

Efficient Generation of Random Signals with Prescribed Probability Distribution and Spectral Bandwidth via Ergodic Transformations

Andre M. McDonald
Modelling and Digital Science Unit
Council for Scientific and Industrial Research
Brummeria, South Africa
email: amcdonald@csir.co.za

Michaël A. van Wyk
School of Electrical and Information Engineering
University of the Witwatersrand
Johannesburg, South Africa

Abstract—A novel random signal generator design that accommodates the specification of both the sample probability distribution as well as the signal bandwidth is presented in this paper. The generator achieves a high degree of computational efficiency through the nonlinear transformation of trajectories produced by a discrete-time dynamical system which has an ergodic map as evolution rule. The ergodic map is designed using a recently proposed solution of the inverse Frobenius–Perron problem that allows for the selection of the map’s invariant distribution as well as its spectral characteristics. The nonlinear transformation is obtained via a novel piecewise polynomial fitting algorithm, which facilitates the approximation of absolutely continuous probability distributions over compact support with greater accuracy than existing techniques. Numerical experiments indicate that the proposed design achieves a reduction in signal generation time of up to 22% compared to a conventional generator, while at the same time using a smaller lookup table, maintaining a comparable level of accuracy, and offering flexibility in the selection of the signal bandwidth. It is concluded that the proposed approach is suitable for signal generation in applications where low computational complexity is a critical requirement.

I. INTRODUCTION

Performance evaluation of signal processing systems via Monte–Carlo simulation involves the processing of sample functions from random signal models of the system environment [1]. During simulation, sample functions are typically produced by random signal generators that offer the flexibility to specify both the probability distribution and power spectrum of the random signal model. These generators are required to maintain an appropriate tradeoff between computational efficiency and accuracy in adhering to the prescribed statistics, as dictated by the context in which the generators are applied.

Computational efficiency is a critical requirement of systems that perform hardware–in–the–loop performance evaluation of operational radar receivers [2]–[4]. Within this context, pseudorandom signals are typically generated by field programmable gate array devices that are referred to as digital radio frequency memory (DRFM) systems. These systems are often required to generate samples at multi–gigahertz rates while maintaining flexibility and exceeding a minimum acceptable level of accuracy. Conventional random signal generators,

which follow the approach of filtering and transforming a Gaussian random process (Fig. 1, top), are sufficiently flexible for use in DRFM systems. However, these generators are computationally inefficient when sampling from several common probability distributions [1]; this leads to a scenario where accuracy has to be sacrificed for efficiency through low order approximation of the prescribed distribution.

This paper presents a novel random signal generator that is suitable for use in applications where efficiency is a critical requirement, but where accuracy and flexibility are to be maintained. The generator’s design (Fig. 1, bottom) follows an alternative approach towards random signal generation that leads to a significant gain in computational efficiency. Instead of using a Gaussian random process generator, the design incorporates a discrete–time dynamical system which has an ergodic map as evolution rule. The map is recursively evaluated to directly produce a uniformly distributed signal with prescribed bandwidth, thereby rendering the conventional generator’s transformation of the Gaussian process to a uniform process unnecessary. The map is a member of the semi–Markov piecewise linear maps, and is designed using a recent solution of the inverse Frobenius–Perron (FP) problem [5], [6] that facilitates selection of the map’s invariant distribution as well as its spectral characteristics.

A memoryless nonlinear transformation is applied to the uniformly distributed samples in order to realize a prescribed absolutely continuous probability distribution over compact support. Whereas the transformation equal to the inverse cumulative distribution function (CDF) F^{-1} of the prescribed distribution induces this distribution, the proposed design uses a piecewise polynomial approximation of F^{-1} to improve computational efficiency. Existing algorithms for fitting polynomials in this context [1], [7] have deficiencies that lead to unnecessary intervals in the approximation; this leads to a larger coefficient lookup table and additional time required to search through the table. These drawbacks are addressed by a novel monotonic cubic polynomial fitting algorithm that produces a more accurate approximation of the inverse CDF over a smaller number of intervals.

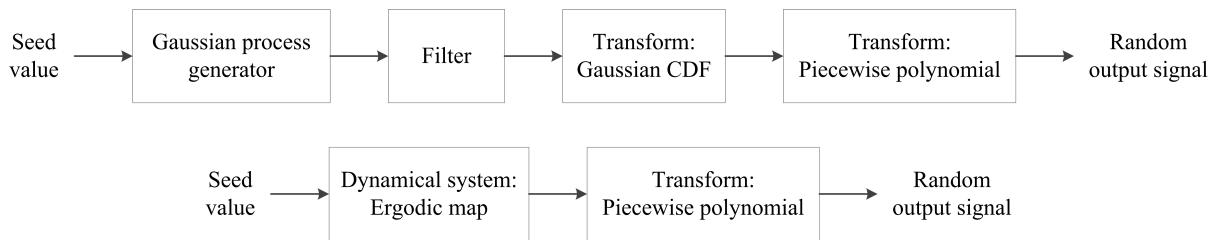


Fig. 1. Conventional (top) and proposed (bottom) random signal generators.

Numerical experiments were carried out to benchmark the performance of the novel random signal generator against a second generator that follows the conventional approach to signal generation. The results indicate that the proposed generator is significantly more efficient, achieving a reduction in generation time of between 18% and 22% for the beta distribution at the same level of accuracy as the conventional generator. The proposed generator also maintains smaller lookup tables (a reduction of up to 33%), while offering flexibility in specifying the signal bandwidth.

The remainder of this paper is set out as follows. Section II presents a survey of random signal generator designs aimed at maximizing efficiency while maintaining acceptable accuracy and flexibility. The fundamentals of random signal generation through the use of ergodic maps are presented in section III. Section IV presents the design of the proposed generator, whereas the results of experimentation are presented in section V. Conclusions are drawn in section VI.

II. LITERATURE SURVEY

Conventional random signal generators that accommodate the specification of both the sample distribution and the power spectrum of the signal follow the filter-and-transformation approach, in which a filter shapes the power spectrum of a Gaussian random process, and a transformation induces the prescribed sample distribution (refer to [1] and Fig. 1, top). The transformation equal to the composition of the Gaussian CDF and the inverse CDF F^{-1} of the prescribed distribution induces this distribution exactly. However, the computational complexity associated with the evaluation of this transformation is unacceptably high in cases where F^{-1} has to be evaluated via numeric search, which includes many common distributions such as the gamma and beta distributions [1].

Several methods aimed at lowering the transformation complexity of conventional generators were investigated in [1], [7], [8]. These methods trade accuracy in reproducing the prescribed distribution for efficiency by constructing approximations of F^{-1} that are computationally less expensive to evaluate. The approximation of F^{-1} with straight lines over disjoint intervals was considered in [1]. Despite the low computational cost of evaluating the approximation over each interval, the accuracy of the approximation improves slowly with respect to the number of intervals. This leads to impractically large coefficient lookup tables and a slower search-based table lookup during generation.

The nonlinear approximation of F^{-1} with piecewise-defined higher-order polynomials over a fixed number of intervals was considered in [8]. In this approximation, the interval boundaries are chosen to ensure that those intervals towards the center of the domain $[0, 1]$ are of uniform width, whereas the width of each successive interval is halved towards the distribution tail. Lagrange interpolation is carried out to obtain a Chebyshev polynomial approximation. This method presents a substantial improvement over linear approximation. However, due to its complexity and an excessive amount of time required to derive the approximation, the method is not considered to be suitable for routine use in approximating arbitrary distributions [1].

The approximation of F^{-1} with third and fifth order polynomials defined piecewise over a variable number of intervals was considered in [1], [7]. The segmentation strategy involves the recursive bisection of the domain until a specified level of accuracy is achieved over each interval. This method produces generators that are computationally efficient; however, its recursive segmentation strategy tends to produce unnecessary intervals while approaching those regions of the domain where the slope of F^{-1} is steep. Furthermore, the method does not guarantee that each polynomial is strictly monotonic — if this is not the case, the straightforward inverse relationship between the approximation and the resulting CDF no longer holds, leading to a loss in accuracy.

III. RANDOM SIGNAL GENERATION VIA ERGODIC MAPS

Consider a nonlinear map $S : \mathcal{I} \rightarrow \mathcal{I}$, where $\mathcal{I} = [0, 1]$ is the unit interval over \mathbb{R} . Let S be measurable over the Borel σ -algebra \mathbb{B} on \mathcal{I} and nonsingular with respect to the Borel measure μ defined on \mathbb{B} . Furthermore, let X_0 denote a random variable on \mathcal{I} with an absolutely continuous distribution and probability density function (PDF) f_0 . The generation of sample functions of the random process $\{X_1, X_2, \dots\}$ defined by the expression $X_{i+1} = S(X_i)$, $i \in \{0, 1, \dots\}$, is considered.

The FP operator $\mathcal{P}_S : L^1 \rightarrow L^1$ associated with the map S characterises the evolution of PDF f_i under the evaluation of S , such that $f_{i+1} = \mathcal{P}_S(f_i)$ [9]. Let S be an ergodic map. It follows that there exists at most one PDF f_S^* that is a stationary density of \mathcal{P}_S — i.e., $f_S^* = \mathcal{P}_S(f_S^*)$. If it exists, the density f_S^* is referred to as the invariant density with respect to S .

A consequence of ergodicity is that sample averages converge almost everywhere to ensemble averages in the limit of infinitely long trajectories generated by measure-preserving

maps [9]. This implies that the sample distributions of trajectories $\{x_1, x_2, \dots\}$ converge to f_S^* asymptotically for almost all x_0 . The proposed approach to random signal generation exploits this property; specifically, a generator with a uniform density f over the domain $[0, 1]$ is realized by designing an ergodic and measure preserving map S such that its FP operator \mathcal{P}_S has the stationary density $f_S^* = f$. Furthermore, by virtue of the fact that the eigenspectrum of \mathcal{P}_S determines the characteristics of the time autocorrelation function [6], a power spectrum with the required characteristics is realized by selecting the eigenspectrum of \mathcal{P}_S during the design phase. Sample functions are generated by evaluating $x_{i+1} = S(x_i)$ for distinct initial values x_0 , which serve as seed values.

The proposed generator uses a map from the class of semi-Markov ergodic maps [10], due to their computational simplicity and flexibility. A \mathcal{Q} -semi-Markov map is defined by a hierarchy of two sets \mathcal{Q} and \mathcal{R} of nonoverlapping intervals that partition the domain \mathcal{I} . Let $\mathcal{Q} = \{Q_1, Q_2, \dots, Q_N\}$, where $Q_n = [q_{n-1}, q_n]$ for $n = 1, 2, \dots, N-1$, $Q_N = [q_{N-1}, 1]$ and $q_0 = 0$. Furthermore, let $\mathcal{R} = \{(R_{j,1})_{j=1}^{l(1)}, (R_{j,2})_{j=1}^{l(2)}, \dots, (R_{j,N})_{j=1}^{l(N)}\}$ such that $Q_n = \cup_{j=1}^{l(n)} R_{j,n}$. The map $S : \mathcal{I} \rightarrow \mathcal{I}$ is \mathcal{Q} -semi-Markov if S restricted to each interval $R_{j,n}$ is monotonic and $S(R_{j,n}) \in \mathcal{Q}$ for all $n = 1, 2, \dots, N$ and $j = 1, 2, \dots, l(n)$.

Any invariant density f_S^* of a piecewise linear and expanding \mathcal{Q} -semi-Markov map is piecewise constant on the intervals of \mathcal{Q} [10]. This result facilitates the restriction of the FP operator \mathcal{P}_S of any map from this class to the space of functions constant on the intervals of \mathcal{Q} . Let PDFs with domain \mathcal{I} that belong to this space be represented by row vectors \mathbf{f} of length N , such that each vector element equals the constant value of the density over the corresponding interval. The FP operator \mathcal{P}_S may now be represented by an N -by- N matrix \mathbf{P}_S (the FP matrix of S), such that $\mathbf{f}_{i+1} = \mathbf{f}_i \mathbf{P}_S$. It follows that the invariant density \mathbf{f}_S^* is equal to the normalized left eigenvector of \mathbf{P}_S that corresponds to the unity eigenvalue.

The eigenvalues of \mathbf{P}_S determine the dominant spectral characteristics of the map [6]; i.e., each eigenvalue corresponds to a distinct spectral mode, where the mode's center frequency is equal to the eigenvalue argument, and the mode's bandwidth is inversely proportional to the eigenvalue magnitude. Hence, both the invariant density and the dominant spectral characteristics of an ergodic semi-Markov map are determined by its FP matrix eigenvalues and eigenvectors.

IV. RANDOM SIGNAL GENERATOR DESIGN

The design of the proposed generator is presented in Fig. 1 (bottom). The generator uses a discrete-time dynamical system which has an ergodic \mathcal{U} -semi-Markov map $S : \mathcal{I} \rightarrow \mathcal{I}$ as evolution rule, where \mathcal{U} is the uniform partition of $\mathcal{I} = [0, 1]$. The recursive evaluation of the evolution rule, starting with an initial (seed) value x_0 , generates a signal with a uniform distribution and with a prescribed bandwidth. The prescribed distribution is realized through a memoryless nonlinear transformation, which is a piecewise-defined polynomial approximation of the distribution's inverse CDF.

The map S is designed using a recently proposed solution of the inverse FP problem [5], [6]. The design process involves the synthesis of a doubly stochastic matrix \mathbf{P} with a prescribed eigenvalue spectrum, and the subsequent derivation of the \mathcal{U} -semi-Markov map S such that its FP matrix \mathbf{P}_S is equal to \mathbf{P} . The doubly stochastic matrix \mathbf{P} is constructed using the recursive Markov state disaggregation algorithm [5] — starting with the elementary one-state Markov chain, the states of the Markov chain are recursively disaggregated (or ‘split’) in a manner such that the resulting states are equiprobable. Following a specified number of rounds of disaggregation, the state transition matrix of the final Markov chain is selected as the matrix \mathbf{P} . The disaggregation algorithm allows for the selection of an additional transition matrix eigenvalue as each state is disaggregated, thereby providing control over the signal bandwidth. The \mathcal{U} -semi-Markov map S is constructed from the entries of \mathbf{P} using the algorithm provided in [6], [10].

The nonlinear transformation is a piecewise-defined cubic polynomial approximation \hat{G} of the inverse CDF of the prescribed distribution. Let $G = F^{-1}$ denote the inverse CDF, where $G : \mathcal{I} \rightarrow [a, b]$. The approximation \hat{G} is defined over the intervals $[x_1, x_2], [x_2, x_3], \dots, [x_K, x_{K+1}]$, where the interval endpoints are selected iteratively, and where $x_1 \triangleq \delta_L$ and $x_{K+1} \triangleq 1 - \delta_R$ such that $0 < \delta_L, \delta_R \ll 1$. The coefficients of each polynomial are selected to ensure that the approximation \hat{G} and its first derivative are respectively equal to G and G' at each interval's endpoints, thereby ensuring that both the PDF and CDF of the resulting distribution are perfectly fitted to the prescribed distribution at these points. The coefficients of $\hat{G}(x)|_{[x_k, x_{k+1}]} \triangleq a_k(x - x_k)^3 + b_k(x - x_k)^2 + c_k(x - x_k) + d_k$ are computed using the expressions

$$a_k = (\Delta x_k)^{-2}[-2m_k + G'(x_k) + G'(x_{k+1})], \quad (1)$$

$$b_k = (\Delta x_k)^{-1}[3m_k - 2G'(x_k) - G'(x_{k+1})], \quad (2)$$

$$c_k = G'(x_k), \quad (3)$$

$$d_k = G(x_k), \quad (4)$$

where $\Delta x_k \triangleq x_{k+1} - x_k$, $\Delta y_k \triangleq G(x_{k+1}) - G(x_k)$ and $m_k \triangleq \Delta y_k / \Delta x_k$, for $k = 1, 2, \dots, K$.

Iterative segmentation is performed in a manner which ensures that the L^1 norm of the difference between \hat{G} and G over the domain \mathcal{I} does not exceed a prescribed error threshold. Let the error $\epsilon(x_k, x_{k+1})$ over the interval $[x_k, x_{k+1}]$ be defined as

$$\epsilon(x_k, x_{k+1}) \triangleq \int_{x_k}^{x_{k+1}} |\hat{G}(x) - G(x)| dx. \quad (5)$$

The interval endpoints are selected to ensure that the normalized error over each interval does not exceed a specified threshold ϵ^* ; i.e., $\epsilon(x_k, x_{k+1}) / \Delta x_k \leq \epsilon^*$ for $k = 1, 2, \dots, K$. Assuming that $a \leq \hat{G}(x) < G(\delta_L)$ for $x < \delta_L$ and $G(1 - \delta_R) < \hat{G}(x) \leq b$ for $x > 1 - \delta_R$, it follows that $\epsilon(0, 1) < \epsilon^* + \epsilon_L^* + \epsilon_R^*$, where $\epsilon_L^* = [\max\{\delta_L, G(\delta_L) - a\}]^2$ and $\epsilon_R^* = [\max\{\delta_R, b - G(1 - \delta_R)\}]^2$. The endpoints δ_L and $1 - \delta_R$ are obtained from the prescribed thresholds ϵ_L^* and ϵ_R^* via numeric search during the first step of segmentation.

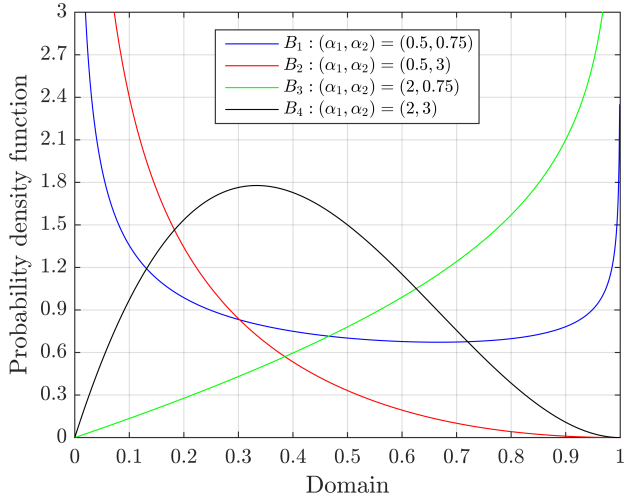


Fig. 2. Probability density functions of beta distributions B_1 to B_4 .

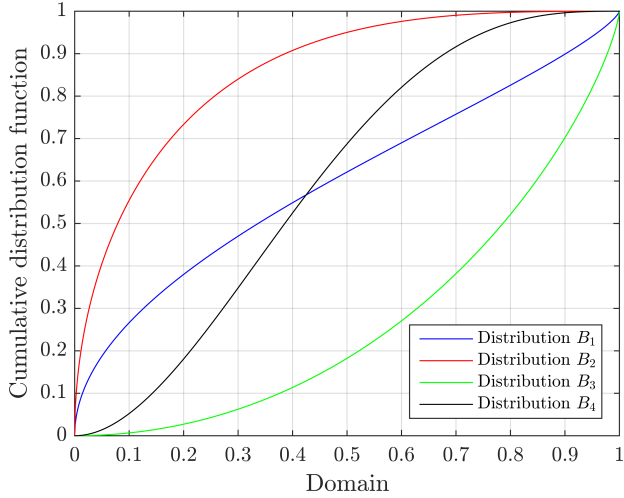


Fig. 3. Cumulative distribution functions of beta distributions B_1 to B_4 .

The right endpoints of the intervals are selected iteratively, starting with the leftmost interval $[\delta_L, x_2]$ and proceeding to the right. The right endpoint x_{k+1} of the interval $[x_k, x_{k+1}]$ is selected as the largest $x \in R_k$, where $R_k \triangleq (x_k, 1 - \delta_R]$, such that the normalized error over the interval $A_k \triangleq [x_k, x]$ does not exceed ϵ^* , and which ensures monotonicity of the polynomial over the interval. Monotonicity is maintained [11] over the interval A_k if $0 \leq \alpha(A_k), \beta(A_k) < 3$, where $\alpha(A_k) \triangleq G'(x_k)/m(x_k, x)$, $\beta(A_k) \triangleq G'(x)/m(x_k, x)$, and $m(x_k, x) \triangleq [G(x) - G(x_k)]/[x - x_k]$. It follows that

$$x_{k+1} = \max\{x \in R_k : \alpha(A_k), \beta(A_k) \in [0, 3), \tilde{\epsilon}(A_k) \leq \epsilon^*\}, \quad (6)$$

where $\tilde{\epsilon}(A_k) \triangleq \epsilon(x_k, x)/(x - x_k)$. Numerical experimentation has led to the conclusion that the value of x_{k+1} may readily be obtained, for smooth probability distributions, via a numeric search over a sufficiently fine grid in R_k .

The proposed design uses additional transformations to achieve a more accurate fit to distributions with heavier tails.

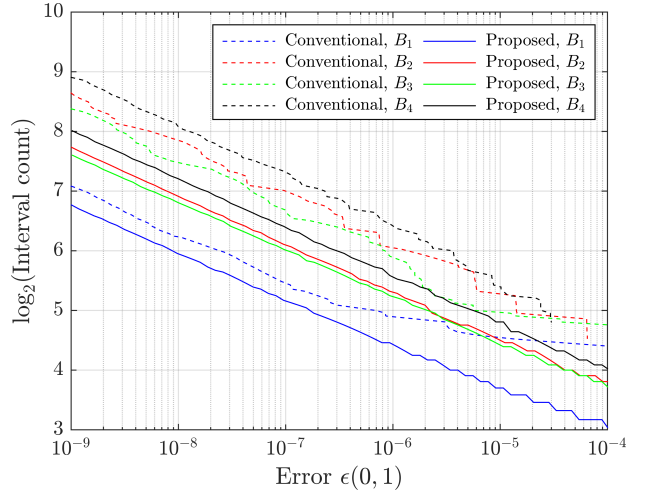


Fig. 4. Proposed and conventional generators' interval count, as a function of the polynomial approximation error $\epsilon(0, 1)$.

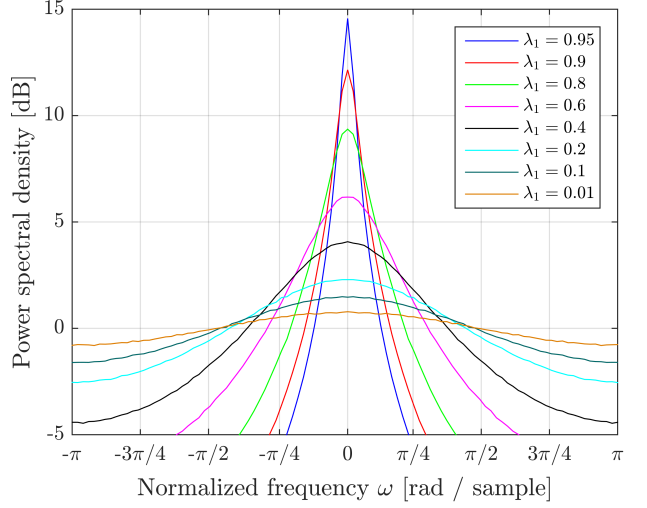


Fig. 5. Power spectra of the proposed generator.

Two intervals $[\delta_L, T_L]$ and $[1 - T_R, 1 - \delta_R]$ are defined over the domain of G , such that these intervals correspond to the left and right tails of the prescribed distribution. Within these intervals, cubic polynomials are fitted over $z \triangleq \log(x)$ to $\log(G(e^z) - a)$ and $\log(b - G(1 - e^z))$ respectively, thereby stretching the domain and range of G . Whereas these transformations facilitate more accurate approximation over fewer intervals, they incur the computational costs of exponentiation and the evaluation of logarithms.

V. NUMERICAL EXPERIMENTS

The performance of the proposed generator was benchmarked against the conventional generator of Fig. 1. The generators were designed to produce beta distributed signals with PDF $f(x, \alpha_1, \alpha_2) = x^{\alpha_1-1}(1-x)^{\alpha_2-1}/B(\alpha_1, \alpha_2)$, $x \in (0, 1)$, where $B(\alpha_1, \alpha_2) \triangleq \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1 + \alpha_2)$ and $\alpha_1, \alpha_2 > 0$ denote the distribution parameters. Four distinct parameter pairs (α_1, α_2) were selected, thereby producing dis-

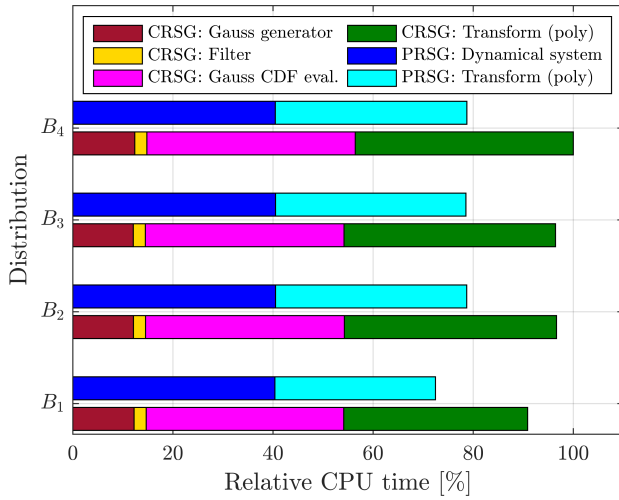


Fig. 6. Relative processing time of the conventional (CRSG) and proposed (PRSG) generators.

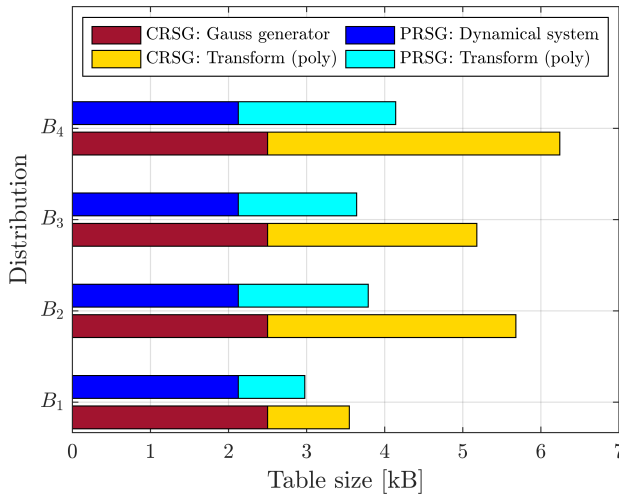


Fig. 7. Table size of the conventional (CRSG) and proposed (PRSG) generators.

distributions B_1 to B_4 (Figs. 2 and 3) with distinct characteristics that facilitate a thorough performance evaluation.

The proposed generator’s ergodic map was designed over the $N = 16$ interval partition \mathcal{U} , and with dominant eigenvalue λ_1 as indicated in Fig. 5. The polynomial approximation was computed with $\epsilon_L^* = \epsilon_R^* = 10^{-16}$ and $\epsilon^* = 10^{-9}$ for the efficiency benchmark. The tail intervals were defined with parameters $T_L = T_R = 10^{-3}$. The conventional generator was designed with the goal of maximizing efficiency, and with the same approximation error as the proposed generator for the efficiency benchmark; the ziggurat Gaussian generator [12] and the Chebyshev approximation of the Gauss CDF [13] were incorporated in the design. The polynomial approximation of the conventional generator was derived using the cubic Hermite interpolation method of [1]. The conventional generator uses a first order IIR filter for spectral shaping.

Fig. 4 compares the number of intervals produced by the

two fitting algorithms in approximating B_1 to B_4 , and while maintaining the same level of accuracy. The novel algorithm outperforms the conventional algorithm and achieves a reduction of up to half the interval count for heavier-tailed distributions B_2 to B_4 ; this leads to smaller lookup tables that are less expensive to search. The novel generator simultaneously accommodates a wide range of signal bandwidths (Fig. 5).

The efficiency benchmark is presented in Fig. 6, whereas the corresponding lookup tables are compared in Fig. 7. The proposed generator achieves a reduction in processing time of 18% to 22% during signal generation. The greatest gain is observed for the heavier-tailed distribution B_4 ; this implies that additional computation in the tail is compensated for by the efficiency gain due to interval count reduction. The novel generator uses smaller tables, with up to 33% size reduction.

VI. CONCLUSIONS

The results indicate that the proposed approach to random signal generation, which involves the use of ergodic maps and piecewise polynomial transformations constructed via a novel fitting algorithm, provides a significant computational efficiency gain and table size reduction compared to the conventional design. This performance gain is achieved without compromising on accuracy and while accommodating the selection of the signal bandwidth. The proposed design is suitable for applications where efficiency is critical.

REFERENCES

- [1] W. Hörmann, J. Leydold, and G. Derflinger, *Automatic Nonuniform Random Variate Generation*. Springer Science & Business Media, 2004.
- [2] K. Olivier, J.E. Cilliers, and M. Du Plessis, “Design and performance of wideband DRFM for radar test and evaluation,” *IET Electronic Letters*, vol. 47, no. 14, pp. 824–825, 2011.
- [3] J.J. Strydom, J.E. Cilliers, A.M. McDonald, and K. Olivier, “Hardware in the loop testing and evaluation of seaborne search radars,” in Proc. *49th Annual AOC International Convention and Symposium*, 2012.
- [4] J.E. Cilliers, A.M. McDonald, J.J. Strydom, and M.A. Van Wyk, “Hardware in the loop (HIL) generation of airborne clutter using a sum of complex sinusoids,” in Proc. *2015 IEEE Radar Conference*, Arlington, VA, 2015, pp. 1605–1609.
- [5] A.M. McDonald and M.A. van Wyk, “Solution of the inverse Frobenius-Perron problem for semi-Markov chaotic maps via recursive Markov state disaggregation,” in Proc. *25th European Signal Processing Conference*, Kos, Greece, 2017, pp. 1604–1608.
- [6] A.M. McDonald and M.A. van Wyk, “Construction of semi-Markov ergodic maps with selectable spectral characteristics via the solution of the inverse eigenvalue problem,” in Proc. *9th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Kuala Lumpur, Malaysia, 2017, pp. 987–993.
- [7] W. Hörmann and J. Leydold, “Continuous random variate generation by fast numerical inversion,” *ACM Transactions on Modeling and Computer Simulation*, vol. 13, no. 4, pp. 347–362, 2003.
- [8] J. H. Ahrens and K.D. Kohrt, “Computer methods for efficient sampling from largely arbitrary statistical distributions,” *Computing*, vol. 26, no. 1, pp. 19–31, 1981.
- [9] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. 2nd ed., Springer-Verlag, 1994.
- [10] P. Góra and A. Boyarsky, “A matrix solution to the inverse Perron-Frobenius problem,” *Proceedings of the American Mathematical Society*, vol. 118, no. 2, pp. 409–414, 1993.
- [11] G. Wolberg and I. Alf, “Monotonic cubic spline interpolation,” in Proc. *Computer Graphics International*, Canmore, Alta., 1999, p. 188–195.
- [12] G. Marsaglia and W. W. Tsang, “The ziggurat method for generating random variables,” *Journal of stat. software*, vol. 5, no. 8, pp. 1–7, 2000.
- [13] W. J. Cody, “Rational Chebyshev approximations for the error function,” *Mathematics of Computation*, vol. 23, no. 107, pp. 631–637, 1969.