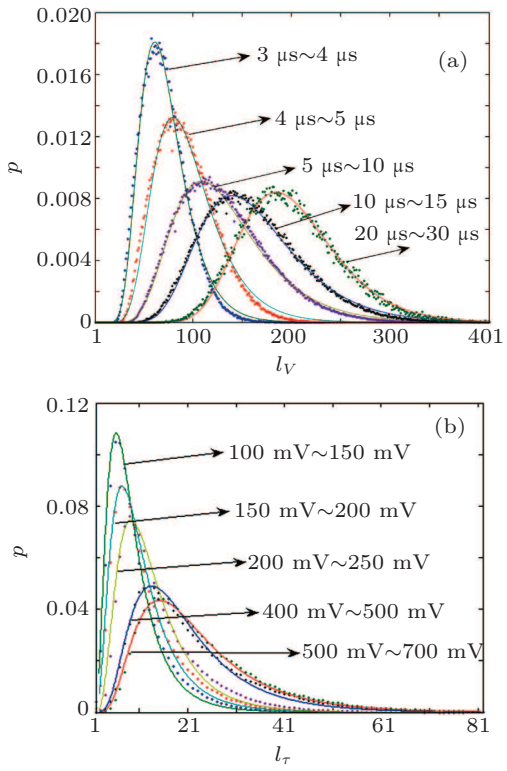


Fig. 7. (color online) Distributions of signal subsets of random particles suspended in the air, with the x axis representing the channel number and the y axis denoting the probability of distribution. In panel (a) the dotted lines represent the measured results of amplitude distributions at different widths, the solid lines denote the fitting results. In panel (b) the dotted lines represent the measured results of width distributions at different amplitudes, the solid lines refer to the fitting results.

As shown in Fig. 7, some basic characteristics of the curves $\tilde{p}(l_V, l_\tau = C)$ and $\tilde{p}(l_\tau, l_V = C)$ could be obtained.

Firstly, the pulse signal distributions do not satisfy central symmetry. The peak value and width of the amplitude distribution increase with the sample width interval increasing. Besides, the peak value and width of the width distribution increase with the sample amplitude interval increasing. Thirdly, the distributions $\tilde{p}(l_V, l_\tau = C)$ and $\tilde{p}(l_\tau, l_V = C)$ in the different subsets are similar in geometric structure. The lognormal distribution fitting results are represented with the solid lines in Fig. 7, described by the following equations:

$$p(l_V, l_\tau = C) = \frac{1}{\Omega_{l_\tau L_V} \sigma_{\ln V} l_V} \exp\left(-\frac{(\ln l_V - \mu_{\ln V})^2}{2\sigma_{\ln V}^2}\right), \quad (19)$$

$$p(l_\tau, l_V = C) = \frac{1}{\Omega_{l_V L_\tau} \sigma_{\ln \tau} l_\tau} \exp\left(-\frac{(\ln l_\tau - \mu_{\ln \tau})^2}{2\sigma_{\ln \tau}^2}\right). \quad (20)$$

As expected, different pulse signal subsets of the count distributions satisfy the lognormal distribution $p(l_V, l_\tau = C)$ and $p(l_\tau, l_V = C)$ stably. The fitting values are shown in Table 6.

The correlation coefficients of all experimental results and fitting results of statistical distribution subsets are all more than 0.995. Moreover, the chi-square distribution fit test of lognormal distribution is done in Table 7.

Table 6. Fitting values of signal subsets of particles suspended in the air.

Amplitude distributions under different widths			Width distributions under different amplitudes		
	Fitting data			Fitting data	
Width scope/μs	Geometric mean $\ln \bar{l}_V$	Geometric standard deviation $\sigma_{\ln V}$	Amplitude scope/mV	Geometric mean $\ln \bar{l}_\tau$	Geometric standard deviation $\sigma_{\ln \tau}$
3–4	4.23	0.34	100–150	2.00	0.595
4–5	4.51	0.36	150–200	2.21	0.594
5–10	4.83	0.38	200–250	2.42	0.565
10–15	5.05	0.33	400–500	2.84	0.555
20–30	5.28	0.25	500–700	2.96	0.501

Table 7. Results of the chi-square distribution fit test.

Sample	Chi-square distribution fit test of lognormal distribution					
	N	$\sum f_i^2/N\hat{p}_i$	χ^2	$k-s-1$	$\chi^2_{0.05}(k-s-1)$	Accept?
3 μs–4 μs	95741	95880.342	139.342	140	168.613	accept
4 μs–5 μs	83473	83624.983	151.983	185	220.991	accept
5 μs–10 μs	221213	221385.77	172.77	296	337.125	accept
10 μs–15 μs	85599	85717.513	118.513	301	342.462	accept
20 μs–30 μs	18289	18383.001	94.001	241	278.213	accept
100 mV–150 mV	107394	107407.901	13.901	50	67.505	accept
150 mV–200 mV	103487	103509.31	22.31	60	79.082	accept
200 mV–250 mV	87959	88006.635	47.635	67	87.108	accept
400 mV–500 mV	55400	55457.323	57.323	89	111.022	accept
500 mV–700 mV	30806	30834.982	28.982	83	105.267	accept

The studies of background noise pulse signal distribution subsets at different amplitude and width intervals are carried out with the above method. The fitting distribution curve, fitting values, and chi-square distribution fit test are shown in Fig. 8 and Tables 8 and 9.

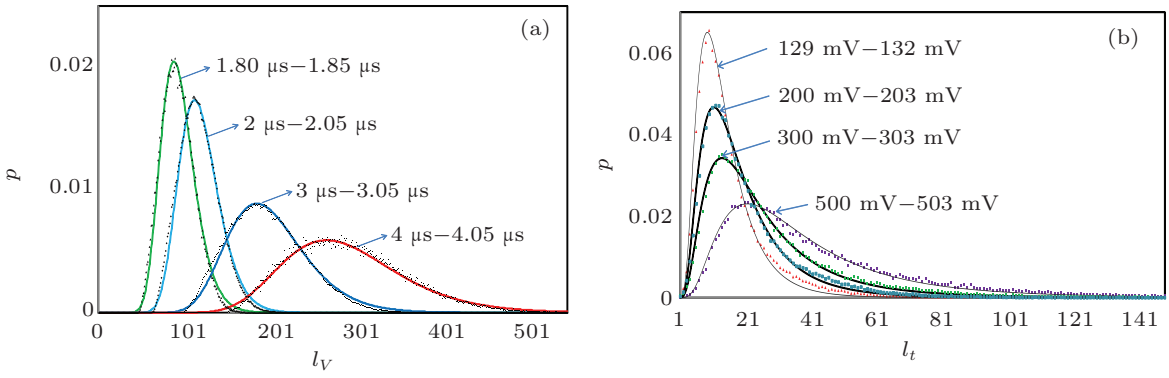


Fig. 8. (color online) Distributions of background noise subsets, with the x axis representing the channel number and the y axis denoting the probability of distribution. Panel (a) shows the amplitude distributions under different widths, with the dotted lines representing the measured results of amplitude distributions at different widths and the solid line denoting the fitting results. Panel (b) shows the width distributions under different amplitudes, with the dotted line representing the measured results of width distributions at different amplitudes and the solid lines denoting the fitting results.

Table 8. Fitting values of signals subsets of background noise.

Amplitude distributions under different widths			Width distributions under different amplitudes		
	Fitting data			Fitting data	
Width scope/μs	Geometric mean $\ln \bar{l}_V$	Geometric standard deviation $\sigma_{\ln V}$	Amplitude scope/mV	Geometric mean$\ln \bar{l}_\tau$	Geometric standard deviation $\sigma_{\ln \tau}$
1.80–1.85	4.565	0.210	129–132	2.55	0.56
2–2.05	4.765	0.200	200–203	2.815	0.62
3.0–3.05	5.280	0.235	300–303	3.08	0.67
4.0–4.05	5.650	0.245	500–503	3.50	0.65

The correlation coefficients of all experimental results and fitting results of statistical subsets are all more than 0.995.

Table 9. Results of chi-square distribution fit test.

Sample	Chi-square distribution fit test of lognormal distribution					
	N	$\sum f_i^2/N\hat{p}_i$	χ^2	$k-s-1$	$\chi^2_{0.05}(k-s-1)$	Accept?
1.80 μs–1.85 μs	100622	100712.83	90.83	116	142.138	accept
2 μs–2.50 μs	101769	101839.54	70.54	140	168.613	accept
3.0 μs–3.05 μs	95358	95462.864	104.864	252	293.248	accept
4.0 μs–4.50 μs	84667	84882.018	215.018	364	409.488	accept
129 mV–132 mV	60797	60825.711	28.711	82	104.139	accept
200 mV–203 mV	57630	57672.098	42.098	98	122.108	accept
300 mV–303 mV	52884	52948.619	64.619	126	153.198	accept
500 mV–503 mV	44299	44371.877	72.877	167	198.154	accept

Finally, the studies show that the signal subsets have a unified form of lognormal distribution not only among the three typical random processes but also between the entirety and the subsets.

5.3. Similarity among random quantity distribution functions

In practical application, as the relation between the two-physical-quantity (V and τ) statistical distributions can be theoretically derived, the other physical quantity statistical distribution can be accurately calculated if another physical quantity statistical distribution has been measured.

From the above experimental results, the random signal distributions match well with the form of the lognormal distribution when the natural numbers are used as the variables by two-parameter model. The distribution functions $p(l_V)$, $p(l_\tau)$, $p(l_V, l_\tau = C)$, and $p(l_\tau, l_V = C)$ have statistically self-similar characteristics.

The total number of the signal channels L_V and L_τ tend to be infinite with the improvement of the measurement accuracy $\Delta V \to 0$ and $\Delta\tau \to 0$, the normalized coefficient of the probability density function tends to be $\sqrt{2\pi}$. Supposing that $p(l_V)\,\mathrm{d}l_V = p(l_\tau)\,\mathrm{d}l_\tau$, the nonlinear relationship $l_V = (bl_\tau)^\alpha$ can be derived as

$$\begin{aligned} p(l_V)\,\mathrm{d}l_V &\equiv \frac{\mathrm{d}l_V}{\Omega\sigma_{\ln V} l_V}\exp\left(-\frac{\left(\ln l_V - \ln \bar{l}_V\right)^2}{2\sigma^2_{\ln V}}\right) \\ &= \frac{\mathrm{d}l_\tau}{\Omega\left(\sigma_{\ln V}/\alpha\right) l_\tau}\exp\left(-\frac{\left(\ln l_\tau + \ln b - \ln \bar{l}_V/\alpha\right)^2}{2\sigma^2_{\ln V}/\alpha^2}\right) \\ &= \frac{\mathrm{d}l_\tau}{\Omega\sigma_{\ln \tau} l_\tau}\exp\left(-\frac{\left(\ln l_\tau - \ln \bar{l}_\tau\right)^2}{2\sigma^2_{\ln \tau}}\right) \\ &\equiv p(l_\tau)\,\mathrm{d}l_\tau. \end{aligned} \tag{21}$$

According to Eq. (21), we can obtain $\alpha = \sigma_{\ln V}/\sigma_{\ln \tau}$ and $b = \exp[(\sigma_{\ln \tau}/\sigma_{\ln V})\ln \bar{l}_V - \ln \bar{l}_\tau]$. With these relations, the amplitude distribution curve could be coincided with the width distribution curve, however, the channel of one distribution curve is a non-natural number sequence.

Besides, when the two independent characteristic quantities V and τ are used to describe the random pulse signal group, Eq. (21) indicates that if the statistical distribution satisfies the lognormal distribution, the corresponding distributions of different characteristic quantities meet the non-linear self-similarity transformation relation within the domain.

The relation among different characteristic quantity distributions is well expressed by

$$\frac{V}{\Delta V}\frac{1}{\bar{l}_V} = \left(\frac{\tau}{\Delta\tau}\frac{b}{\bar{l}_\tau}\right)^{\sigma_{\ln V}/\sigma_{\ln \tau}}. \tag{22}$$

Besides, all the above measured results can be described by Eqs. (7) and (22). Equation (22) is also the non-integeral dimension fractal relation of the statistical distributions of different random laser scattering pulse signal groups, where the fractal dimension corresponds to the ratio of α for the distributions of different characteristic quantities.

6. Conclusions

In this paper, we proposed the random process model theoretically which clearly provides the relation between mea-

sured results and the different random factors in the measuring system. Based on this proposed model, the constraint relation of the statistical distribution function of the measured result is derived. Moreover, we have also introduced the geometric mean, which is important in a random study as a statistical parameter and should be positive, corresponding to the positive property of measured results. According to the theoretical analysis, we have studied the distribution functions of measured results in the same experimental system of three typical objects including particles suspended in the air, standard particles, and background noise. We have found that statistical amplitude and width distribution functions of all of these three typical objects match well with lognormal distribution when natural numbers are used as the variables using a two-parameter model testified by chi-square and correlation coefficient. And this uniform distribution corresponds to the randomness of the measurement process. The comparison between lognormal and normal distribution fitting results shows that lognormal distribution has more advantages in describing random measured results. Besides, the subset of these measuring objects is measured, the results indicate a lot of relations such as the relation between amplitude and width distributions, the relation between global and local distributions, and the relation between subsets of amplitude, and width distributions of the same object satisfy the non-linear self-similarity transformation relation. The research provides a new way for describing a random process with lognormal distribution and this enables accurate analysis of random results.

References

[1] Gaddum J H 1945 *Nature* **156** 463
[2] Ahrens L H 1954 *Geochimica et Cosmochimica Acta* **6** 121
[3] Aitchison J and Brown J A C 1957 *The Logfiormal Distribution* (Cambridge: Cambridge University Press)
[4] Spratt J and John S 1969 *J. Surg. Res.* **9** 151
[5] Baker L, Giancola A J and Allahdadi F 1992 *J. Appl. Phys.* **72** 2724
[6] Peter A M K 2010 *Meas. Sci. Technol.* **21** 105501
[7] Limpert E, Stahel W A and Abbt M 2001 *BioScience* **51** 341
[8] Chen J, Li S and Ma H R 2007 *Chin. Phys.* **16** 2600
[9] Granqvist C G and Buhrman R A 1976 *J. Appl. Phys.* **47** 2200
[10] Park K T, Park D, Lee S G and Hwang J 2009 *J. Aerosol. Sci.* **40** 858
[11] Baker L, Giancola A J and Allahdadi F 1992 *J. Appl. Phys.* **72** 2724
[12] Yang J, Bian B M, Peng G, Yan Z G and Li Z H 2010 *Acta Phys. Sin.* **59** 7713 (in Chinese)
[13] Yang J, Bian B M and Wang C Y 2011 *Sensor Letters* **9** 1556
[14] Lin Y L, Yan Z G, Wang C Y and Bian B M 2012 *Acta Phys. Sin*. **61** 100505 (in Chinese)