A High-Resolution Single-Photon Arrival-Time Measurement With Self-Antithetic Variance Reduction in Quantum Applications: Theoretical Analysis and Performance Estimation

TONY WU, (Graduate Student Member, IEEE),
AND TZU-CHIEN HSUEH, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA
Corresponding author: Tzu-Chien Hsueh (e-mail: tzhsueh@eng.ucsd.edu).
This work was supported by the National Science Foundation under Grant 2045935.

ABSTRACT An almost all-digital time-to-digital converter (TDC) possessing subpicosecond resolutions, scalable dynamic ranges, high linearity, high noise immunity, and moderate conversion rates can be achieved by a random sampling-and-averaging (RSA) approach with the self-antithetic variance reduction (SAVR) technique for time-correlated single-photon counting (TCSPC) quantum measurements. This article presents detailed theoretical analysis and behavior-model verifications of the SAVR technique to effectively enhance the conversion rate of an asynchronous RSA-based TDC by more than 62 times with 7% power overhead. In addition, the proposed performance estimation methodology for SAVR can greatly improve the computation efficiency during the system-level design and reduce the read-out circuit complexity in the silicon-photonics RSA-based TCSPC realization.

INDEX TERMS Antithetic variate, correlated random variable, joint probability density function, Monte Carlo method, quantum probability amplitude, single-photon counting, stochastic random sampling, time-domain modulo operation, time-to-digital converter (TDC), variance reduction (VR).

I. INTRODUCTION
Time-correlated single-photon counting (TCSPC) [1], [2], [3], [4], [5] has become the key functionality in a variety of emerging quantum technology, including quantum imaging/sensing [6], [7], [8], quantum-state preparations [9], [10], [11], [12], quantum cryptography [13], [14], [15], positron emission tomography [16], [17], time-resolved spectroscopy [18], fluorescence-lifetime imaging [19], [20], molecular imaging, live-cell/tissue microscopy [2], free-space time-of-flight measurements [21], and light detection-and-ranging (LiDAR) [6], [22]. One example of quantum-state detections shown in Fig. 1 exploits TCSPC to measure the quantum state of a light beam, which is a “vector” but very different from the vector defined in classical physics. Under the Dirac notation, a normalized quantum state $|\psi\rangle$ can be expressed by the superposition of the horizontal polarization basis vector $|H\rangle$ and vertical polarization basis vector $|V\rangle$ in the Hilbert space [10] as follows:

$$|\psi\rangle = c_H |H\rangle + c_V |V\rangle, \quad |c_H|^2 + |c_V|^2 = 1 \quad (1)$$

where $c_H$ and $c_V$ are complex coefficients, and the sum of their squared magnitudes is “1” in this orthonormal basis state-expression. More importantly, $c_H$ and $c_V$ are referred to as complex probability amplitudes of the polarization basis vectors, $|H\rangle$ and $|V\rangle$, respectively. The physical meaning of these coefficients is that the squared magnitude of the coefficient, multiplying a particular polarization basis vector, is equal to the probability of photons occurring in that polarization after many single-photon measurements, which can be performed by the apparatus with TCSPC shown in Fig. 1(a); whenever TCSPC$_H$ receives a single photon from the $|H\rangle$ output of the light polarizer at a certain time of $\Delta t_H$, the circuit in TCSPC$_H$ adds a one on the accumulated number $N_H(\Delta t_H)$ and records the quantized...
Δt_H, i.e., a time-to-digital conversion (TDC) process, for the received photon. Similarly, TCSPECV accumulates N_V(Δt_V) and records the quantized Δt_V for the received photons from the |V⟩ output of the light polarizer. Thus, the probability of the photon in each polarization can be obtained as follows:

\[
P(H|\psi) = \frac{|c_H|^2}{N_H(\Delta t_H) + N_V(\Delta t_V)}
\]

\[
P(V|\psi) = \frac{|c_V|^2}{N_H(\Delta t_H) + N_V(\Delta t_V)}.
\]

(2)

Note that the notations of the probabilities are conditioned on the measured quantum state |ψ⟩. Meanwhile, since the quantum state is not affected by the reference phase of its complex coefficients, usually the phase of c_H can be safely set to zero, and the phase of c_V, φ, can be obtained by performing two more quantum-state measurements, as shown in Fig. 1(b) and (c):

\[
P(+45|\psi) = \frac{1}{2} + |c_H| \cdot |c_V| \cdot \cos\phi
\]

\[
= \frac{N_{+45}(\Delta t_{+45})}{N_{-45}(\Delta t_{-45}) + N_{+45}(\Delta t_{+45})}
\]

\[
P(-45|\psi) = \frac{1}{2} + |c_H| \cdot |c_V| \cdot \sin\phi
\]

\[
= \frac{N_{-45}(\Delta t_{-45}) + N_{+45}(\Delta t_{-45})}{N_L(\Delta t_L) + N_R(\Delta t_R)}
\]

(3)

where (+45, −45) and (L, R) are the other two pairs of the orthonormal polarization basis vectors. With these three conditional probability measurements, the complex probability amplitudes, c_H and c_V, of the quantum state can be fully determined from (2) and (3) [11]. In other words, the probability measurements performed by the TCSCPC system are essential in order to determine a quantum state.

Consequently, the TDC specifications in the TCSCPC systems are required to achieve picosecond fine resolutions, submicrosecond dynamic ranges, high linearity and fast conversion rates under aggressive silicon-area, voltage, and power constraints of modern complementary metal-oxide semiconductor (CMOS) integrated-circuit process technology. The state-of-the-art TDC designs can satisfy some of these specifications, however, with tradeoffs among the other performance metrics [1], [2], [3], [4], [6], [7], [8], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44]. To simultaneously accommodate all performance requirements for the majority of quantum applications, a TCSCPC system proposed in [45] incorporates the digital random sampling-and-averaging (RSA) technique [46], [47], [48], [49] into a two-step TDC procedure to enable high-resolution time-interval measurements without compromising the performance metrics among accuracy, dynamic range, linearity, and power/area efficiency. However, the slow conversion rate of the RSA technique has greatly limited the broadness of its applications especially for high frame-rate/fill-factor quantum imaging and ranging systems [6], [7], [8], [16], [17], [18], [19], [20], [21], [22]. Fortunately, there is a concept of variance reduction (VR), which has been broadly used in the fields of applied mathematics and financial engineering [51], [52] to reduce time consumption of the Monte Carlo methods in derivatives pricing and risk management [52]. Based on the similar concept, this article proposes a simple circuit implementation for the self-antithetic variance reduction (SAVR) technique [50] to effectively suppress the quantization-noise power or equivalently enhance the conversion rate of an asynchronous RSA-based TCSCPC system by automatically introducing autocorrelations into the sampled random variable during the TDC process.

To comprehensively describe a practical RSA measurement with the SAVR technique, this article derives the mathematical expressions of the theoretical variance and its approximation form for highly efficient performance-predictions, which are all experimentally verified by the simulations. Meanwhile, the methodology of converting the mathematical models of the RSA with SAVR technique into an almost all-digital and low-power implementation is elaborated by a circuit-level example with a guideline of setting the circuit parameters. Compared to an ordinary RSA measurement [45], enabling the SAVR technique can enhance the conversion rate by more than 62 times with about 7% power overhead based on the simulation results.

The rest of this article is organized as follows. The circuit-and-system level overview of an RSA-based TCSCPC system with the SAVR technique is introduced in Section II. The required fundamentals of RSA and parameter definitions are summarized in Section III. The probability principles, theoretical variances, performance estimation methodology, and behavior-model simulations of the RSA with SAVR technique are derived and presented in Section IV. Finally, Section V concludes this article.

II. CIRCUIT AND SYSTEM OVERVIEW

The block diagram of the RSA-based TCSCPC system with the SAVR techniques is shown in Fig. 2(a). The silicon-photonic interface is similar to the system described in [45], where each detection pixel includes a single-photon avalanche diode (SPAD) with the quenching/clamping circuits [3], [6], [7], [8] for the optical-to-electrical power

\[
(Figure 1. Three quantum-state measurements using TCSCPC for the orthonormal polarization basis vectors of (a) |H⟩ and |V⟩, (b) |−45⟩ and |+45⟩, and (c) |L⟩ and |R⟩.)
\]
domain transition followed by a silicon-photonics analog front-end (AFE) [53] and high-bandwidth CMOS pulse generator to convert the received single photons to event-triggered electrical voltage pulses. For the high-accuracy time-interval measurement, the timing of \( T_{\text{START}} \) is set by the START pulse from a specific single-photon detection pixel [2] while the multiplexer for the STOP pulse can select the timing of \( T_{\text{STOP}} \) from either the other specific single-photon detection pixel [2] or the system input clock \( CK_{\text{IN}} \) [3], [4], [6], [7], [8]. In any configuration, the time interval \( \Delta t \) between \( T_{\text{START}} \) and \( T_{\text{STOP}} \) is the primary quantity under the measurement, as shown in Fig. 2(a). The time-to-amplitude conversion (TAC) circuit, containing a current source and capacitor banks with low-resolution static controls for coarse dynamic-range/conversion-gain tunability, converts the time-interval information into a constant dc voltage [45] buffered by the variable-gain amplifier (VGA) offering noise rejection and driving capabilities with additional tunability.

The TDC mechanism, which is the main focus of this article, is illustrated in the lower half of Fig. 2(a). The two identical voltage-controlled delay lines (VCDL) are both driven by the input clock \( CK_{\text{IN}} \), so the clock periods of \( CK_1 \) and \( CK_2 \) are equal to the period of \( CK_{\text{IN}} \), \( T \), but the time-domain delays of \( CK_1 \) and \( CK_2 \) are functions of the dc voltages \( V_{\text{DD}} \) and \( V_{\text{VGA}} \), respectively; and the periodic delta \( \tau \) between their delays represents the scaled version of \( \Delta t \), as described in [45]. After merging \( CK_1 \) and \( CK_2 \) by the rising clock-edge combiner, the resulting clock \( CK_{\tau} \) maintains the \( T \) periodicity while its positive duty-cycle \( \tau/T \) is the primary quantity under the RSA measurement, as shown in Fig. 3.

To perform the random sampling process, a free-running ring-based digitally controlled oscillator (DCO) generates an asynchronous clock, \( CK_{\text{DCO}} \), to sample the waveform of \( CK_{\tau} \) through a 1-bit D flip–flop (DFF). At the meantime, the randomness of \( CK_{\text{DCO}} \) sampling-instants is mainly accomplished by a digital pseudorandom-binary-sequence generator (PRBS Gen.) in Fig. 2(a) to dynamically modulate the DCO period and to ensure that the sampling probability density function (PDF) can satisfy the RSA criteria for the sampling outcome \( Y \) to be the random variable described in [45] or Section III. Finally, the averaging and noise-filtering
processes are executed by the data and cycle accumulators (ACC.) to first count the numbers of ones \(N_Y\) and \(N_{DCO}\) at the outputs of the DFF and DCO, respectively; then, the final result per measurement \(\tilde{Y}\) (mean value of all sampled \(Y\)) is basically the ratio of the counter outputs \(N_Y/N_{DCO}\) [45]. At this point, the process of an asynchronous RSA-based TDC has been completed.

To enhance the conversion rate of RSA, the SAVR technique can be realized on the top of asynchronous RSA by simply adding extra static coarse controls for the reconfigurabilities of the DCO frequency and PRBS energy-level. The detailed circuit and reconfigurable schemes of the DCO are shown in Fig. 2(b), including a 7-bit PRBS generator at the bottom and a four-stage pseudodifferential inverter-based DCO at the middle with static push-pull resistance-capacitance controls for SAVR. This PRBS generator produces an independent and identically distributed (I.I.D.) 7-bit binary control code, PRBS<6:0>, to dynamically modulate the DCO period through the digital-transistor-capacitor banks in the rate of \(CK_{DCO}\) as follows:

\[
T_{DCO,n} = T_{DCO,MIN} + \Delta T_{PRBS,n} \\
T_{DCO,MIN} \leq T_{DCO,n} \leq T_{DCO,MIN} + \Delta T_{PRBS,MAX}
\]

where “\(n\)” is the sample index from 1 to \(N_{DCO}\); \(T_{DCO,n}\) is the \(n\)th period of the DCO; \(\Delta T_{PRBS,n}\) is the \(n\)th DCO period extension controlled by PRBS<6:0>; \(T_{DCO,MIN}\) is the minimum DCO period when \(\Delta T_{PRBS,n} = 0\); \(\Delta T_{PRBS,MAX}\) is the maximum DCO period extension, which sets the span of the time-domain sampling PDF of a “single” \(CK_{DCO}\) sampling edge, \(f_{DCO,1}(t)\), as shown in Fig. 3. Intuitively, the nonlinearity of the binary-control scheme and only 127 (= \(2^7 - 1\)) possible digital-to-period conversion steps would cause poor uniform-distributed sampling PDFs, which are not only discrete but also unequally spaced. Fortunately, because of the phase-noise accumulation property of ring oscillators [54], inherent circuit/device noise [54], central limit theorem [55], and modulo-\(T\) circular convolution theorem [55], [56], this low-cost implementation can equivalently offer extremely high resolutions and continuous sampling PDFs, as discussed in [45].

The low-power control schemes shown in Fig. 2(b) for enabling SAVR are the static reconfigurabilities of \(\Delta T_{PRBS,MAX}\) and \(T_{DCO,MIN}\) highlighted in pink and green, respectively. To statically adjust \(\Delta T_{PRBS,MAX}\) (pink), the thermometer code \(S<6:0>\) can force the consecutive most-significant-bits (MSBs) of PRBS<6:0> to zeros, which accordingly shut down the MSB capacitor banks and only leave the residual activated least significant bits (LSBs) of PRBS<6:0> to modulate \(\Delta T_{PRBS,n}\). In other words, the range of \(\Delta T_{PRBS,n}\), i.e., \(0 \leq \Delta T_{PRBS,MAX}\), can be scaled down in powers of two based on the number of zeros in \(S<6:0>\) with a constant LSB step size; this coarse scaling approach correspondingly reduces the total number of the digital-to-period conversion steps. Therefore, the other control knob is employed by universally tuning the common drain-source dc voltage of the capacitor banks through a power/area efficient R-2R resistive ladder digital-to-analog converter [57], i.e., R-2R DAC in Fig. 2(b), without an analog voltage buffer due to the pseudodifferential DCO architecture. Using the digital transistors to perform this varactor-based capacitance tunability [54] can moderately scale down the LSB step size with negligible power overhead to partially maintain the total number of digital-to-period conversion steps when \(S<6:0>\) heavily scales down \(\Delta T_{PRBS,MAX}\). It is important to note that the reconfigurability of \(\Delta T_{PRBS,MAX}\) is mainly for demonstrating the effect of SAVR; the theoretical and simulation results in Section IV both confirm that a smaller \(\Delta T_{PRBS,MAX}\) offers more SAVR, but this is limited by the achievable minimum LSB capacitance in CMOS process technology. In any case, the gaps among the LSB steps in the DCO sampling PDFs are filled by the accumulation of circuit/device noise as mentioned. On the other hand, to statically adjust \(T_{DCO,MIN}\) [see green in Fig. 2(b)], the primary inverter stages of the DCO are implemented in a cascode structure, so the gate terminals of the partial tail P/N transistors can be biased through the replica of the inverter stage [54] and controlled by a low-power current DAC, as shown in Fig. 2(b). At first glance, the implementations of these static reconfigurabilities are simple and ordinary, which indeed indicates the low power/area overhead of this SAVR technique. However, the theory and concept in behind are quite complicated and unintuitive; the detailed analysis is elaborated in Section IV.

All inherent phases of the DCO can be exploited to simultaneously sample the information signal \(CK_T\) and linearly improve the conversion rate of each RSA measurement. To minimize the complexity, each DCO clock phase can actually manage its own asynchronous RSA process (DFF and ACC.) with its own clock (or phase) domain, as shown at the top of Fig. 2(b); once the common cycle accumulator reaches the designated sampling number \(N_{DCO}\), an indicator is asynchronously sent to all data accumulators to stop individual counting processes all together, and then a final summer only needs a one-time operation to add all counter results in total. Note that the theoretical analysis in this article only focuses on one of these parallel RSA processes since they are basically identical and just occurring simultaneously. More importantly, the correlation induced by SAVR only exists among all sequential samples of each individual random variable, \(Y\) (i.e., at each DFF output), generated by its own DCO sampling phase, not among multiple DFF outputs. Therefore, the theoretical analysis only considers one set of \(Y, CK_{DCO}, N_Y,\) and \(N_{DCO}\), as shown at the top-right corner of Fig. 2(b).

### III. RANDOM SAMPLING-AND-AVERAGING OVERVIEW

This section briefly summarizes the parameter definitions and important results from [45], which are essential information and background for the RSA with SAVR technique.

In the asynchronous RSA-based TCSPC system of this article, \(Y\) is the primary random variable; \(E[Y]\) is the expectation of the random variable; \(Y_n\) is the \(n\)th experimental
sample of $Y$; $N_{\text{DCO}}$ is the total number of the samples; $\bar{Y}$ is the mean of $Y_n$ and the primary result per RSA measurement, which is also the Monte Carlo estimate [52] of the “ideal” expectation $E[Y] = E[\bar{Y}]$. Therefore, the theoretical variance of $\bar{Y}$, $\text{Var}[\bar{Y}]$, is the performance metric of the measurement, not $\text{Var}[Y]$. One example of the asynchronous RSA sampling processes is shown in Fig. 3, where the waveforms of $CK\tau$ and $CK_{\text{DCO}}$ are assumed to have coincident rising edges at $t = 0$ for the sake of simplicity. With the parameter definitions in (4) and the phase-noise accumulation property of ring oscillators [54], the $n$th absolute sampling time $t_{\text{SAMP},n}$, as illustrated in Fig. 3, can be generalized as follows:

$$
t_{\text{SAMP},n} = \sum_{k=1}^{n} T_{\text{DCO},k} = n \cdot T_{\text{DCO,MIN}} + \sum_{k=1}^{n} \Delta T_{\text{PRBS},k}. \tag{5}
$$

Each $t_{\text{SAMP},n}$ contains the deterministic term $n \cdot T_{\text{DCO,MIN}}$ and stochastic term due to the DCO phase-noise accumulation, which describes the uncertainty of each sampling instance and can only be represented by a PDF $f_{d\text{DCO},n}(t)$. Therefore, the $n$th $CK_{\text{DCO}}$ rising edge occurs randomly but is confined within the distribution span and density magnitude of its own PDF, i.e., light-red areas in Fig. 3. More importantly, the stochastic term of each $t_{\text{SAMP},n}$ is the accumulation of “$n$” samples of an I.I.D. random variable (i.e., $\Delta T_{\text{PRBS},k}$, $k = 1$ to $n$) created by the PRBS generator for “$n$” times, as shown in the second term of (5); equivalently, the PDF of the $n$th DCO sampling instant, $f_{d\text{DCO},n}(t)$, is the convolution result of total “$n$” fundamental PDFs, $f_{d\text{DCO},1}(t)$, from the PRBS generator based on the convolution theorem [55]. Note that the fundamental PDF, $f_{d\text{DCO},1}(t)$, has a $\Delta T_{\text{PRBS,MAX}}$ distribution span and constant $1/\Delta T_{\text{PRBS,MAX}}$ density magnitude, as shown in Fig. 3. When $n >> 1$, the central limit theorem guarantees that $f_{d\text{DCO},n}(t)$ converges to a Gaussian distribution with a wide distribution span $n \cdot \Delta T_{\text{PRBS,MAX}}$, as shown at the bottom right of Fig. 3 and top of Fig. 4(a) regardless of the sampling PDF, $f_{d\text{DCO},1}(t)$, from the PRBS generator.

Because of the periodicity of $CK\tau$, the entire distribution span of $f_{d\text{DCO},n}(t)$ is automatically segmented and compressed into a [0, $T$] duration and equivalently converted into a modulo-$T$ random sampling PDF, $f_n(t)$, which still follows the convolution theorem but shall be mathematically expressed by a circular convolution $C\text{Conv}[]$, due to the modulo-$T$ operation [56] as follows:

$$
f_i(t) = C\text{Conv} \left[ f_{d\text{DCO},1}(t) \cdot \delta(t) , T \right], \quad f_n(t) = C\text{Conv} \left[ f_{n-1}(t) \cdot f_1(t) , T \right], \quad n > 1 \tag{6}
$$

where $\delta(t)$ is the unit impulse; the “$i$” of $f_i(t)$, $f_{n-1}(t)$, and $f_n(t)$ is the modulo-$T$ time-domain variable within $[0, T)$, but the “$i$” of $f_{d\text{DCO},1}(t)$ and $\delta(t)$ is the absolute time-domain variable referenced to $t = 0$. Based on (6), $f_1(t)$ plays as not only the

FIGURE 4. (a) Concept of the modulo-$T$ operation and density-magnitude convergence of a random sampling PDF in the asynchronous RSA implementation [45]. (b) Development examples of the conditional joint sampling PDFs of a $[Y_n, Y_s]$ pair from “$n$” to “$k$” = (n + 1), (n + 2), and (n + 3), where $r = T/2$, $\mod(T_{\text{DCO,MIN}}, T) \approx T/2$, and $\Delta T_{\text{PRBS,MAX}} \approx T/4$. 

Wu and Hsueh: HIGH-RESOLUTION SINGLE-PHOTON ARRIVAL-TIME MEASUREMENT

IEEE Transactions on Quantum Engineering

VOLUME 3, 2022

5500415
Bernoulli [55] random variable as each $Y_n$ only has two possible outcomes Logic-1 or Logic-0. In sum, an asynchronous RSA measurement can trade in a larger number of samples, $N_{DCO}$, which is equivalent to a longer measurement time or slower conversion rate, for achieving a higher accuracy.

In this article, the experimental variances obtained from simulation data are required to verify the theoretical derivations and can be expressed as follows [45], [52], [58]:

$$\text{Var}[\bar{Y}] \approx \frac{\sum_{m=1}^{N_{\text{EXP}}}(\bar{F}_m - \frac{1}{N_{\text{EXP}}-1} \cdot \sum_{n=1}^{N_{\text{EXP}}} F_n)^2}{N_{\text{EXP}} - 1}.$$  \hfill (10)

The accuracy of this experimental variance depends on the number of $\bar{Y}$, $N_{\text{EXP}}$, obtained from experiments or simulations. Note that, in realistic RSA measurements under a certain accuracy requirement with the settings of $N_{\text{DCO}}$, $T_{\text{DCO,MIN}}$, $\Delta T_{\text{PRBS,MAX}}$, and $T$, only a single $\bar{Y}$ is required to represent one measurement result [45].

### IV. SELF-ANTITHETIC VARIANCE REDUCTION

To comprehensively elaborate SAVR and its realization methodology in an asynchronous RSA system, the following analysis process can be started from formulating the variance of a Monte Carlo estimate in general as follows:

$$\text{Var}[\bar{Y}] \approx \text{Var}[\sum_{n=1}^{N_{\text{DCO}}} Y_n] = \frac{\sum_{n=1}^{N_{\text{DCO}}} \text{Cov}[Y_n, Y_k]}{N_{\text{DCO}}^2} = \frac{\sum_{n=1}^{N_{\text{DCO}}} \sum_{k=1}^{N_{\text{DCO}}} \text{Cov}[Y_n, Y_k]}{N_{\text{DCO}}^2} = \frac{1}{N_{\text{DCO}}} (\sum_{n=1}^{N_{\text{DCO}}} \text{Var}[Y_n] + 2 \cdot \sum_{n=1}^{N_{\text{DCO}}-1} \sum_{k=n+1}^{N_{\text{DCO}}} \text{Cov}[Y_n, Y_k])$$  \hfill (11)

where $\text{Cov}[Y_n, Y_k]$ is the pairwise covariance of any two samples $Y_n$ and $Y_k$: both “n” and “k” are the sample indexes. When $n = k$, $\text{Cov}[Y_n, Y_k] = \text{Var}[Y_n]$, i.e., the variance of the nth sample, otherwise $\text{Cov}[Y_n, Y_k] = \text{Cov}[Y_k, Y_n]$, which is a symmetric covariance matrix. If $Y_1, Y_2, \ldots$ and $Y_{N_{\text{DCO}}}$ are all pairwise independent, then $\text{Cov}[Y_n, Y_k] = 0$ for all $n \neq k$ in (11), and the pairwise covariance sum [the second term in the fifth line of (11)] is zero, so (11) becomes the I.I.D. situation, as shown in (9). The key idea of SAVR is to create nonzero correlations among the samples, i.e., $Y_1, Y_2, \ldots$ and $Y_{N_{\text{DCO}}}$, of the random variable $Y$, so that the pairwise covariance sum can be negative, and then overall variance $\text{Var}[\bar{Y}]$ per RSA measurement can be effectively less than the variance of the I.I.D. situation in (9).
A. THEORETICAL CONCEPT AND PERFORMANCE ESTIMATION

According to the detailed analysis about the covariances of the adjacent samples Cov\([Y_n, Y_{n+1}]\) under three \(\Delta T_{\text{PRBS,MAX}}\) scenarios (i.e., \(\Delta T_{\text{PRBS,MAX}} < \tau, \tau = \tau, \) or \(\tau > T\)) in [45], the conclusion shows Cov\([Y_n, Y_{n+1}]\) can have pronounced nonzero values when \(\text{Mod}[\Delta T_{\text{PRBS,MAX}}, T] < \tau\) due to the nonuniform conditional joint PDFs within their own integral time-intervals \([0, \tau]\) and \([\tau, T]\). In this article, the case of nonzero covariances between adjacent samples is further extended to the deduction of any pairwise correlation between \(Y_n\) and \(Y_k\) due to the fact of the DCO phase-noise accumulation property reflected by (5) and (6). Thus, if the technique can take advantage of these nonzero covariances, Cov\([Y_n, Y_k]\), and assure the pairwise covariances sum is negative, then VR can be successfully performed.

To find Cov\([Y_n, Y_k]\), all conditional joint PDFs of any \([Y_n, Y_k]\) pair have to be first formulated, where \(n = 1\) to \(N_{\text{DCO}} - 1\) and \(k = (n + 1)\) to \(N_{\text{DCO}}\) are sufficient to cover all covariance elements in the symmetric covariance matrix. Though there are only four possible binary combinatorial outcomes of a certain \([Y_n, Y_k]\) pair, its conditional joint PDFs shall also cover all possible binary combinatorial outcomes from \(Y_{n+1}\) to \(Y_{k-1}\) because the accumulation property and convolution theorem described in (6) form the chain of correlations from each specific \(Y_n\) through \(Y_{n+1}, Y_{n+2}, \ldots, Y_{k-2}\), and \(Y_{k-1}\) all the way to each specific \(Y_k\). Thus, the conditional joint PDFs of a \([Y_n, Y_k]\) pair under all possible conditions of \([Y_n, \ldots, Y_{k-1}]\) shall be generalized as follows:

\[
f_{n,k}(t) \bigg|_{Y_{n-1}, \ldots, Y_{k-1} = D2B[q]} = CConv \left[ f_{n,k-1}(t) \bigg|_{Y_{n-1}, \ldots, Y_{k-1} = D2B[q]} , f_1(t) , T \right]
\]

(12)

where \(q = 0, 1, \ldots, \left(2^k - n \right)\); all possible binary combinatorial conditions of \([Y_n, \ldots, Y_{k-1}]\) are represented by the corresponding decimal numbers “\(q\)” and decimal-to-binary operators D2B\([-\)] for the sake of simplicity. In Fig. 4(b), the PDF curves and annotations highlighted in black represent the examples of (12), including \(k = (n + 1), (n + 2),\) and \((n + 3)\), which follows the convolution theorem under the conditions of \([Y_n, \ldots, Y_{k-1}]\). Since the location of “\(\tau\)” within the modulo-\(T\) time interval \([0, T]\) defines the PDF boundaries for the probability of \(Y_k\) to be Logic-1 or Logic-0, the conditional joint PDF of a \([Y_n, Y_k]\) pair under all possible conditions of \([Y_n, \ldots, Y_k]\) can be obtained by forcing one part, i.e., \([0, \tau]\) or \([\tau, T]\), of the PDF in (12) to zero to satisfy one of the possible \(Y_k\) conditions as follows:

\[
f_{n,k}(t) \bigg|_{Y_{k} = n} = \begin{cases} f_{n,k}(t) \bigg|_{Y_{n-1}, \ldots, Y_{k-1} = D2B[q]} & 0 \leq t < \tau \\ 0, & \tau \leq t < T \end{cases}
\]

(13)

In other words, each conditional joint PDF under the conditions of \([Y_n, \ldots, Y_{k-1}]\) in (12) can diversify into two conditional joint PDFs under the conditions of \([Y_n, \ldots, Y_k]\), as shown in (13) and (14), which are highlighted in blue and orange, respectively, in Fig. 4(b). Note that the distribution profiles of all conditional joint PDFs are convoluted functions of \(\tau\) and \(f_1(t)\), and \(f_1(t)\) itself is a function of \(\text{Mod}[T_{\text{DCO,MIN}}, T] \) and \(\Delta T_{\text{PRBS,MAX}}\). Therefore, the plots shown in Fig. 4(b), where \(\tau = T/2, \text{Mod}[T_{\text{DCO,MIN}}, T] = T/2,\) and \(\Delta T_{\text{PRBS,MAX}} = T/4\), can be very different if the values of these parameters are set differently; i.e., these parameters dominantly affect the behaviors of the pairwise covariances, Cov\([Y_n, Y_k]\), which now can be generalized as follows based on the information of all conditional joint PDFs of a \([Y_n, Y_k]\) pair from (13) and (14):

\[
\text{Cov}[Y_n, Y_k] = \int_0^T \left( y_n(t) - P_1 \right) \cdot \left( y_k(t) - P_1 \right) \cdot f_{n,k}(t) \cdot dt
\]

(14)

As shown in (15), Cov\([Y_n, Y_k]\) can be grouped into four terms based on the binary combinatorial outcomes of a \([Y_n, Y_k]\) pair, and each term is the summation of \(2^k - n\) integrals of the conditional joint PDFs under all possible conditions of \([Y_n, \ldots, Y_{k-1}]\). Thus, the total number of summation terms or conditional joint PDFs in (15) is \(2^k - n\) for each Cov\([Y_n, Y_k]\). In addition, each conditional joint PDF, \(f_{n,k}(t)\)\([Y_n, Y_{k-1}] = D2B[q]\), in (15) is obtained by \((k - n)\) times of the modulo-\(T\) circular convolutions shown in (12). Overall, the computation effort of (15) is dominated by total \((k - n)\) modulo-\(T\) circular convolutions for each Cov\([Y_n, Y_k]\). On top of this, an overall variance \(\text{Var}[T]\) calculation, requires total \((N^2_{\text{DCO}} - N_{\text{DCO}})/2\) of Cov\([Y_n, Y_k]\) elements in the whole covariance summation, as shown in fifth line of (11), and usually \(N_{\text{DCO}}\) is in the range of \(2^8\) to \(2^{18}\).
Under the asynchronous RSA setting of 256, and \( \text{Var}(\text{red}) \), multiple are set to \( k = 1, 128 \) (\( \approx N/T \approx 0.5 \approx 0.5 \)), 1, 128, 1, and 256 across \( k = 1 \) to \( N_{\text{DCO}} \) when \( \Delta T_{\text{PRBS,MAX}} \approx T/8 \), 1, and 256, respectively, and therefore, \( k - \text{Var} \approx \text{N}_{\text{DCO}} \), \( \text{Cov} \approx 0.5 \), 1, 128, and 256 across can be observed by plotting \( \text{Mod}(n) \). The theoretical variance calculation is elaborated by specific examples in an asynchronous RSA system, and then general cases can be further summarized.

The following two examples have common parameter setups: the time-domain quantity under the RSA measurement, \( \tau/T = 0.5 (\approx P_1), N_{\text{DCO}} = 256 \), and \( \text{Mod}(T_{\text{DCO,MIN}}, T) \approx T/2 \). The only difference is that their \( \Delta T_{\text{PRBS,MAX}} \) are set to \( T/8 \) and \( T/16 \) individually through the static PRBS energy-level control shown in Fig. 2. The correlation function between \( Y_k \) and \( Y_\bar{k} \) can be observed by plotting \( \text{Cov}(Y_n, Y_k) \) of each “n” across all possible “k” (\( = 1 \) to \( N_{\text{DCO}} \), as shown in Fig. 5(a) and (b), where only the cases of \( n = 1, 128 (N_{\text{DCO}}/2) \), and 256 (\( N_{\text{DCO}} \)) are included with \( N_{\text{EXP}} = 2^{13} \). For the purpose of reference, the I.I.D. scenario (i.e., \( \Delta T_{\text{PRBS,MAX}} = T \)) verified in [45] is also included in Fig. 5, where the I.I.D. correlation functions (black) are basically zero for all \( n \neq k \) and have the nonzero value, 0.25 (\( = P_1 - P_0 \)), at \( n = k \). From these examples of \( \Delta T_{\text{PRBS,MAX}} < T \) (red), multiple attributes can be observed: first, \( \text{Cov}(Y_n, Y_k) \) of each “n” is self-symmetrical with respect to “k = n,” e.g., \( \text{Cov}(Y_{128}, Y_{127}) \approx \text{Cov}(Y_{128}, Y_{129}) \); second, because of the symmetric covariance matrix and \( \text{Cov}(Y_n, Y_k) = \text{Var}(Y_n) = P_1 - P_0 \) when \( k = n \), all correlation functions have the same profile but shift along the k-axis based on the value of “n,” e.g., \( \text{Cov}(Y_1, Y_k), \text{Cov}(Y_{128}, Y_k), \) and \( \text{Cov}(Y_{256}, Y_k) \) shift their peaks to \( k = 1, 128 \), and 256, respectively, and therefore, \( \text{Cov}(Y_n, Y_k) = \text{Cov}(Y_{N_{\text{DCO}} - n - 1}, Y_{N_{\text{DCO}} - k + 1}); \) third, the correlation is enhanced (longer correlation tails) with decreasing \( \Delta T_{\text{PRBS,MAX}} \); fourth, the correlation degrades with increasing \( \text{LCMPs} \) of the sign-alternation patterns and correlation-envelope periods of the two \( \Delta T_{\text{PRBS,MAX}} \) scenarios. The theoretical variance calculation can be greatly simplified by grouping the effect of correlation-envelope period which guarantees the cancellation in the total covariance sum, i.e., effective VR, though the amplitude of the correlation-envelope attenuates with \( k - n \). The zoom-in versions of the correlation functions at the top of Fig. 5(a) and (b), respectively, highlight the least-common-multiple periods (LCMPs) of the sign-alternation patterns and correlation-envelope periods of the two \( \Delta T_{\text{PRBS,MAX}} \) scenarios. The theoretical variance calculation can be greatly simplified by grouping the effect of covariance cancellation into single or multiple LCMPs instead of considering all the covariance elements.

The effect of this SAVR technique can be more comprehensively demonstrated by the single-dimensional covariance sums and the overall (i.e., two-dimensional) covariance sum \( \text{Var}(\tilde{Y}) \). The single-dimensional covariance sum is defined as the summation of all \( \text{Cov}(Y_n, Y_k) \) from \( k = 1 \) to \( N_{\text{DCO}} \) with respect to each “n,” e.g., the summation of all red data points in the \( \text{Cov}(Y_{128}, Y_k) \) plot is the single-dimensional covariance sum at \( n = 128 \). The single-dimensional covariance sums of the uncorrelated (I.I.D.) and two correlated examples are shown in Fig. 5(c) whose horizontal-axis variable

**FIGURE 5.** Under the asynchronous RSA setting of \( \text{Mod}(T_{\text{DCO,MIN}}, T) \approx T/2, \tau/T = 0.5, N_{\text{DCO}} = 256, \) and \( N_{\text{EXP}} = 2^{13} \), (a) correlation functions of \( n = 128, 1, \) and 256 across \( k = 1 \) to \( N_{\text{DCO}} \) when \( \Delta T_{\text{PRBS,MAX}} \approx T/8 \), (b) correlation functions of \( n = 128, 1, \) and 256 across \( k = 1 \) to \( N_{\text{DCO}} \) when \( \Delta T_{\text{PRBS,MAX}} \approx T/16 \), and (c) single-dimensional covariance sums and partial covariance sums of RSA w/o SAVR (\( \Delta T_{\text{PRBS,MAX}} \approx T/8 \) and w/ SAVR (\( \Delta T_{\text{PRBS,MAX}} \approx T/8 \) and \( T/16 \)).
is “n,” and \( n = 1 \) to \( N_{\text{DCO}} \). The single-dimensional covariance sums of the I.I.D. example (black dots) are always 0.25 for all “n” since each “n” only has its nonzero covariance \( \text{Var}[Y_n] = 0.25 \) at \( n = k \). On the other hand, the single-dimensional covariance sums of the two correlated examples (red dots) are all smaller than 0.25 across all “n” in Fig. 5(c) because the periodical sign-alternation patterns of their correlation functions are canceling the power of \( \text{Var}[Y_n] = 0.25 \) at \( n = k \), as shown in Fig. 5(a) and (b). This is the reason for naming this technique as self-antithetic VR. Finally, the overall variance of RSA can be obtained by the summation of all single-dimensional covariance sums divided by \( N_{\text{DCO}}^2 \), as shown in the third line of (11). According to the fact shown in Fig. 5(c), the overall variances (sums of red dots) are reduced from the value of the I.I.D. variance (sums of black dots). Equivalently, the first term in the fifth line of (11), i.e., \( P_1 \cdot P_0/N_{\text{DCO}} \), equals the I.I.D. variance without any VR, and the second term in the fifth line of (11) is negative to successfully perform VR.

The analysis results demonstrated in Fig. 5 offer further insights: first, a smaller \( \Delta T_{\text{PRBS-MAX}} \), i.e., a narrower fundamental sampling PDF \( f_1(t) \) creates stronger correlations across all sampling points (i.e., longer correlation tails), longer LCMPs (e.g., 16 points for \( T/8 \), 32 points for \( T/16 \)), and eventually more variance reduction; second, though the sign of each single-dimensional covariance sum can be either positive or negative with VR, the average of all single-dimensional covariance sums is always above zero, as shown in Fig. 5(c), which verifies the principles of nonnegative variances and finite measurement resolutions; third, more importantly, the computation efficiency of each theoretical single-dimensional covariance sum and overall variance can be greatly improved by only summing the covariances surrounding the “\( k = n \)” within a few LCMPs as follows:

\[
\text{If } \Delta T_{\text{PRBS-MAX}} = \frac{2 \cdot T}{\text{LCMP}} \text{ and } r \cdot \text{LCMP} + 2 \leq n \leq \frac{2 \cdot N_{\text{DCO}} - r \cdot \text{LCMP} + 2}{2},
\]

\[
E \left[ \sum_{k=1}^{N_{\text{DCO}}} \text{Cov}[Y_n, Y_k] \right] \approx \frac{1}{N_{\text{DCO}}} \cdot \sum_{k=1}^{N_{\text{DCO}}-1} \text{Cov}[Y_{(n+\text{LCMP})}, Y_k]. \tag{16}
\]

where “\( r \)” is the number of LCMP included in the approximation of (16). In Fig. 5(c), the red dots represent the accurate single-dimensional covariance sums; the blue dots represent the partial single-dimensional covariance sums formulated by the right-hand side of (16) with \( r = 1 \) and \( k = (n - \text{LCMP}/2) \) to \( (n + \text{LCMP}/2 - 1) \). Although the red and blue dots at each “n” contain certain amounts of delta, the distribution of the blue dots is relatively constant and basically the average of the red dots except the values of “n” approach 1 or \( N_{\text{DCO}} \) since the correlation functions of these “n” have significantly unbalanced correlation tail lengths, e.g., \( \text{Cov}[Y_1, Y_k] \) and \( \text{Cov}[Y_{256}, Y_1] \) in Fig. 5(a) and (b). Therefore, any partial single-dimensional covariance sum (blue dot) at any “n” far away from 1 or \( N_{\text{DCO}} \) is sufficient for approximating the average of the accurate single-dimensional covariance sums formulated by the left-hand-side of (16). For example, by plugging (16) into the third line of (11) with “\( n = N_{\text{DCO}}/2 \),” the overall variance can be approximated as follows:

\[
\text{Var}[\hat{Y}] = \frac{E}{N_{\text{DCO}}} \left[ \sum_{k=1}^{N_{\text{DCO}}} \text{Cov}[Y_n, Y_k] \right] \approx \frac{1}{N_{\text{DCO}}} \cdot \sum_{k=1}^{N_{\text{DCO}}-1} \text{Cov}[Y_{(n+\text{LCMP})}, Y_k]. \tag{17}
\]

Note that the \( \text{Cov}[Y_n, Y_k] \) in (16) and (17) is referring to its theoretical definition in (15) and annotated as blue circles in Fig. 5(a) and (b), while the all red dots in Fig. 5 are obtained from the statistical simulation results. And, the approximation errors in (16) and (17) can be improved by extending the number of LCMP, i.e., “\( r \)” included in the summation operators.

As discussed earlier, if the theoretical computation effort is dominated by the modulo-\( T \) circular convolutions, the computation efficiency improvement from (11) to (17) can be evaluated by the ratios between their operation numbers of modulo-\( T \) circular convolutions. The covariance matrix in the case of \( \Delta T_{\text{PRBS-MAX}} \approx T/8 \) (LCMP = 16) with \( N_{\text{DCO}} = 256 \) is shown in Fig. 6(a). Since the correlation tails of this example are roughly vanished when \( k - n > N_{\text{DCO}}/2 \), as shown in Fig. 5(a), only the covariance elements circled in black need to be considered to estimate the computation efforts for both (11) and (17). For (11), because the \( \text{Cov}[Y_n, Y_k] \) of each “n” is self-symmetrical with respect to “\( k = n \),” the covariance elements circled in dashed-red are identical to those circled in solid-red, and covariance elements circled in green are the images of those circled in red. With the property of this symmetric covariance matrix, i.e., \( \text{Cov}[Y_n, Y_k] = \text{Cov}[Y_k, Y_n] \), all covariance elements circled in black can be fully obtained by only calculating the covariance elements circled in solid-red, as shown in the upper half of Fig. 6(a), and the required number of modulo-\( T \) circular convolutions in (11) is expressed in the numerator of

\[
\sum_{k=1}^{N_{\text{DCO}}-1} \left( k - \frac{N_{\text{DCO}}}{2} \right) \cdot 2^{(k - \frac{N_{\text{DCO}}}{2} + 1)}. \tag{18}
\]

On the other hand, to cover two LCMPs for (17), only the covariance elements within one LCMP circled in blue shown in the lower half of Fig. 6(a) need to be calculated, and the other LCMP circled in orange is again the image of those circled in blue, so the required number of modulo-\( T \) circular convolutions in (17) is expressed in the denominator of (18). The ratio of modulo-\( T \) circular convolution numbers between (11) and (17) under this specific example is about \( 2.18 \times 10^{34} \). This incredible “computation effort reduction” (NOT variance reduction), even just for \( N_{\text{DCO}} = 2^8 \) and two
under the asynchronous RSA setting of $\Delta_{\text{DCO}} = 0.75$, $N_{\text{EXP}} = 2^{13}$, (a) required covariance elements for complete (top) and approximate (bottom) overall variance calculations when $\text{Mod}[\text{DCO,MIN}, T] \approx T/2$ and $\Delta_{\text{PRBS,MAX}} \approx T/8$, (b) correlation functions of $n = 128$ across $k = 1$ to $N_{\text{DCO}}$ when $\tau/T = 0.2, 0.3, 0.7,$ and $0.8$, $\text{Mod}[\text{DCO,MIN}, T] \approx T/2$, and $\Delta_{\text{PRBS,MAX}} \approx T/16$, and (c) correlation functions of $n = 128$ across $k = 1$ to $N_{\text{DCO}}$ when $\tau/T = 0.5$, $\text{Mod}[\text{DCO,MIN}, T] \approx 0.75$-$T$, and $\Delta_{\text{PRBS,MAX}} \approx T/2$ and $T/8$.

**B. PRACTICAL CONSIDERATION AND SIMULATION RESULTS**

In realistic applications, the SAVR technique has to be effectively applicable to a wide dynamic range of $\tau/T$. Four more examples of the correlation functions for the variates of $\tau/T$ ($= 0.2, 0.3, 0.7,$ and $0.8$) are shown in Fig. 6(b) to illustrate the effects of SAVR with respect to the time intervals under the asynchronous RSA measurements. All parameter settings in Fig. 6(b) are consistent with those in Fig. 5(b) except the different values of $\tau/T$, so their correlation functions exhibit similar correlation tails, LCMPs, and sign-alternation patterns; however, in Fig. 6(b), the amplitudes and upper/lower swings of the covariance envelopes are basically attenuated and distorted with $|\tau/T - 0.5|$. Especially, the unbalanced upper/lower envelope swings around the central covariances (at $n = k$) degrades the efficiency of SAVR when $\tau/T$ is moving away from 0.5.

The variance of asynchronous RSA versus dynamic range with different SAVR settings is shown in Fig. 7(a) where $\text{Mod}[\text{DCO,MIN}, T] \approx T/2$, $N_{\text{DCO}} = 2^{14}$, and $N_{\text{EXP}} = 2^{11}$. Multiple important conclusions have been verified and demonstrated in this figure. First, a stronger VR can be statically enabled by decreasing the PRBS time-domain noise power (or energy level), i.e., $\Delta_{\text{PRBS,MAX}}$ [or the PDF span of $f_1(t)$] to extend the covariance cancellation effect (or widening the Cov[$Y_n$, $Y_k$] envelopes and correlation tails) without any extra hardware/circuit cost required to enhance the resolution of the PRBS generator since the infinite phase resolution has been taking care of by the inherent circuit/system noise accumulations through the DCO [45], [54]. Second, the efficiency of SAVR degrades with the increase of $|\tau/T - 0.5|$ as discussed, but the quantization-noise power of the RSA measurement $\text{Var}[P]$ stays relatively constant across the entire dynamic range. Third, the blue dots are the theoretical variance approximations based on (17) with covering at least two LCMPs ($r \geq 2$) for different $\Delta_{\text{PRBS,MAX}}$ scenarios. Even on a linear scale, the theoretical approximations (blue dots) from (17) match well with the statistical
results (red dots) from the RSA experimental simulations; this verifies the contribution of (17), which can greatly improve the computation efficiency without compromising the estimation accuracy.

In Fig. 7(c), the variances of asynchronous RSA are plotted as functions of $N_{DCO}$ on the dBW scale with different SAVR settings at $\tau/T = 0.5$ with $N_{EXP} = 2^8$. Similar to the cases of the oversampling synchronous RSA in [45], the weak law of large numbers still holds for the asynchronous RSA regardless of the VR technique; also, both theoretical approximation (blue lines) and simulation (red dots) results are well aligned and perform consistent variance degradations at $-3$ dBW per octave of $N_{DCO}$ or equivalently $-6$ dBW per octave of effective number of bits (ENO) [45]. More importantly, the variances at $\tau/T = 0.5$ exhibit $-3$ dBW per octave of the $\Delta T_{PRBS,MAX}$ divisor for all $N_{DCO}$ in both Fig. 7(a) and (c).

The key contribution of SAVR is not only to improve the ENOB, which can be obtained by increasing $N_{DCO}$ anyway, but more significantly improves the conversion rate of the asynchronous RSA measurement. For example, to achieve the same variance of $-60$ dBW in Fig. 7(c), when $\Delta T_{PRBS,MAX} \approx T$ (SAVR disabled), it requires $N_{DCO} = 2^{12}$ and $T_{DCO,AVG} = (T_{DCO,MIN} + \Delta T_{PRBS,AVG}) \approx (T/2 + T/2) = T$; on the other hand, when $\Delta T_{PRBS,MAX} \approx T/32$ (SAVR enabled), it requires $N_{DCO} = 2^{13}$ and $T_{DCO,AVG} = (T_{DCO,MIN} + \Delta T_{PRBS,AVG}) \approx (T2/2 + T/64) = 0.516-T$, so the conversion rate, i.e., $1/(N_{DCO} \cdot T_{DCO,AVG})$, is improved by 62 times when the SAVR technique is enabled. Note that this example is at $\tau/T = 0.5$, which has the worst quantization-noise power when SAVR is disabled (I.I.D.) but maximum variance reduction across the entire $\tau/T$ dynamic range, as shown in Fig. 7(a). Besides, the required setting of $Mod[T_{DCO,MIN}, T] \approx 0.5-T$ gives the flexibility of design options for $T_{DCO,MIN}$ including $0.5-T$, $1.5-T$, $2.5-T$, and so on. If the DCO frequency has reached the limit due to circuit bandwidth or power limitations, then proportionally extending $\tau$ and $T$ together can be an alternative way to achieve the fastest sampling option ($T_{DCO,MIN} \approx 0.5-T$), which can be simply accomplished by statically scaling the conversion gains of the analog circuits, i.e., $K_{TAC}$, $K_{VGA}$, and $K_{DL}$, through the inherent calibration capability of the RSA measurement system [50]. Meanwhile, the nonlinearities and offsets due to analog-circuit nonidealities and mismatches can be precalibrated by RSA itself without any assistance from extra hardware [45], [49], [50].

Regarding the power overhead, the DCO power consumption with versus without SAVR includes four major factors in the comparison: first, the DCO average frequency ($f_{DCO,AVG} \approx 1/T_{DCO,AVG}$) is almost doubled, i.e., $1/(0.516-T)$ versus $1/T$; second, the per-stage capacitance load ($C_{L}$) controlled by the PRBS generator is smaller since it is dominantly scaled with $\Delta T_{PRBS,MAX}$, i.e., $T/32$ versus $T$; third, the power of the PRBS digital circuit is doubled with the DCO average frequency, i.e., 0.5 mW versus 0.25 mW; fourth, additional dc power, i.e., 0.2 mW, of the R-2R and current DACs is required for the DCO static reconfigurations plotted by the green and pink lines in Fig. 2(b). Therefore, according to the simulation results and power estimations above ($C_{L} \cdot V_{DD} \cdot f_{DCO,AVG} + \text{PRBS power} + \text{dc power}$), the overall DCO power roughly stays the same with and without SAVR, i.e., $(2.4 + 0.5 + 0.2) \text{ mW} \text{ versus} (2.75 + 0.25 + 0) \text{ mW} = 3.1 \text{ mW} \text{ versus} 3 \text{ mW}$. Meanwhile, the power consumptions of the TAC, VGA, VCDL, and edge-combiner are all independent from the SAVR technique, but the dynamic power of the DFF and clock buffers are scaled up with the DCO average frequency. Thus, the TDC power numbers ($\text{TAC} + \text{VGA} + \text{VCDL} + \text{edge-combiner} + \text{DFF} + \text{clock buffer}$) with and without SAVR are 1.5 mW and 1.3 mW, respectively. Overall, the total RSA-based TDC power, i.e., $\text{DCO} + \text{TDC} = 4.6 \text{ mW} \text{ versus} 4.3 \text{ mW}$, is increased only by 7% after enabling SAVR for a 62 times conversion rate enhancement since the DCO power is roughly independent from this technique as discussed. Note that the power estimations abovementioned include one clock generator (i.e., DCO) and one TDC, but in real TCSPC applications one clock generator with proper
clock distribution can support multiple TDCs in the pixel array.

Regarding the area overhead of the SAVR technique, the additional hardware includes the R-2R and current DACs for the DCO static reconstructions plotted by the green and pink lines in Fig. 2(b), which add an extra $25 \mu m \times 25 \mu m$ overhead on the top of the DCO active area in a 22-nm CMOS process technology.

Based on the theoretical analysis and experimental simulations of the SAVR technique so far, two practical concerns may be raised. One is about the limitation of SAVR or its conversion rate enhancement. Because it seems that a smaller $\Delta T_{PRBS,MAX}$ (or noise energy level) offers more SAVR, and if the minimum achievable $\Delta T_{PRBS,MAX}$ is dominated by the single LSB capacitance, which is process-technology dependent as mentioned in Section II, modulated by a 1-bit PRBS, then the system could simply rely on the noise accumulation of inherent thermal noise from the DCO circuit. This idea is theoretically valid, but accumulating such low-energy noise meanwhile induces two issues to offset the benefit of SAVR: first, the sampling PDF $f_{s}(t)$ now requires “n” to be extremely large to form a uniform PDF, as shown in Fig. 4(a) [45], and satisfy the expectation in (7); second, now the values of LCMP also become extremely large for effective SAVR due to its reciprocal relation with the noise energy level shown in (16). Both cause each RSA measurement to require a very large number of $N_{DCO}$ in total, so SAVR turns out to even slow down the conversion rate in this case. The other concern is about whether variance addition (VA) can occur due to a combinational setup of $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$, and $T$ deviating from the one for VR. The answer can be elaborated by an example shown in Fig. 6(c) having the identical parameter settings as those in Fig. 5 ($\tau/T = 0.5$, $N_{DCO} = 2^8$, $N_{EXP} = 2^{13}$) except $Mod[T_{DCO,MIN}, T] \approx 0.75\cdot T$ instead. When $\Delta T_{PRBS,MAX} \approx T/8$ shown in the lower half of Fig. 6(c), the correlation function exhibits a different sign-alternation pattern and envelope from these shown in Fig. 5(a), but the single-dimensional covariance sum is still less than that of the I.I.D. case (i.e., 0.25); equivalently this is VR. However, when $\Delta T_{PRBS,MAX} \approx T/2$ shown in the upper half of Fig. 6(c), the correlation function contains all positive values, so definitely the single-dimensional covariances and overall variance are larger than those of the I.I.D. case; equivalently this is VA. The asynchronous RSA variance versus dynamic range plots shown in Fig. 7(a) and (b) contains identical parameter settings except $Mod[T_{DCO,MIN}, T] \approx T/2$ and 0.75-$T$, respectively, for the purpose of the side-by-side comparison. Any quantization-noise power lower than the I.I.D. variance curve performs VR otherwise VA. The unfavorable VA occurs whenever the coincident interactions between $Mod[T_{DCO,MIN}, T]$ and $\Delta T_{PRBS,MAX}$ always generates positive pairwise covariances as the example of simultaneously $Mod[T_{DCO,MIN}, T] \approx 0.75\cdot T$ and $\Delta T_{PRBS,MAX} \approx T/2$ shown in Fig. 6(c) and (b). In other words, there could be several possible combinations of $Mod[T_{DCO,MIN}, T]$ and $\Delta T_{PRBS,MAX}$ which can cause severe variance additions.

And, this is the primary downside of the SAVR technique associated with the upside of high conversion rate improvement and negligible extra power cost for an asynchronous RSA measurement system. Fortunately, the variance approximation, i.e., (17), developed in this research can efficiently predict the quantization noise power under any combination of the signal and circuit parameters, including all $\tau/T$, $N_{DCO}$, $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$, and $T$. In sum, the verified results from both RSA simulation and theoretical approximation conclude that SAVR can offer favorable variance reduction under the minimum requirement of having $Mod[T_{DCO,MIN}, T] \in [0.25\cdot T, 0.75\cdot T]$ with $\Delta T_{PRBS,MAX} \leq T/4$. Also, note that $Mod[T_{DCO,MIN}, T] \approx T/2$ with $T/32 < \Delta T_{PRBS,MAX} \leq T/4$ can offer the maximum flattened variance distribution across the entire $\tau/T$ dynamic range.

Multiple case studies in this section are summarized in Fig. 7(c), including I.I.D., VR, and VA scenarios as discussed. Note that each red dot for a certain $N_{DCO}$ appropriately on the I.I.D. theoretical variance line is obtained from (10) by averaging $2^T (= N_{EXP})$ gray dots in Fig. 7(c). All theoretical approximation (i.e., blue lines) and simulation (i.e., red dots) results are well aligned and all follow the weak law of large numbers to perform a consistent variance degradation at $-3$ dBW per octave of $N_{DCO}$.

V. SUMMARY AND FUTURE WORK

The key signal and circuit parameters are summarized below for enabling the RSA with SAVR technique: $\tau$ (converted from $\Delta t$) is the quantity under each measurement; $T$ (converted from $\Delta_{MAX}$) is the dynamic range under each measurement; $T_{DCO,MIN}$ is the deterministic portion of each DCO period; $\Delta T_{PRBS,MAX}$ is the PDF span of the stochastic portion of each DCO period; the relations among $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$, and $T$ are the key parameters to perform effective SAVR; $N_{DCO}$ is the total number of the samples; $N_{EXP}$ is not involved in the circuit implementation but required for the software/lab experiments to statistically verify the theoretical formulas and the performance estimations of the RSA-based technique.

The comparison of the RSA-based TCSPC systems in [45] and this article is summarized in Table 1 based on the circuit simulations along with the silicon measurement results of multiple state-of-the-art TDC implementations. Note that the primary focus of this article is to comprehensively elaborate the probability theories and implementable mathematical models of the SAVR technique, so the intent of this comparison is to cover a bigger picture of the modern TDC evolution in standard CMOS process nodes, not to show unfair competition by taking advantage of simulation results.

Overall, the SAVR technique enhances the conversion rate by 62 times (up to 120 kHz) with 7% power overhead compared to those of the ordinary RSA-based TDC in [45]. This improvement is also reflected by the energy efficiency of 3.1 pJ/step. To further broaden the potential of RSA in the quantum applications, a couple of VR-related techniques are under investigation. One is to stretch the limit of SAVR, i.e.,
TABLE 1. TDC Technique Comparison and Summary

| Work | TCAS-I-22 | TCAS-I-22 | This Work | JSSC’14 | JSSC’18 | [61] | [62] Access’21 | [63] TCAS-I-22 |
|---|---|---|---|---|---|---|---|---|
| Technology | 22 nm | 22 nm | 22 nm | 90 nm | 65 nm | 40 nm | 28 nm | 65 nm |
| Technique | Syn. RSA | Asyn. RSA | Asyn. RSA w/ SAVR | SRO | CT ΔE | Flash ΔE | CCRO | URF and Vernier |
| Clock Gen. | PLL and PI | DCO | DCO | No Info. | PLL | No Info. | Ring Osc | DLE |
| Clock Power | 25 mW | 3 mW | 3.1 mW | No Info. | No Info. | No Info. | No Info. | No Info. |
| Sampling Frequency | 4 GS/s | 4 GS/s | 7.8 GS/s | 1/TDC_dalo | 750 Ms/s | 250 Ms/s | 50 Ms/s | 125 Ms/s |
| Sampling Phases | 8 | 8 | 8 | N/A | N/A | N/A | N/A | 126 |
| Dynamic Range (DR) | | | | | | | | |
| ENOB (Fixed) | 10 | 16 @ NOCO = 2^16 | 14 @ NOCO = 2^18 | 16 @ NOCO = 2^19 | 16 @ NOCO = 2^20 | 16 @ NOCO = 2^21 | 16 @ NOCO = 2^22 | 16 |
| Effective Resolution | 9.76 ps @ 976 ps | 0.61 ps @ 61 ps | 0.61 ps @ 61 ps | 0.315 ps | 0.182 ps | 0.147 ps | 125 Ms | 91.4 ps |
| Conversion Rate (CVR) | 32 kHz | 12 ENOB | 12 ENOB | 120 kHz | 12 ENOB | 2 MHz | 2 MHz | 5 MHz |
| TDC Power | 1.3 mW | 1.5 mW | 1.5 mW | 8.4 mW | 1.32 mW | 12.3 mW | 51.4 mW |
| TDC FoM ** | 39.7 ps/step | 12 ENOB | 159 ps/step | 3.3 ps/step | 12 ENOB | 0.08 ps/step | 0.47 ps/step | 0.16 ps/step | 1.18 ps/step | 1.88 ps/step |
| TDC Area | 0.01 mm² | 0.01 mm² | 0.01 mm² | 0.02 mm² | 0.055 mm² | 0.08 mm² | 0.078 mm² | 0.15 mm² |
| Linearity Calibration | R 1 Self-Calibration | R 2 Self-Calibration | R 4 Self-Calibration | Off-Chip Self-Correction | No Info. | Off-Chip Self-Correction | One-Chip Self-Correction | Two-Chip Self-Correction |
| Post Noise Filtering | 1-bit Digital Accumulator | 1-bit Digital Accumulator | 1-bit Digital Accumulator | Digital Phase Filter | PLL Loop Filter | High-order Digital Filter | No Info. | No Info. |
| Expect Eq. | (10) in [45] | (15) in [45] | (15) in [45] | (15) in [45] | (15) in [45] | (15) in [45] | (15) in [45] | (15) in [45] |
| Variance Eq. | [11] in [45] | [11] in [45] | [11] in [45] | [11] in [45] | [11] in [45] | [11] in [45] | [11] in [45] | [11] in [45] |
| Parameters at T = 250 ps | OSR = 1 | T_Dcmax = T/2 | T_Dcmax = T/2 | T_Dcmax = T/2 | T_Dcmax = T/2 | T_Dcmax = T/2 | T_Dcmax = T/2 | T_Dcmax = T/2 |

* Effective Resolution = DR/(2^ENOB). ** TDC FoM = (TDC Power)/(2 BW 2^SNR) = (TDC Power)/(CVR 2^SNR).

scaling down the ΔPRBS,MAX and/or LSB capacitance for the DCO frequency modulation toward the limit of advanced CMOS process technology, which can pretty much offer additional 4 times conversion rate improvement without any extra power consumption. The other is to utilize a well-controlled VR technique with the downside of higher power/area overhead [50]. The forthcoming research and silicon-photonic realization will take the complementary benefits of these VR techniques to boost the conversion rate of RSA up to the range of MHz and accordingly to achieve sub-pJ/step energy efficiency.

REFERENCES

[1] W. Becker, Advanced Time-Correlated Single Photon Counting Techniques. Berlin, Germany: Springer, 2005, doi: 10.1007/3-540-28882-1.

[2] W. Becker, The Bh TCSPC Handbook, 7th ed. Berlin, Germany: Becker & Hickl GmbH, 2017. [Online]. Available: https://www.becker-hickl.com/literature/documents/flim/the-bh-tcspc-handbook/

[3] B. Markovic, S. Tisa, F. A. Villa, A. Tosi, and F. Zappa, “A high-linearity, 17 ps precision time-to-digital converter based on a single-stage Vernier delay loop fine interpolation,” IEEE Trans. Circuits Syst. I, Regular Papers, vol. 60, no. 3, pp. 557–569, Mar. 2013, doi: 10.1109/TCSI.2012.2215737.

[4] N. A. W. Dutton et al., “A time-correlated single-photon-counting sensor with 14GS/s histogramming time-to-digital converter,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, 2015, pp. 1–3, doi: 10.1109/ISSCC.2015.2535403.

[5] M. Beck, “Comparing measurements of g^2(0) performed with different coincidence detection techniques,” J. Opt. Soc. Amer. B, vol. 24, no. 12, pp. 2972–2978, Dec. 2007, doi: 10.1364/JOSAB.24.002972.

[6] C. H. Bennett, “Quantum cryptography using any two nonorthogonal states,” Phys. Rev. Lett., vol. 68, pp. 3121–3124, May 1992, doi: 10.1103/PhysRevLett.68.3121.

[7] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, “A 128 × 128 single-photon-image sensor with column-level 10-bit time-to-digital converter array,” IEEE J. Solid-State Circuits, vol. 43, no. 12, pp. 2977–2989, Dec. 2008, doi: 10.1109/JSSC.2008.2082634.

[8] M. Wahl, “Time-correlated single-photon counting,” PicoQuant GmbH, Berlin, Germany, Tech. Note, 2014. [Online]. Available: https://www.picquant.com/images/uploads/page/files/7253/technote_tcspc.pdf

[9] F. Villa et al., “CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3-D time-of-flight,” IEEE J. Sel. Topics Quantum Electron., vol. 20, no. 6, pp. 364–373, Nov./Dec. 2014, doi: 10.1109/JSTQE.2014.2342197.

[10] U. Leonhard, Measuring the Quantum State of Light. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[11] C. Spee, J. I. de Vicente, and B. Kraus, “Remote entanglement preparation,” Phys. Rev. A, vol. 88, Jul. 2013, Art.no.010305, doi: 10.1103/PhysRevA.88.010305.

[12] M. Hillery, V. Bužek, and A. Berthiaume, “Quantum secret sharing,” Phys. Rev. A, vol. 59, no. 3, pp. 1829–1834, Mar. 1999, doi: 10.1103/PhysRevA.59.1829.

[13] L. H. C. Braga et al., “A fully digital 8×16 SiPM array for PET applications with per-pixel TDCs and real-time energy output,” IEEE J. Solid-State Circuits, vol. 49, no. 1, pp. 301–314, Jan. 2014, doi: 10.1109/JSSC.2013.2284351.

[14] A. S. Yousif and J. W. Haslett, “A fine resolution TDC architecture for next generation PET imaging,” IEEE Trans. Nucl. Sci., vol. 54, no. 5, pp. 1829–1834, Mar. 2007, doi: 10.1109/TNS.2007.910435.
[19] C. Veerappan et al., “A 160 × 128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter,” in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2011, pp. 312–314, doi: 10.1109/ISSCC.2011.5746333.

[20] R. M. Field, S. Reavol, and K. L. Shepard, “A 100 fps, time-correlated single-photon-counting-based fluorescence-lifetime imager in 130 nm CMOS,” *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 867–880, Apr. 2014, doi: 10.1109/JSSC.2013.2329777.

[21] M. Perenzenzi, D. Perenzenzi, and D. Stoppa, “A 64 × 64-pixel digital silicon photomultiplexer direct TOF sensor with 100-MPhotons/s/pixel background rejection and imaging/altitude mode with 0.14% precision up to 6 km for spacecraft navigation and landing,” *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 151–160, Jan. 2017, doi: 10.1109/JSSC.2016.2623635.

[22] H. Seo et al., “A 36-channel SPAD-integrated scanning LiDar sensor with multi-event histogramming TDC and embedded interference filter,” in *IEEE Symp. VLSI Circuits*, 2020, pp. 1–2, doi: 10.1109/VLSICircuits18222.2020.9162807.

[23] T. Rahkonen and J. Kostomavaara, “The use of stabilized CMOS delay lines for the digitization of short time intervals,” *IEEE J. Solid-State Circuits*, vol. 28, no. 8, pp. 887–894, Aug. 1993, doi: 10.1109/23.1213352.

[24] K. Karadamoglou, N. P. Paschalidis, E. Sarris, N. Stamatopoulos, G. Kottaras, and V. Paschalidis, “An 11-bit high-resolution and adjustable-range CMOS time-to-digital converter for space science instruments,” *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 214–222, Jan. 2004, doi: 10.1109/JSSC.2003.817263.

[25] Y. Liu et al., “Multi-stage pulse shrinking time-to-digital converter for time interval measurements,” in *Proc. Eus. Microw. Integr. Circuits Conf.*, Munich, Germany, 2007, pp. 267–270, doi: 10.1109/EUCW.2007.4404108.

[26] C. Ljeslin, J. Christiansen, A. Marchioro, and O. Klingheim, “An integrated 16-channel CMOS time to digital converter,” *IEEE Trans. Nucl. Sci.*, vol. 41, no. 4, pp. 1104–1108, Aug. 1994, doi: 10.1109/23.292866.

[27] K. A. Ditkowsky, S. Vemulapalli, P. Vallur, J. Wallberg, and P. T. Bal-...
[58] G. Upton and I. Cook, *A Dictionary of Statistics*, 3rd ed. Oxford, U.K.: Oxford Univ. Press, 2014.

[59] A. Elshazly, S. Rao, B. Young, and P. K. Hanumolu, “A noise-shaping time-to-digital converter using switched-ring oscillators—Analysis, design, and measurement techniques,” *IEEE J. Solid-State Circuits*, vol. 49, no. 5, pp. 1184–1197, May 2014, doi: 10.1109/JSSC.2014.2305651.

[60] M. B. Dayanik and M. P. Flynn, “Digital fractional-N PLLs based on a continuous-time third-order noise-shaping time-to-digital converter for a 240-GHz FMCW radar system,” *IEEE J. Solid-State Circuits*, vol. 53, no. 6, pp. 1719–1730, Jun. 2018, doi: 10.1109/JSSC.2018.2806929.

[61] Y. Wu, P. Lu, and R. B. Staszewski, “A time-domain 147fs RMS 2.5-MHz bandwidth two-step flash-MASH 1-1-1 time-to-digital converter with third-order noise-shaping and mismatch correction,” *IEEE Trans. Circuits Syst. I, Regular Papers*, vol. 67, no. 8, pp. 2532–2545, Aug. 2020, doi: 10.1109/TCSI.2020.2983581.

[62] O. Järvinen et al., “A 100–750 MS/s 11-Bit time-to-digital converter with cyclic-coupled ring oscillator,” *IEEE Access*, vol. 9, pp. 48147–48156, 2021, doi: 10.1109/ACCESS.2021.3068838.

[63] W. Zhang, R. Ma, X. Wang, H. Zheng, and Z. Zhu, “A high linearity TDC with a united-reference fractional counter for LiDAR,” *IEEE Trans. Circuits Syst. I, Regular Papers*, vol. 69, no. 2, pp. 564–572, Feb. 2022, doi: 10.1109/TCSI.2020.3045731.

Tony Wu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electrical and computer engineering in 2019 and 2021, respectively, from the University of California, San Diego, CA, USA, where he is currently working toward the Ph.D. degree.

Since 2020, he has been a Research and Teaching Assistant with Integrated Communication Circuits Lab, University of California, San Diego. His research subjects include mixed-signal circuits and systems with an emphasis on high-speed and low-power digital designs.

Tzu-Chien Hsueh (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taiwan, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Los Angeles, CA, USA, in 2010.

From 2001 to 2006, he was a Mixed-Signal Circuit Design Engineer in Hsinchu, Taiwan. From 2010 to 2018, he was a Research Scientist with Intel Lab Signaling Research and an Analog Engineer with the Intel I/O Circuit Technology, Hillsboro, OR, USA. Since 2018, he has been an Assistant Professor in electrical and computer engineering with the University of California, San Diego, CA, USA. His research interests include wireline electrical/optical transceivers, clock-and-data recovery, data-conversion circuits, on-chip performance measurements/analyzers, and digital/mixed signal processing techniques.

Dr. Hsueh was a recipient of multiple Intel Division and Academy Awards from 2012 to 2018, the 2015 IEEE JOURNAL OF SOLID-STATE CIRCUITS Best Paper Award, and the 2020 NSF CAREER Award. He was with the Patent Committee for Intel Intellectual Property (Intel IP) and the Technical Committee for Intel Design and Test Technology Conference from 2016 to 2018. Since 2018, he has been the Technical Program Committee for IEEE Custom Integrated Circuits Conference and a Guest Associate Editor for IEEE SOLID-STATE CIRCUITS LETTERS.