

# Total Rejection Sampling and the Reduction of the Wave Function

Noureddine Kermiche\*

Western Digital Corporation, Irvine, CA 926121

\***Corresponding author:** Noureddine Kermiche, Western Digital Corporation, Irvine, CA 92612, E-mail: noureddine.kermiche@wdc.com

**Received date:** 22-July-2023, Manuscript No. tspa-23-107715; **Editor assigned:** 24-July-2023, Pre-QC No. tspa-23-107715 (PQ); **Reviewed:** 10-August-2023, QC No. tspa-23-107715 (Q); **Revised:** 12-August-2023, Manuscript No. tspa-23-107715 (R); **Published:** 31-August-2023, DOI. 10.37532/2320-6756.2023.11(8).367

## Abstract

We propose an exact new sampling algorithm using classical statistical theory. The proposed method will process an infinite amount of rejected samples in parallel before accepting one exact single sample from a target distribution using a winner-take-all mechanism. The algorithm is applied to sampling from general distributions in addition to sampling from the wave function density distribution which lies at the heart of the measurement problem in quantum physics.

*Keywords: Exact sampling; Rejection sampling; Reduction of the wave function; Quantum measurement problem*

## I. Introduction

Assuming that the reduction of the quantum wave function is a real physical process implies that we must believe that nature has found the perfect algorithm for statistical sampling from density distributions. In the double slit experiment, the locations of the scintillations appearing on the photographic plate are independent random samples drawn from a predictable density distribution derived from deterministic equations of interference patterns of propagating waves. This mysterious sampling algorithm appears to be instantaneous, does not waste extra energy, and does not suffer from the curse of dimensionality which plagues all known classical sampling algorithms. And regardless of whether the reduction of the wave function is a real physical process or not, this sampling ability is what gives quantum computers their potential real advantage over their classical counterparts.

The mysterious sampling algorithm that nature has apparently found cannot be duplicated using known classical sampling algorithms. Sampling during the reduction of the wave function is apparently 100% efficient since we always get 1 scintillation on the photographic plate for every wave-packet. The 100% efficiency requirement eliminates classical rejection sampling algorithms as a tentative explanation since rejection sampling with its rejection/acceptance rule is never 100% efficient [1]. The locations from a sequence of scintillations are completely independent from each other which also eliminates all sequential sampling algorithms such as Markov Chain Monte Carlo (MCMC) [2]. Quantum sampling due to the reduction of the wave function is also an exact sampling algorithm which eliminates all approximate classical sampling algorithms as possible models.

In this study, we construct a new classical sampling algorithm that tries to mimic quantum sampling. We introduce a new way of performing exact sampling from a target density distribution that works at any dimension. We show that emulating quantum sampling can be achieved by accepting one sample out of an infinite amount of rejected samples that are processed all at once. And

**Citation:** Kermiche N. Total Rejection Sampling and the Reduction of the Wave Function. J. Phys. Astron.2023;11(8):367.

©2023 Trade Science Inc.

while the processing of the infinite number of rejection samples uses only local information, choosing the winner-take-all (WTA) sample requires a non-local (global) optimization operation. Physical implementation of the WTA mechanism, as well as the noise sources that power the algorithm, are also discussed.

The algorithm is applied to sampling from general distributions and is also applied to sampling from a wave function density distribution. This study infers by analogy that it is possible that the quantum sampling algorithm is 0% efficient instead of the apparent 100% efficiency because an infinite number of samples are rejected every time a single sample is accepted from the target distribution. The appearance of 100% efficiency could be due to processing an infinite amount of non-local information both in parallel and all at once. However, this claim is contingent on assuming that the reduction of the wave function is a real physical process as mentioned previously and not simply some good “cookbook” recipe [3, 4].

## II. Total Rejection Sampling

Instead of using a classical random sampling strategy, where the sampling algorithm test, accept, or reject one sample at a time, we propose a new approach where all tentative samples are accessible at the same time. We will show that a sample that is accepted from all these samples, using a minimization criterion, will be an exact sample from the target distribution.

### Total rejection sampling algorithm

**Theorem 1:** Given a continuous probability density function  $f(x)$  where  $x \in R^n$  for  $n \geq 1$  that is known up to a multiplicative constant, and given a uniformly distributed random function  $\{U(x), 0 \leq U(x) \leq 1, x \in R^n\}$ , where the random variables  $U(x)$  for different  $x$  values are identical and independently distributed, then the random variable obtained by finding the location  $x$  of the

minimum in  $R^n$  of the ratio  $\frac{U(x)}{f(x)}$  has a probability density function  $\frac{f(x)}{\int_{R^n} f(y)dy}$ .

**Theorem 2:** Given a probability density function  $f(x)$  where  $x \in R^n$  for  $n \geq 1$  that is known up to a multiplicative constant, and given a positive random function  $\{X(x), 0 \leq X(x) < \infty, x \in R^n\}$  where the random variables  $X(x)$  for different  $x$  values are independently distributed over  $x$  with identical distribution density  $f_x(y) = ae^{-ay}$  with  $a > 0$  and  $y \in [0, \infty]$ , then the random

variable obtained by finding the location  $x$  of the minimum in  $R^n$  of the ratio  $\frac{X(x)}{f(x)}$  has a probability density function  $\frac{f(x)}{\int_{R^n} f(y)dy}$ .

Proofs for both theorems can be found in the Appendix.

What the two theorems are advocating is the least efficient sampling algorithm possible. To extract one exact sample from a target distribution, we need to evaluate the distribution everywhere and we also need a “white noise” to be generated everywhere. The ratio between the noise and the target distribution is locally evaluated but the minimum operation to find the sample location is non-local (global).

Finding a physical process that can implement the proposed algorithms is a tall task due to the requirements of collecting information everywhere, the non-locality of the minimum operation, and the fact that pure “white noises” are ideal constructs that do not exist in nature.

Besides the uniform white noise used in Theorem 1 and the exponential white noise in Theorem 2, we could not generalize the proposed algorithm to other noise sources. It is possible that there are other distributions that could be used but computer simulation of many examples using other distributions proved that a general white noise will not work with the proposed algorithm.

In the experimental section of this study, we will show that a good approximation of a target distribution can be achieved using a small finite number of tentative samples and that the generated distribution will converge to the ideal target distribution as the number of tentative rejected samples increases to all  $x \in R^n$ .

**Noise sources**

An example of a probability distribution that can be used for Theorem 2 is the sum or the squares of 2 normal random variables.

For normal distributions with 0 means and unit variances, the distribution is called a Chi-Squared distribution  $\chi^2_2(x) = \frac{1}{2} e^{-\frac{x}{2}}$ .

Chi-Squared distribution random variables can be obtained using the following method that combines the Central Limit Theorem and Fourier series of white phase noise with uniform distribution. The initial application of such an idea to generate Gaussian white noise from uniform random phases was proposed in [5]. A complete mathematical treatment of this result can be found in [6].

**Theorem 3:** given  $n$  independent and uniformly distributed random angles  $\{\theta_j, 0 \leq \theta_j \leq 2\pi, j = 1, \dots, n\}$ , and  $i = \sqrt{-1}$ , and as  $n \rightarrow \infty$ , the real and imaginary parts of the normalized discrete Fourier transform  $S_k$  of the random phases,

$$S_k = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{i\theta_j} e^{i2\pi \frac{(k-1)(j-1)}{n}} \quad \text{for } k=1, \dots, n$$

will converge to independent identically distributed normal random variables with mean 0 and variance  $\frac{1}{\sqrt{2}}$ , and  $|S_k|^2$  will converge to an independent identically Chi-Squared exponential distribution  $\chi(x) = e^{-x}, x \geq 0$ .

The limit exponential distribution labeled  $|S(x)|^2$  for  $n \rightarrow \infty$  is going to be used for the rest of this study. In the following we will focus on applying Theorem 2 assuming that the white noise is generated using Theorem 3.

**Reduction of the Wave Function**

**Other ratio options and the measurement problem**

Finding the location of the minimum in  $R^n$  of the ratio  $\frac{X(x)}{f(x)}$  will give the same statistics as finding the maximum of the ratio  $\frac{f(x)}{X(x)}$  or finding the maximum in  $R^n$  of the ratio:

$$\frac{1}{1 + \frac{X(x)}{f(x)}} = \frac{f(x)}{f(x) + X(x)}$$

This new ratio is bounded between 0 and 1. For numerical or physical implementations of the algorithm, this bounded ratio is preferable, but it is not necessary.

Using the definition of probability from quantum mechanics given a unitary wave function  $\psi(x)$  and the white noise  $S(x)$  from Theorem 3, the bounded ratio is:

$$R(x) = \frac{|\psi(x)|^2}{|\psi(x)|^2 + |S(x)|^2}$$

Theorem 2 guarantees that the random location that maximized this ratio will have a distribution  $|\psi(x)|^2$  which is exactly what the quantum measurement problem is.

### Physical collapse models

To address physical implementation of the proposed method when sampling from the wave function density, we start by looking at existing physical collapse models of the reduction of the wave function. In most physical collapse models [7], the reduction of the wave function  $\psi(x)$  occurs randomly in space and at random times. The reduction is modeled as Gaussian “hits” at some random location  $x_i$ . The wave function undergoes a local shrinking using the multiplication,

$$\psi_i(x) = \psi(x) \exp\left(-\frac{(x-x_i)^2}{2d^2}\right)$$

where  $d$  is a new constant of nature. And because the multiplication invalidates the unitary property of the wave function, the wave function undergoes a non-local normalization:

$$\psi(x) = \frac{\psi_i(x)}{|\psi_i(x)|_2^2}.$$

These models have a circular argument problem since  $x_i$  are sampled from the same  $\psi(x)$  using the probability density  $|\psi(x)|^2$ . These models do not attempt to explain the mysterious quantum sampling algorithm but merely tries to use the same mechanism to explain it. The main criticism of these models is based on an economic principle: if nature knows how to sample from  $|\psi(x)|^2$  anytime and anywhere, then why waste samples in empty space with all the added complications such as energy non-conservation and introduce new constants of nature instead of sampling only once at the photographic plate?

### Physical implementation of winner-take-all mechanism

Given the ratio that was proposed previously with the previously defined white noise  $S(x)$ :

$$R(x) = \frac{|\psi(x)|^2}{|\psi(x)|^2 + |S(x)|^2}$$

and as stated earlier, Theorem 2 guarantees that the random location that maximizes this random ratio will have a distribution  $|\psi(x)|^2$ . If we define  $\psi_0(x) = \psi(x)$  and  $S_0(x) = S(x)$  for time step 0, the non-local winner-take-all operation can easily be replaced by the following recursive relation at time step  $i$  given the ratio at time step 0:

$$\psi_{i+1}(x) = \psi_i(x) \frac{|\psi_0(x)|^2}{|\psi_0(x)|^2 + |S_0(x)|^2}$$

followed by the normalization step

$$\psi_{i+1}(x) = \frac{\psi_{i+1}(x)}{|\psi_{i+1}(x)|_2^2}.$$

The sequence of wave functions will converge to the Dirac delta function at the same maximum pointer with distribution  $|\psi(x)|^2$  which solves the measurement problem without the circular argument seen in the previous physical collapse model. However, we did not solve the non-locality issue because we just replaced the original non-local maximum operation with another non-local normalization. The new sampling proposal based on Theorems 1 and 2 is non-local by construction and finding a local shortcut may not be possible. The issue of non-locality could be similar the dilemma that is encountered every time we try to explain quantum physics using classical approaches [8].

Trying to use some function of  $R(x)$  as a non-linear random potential in the Schrodinger equation to induce reduction of the wave

function could be an alternative idea in future studies to see if we can relax the non-local constraints of Theorems 1 and 2.  $S(x)$  being a white noise is another complication since pure white noises, as stated previously, are ideal constructs that do not exist in nature. We could not find a way to make the algorithm work with colored noises but that does not mean that it cannot be done with some new extensions to the ideas proposed in this study.

### Numerical Example for Sampling from General Distributions

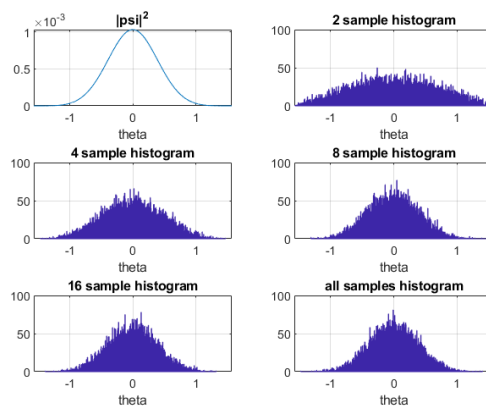
In the following we are going to use Theorem 1 to sample from the following function  $\psi(\theta)$ . The function is derived from the interference pattern of  $N$  slits with wavelength  $\lambda$ . We set  $a = \lambda$  as the width of each slit and  $d = 4\lambda$  is the distance between the slits. The function  $\psi(\theta)$  is defined for angle  $\theta$  where  $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ .

$$\psi(\theta) = I_0 \frac{\sin\left(\pi \frac{a}{\lambda} \sin(\theta)\right) \sin\left(\pi \frac{Nd}{\lambda} \sin(\theta)\right)}{\pi \frac{a}{\lambda} \sin(\theta) N \sin\left(\pi \frac{d}{\lambda} \sin(\theta)\right)}$$

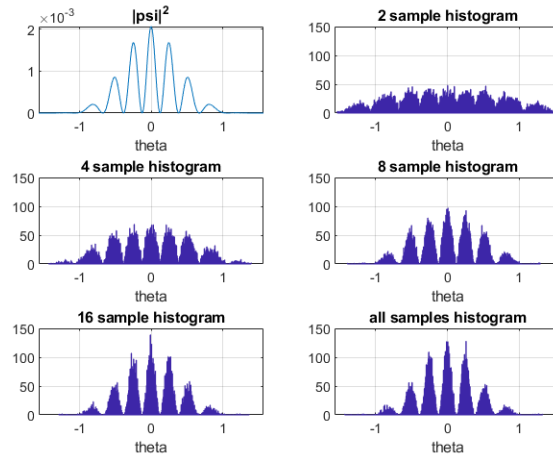
The sharpness of the distribution  $|\psi(\theta)|^2$  is controlled by changing the number of slits  $N$ .  $I_0$  is a normalization constant.  $|\psi(\theta)|^2$  is evaluated using a small  $\theta$  step size (0.0001).

Instead of using all the samples of  $\theta$  as stated in Theorem 1 and construct a histogram of the location that minimizes the ratio  $\frac{U(\theta)}{|\psi(\theta)|^2}$ , we tried to find the minimum ratio using only 2, 4, 8, and 16 random samples in the range  $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ . The quality of the obtained histogram is compared against the asymptotic condition where one sample is picked out of all samples in the range  $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ .

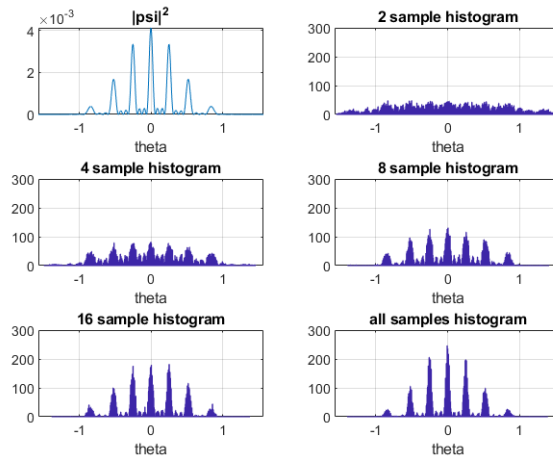
**FIG.1-3** show clearly that the generated distribution converges to the target distribution as the numbers of samples changes from 2 to 16 and as the number of samples approach the asymptotic condition of Theorems 1 where all samples are used.



**FIG.1. A soft distribution for  $N = 1$ . The generated distribution is a good approximation to the target distribution after only 8 samples are used.**



**FIG. 2. A less soft distribution for  $N = 2$ . The generated distribution is closer to the target distribution after 16 samples are used. The number of samples to approximate the target distribution is larger than the softer distribution with  $N=1$  as seen in FIG. 1.**



**FIG. 3 A sharp distribution for  $N=4$ . The distribution is close to the target distribution if we use a larger number of samples than the 16 seen in FIG. 2. The number of samples required to approximate the target distribution is larger for distributions with sharp and isolated peaks.**

Even with a limited number of rejected samples, the efficiency of the proposed sampling is not necessarily better than existing rejection sampling algorithms. However, the new sampling algorithm has an advantage because a sample is always returned after a fixed number of rejections. The quality of the distribution improves as the number of samples increase since we know that the new sampling algorithm is an exact sampling algorithm in the asymptotic limit where all samples  $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ . are used to generate tentative rejection samples.

**Discussion**

We proposed a new way of looking at algorithmic sampling from a target distribution. We presented numerical examples to show

how the proposed method can be applied to practical problems. However, applying the proposed algorithm to sampling difficult high-dimensional and sharp distributions is still a work in progress.

The proposed algorithm was applied to the measurement problem in quantum physics. We do not know if the quantum measurement problem is a real physical process or not. And even if we assume that such a problem is real, we do not know if the present study contributed to addressing the problem. However, we did introduce a new classical sampling algorithm that tries to mimic quantum sampling. The proposed algorithm is powered by white noise and uses non-local information. The algorithm physical limitations were highlighted. However, and despite the limitations, we were able to use the algorithm to address the circular argument problem that plagues most collapse theories where the random location of the “hits” that cause the reduction of the wave function are samples drawn from the same collapsing wave function that the collapse algorithm is trying to explain.

We believe that this study could open new avenues for looking at a vexing problem that has been staring at us for so long: if the reduction of the wave function is real then how does nature know how to sample from the wave function density?

## Appendix

### Proof of theorem 1

We start by analyzing discrete distributions before generalizing to continuous distributions. Assume we have a discrete binary distribution that takes a pointer value 1 with probability  $p_1$  and pointer value 2 with probability  $p_2$  with  $p_1 + p_2 = 1$ .  $U_1$  and  $U_2$  are independent and identically uniformly distributed random variables between 0 and 1. A new random variable  $Y$  is set to value 1 if  $\frac{U_1}{p_1} \leq \frac{U_2}{p_2}$  and it is set to value 2 otherwise. Using the independence property of  $U_1$  and  $U_2$ , a quick calculation will show that the probability that  $Y$  is equal to value 1 is:

$$P(Y = 1) = \int_0^1 \left[ \int_0^1 \min\left(1, \frac{p_2}{p_1} u_1\right) du_2 \right] du_1$$

Equivalently, we have,

$$P(Y = 1) = \int_0^1 \left( 1 - \min\left(1, \frac{p_2}{p_1} u_1\right) \right) du_1 \tag{1}$$

And finally:

$$P(Y = 1) = 1 - \frac{p_1}{2p_2} \text{ if } p_1 \leq p_2$$

or

$$P(Y = 1) = \frac{p_2}{2p_1} \text{ if } p_1 > p_2$$

The reader can verify that by symmetry that the integration operation will result in the following probability for pointer 2:

$$P(Y = 2) = 1 - P(Y = 1)$$

Now we are going to assume that the discrete distribution can take  $m$  possible values where  $\sum_{j=1}^m p_j$  is equal to some unknown constant. We could set the constant to 1 but the probability normalization is not necessary for the following calculations to yield

normalized pointer probabilities. The probability of the random variable  $Y$  is equal to a pointer  $k$  between 1 and  $m$  where

$\frac{U_k}{p_k} \leq \frac{U_j}{p_j}$  for all  $j \neq k$  can be derived by generalizing (1) from the binary pointer case to  $m$  pointers:

$$P(Y = k) = \int_0^1 \prod_{j \neq k} \left( 1 - \min \left( 1, \frac{p_j}{p_k} u \right) \right) du \tag{2}$$

Using the change of variables  $u = \frac{z}{p_k}$ , we can write (2) as:

$$P(Y = k) = p_k g_k \tag{3}$$

with

$$g_k = \int_0^{p_k} \prod_{j \neq k} (1 - \min(1, p_j z)) dz$$

We are going to show that for large  $m$ , the term  $g_k$  is a constant that does not depend on  $p_k$ . We use the Leibniz integral rule to compute the derivative function of  $g_k$  with respect to  $p_k$ :

$$\frac{\partial g_k}{\partial p_k} = -\frac{1}{2 p_k^2} \prod_{j \neq k} \left( 1 - \min \left( 1, \frac{p_j}{p_k} \right) \right) \tag{4}$$

For large  $m$ , the product in (4) approaches the infinite product,

$$\prod_{y \in R^n} \left( 1 - \min \left( 1, \frac{f(y)}{f(x)} \right) \right)$$

Since  $f(x)$  is continuous and for all  $x \in R^n$ , we can always find  $y \in R^n$  such that  $\frac{f(y)}{f(x)} \geq 1$ . This means that there is always a term in the infinite product that is identically equal to zero. Thus, and for large  $m$ , the derivative function (4) is identically equal to 0 and  $g_k$  is a constant function. And since (2) is a normalized probability, we have for large  $m$ ,

$$g_k = \frac{1}{\sum_{j=1}^m p_j}$$

And (3) can be written as,

$$P(Y = k) = \frac{p_k}{\sum_{j=1}^m p_j} \tag{5}$$

The extension of (5) to a probability density where the random variable  $Y$  takes a continuous pointer  $x \in R^n$  with  $\frac{U(x)}{f(x)} \leq \frac{U(y)}{f(y)}$  for all  $y \neq x$  is,

$$p(Y = x) = \frac{f(x)}{\int_{R^n} f(y) dy}$$

This completes the proof.

**Proof of theorem 2**

As we did in proving Theorem 1, we start the proof by analyzing discrete distributions before generalizing to continuous distributions. Assume we have a discrete binary distribution that takes a pointer value 1 with probability  $p_1$  and pointer 2 with



probability  $p_2$  with  $p_1 + p_2$  not necessarily normalized to 1.  $X_1$  and  $X_2$  are independent and identically positive distributed random variables with exponential density  $f_X(x) = ae^{-ax}$  with  $a > 0$ . A new random variable  $Y$  is set to value 1 if  $\frac{X_1}{p_1} \leq \frac{X_2}{p_2}$  and it is set to value 2 otherwise. Using the independence property of  $X_1$  and  $X_2$ , a quick calculation will show that the probability that  $Y$  is equal to value 1 is:

$$P(Y = 1) = \int_0^\infty f_X(u_1) \left[ \int_{\frac{p_2 u_1}{p_1}}^\infty f_X(u_2) du_2 \right] du_1$$

Equivalently, we have,

$$P = (Y = 1) = a^2 \int_0^\infty e^{-au_1} \left[ \int_{\frac{p_2 u_1}{p_1}}^\infty e^{-au_2} du_2 \right] du_1 \tag{6}$$

The integration yields the normalized probabilities:

$$P(Y = 1) = \frac{p_1}{p_1 + p_2}$$

And

$$P(Y = 2) = \frac{p_2}{p_1 + p_2} \tag{7}$$

Now we are going to assume that the discrete distribution can take  $m$  possible values where  $\sum_{j=1}^m p_j$  is equal to some unknown constant. We could set the constant to 1 but the probability normalization is not necessary for the following calculations to yield normalized pointer probabilities. The probability of the random variable  $Y$  is equal to a pointer  $k$  between 1 and  $m$  where

$\frac{X_k}{p_k} \leq \frac{X_j}{p_j}$  for all  $j \neq k$  can be derived by generalizing (6) from the binary pointer case to  $m$  pointers:

$$P(Y = k) = a^m \int_0^\infty e^{-au} \prod_{j \neq k} \left( \int_{\frac{p_j u}{p_k}}^\infty e^{-au_j} du_j \right) du$$

The development of the integrals yields the following normalized probability result for any number of pointers,

$$P(Y = k) = \frac{p_k}{\sum_{j=1}^m p_j} \tag{8}$$

The extension of (8) to a probability density where the random variable  $Y$  takes a continuous pointer  $x \in R^n$  with  $\frac{X(x)}{f(x)} \leq \frac{X(y)}{f(y)}$  for all  $y \neq x$  is,

$$p(Y = x) = \frac{f(x)}{\int_{R^n} f(y) dy}$$

Which completes the proof.

A comprehensive proof should involve measure theory when going from discrete to continuous random variables and from a

discrete to a white noise that is defined everywhere. However, the applications of both theorems to numerical sampling from various target distributions provided the practical validity of both theorems.

## REFERENCES

1. CPSC 540: Machine Learning
2. An Introduction to MCMC for Machine Learning
3. Leggett AJ. Probing quantum mechanics towards the everyday world: where do we stand?. *Physica Scripta*. 2002;2002(T102):69.
4. Leggett AJ. Testing the limits of quantum mechanics: motivation, state of play, prospects. *J. Phys.: Condens. Matter*. 2002;14(15):R415.
5. C. Lanczos, B. Gellai., "Fourier Analysis of Random Sequences," *Comput. Math. appl.*1975; 1(3-4): 269-76.
6. Peligrad M, Wu WB. Central limit theorem for Fourier transforms of stationary processes.
7. Ghirardi G.C., Rimini A., Weber T. Unified dynamics for microscopic and macroscopic systems. *Phys. rev. D*. 1986;34(2):470.
8. Bell JS. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press. 1987, 2004. 17. On the impossible pilot wave.