An Alternative Method for Generating Almost Infinite Sequence of Gaussian Variables

Nyah C. Temaneh, F. A. Phiri, E. Ruhunga

Abstract**—**Most of the well known methods for generating Gaussian variables require at least one standard uniform distributed value, for each Gaussian variable generated. The length of the random number generator therefore, limits the number of independent Gaussian distributed variables that can be generated meanwhile the statistical solution of complex systems requires a large number of random numbers for their statistical analysis. We propose an alternative simple method of generating almost infinite number of Gaussian distributed variables using a limited number of standard uniform distributed random numbers.

Keywords**—**Gaussian variable, statistical analysis, simulation of Communication Network, Random numbers.

I. Introduction

THE development and design of a simulation tool for the planning and optimization of radio communication systems is a multi-dimensional process due to the large number of different design requirements and systems parameters and the degree of randomness involved. Such planning tools considerably simplify and increase the effectiveness of the designing, planning and optimization of radio communication systems. Amongst the tasks involved are the simulation of the electromagnetic environment and the estimation of its effect on radio communication devices by mathematical modelling methods. Monte Carlo simulation has become one of the most important tools in the field of science and engineering [1] for the statistical analysis of large and complex systems. The Monte-Carlo simulation provides the most accurate predictions in network planning and optimization tools through the utilization of Monte Carlo algorithms, thus providing reliable and credible results [1] – [4]. The Spectrum Engineering Advance Monte Carlo Analysis Tool – SEAMCAT [5] is an example of a software tool based on the Monte-Carlo simulation method. This tool permits statistical modeling of different radio interference scenarios for performing sharing and compatibility studies between radio communication systems in the same or adjacent frequency bands.

The Monte Carlo simulation methodology relies on a good source of numbers that appear to be random. Methods for producing pseudorandom numbers and transforming those numbers to simulate samples from various distributions are among the most important topics in statistical computing [6]. One of the distribution from which samples are often required is the Gaussian distribution.

Amongst the many ways for generating the value of a Gaussian (normally distributed random) variable include:

1. Inverse transform sampling method which takes a sample from the standard uniform distribution and maps it to a normally distributed sample by using the inverse probability integral.
2. The Box–Muller transform method which generates pairs of independent standard normally distributed random numbers, given a source of uniformly distributed random numbers.
3. The acceptance-rejection method in which some of the total input uniformly distributed random numbers generated is thrown away.
4. The use of a sum of large number of independent uniform distributed random variables (this sum according to the law of large numbers approximates to a normal distribution).

Most of the well known methods for generating the value of a Gaussian variable described above are different in procedure but have in common the fact that each Gaussian value generated, requires at least one sample from the standard uniform distribution. The length [6] (the number of random samples that can be generated by a random number generator without repetition) of the random number generator therefore, limits the number of independent Gaussian distributed variables that can be generated. At the same time the statistical solution of complex problems e.g. the planning and optimization of a radio communication system like a mobile cellular communication network [2, 4] require a large number of random numbers for their statistical analysis. Therefore the use of the existing methods of generating a normally distributed variable is limited.

In this paper, we propose an alternative simple method of generating two sets of almost infinite number of Gaussian distributed variables x_k and y_k on the bases of the complex

Nyah C. Temaneh is with the Electronics and Telecommunications Engineering Department, University of Namibia P.O.Box 3624 Ongwediva, Namibia (phone: +264652324111; fax: +264652324071; e-mail: cntemaneh@ unam.na).

F. A. Phiri is with the Electronics and Telecommunications Engineering Department, University of Namibia P.O.Box 3624 Ongwediva, Namibia (e-mail: fphiri@ unam.na).

E.Ruhunga is with the Electronics and Telecommunications Engineering Department, University of Namibia P.O.Box 3624 Ongwediva, Namibia (e-mail: eruhunga @ unam.na).

exponential $\exp(j2\pi k)$; where, $k = 1, 2, ...$ and using only a limited number of standard uniform distributed random numbers.

In section II, the proposed algorithm for generating the Gaussian distributed variables and the investigation of the important properties necessary to establish the Gaussian nature of the generated data are presented. The next sections will be dedicated to the investigations of the properties of normality as mentioned in this section. In section III, we investigate the correlation between the two sets of data x_k and y_k and its dependence on the sample seize N where we found that the two sets of data are uncorrelated. In section IV, we investigate the behaviour of the two sets of data as the sample seize N is increased towards infinity. By constructing the histograms of the data and calculating the mean value, the standard deviation for each set of data and the co-variation matrices were obtained for different sample sizes. In section V, we conclude on the results obtained.

II. Proposed Algorithm For Generating Gaussian Distributed Variables

Any complex number has a "real" (x) and "imaginary" ($j\,y$) part, and is expressed as $Z = x + j\,y$. The imaginary part is the square root of a negative number, which is really a non-existent number. Complex numbers can be used to define a two dimensional variable. The representation of a complex number requires two orthogonal axes. It can also be visualized as a point on the complex number plane, or as a vector originating at the origin and terminating at the point. The complex exponential $\exp(j2\pi k)$ can be expressed in the form $\exp(j2\pi k) = \cos(2\pi\,\mathrm{k}) + j\sin(2\pi\,\mathrm{k})$, where, $k = 1, 2, \ldots$.

A sequence of complex random variables Z_k, can be formed using the complex exponential $\exp(j2\pi k)$ and a sequence of numbers X_i $(i = 1, 2, 3, \ldots, n)$ which are independent and standard uniformly distributed as

$$
\begin{aligned}
Z_k &= \sqrt{\frac{2}{n}}\cdot\left[\sum_{i=1}^{n}\exp\left(j\,2.\pi.k.X_i\right)\right] \\
&= \sqrt{\frac{2}{n}}\sum_{i=1}^{n}\left[\cos\left(2\pi k X_i\right) + j\sin\left(2\pi k X_i\right)\right] \qquad (1) \\
&= x_k + j\,y_k
\end{aligned}
$$

Therefore, it is possible to assign a value for n (example, between 10 and 50) and then generate a sequence of standard uniform distributed random numbers X_i then, performing calculations according to the algorithm (Fig. 1) based on expression (1), we obtain two random sequences x_k and y_k. In Fig.1, the random number generator (RNG) provides the standard uniform distributed variable X.

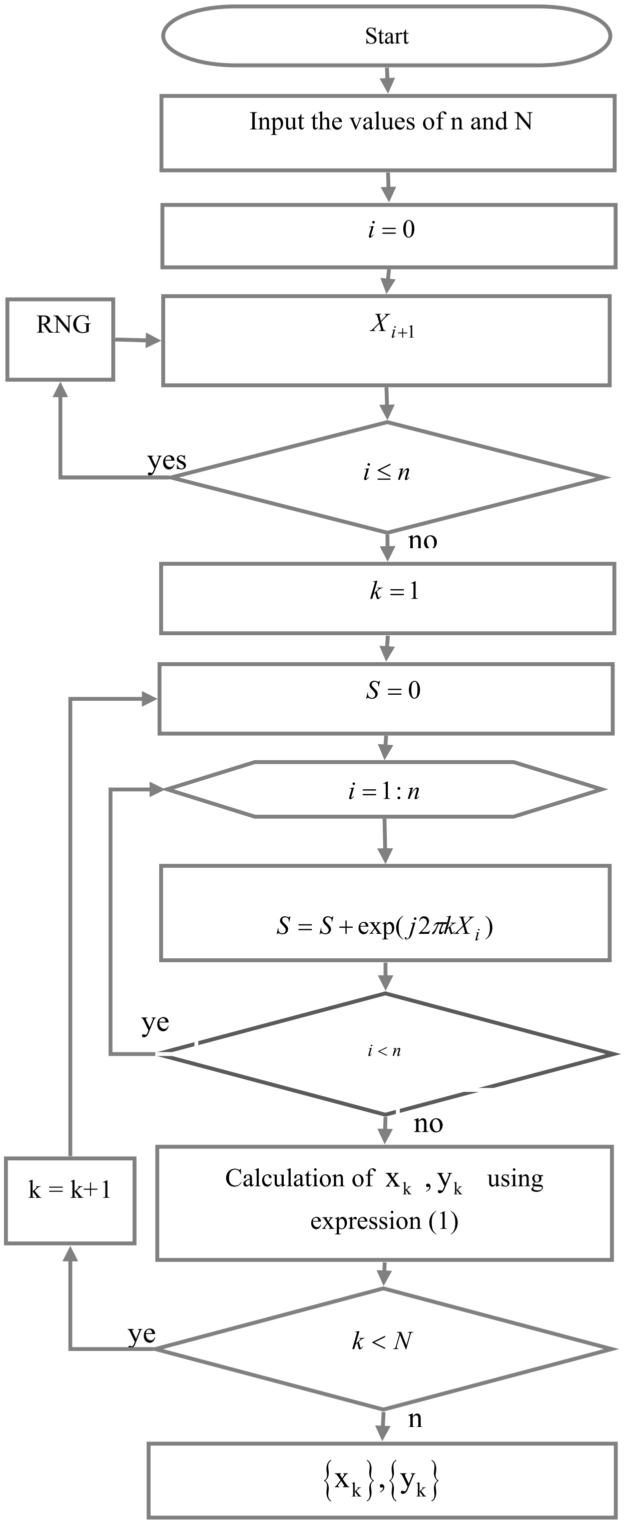

Fig. 1 Algorithm for the generation of independent Gaussian sequences x_k and y_k based on the expression (1)

In order to establish that the random sequences x_k and y_k are each normally distributed, we have to investigate the following important properties:

1. That the sequences of random variables x_k, y_k are uncorrelated.
2. That as the length of each of the sequences x_k and y_k, turns to infinity, the random variables are independent and each asymptotically normal (central limit theorem).
3. That the mathematical expectations of each of the random sequences x_k and y_k equals zero and the co-variation matrix equals unity.

The next sections will be dedicated to the investigations of the above properties for the normality of x_k and y_k.

III. Correlation Between Random Sequences x_k and y_k

The correlation coefficient (r) is a statistical calculation that is used to examine the relationship between two sets of data, in this case the sequences x_k and y_k. The value of the correlation coefficient will always be between 1 and -1 and it tells us about the strength and the nature of the relationship. A correlation coefficient that has a value of exactly 1 or -1 would be a perfectly straight line when the data sets are plotted on a graph. Values of r that are close to 0 tell us there is no relationship between the sets of data x_k and y_k. Positive correlation coefficients tell us that when one variable is increased, the other variable will increase as well. Negative correlation coefficients tell us that as one of the variables is increased, the other variable will decrease.

The correlation coefficient r will be calculated according to the expression [7]:

$$r = \frac{\sum_{k=1}^{N} x_k y_k - \frac{1}{N}\left(\sum_{k=1}^{N} x_k . \sum_{k=1}^{N} y_k\right)}{\sqrt{\left[\sum_{k=1}^{N} x_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N} x_k\right)^2\right] \times \left[\sum_{k=1}^{N} y_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N} y_k\right)^2\right]}}$$

With the value of n taken as 40, the sequences of data x_k and y_k were each generated according to expression (1) above for values of $k = 1, 2, ... N$, where N is the length (sample or data size) of each of the sequences and was taken to be between 10000 and 100000 samples. The correlation coefficient r as a function of generated data size N was obtained and graphically (Fig. 2).

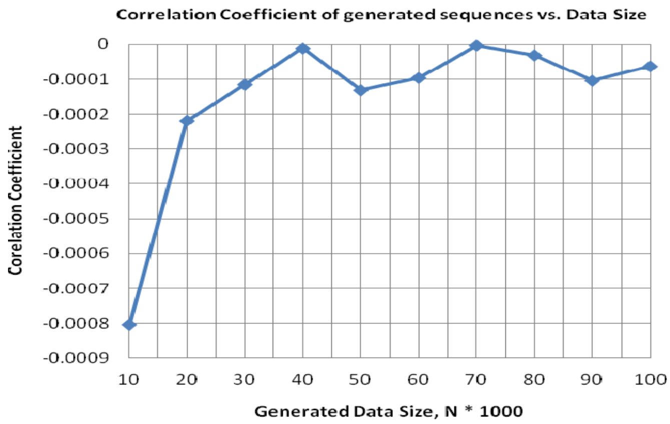

Fig. 2 Correlation Coefficient vs. generated Data Size

From the results in Fig. 1, it is noted that the correlation coefficient r fluctuates in the neighborhood of zero for all the investigated values of the generated data size N. This is a very strong indication that the random numbers x_k and y_k thus generated are uncorrelated.

IV. The Behaviour Of The Random Sequences x_k And y_k AS $N \to \infty$

The random sequences x_k and y_k were generated as before according to the proposed algorithm based on expression (1) for values of data size N between 10000 and 50000 samples. The histogram plot for both generated sequences x_k and y_k for the different values of generated data size N were obtained (Fig. 3).

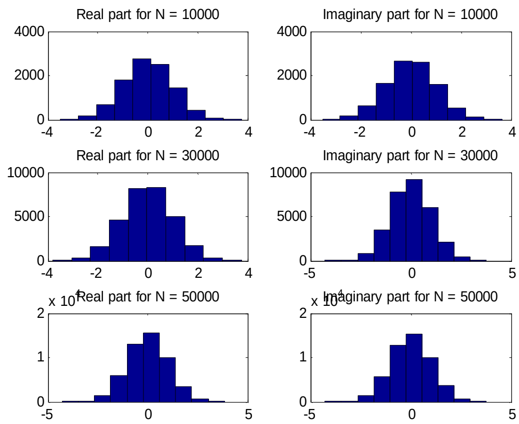

Fig. 3 Histogram for generated random sequences x_k and y_k for different values of N

For each data size, the mean values ($\overline{x_k}$, $\overline{y_k}$), standard deviations $(\sigma_{x_k}, \sigma_{y_k})$ and the co-variation matrix $(\text{cov}_x, \text{cov}_y)$ for each sequence were obtained (Table 1-3) and graphically represented (Fig. 4 -6) respectively.

TABLE I
THE MEAN OF THE GENERATED RANDOM SEQUENCES x_k AND y_k

N		10000	30000	50000
x_k	$\overline{x_k}$	-0.000997	-0.000158	-8.22e-05
y_k	$\overline{y_k}$	0.0006618	8.4545e-05	4.0099e-05

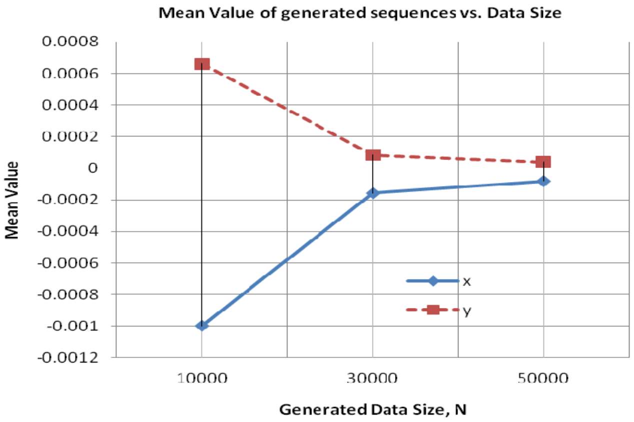

Fig. 4 The MEAN value of the of generated random sequences x_k and y_k as a function of N

TABLE II
THE STANDARD DEVIATION OF THE GENERATED RANDOM SEQUENCES x_k AND y_k

N		10000	30000	50000
x_k	σ_{x_k}	0.9991	0.9973	0.9984
y_k	σ_{y_k}	1.00082	0.9980	0.9988

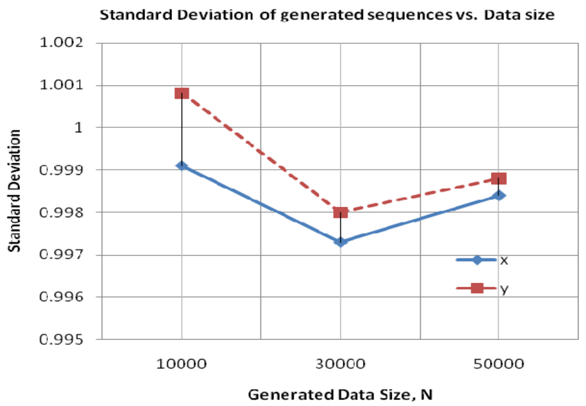

Fig. 5 The Standard deviation of the generated random sequences x_k and y_k as a function of N

TABLE III
THE COVARIANCE MATRIX OF THE GENERATED RANDOM SEQUENCES x_k AND y_k

N		10000	30000	50000
x_k	cov_x	0.9983	0.9946	0.9969
y_k	cov_y	1.0016	0.9961	0.9976231

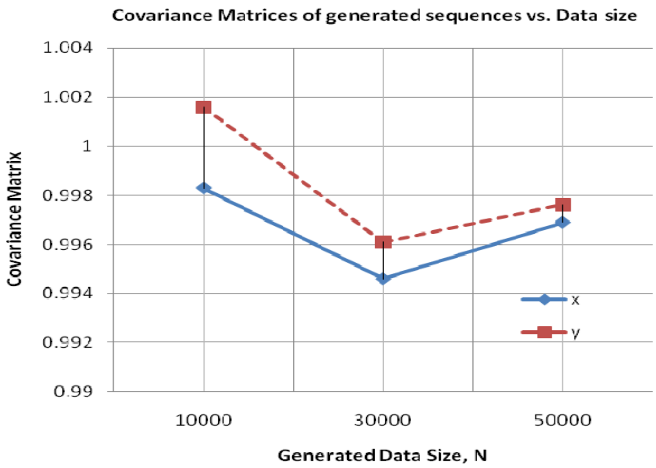

Fig. 6 The Covariance matrix of the generated random sequences x_k and y_k as a function of N

V. Discusion and Analysis of Results

From the histogram (Fig. 3), it can be seen that as the sample size N increases, the random sequences x_k and y_k generated by the proposed algorithm approximate to a normal distribution. The results in Table 1, Table 2 and Table 3 show that as the sample size N increases, the mean value of the sequences x_k and y_k is approximately zero, the standard deviation and covariance matrix each approximate to one respectively. Therefore, the algorithm (Fig. 1) based on the expression (1) can be used to generate independent random sequences x_k and y_k each having an approximate standard normal distribution (mean of zero and standard deviation of one) as k increases.

The length of each sequence N is defined as $N = (10^p - 1)$, where p is the number of decimal digits of the random number X. If $k > 10^p - 1$, the values of Z_k (x_k and y_k) starts to repeat i.e. $Z_{k+N} = Z_k \Rightarrow x_{k+N} = x_N$ and $y_{k+N} = y_N$.

Increasing the value of p by one, the value of N increases by $\frac{10^p.10}{10^p} \approx 10\,times$. Therefore the length of the random sequence can be increased considerable by increasing the number of decimal digits of the random number X.

VI. CONCLUSION

The implementation of the proposed simple method of generating almost infinite sequence of normally distributed variables and the analysis of the results showed that indeed an almost infinite sequence of random numbers distributed normally can be achieved using only a limited number of standard uniform distributed random numbers.

This method will be of interest to those who for example use simulation for the solution of complex system involving a high degree of randomness like in cellular mobile communication networks.

REFERENCES

[1] James E. Gentle. Random Number Generation and Monte Carlo Methods, Series: Statistics and Computing. 2nd ed. 2003. Corr. 2nd printing, 2003, XV, 300 p., Hardcover. ISBN: 978-0-387-00178-4

[2] Nyah. C. Temaneh, "Monte-Carlo Technique Estimation of a Probability of Intermodulation Interference in a Cellular Wireless Communication Network", Proceedings 2010 IEEE Region 8 International Conference on Computational Technologies in Electrical and Electronics Engineering, Irkutsk, Russia 2010, pp. 329 – 334.

[3] Nyah. C. Temaneh, K. E. Vinogradov, A. N. Krenev, "Monte-Carlo based estimation of probabilistic characteristics of signal to noise ratio with GSM 900 cellular communication network as case study (in Russian)", in Proceedings of IX international scientific – technical conference on Radiolocation, Navigation and Communication, Voronesh (Russia), vol. 2, 2005, pp. 1182 – 1188.

[4] Nyah. C. Temaneh, "Estimation of a probability of interference in a cellular communication network using the Monte Carlo Technique." Proceedings of the Southern African Telecommunications and Networks Conference, SATNAC 2009, Swaziland, September 2009.

[5] ERC Report 68,"Monte-Carlo Simulation Methodology for the use in sharing and compatibility studies between different radio services or systems" Naples, February 2000

[6] D. H. Lehmer. Mathematical methods in large-scale computing units. In Proc. 2nd Sympos. On Large Scale Digital Calculating Machinery, Cambridge, MA, 1949, PP. 141-146, Cambridge, MA, 1951. Harvard University Press.

[7] Gurskiy E. I., Probability theory with elements of mathematical statistics. Moscow, 1971.

Nyah C. Temaneh (M'2009) was born in Cameroon. He received the Master of Science (MSc.) in Engineering from the Faculty of Radio communication, Radio broadcasting and Television at the Ukrainian State Telecommunications Academy – Odessa in 1995 and a Ph.D. degree in Telecommunications from the Yaroslavl State University – Russia in 2007. Presently he is Senior lecturer at the University of Namibia. His research interest is on Telecommunication's network performance simulation models, for network management, planning, optimisation and analysis including electromagnetic compatibility analysis.