A High-Accuracy Single-Photon Time-Interval Measurement in Mega-Hz Detection Rates with Collaborative Variance Reduction: Theoretical Analysis and Realization Methodology

Ruoman Yang, *Graduate Student Member, IEEE*, Tony Wu, *Graduate Student Member, IEEE*, and Tzu-Chien Hsueh, *Senior Member, IEEE*

Abstract—An almost all-digital time-to-digital converter possessing sub-picosecond resolution, scalable dynamic range, calibratable linearity, high noise-immunity, and fast conversion-rates can be achieved by a stochastic random sampling-and-averaging approach with the proposed collaborative variance reduction (VR) technique for a wide range of time-correlated single-photon counting applications. This paper presents detailed theoretical analysis and behavior-model verifications of both self-antithetic and control-variate VR techniques to enhance the conversion-rate of an asynchronous RSA-based TDC up to 1.5 MHz with 12-ENOB accuracy, 0.36-pJ/step energy efficiency, and 23% power overhead. Also, the conversions of the mathematical closed-form expressions into digital signal-processing implementations are derived and demonstrated for the forthcoming silicon-photonics integrated-circuit realization.

Index Terms—antithetic variate, auto-correlation, control variate, cross-correlation, single-photon counting, stochastic random sampling, time-domain modulo operation, time-to-digital converter, variance reduction.

I. INTRODUCTION

TIME-correlated single-photon counting (TCSPC) systems [1]–[3], which contain time-to-digital conversion (TDC) integrated circuits, have become the key functionality in a variety of emerging quantum technology. The state-of-the-art TDC designs all have pros and cons in certain performance aspects; therefore, depending on the emphasis on speed or resolution, the required TDC specifications can be roughly categorized into two major areas of the TCSPC applications.

First, quantum imaging/sensing [4]–[6], time-resolved spectroscopy [7], [8], positron emission tomography (PET) [8], [9], fluorescence-lifetime imaging (FLIM) [10], [11], time-of-flight (TOF) sensing [12], [13], and light detection-and-ranging (LiDAR) [13]–[15] primarily exploit high-speed and small-area TDC techniques with the downsides of lower resolution, lower

accuracy and higher clock-generation power [5], [6], [9], [10], [16]–[26]. Second, quantum-bit-state probability amplitude measurements [27]–[30], quantum cryptography [31]–[33], molecular imaging [2], [34], and live-cell/tissue microscopy [2], [35], [36] mainly employ high-resolution TDC techniques with the downsides of lower conversion-rates, higher calibration complexity, and high-order digital filtering [1], [2], [37]–[45]. In the long run of quantum-technology development, the demand for supporting both high-speed and high-resolution with low power/area consumption will be the common direction of all TCSPC applications.

Therefore, this paper introduces a two-step TDC architecture [1], [2], [39] incorporating the concept of variance reduction (VR) [46], [47] into the random sampling-and-averaging (RSA) technique [48]–[53] to realize a unified RSA-based TDC architecture for both categories of high-speed and high-resolution TCSPC applications. To achieve this goal, the slow conversion-rate issue of asynchronous RSA addressed in [53] can be firstly alleviated by self-antithetic variance reduction (SAVR) with negligible power overhead but relatively high sensitivity to circuit/signal variations. Then, a simple digital circuit implementation for the control-variate variance reduction (CVVR) technique is secondly exploited to improve the performance consistency of SAVR and to form the collaborative variance reduction (CoVR = SAVR + CVVR) for further quantization-noise power suppression or, equivalently, conversion-rate enhancement. To comprehend the feasibility of the proposed VR techniques, this paper derives the theoretical expectations, variances, and correlation coefficients, which are all experimentally verified by the Monte Carlo simulations [46]. Meanwhile, the conversions of the mathematical closed-form expressions/models into digital signal-processing implementations are elaborated in this paper as well. Compared to an ordinary RSA-based TDC in [53], enabling the CoVR technique can boost the conversion-rate (or detection-rate per

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: tzhsueh@eng.ucsd.edu).
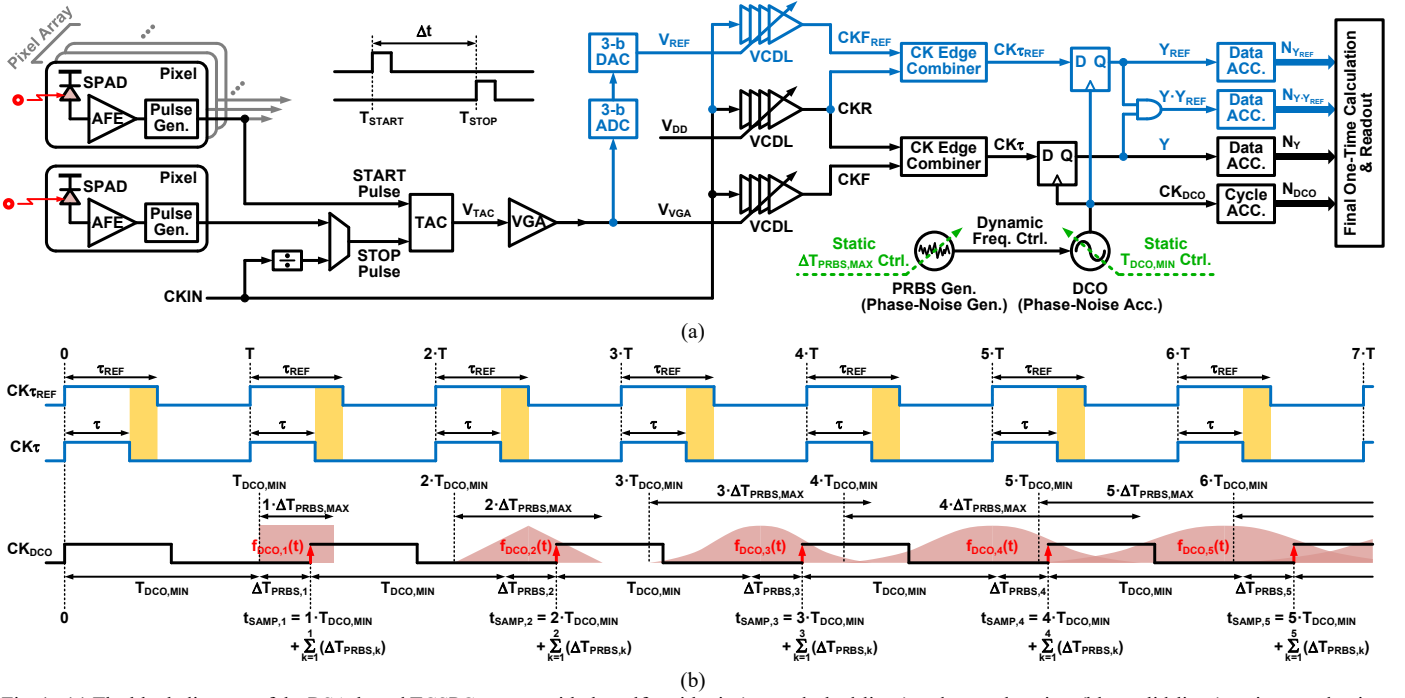
Fig. 1. (a) The block diagram of the RSA-based TCSPC system with the self-antithetic (green dashed-lines) and control-variate (blue solid-lines) variance reduction techniques. (b) The relation among CKτ, CKτ$_{REF}$, and sampling PDFs of CK$_{DCO}$ of the asynchronous RSA process in the absolute time-domain associated with the nature of the DCO phase-noise accumulation property.

pixel) up to 1.5 MHz with 12-effecitve-number-of-bits (ENOB) accuracy, 0.36-pJ/step energy efficiency, and 23% power overhead based on the simulations.

The remainder of the paper is organized as follows. The signal-and-circuit overview of an RSA-based TCSPC system with the SAVR and CVVR techniques is introduced in Section II. The fundamentals of RSA are summarized in Section III. The probability principles, theoretical analyses, and behavioral model simulations of the RSA with SAVR, CVVR, and CoVR techniques are presented in Section IV, V and VI, respectively. The conclusion and future work are summarized in Section VII.

II. SYSTEM-LEVEL OVERVIEW

The block diagram of the RSA-based TCSPC system with the SAVR and CVVR techniques is shown in Fig. 1(a). Each detection pixel contains a single-photon avalanche diode (SPAD), silicon-photonics analog front-end (AFE) [5]–[7], [16], [54], 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$_{START}$ 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$_{STOP}$ from either the other specific single-photon detection pixel [2] or the system input clock, CKIN, [4]–[7], [16]. In any configuration, the time-interval, Δt, between T$_{START}$ and T$_{STOP}$ is the primary quantity under the measurement as shown in Fig. 1(a). The time-to-amplitude conversion (TAC) circuit, containing a tunable current source and capacitor banks for dynamic-range (DR) scalability, converts the one-time captured time-interval, Δt, into a constant DC voltage, V$_{TAC}$ [53],

buffered by the variable-gain amplifier (VGA) for additional noise-rejection and driving capabilities.

The TDC mechanism is primarily illustrated at the right-hand side of Fig. 1(a), where the building blocks and signals plotted in black perform ordinary asynchronous RSA processes [53]; those plotted by the green dashed-lines and blue solid-lines represent the additional circuits for enabling the SAVR and CVVR techniques, respectively. The three identical voltage-controlled delay lines (VCDL) are all driven by CKIN, so the clock periods of CKR, CKF, and CKF$_{REF}$ are all equal to the period of CKIN, T, but the delays of CKR, CKF, and CKF$_{REF}$ are functions of the DC voltages, V$_{DD}$, V$_{VGA}$, and V$_{REF}$, respectively, where V$_{REF}$ is the maximum-likelihood estimate [55] of V$_{VGA}$ and obtained by one-time analog-to-digital and digital-to-analog conversions, i.e., the ADC-DAC pair in Fig. 1(a), after V$_{TAC}$ is settled for the RSA process. Therefore, the periodic delta, τ, between CKR and CKF represents the scaled version of Δt and is a function of (V$_{DD}$ − V$_{VGA}$) through the calibratable conversion gains of the TAC, VGA, and VCDL (i.e., K$_{TAC}$, K$_{VGA}$, and K$_{DL}$, respectively). After merging CKR and CKF by one of the rising edge-combiners, the resulting clock, CKτ, maintains the T periodicity while its duty-cycle, τ/T, carries the Δt information and becomes the primary quantity under the RSA process. By following the same mechanism, CKτ$_{REF}$ also maintains the T periodicity while its duty-cycle, τ$_{REF}$/T, is a function of (V$_{DD}$ − V$_{REF}$) and represents the maximum-likelihood estimate of τ/T under the circumstance of the pre-set resolution of the ADC-DAC pair.

To perform the random sampling process, a free-running ring-based digitally-controlled oscillator (DCO) generates an asynchronous clock, CK$_{DCO}$, to simultaneously sample the waveforms of CKτ and CKτ$_{REF}$ through two individual 1-bit D

flip-flops (DFF) as shown in Fig. 1(a). In the meantime, the randomness of each CK_{DCO} sampling-instant is mainly accomplished by a digital pseudo-random-binary-sequence generator (PRBS Gen.) to dynamically modulate the DCO period, T_{DCO}, and to ensure that the sampling probability density function (PDF) can satisfy the RSA criteria for the DFF outcomes, Y and Y_{REF}, to be the 1-bit binary random variables described in Section III. A 1-bit multiplier, i.e., an AND-gate, is required to generate the bit-on-bit product of the two simultaneously sampled random variables, $(Y \cdot Y_{REF})$, in the rate of CK_{DCO} for CVVR. Finally, the averaging process is executed by the data and cycle accumulators (ACC.) to first count the numbers of Logic-1s, N_Y, $N_{Y_{REF}}$, $N_{Y \cdot Y_{REF}}$, and N_{DCO}, at the outputs of the DFFs, AND-gate and DCO, respectively; then, the final result per RSA process can be obtained by one-time arithmetic calculations based on these accumulator outputs. At this point, the process of an asynchronous RSA-based TDC has been completed regardless of the VR techniques being enabled or not. At first glance, the entire system and process seem simple, which indicates the superior power/area efficiency of the RSA with VR techniques; however, the theory and concept in behind are quite complicated and unintuitive.

III. RANDOM SAMPLING-AND-AVERAGING OVERVIEW

In the asynchronous RSA process shown in Fig. 1(a), Y and Y_{REF} are the random variables; $E[Y]$ and $E[Y_{REF}]$ are the expectations of Y and Y_{REF}, respectively; Y_n and $Y_{REF,n}$ are the n-th samples of Y and Y_{REF}, respectively; N_{DCO} is the total number of the samples; $\overline{Y}$ is the mean of Y_n and represents the primary result per RSA process. In other words, $\overline{Y}$ is the Monte Carlo estimate [46] of the "ideal" expectations, $E[\overline{Y}] = E[Y]$; similarly, $\overline{Y}_{REF}$, $Y_{REF,n}$, and $E[\overline{Y}_{REF}] = E[Y_{REF}]$ follow the same relations. Therefore, the variances, $Var[\overline{Y}]$ and $Var[\overline{Y}_{REF}]$, represent the total noise powers within the Nyquist-frequency bandwidth of the RSA process, not $Var[Y]$ and $Var[Y_{REF}]$. One example of the asynchronous RSA sampling processes is shown in Fig. 1(b). For the sake of simplicity, the asynchronous sampling clock, CK_{DCO}, is assumed to have a coincident rising edge at t = 0 with those of the deterministic clocks, $CK\tau$ and $CK\tau_{REF}$. Based on the phase-noise accumulation property of ring-oscillators [56], the n-th absolute sampling time, $t_{SAMP,n}$, illustrated in Fig. 1(b), can be generalized as follows:

$$t_{SAMP,n} = \sum_{k=1}^{n} T_{DCO,k} = n \cdot T_{DCO,MIN} + \sum_{k=1}^{n} \Delta T_{PRBS,k} \quad (1)$$

where $T_{DCO,k}$ is the k-th period of the DCO; $\Delta T_{PRBS,k}$ is the k-th DCO period extension controlled by the PRBS generator; $T_{DCO,MIN}$ is the minimum DCO period when $\Delta T_{PRBS,k} = 0$; $\Delta T_{PRBS,MAX}$ in Fig. 1(b) is the maximum DCO period extension which sets the span of the time-domain sampling PDF of a "single" CK_{DCO} sampling edge, $f_{DCO,1}(t)$. Each $t_{SAMP,n}$ contains the deterministic term, $n \cdot T_{DCO,MIN}$, and stochastic term, which describes the uncertainty of each sampling instant and can only be represented by a PDF, $f_{DCO,n}(t)$. Therefore, the n-th CK_{DCO}

rising edge occurs randomly but is confined within the distribution span and density magnitude of its own PDF, i.e., the light-red areas in Fig. 1(b). More importantly, the stochastic term of $t_{SAMP,n}$ is the accumulation of "n" samples of an I.I.D. random variable (i.e., $\Delta T_{PRBS,k}$, k = 1 to n) created by the PRBS generator for "n" times as shown in the 2nd term of (1); equivalently, the PDF of the n-th DCO sampling instant, $f_{DCO,n}(t)$, is the convolution result of total "n" fundamental PDFs, $f_{DCO,1}(t)$, from the PRBS generator based on the Convolution Theorem [57]. Note that the fundamental PDF, $f_{DCO,1}(t)$, has a constant $1/\Delta T_{PRBS,MAX}$ density magnitude and $\Delta T_{PRBS,MAX}$ distribution span as mentioned. However, when n >> 1, the Central Limit Theorem [57] guarantees that $f_{DCO,n}(t)$ converges to a Gaussian distribution with a wide distribution span, $n \cdot \Delta T_{PRBS,MAX}$, as shown at the top of Fig. 2(a) regardless of the sampling PDF, $f_{DCO,1}(t)$, from the PRBS generator.

Because of the periodicity of $CK\tau$, the entire distribution span of $f_{DCO,n}(t)$ 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, $CConv[\cdot]$, due to the modulo-T operation [53], [58]:

$$f_1(t) = CConv[f_{DCO,1}(t), \delta(t), T]$$
$$f_n(t) = CConv[f_{n-1}(t), f_1(t), T], \quad n > 1 \quad (2)$$

where $\delta(t)$ is the unit impulse; the "t" of $f_1(t)$, $f_{n-1}(t)$, and $f_n(t)$ is the modulo-T time-domain variable within [0, T], but the "t" of $f_{DCO,1}(t)$ and $\delta(t)$ is the absolute time-domain variable referenced to t = 0. Based on (2), $f_1(t)$ plays as not only the PDF of the 1st sampling instant but also the fundamental PDF element to obtain any $f_n(t)$ from $f_{n-1}(t)$. According to [53], as $f_{DCO,n}(t)$ converges to a Gaussian PDF with increasing "n", $f_n(t)$ converges to a uniformly distributed PDF with a constant density magnitude 1/T across the [0, T] distribution span. In other words, for all "n" >> 1, $f_n(t)$ becomes an "identically distributed" PDF and independent from the parameters of $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$, and even "n" as illustrated in Fig. 2(a). Therefore, the expectations of the RSA processes also converge and can be expressed by a continuous one-dimensional geometric probability [57] format shown below, and two examples are illustrated in Fig. 2(b).

$$E[\overline{Y}] = \lim_{N_{DCO}\to\infty} \frac{\sum_{n=1}^{N_{DCO}} Y_n}{N_{DCO}} = \lim_{N_{DCO}\to\infty} \frac{N_Y}{N_{DCO}}$$

$$= E[Y] = \int_0^T y_n(t) \cdot f_n(t) \cdot dt$$

$$= \int_0^\tau \frac{1}{T} \cdot dt + \int_\tau^T \frac{0}{T} \cdot dt = \frac{\tau}{T} = P_1 \quad (3)$$

$$E[\overline{Y}_{REF}] = \lim_{N_{DCO}\to\infty} \frac{\sum_{n=1}^{N_{DCO}} Y_{REF,n}}{N_{DCO}} = \lim_{N_{DCO}\to\infty} \frac{N_{Y_{REF}}}{N_{DCO}}$$

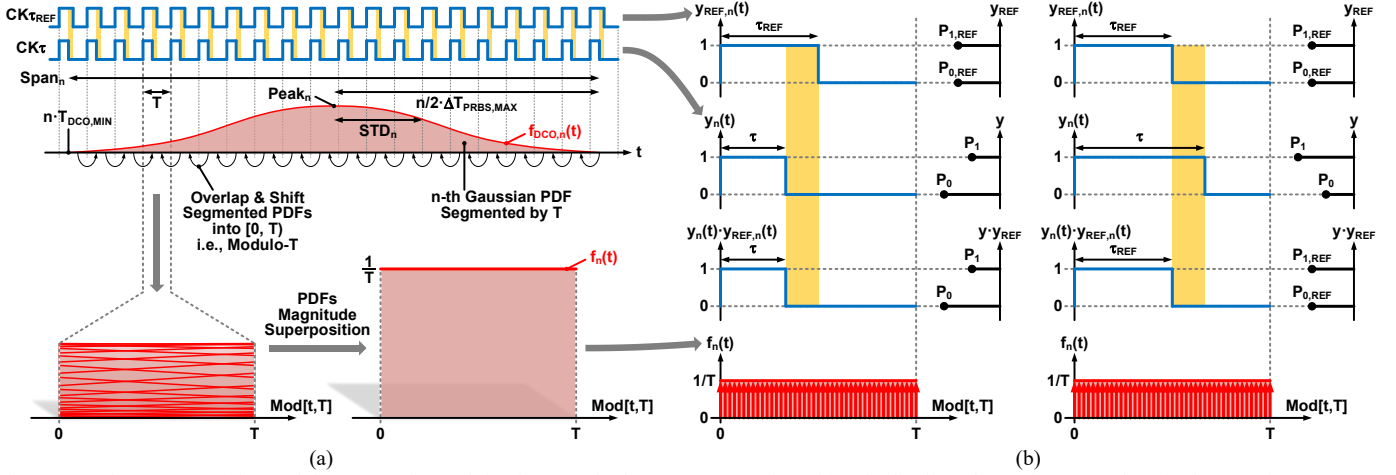$$= E[Y_{REF}] = \int_0^T y_{REF,n}(t) \cdot f_n(t) \cdot dt$$

Fig. 2. (a) The concept of the modulo-T operation and density-magnitude convergence of a uniformly distributed sampling PDF, $f_n(t)$, in the asynchronous RSA process [53]. (b) The examples of the n-th random sampling process (time-domain one-dimensional geometric probability density functions) of the random variable Y, Y_{REF}, and $(Y \cdot Y_{REF})$ with their corresponding binary-domain Bernoulli probability functions.

$$= \int_0^{\tau_{REF}} \frac{1}{T} \cdot dt + \int_{\tau_{REF}}^T \frac{0}{T} \cdot dt = \frac{\tau_{REF}}{T} = P_{1,REF} \quad (4)$$

$$E[Y \cdot Y_{REF}] = \lim_{N_{DCO} \to \infty} \frac{\sum_{n=1}^{N_{DCO}} (Y_n \cdot Y_{REF,n})}{N_{DCO}} = \lim_{N_{DCO} \to \infty} \frac{N_{Y \cdot Y_{REF}}}{N_{DCO}}$$

$$= \int_0^T y_n(t) \cdot y_{REF,n}(t) \cdot f_n(t) \cdot dt$$

$$= \begin{cases} \int_0^\tau \frac{1}{T} \cdot dt = \frac{\tau}{T} = P_1, & if \ \tau \leq \tau_{REF} \\ \int_0^{\tau_{REF}} \frac{1}{T} \cdot dt = \frac{\tau_{REF}}{T} = P_{1,REF}, & if \ \tau > \tau_{REF} \end{cases} \quad (5)$$

where $y_n(t)$ and $y_{REF,n}(t)$ are the modulo-T waveform of $CK\tau$ and $CK\tau_{REF}$ simultaneously sampled by the modulo-T sampling PDF, $f_n(t)$, of CK_{DCO}; $P_1 (= 1 - P_0)$ is the probability of obtaining a Y_n as Logic-1, and P_0 as Logic-0. $P_{1,REF} (= 1 - P_{0,REF})$ and $Y_{REF,n}$ shown in (4) follow the same relation. Based on the approach of generating the waveforms of $CK\tau$ and $CK\tau_{REF}$, $y_n(t)$ and $y_{REF,n}(t)$ have identical rising edges while the difference between them, which is highlighted in yellow and set by the timings of their falling edges (i.e., τ and τ_{REF}) as shown in Fig. 2(b), determines the degree of cross-correlation [55] between Y and Y_{REF} for CVVR. Without diving into the details of CVVR discussed in Section V, this section prepares the expectation of $(Y \cdot Y_{REF})$ in (5), which is the key of extracting the cross-correlation between Y and Y_{REF} in a power-efficient manner indicated in Fig. 2(b), i.e., $E[Y \cdot Y_{REF}]$ can be simply obtained by using $f_n(t)$ to sample $y_n(t) \cdot y_{REF,n}(t)$.

The time-interval, Δt, under each TDC process can be obtained by the accumulator outputs, $\bar{Y} = N_Y / N_{DCO} \approx \tau/T$ in (3), with pre-calibrated circuit parameters, K_{TAC}, K_{VGA}, and K_{DL}:

$$\Delta t = \frac{\tau}{K_{TAC} \cdot K_{VGA} \cdot K_{DL}} \approx \frac{T}{K_{TAC} \cdot K_{VGA} \cdot K_{DL}} \cdot \frac{N_Y}{N_{DCO}} \quad (6)$$

Any RSA-based TDC result, $\bar{Y}$, is a Monte Carlo estimate, so the theoretical variances can be verified by experimental data with the statistical process shown below [46], [59]:

$$Var[\bar{Y}] \approx \frac{1}{N_{EXP} - 1} \cdot \sum_{m=1}^{N_{EXP}} \left(\bar{Y}_m - \frac{1}{N_{EXP} - 1} \cdot \sum_{p=1}^{N_{EXP}} \bar{Y}_m \right)^2 \quad (7)$$

This verification accuracy depends on the number of $\bar{Y}$, N_{EXP}, obtained from experiments or simulations. In a realistic RSA-based TDC under a certain accuracy requirement with the settings of N_{DCO}, $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$ and T, only a single $\bar{Y}$ is necessary and sufficient to represent one TDC result.

IV. SELF-ANTITHETIC VARIANCE REDUCTION

The purpose of utilizing VR in the RSA-based TDC is to trade the accuracy improvement (or the amount of variance reduction) for accelerating the conversion-rate of RSA, which heavily relies on the numbers of samples to achieve high accuracy measurements. The idea of SAVR can be described by formulating the general variance of an RSA process, $Var[\bar{Y}]$, which is the summation of all pairwise covariances (i.e., auto-covariance sum [55]) among all N_{DCO} samples of the random variable, Y, with the weak law of large numbers [57]:

$$Var[\bar{Y}] = Var\left[\frac{\sum_{n=1}^{N_{DCO}} Y_n}{N_{DCO}} \right] = \frac{\sum_{n=1}^{N_{DCO}} \sum_{k=1}^{N_{DCO}} Cov[Y_n, Y_k]}{N_{DCO}^2}$$

$$= \frac{\sum_{n=1}^{N_{DCO}} Var[Y_n]}{N_{DCO}^2} + \frac{2 \cdot \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} Cov[Y_n, Y_k]}{N_{DCO}^2}$$

$$= \begin{cases} \frac{P_1 \cdot P_0}{N_{DCO}} = \frac{Var[Y]}{N_{DCO}}, & if \ \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} Cov[Y_n, Y_k] = 0 \\ Var[\bar{Y}_{SA}] < \frac{Var[Y]}{N_{DCO}}, & if \ \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} Cov[Y_n, Y_k] < 0 \end{cases} \quad (8)$$

where $Cov[\cdot]$ is the covariance operator. As shown in (8), if Y is an independent and identically distributed (I.I.D.) random variable, all pairwise covariance terms are zero when $n \neq k$; then
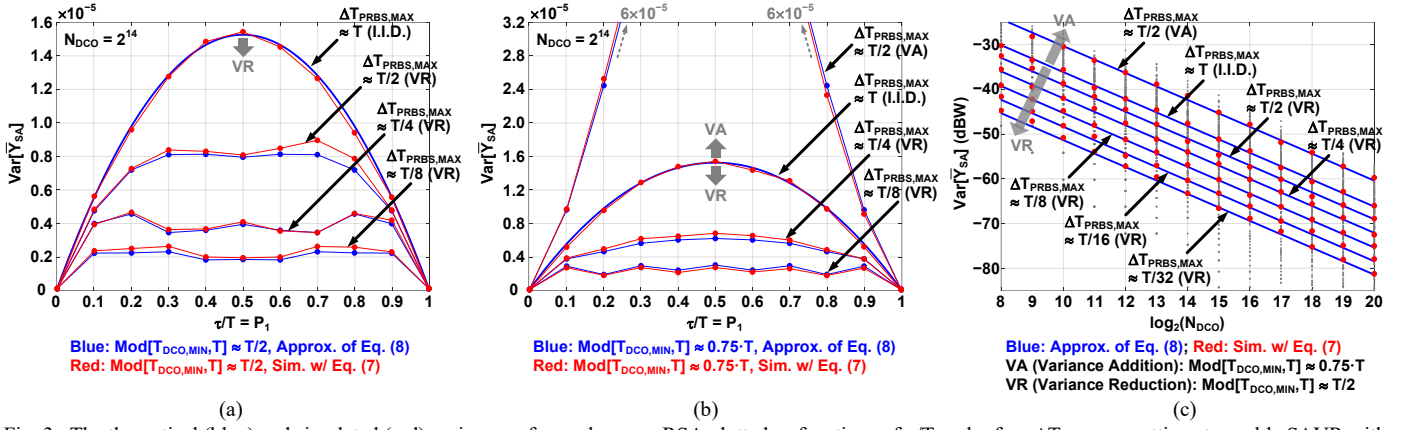
Fig. 3. The theoretical (blue) and simulated (red) variances of asynchronous RSA plotted as functions of τ/T under four $\Delta T_{PRBS,MAX}$ settings to enable SAVR with $N_{DCO} = 2^{14}$ and $N_{EXP} = 2^{11}$, when (a) Mod[$T_{DCO,MIN}$, T] $\approx$ T/2; (b) Mod[$T_{DCO,MIN}$, T] $\approx$ 0.75·T. (c) The theoretical (blue) and simulated (red) variances of asynchronous RSA plotted as functions of N_{DCO} with $\tau/T = 0.5$ and $N_{EXP} = 2^8$, including the I.I.D., VR, and VA scenarios based on the settings of $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$, and T.

the variance equals $P_1 \cdot P_0/N_{DCO}$ [53], [57], which is only contributed by the power summation of each individual sample, i.e., the 1st term in the 2nd line of (8). On the other hand, if there are correlations among the N_{DCO} samples of the random variable, Y, and the pairwise covariance sum of all $n \neq k$ is negative, i.e., the 2nd term in the 2nd line of (8), now the variance can be smaller than that of the I.I.D. situation, i.e., Var[$\bar{Y}_{SA}$] < $P_1 \cdot P_0/N_{DCO}$, which is the main idea of SAVR.

To induce the non-zero pairwise covariance sum when $n \neq k$, one of the important results in [53] of creating a non-I.I.D. random variable for an asynchronous RSA process can be exploited by scaling down the distribution span, $\Delta T_{PRBS,MAX}$, of the fundamental sampling PDF, $f_1(t)$ in (2), which can be implemented by the static PRBS-length control shown in Fig. 1(a). Meanwhile, this non-zero pairwise covariance sum must be negative as well, otherwise the overall effect becomes unfavorable variance addition (VA). To ensure the effective VR, the deterministic term of $f_1(t)$, Mod[$T_{DCO,MIN}$, T], has to be specified and can be implemented by the static DCO frequency control shown in Fig. 1(a) as well. The variances of the RSA with SAVR technique vs. normalized DR are plotted in Fig. 3(a) where Mod[$T_{DCO,MIN}$, T] $\approx$ T/2, $N_{DCO} = 2^{14}$, and $N_{EXP} = 2^{11}$. A couple of important conclusions have been demonstrated. First, a stronger VR can be statically enabled by decreasing the PRBS time-domain noise power, i.e., $\Delta T_{PRBS,MAX}$, without extra hardware to enhance the resolution of the PRBS generator since the infinite phase resolution has been taken care of by the inherent circuit/system noise accumulations through the DCO [53], [56]. Second, the efficiency of SAVR degrades with the increase of $|\tau/T - 0.5|$, but the quantization-noise power, Var[$\bar{Y}_{SA}$], stays relatively constant across the entire DR when $\Delta T_{PRBS,MAX} < T/8$. On the other hand, an example of unfavorable VA is demonstrated in Fig. 3(b) when the coincident interactions between Mod[$T_{DCO,MIN}$, T] and $\Delta T_{PRBS,MAX}$ always generates positive pairwise auto-covariances during the RSA process. Overall, SAVR can offer favorable variance reduction and consistent quantization-noise power under the requirement of setting Mod[$T_{DCO,MIN}$, T] $\in$ [0.25·T, 0.75·T] with $\Delta T_{PRBS,MAX} \leq$ T/4 based on the simulation results of sweeping the value of $T_{DCO,MIN}$ with respect to T,

which indicates the sensitivity of SAVR to the circuit/signal variations; this issue can be effectively mitigated by the CVVR technique discussed in Section VI.

In Fig. 3(c), the variances of the RSA with SAVR technique are plotted as functions of N_{DCO} on the dBW scale under different settings at the worst-case quantization-noise condition ($\tau/T = 0.5$) with $N_{EXP} = 2^8$. All cases follow the weak law of large numbers and perform variance degradations at −3 dBW per octave of N_{DCO} or equivalently −6 dBW per octave of ENOB [53]. 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. 3(a) and 3(c). The key contribution of SAVR is the conversion-rate enhancement with a small amount of power overhead. For example, to achieve the same variance of −60 dBW in Fig. 3(c), when $\Delta T_{PRBS,MAX} \approx$ T (SAVR disabled), it requires $N_{DCO} = 2^{18}$ 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 $T_{DCO,AVG} = (T_{DCO,MIN} + \Delta T_{PRBS,AVG}) \approx$ (T/2 + T/64) = 0.516·T, so the conversion-rate, $\propto 1/(N_{DCO} \cdot T_{DCO,AVG})$, is improved by 62×. About the power consumption with vs. without SAVR, the DCO power includes three 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) vs. 1/T; second, the per-stage capacitance load (C_L) controlled by the PRBS generator is much smaller since it is dominantly scaled with $\Delta T_{PRBS,MAX}$, i.e., T/32 vs. T; third, additional DC power is required for DCO static reconfigurations plotted by the green dashed-lines in Fig. 1(a). According to the simulation result and power estimation above ($C_L \cdot V_{DD}^2 \cdot F_{DCO,AVG}$ + DC power), the overall DCO power roughly stays the same with and without SAVR, i.e., 3.1 mW vs. 3 mW. Meanwhile, the power consumptions of the TAC, VGA, VCDL and edge-combiner are also 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 (TAC + VGA + VCDL + edge-combiner + DFF + clock buffer) with and without SAVR are 1.5 mW and 1.3 mW, respectively. Overall, the total RSA-based TDC power (i.e., DCO + TDC) is increased only by 7% after enabling SAVR for 62× conversion-rate enhancement.

V. CONTROL-VARIATE VARIANCE REDUCTION

The method of CVVR [46] utilizes information about the errors in estimates of known quantities to reduce the variance in an estimate of an unknown quantity under asynchronous RSA processes. It is important to note that both SAVR and CVVR in this paper reduce the quantization-noise power by strategically creating correlations. However, the correlations of SAVR exist among the samples of a random variable, Y, i.e., $[Y_1, Y_2, …, Y_{N_{DCO}}]$, so the correlations in SAVR are auto-correlations [55], and its overall variance is quantified by an auto-covariance sum as mentioned in (8); meanwhile, the correlations of CVVR exist among the samples of at least two random variables, Y and Y_{REF}, i.e., $[Y_1, Y_2, …, Y_{N_{DCO}}]$ and $[Y_{REF,1}, Y_{REF,2}, …, Y_{REF,N_{DCO}}]$, so the correlations in CVVR are cross-correlations [55], and its overall variance is quantified by a cross-covariance sum as elaborated in this section.

For the sake of simplicity, the theoretical analysis of CVVR starts with the assumption that Y and Y_{REF} are individual I.I.D. random variables, i.e., SAVR is disabled, and only CVVR is considered in this section. The first step is to examine the process of sampling $CK\tau$ and $CK\tau_{REF}$ by CK_{DCO} simultaneously as illustrated in Fig. 2(b) to create a cross-correlation between Y and Y_{REF} only at the sampling instants of "n = k":

$$Cov[Y_n, Y_{REF,k}] = E[Y_n \cdot Y_{REF,k}] - E[Y_n] \cdot E[Y_{REF,k}]$$
$$= \begin{cases} Cov[Y, Y_{REF}] \neq 0, & if \ n = k \\ 0, & if \ n \neq k \end{cases} \quad (9)$$

where "n" and "k" are both from 1 to N_{DCO}. Meanwhile, the cross-covariance between the outcomes of the parallel RSA processes, $\bar{Y}$ and $\bar{Y}_{REF}$, can be generalized as follows:

$$Cov[\bar{Y}, \bar{Y}_{REF}]$$
$$= E[\bar{Y} \cdot \bar{Y}_{REF}] - E[\bar{Y}] \cdot E[\bar{Y}_{REF}]$$
$$= \frac{\sum_{n=1}^{N_{DCO}} E[Y_n \cdot Y_{REF,n}]}{N_{DCO}^2} + \frac{2 \cdot \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} E[Y_n \cdot Y_{REF,k}]}{N_{DCO}^2}$$
$$- E[Y] \cdot E[Y_{REF}] \quad (10)$$

Since Y and Y_{REF} are individual I.I.D. random variables, the cross-covariance can be simplified based on (9) as follows:

$$Cov[\bar{Y}, \bar{Y}_{REF}]$$
$$= \frac{N_{DCO} \cdot E[Y \cdot Y_{REF}]}{N_{DCO}^2} + \frac{2 \cdot \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} E[Y_n] \cdot E[Y_{REF,k}]}{N_{DCO}^2}$$
$$- E[Y] \cdot E[Y_{REF}]$$
$$= \frac{E[Y \cdot Y_{REF}] - E[Y] \cdot E[Y_{REF}]}{N_{DCO}} = \frac{Cov[Y, Y_{REF}]}{N_{DCO}} \quad (11)$$

As shown in (11), sampling $CK\tau$ and $CK\tau_{REF}$ by CK_{DCO} simultaneously as illustrated in Fig. 2(b) also creates the cross-correlation between $\bar{Y}$ and $\bar{Y}_{REF}$. Because of the same rising edges, clock period, and random sampling PDFs, $f_n(t)$, the only difference between $y_n(t)$ and $y_{REF,n}(t)$ is their duty-cycles, which is the factor causing the different outcomes between the data sequences of Y and Y_{REF} without considering any mismatch

between the sampling DFFs. By plugging the expectations in (3), (4), and (5) into (11), the cross-covariance becomes:

$$Cov[\bar{Y}, \bar{Y}_{REF}]$$
$$= \frac{Cov[Y, Y_{REF}]}{N_{DCO}} = \frac{\sigma_Y \cdot \sigma_{Y_{REF}} \cdot \rho_{Y,Y_{REF}}}{N_{DCO}} = \sigma_{\bar{Y}} \cdot \sigma_{\bar{Y}_{REF}} \cdot \rho_{\bar{Y},\bar{Y}_{REF}}$$
$$= \begin{cases} \dfrac{P_1 - P_1 \cdot P_{1,REF}}{N_{DCO}} = \dfrac{P_1 \cdot P_{0,REF}}{N_{DCO}}, & if \ \tau \leq \tau_{REF} \\[2mm] \dfrac{P_{1,REF} - P_1 \cdot P_{1,REF}}{N_{DCO}} = \dfrac{P_0 \cdot P_{1,REF}}{N_{DCO}}, & if \ \tau > \tau_{REF} \end{cases} \quad (12)$$

where σ_Y, $\sigma_{Y_{REF}}$, $\sigma_{\bar{Y}}$, and $\sigma_{\bar{Y}_{REF}}$ are standard deviations of Y, Y_{REF}, $\bar{Y}$, and $\bar{Y}_{REF}$, respectively; $\rho_{Y,Y_{REF}}$ is the correlation coefficient [57] between Y and Y_{REF}, and $\rho_{\bar{Y},\bar{Y}_{REF}}$ between $\bar{Y}$ and $\bar{Y}_{REF}$. Since Y and Y_{REF} are individually I.I.D., their standard deviations are the square roots of their variances under the I.I.D. condition shown in (8). Therefore, the correlation coefficients can be found from (12) and (8):

$$\rho_{\bar{Y},\bar{Y}_{REF}} = \rho_{Y,Y_{REF}} = \begin{cases} \sqrt{\dfrac{P_1 \cdot P_{0,REF}}{P_0 \cdot P_{1,REF}}}, & if \ \tau \leq \tau_{REF} \\[3mm] \sqrt{\dfrac{P_0 \cdot P_{1,REF}}{P_1 \cdot P_{0,REF}}}, & if \ \tau > \tau_{REF} \end{cases} \quad (13)$$

Based on (13), the correlation coefficients are functions of τ and τ_{REF} since $P_1 = 1 - P_0 = \tau/T$ and $P_{1,REF} = 1 - P_{0,REF} = \tau_{REF}/T$ as shown in (3) and (4). Thus, the degree of correlation between Y and Y_{REF} is determined by the amount of overlap between the waveforms of the $CK\tau$ and $CK\tau_{REF}$, which is proven by (13).

Once confirming the cross-correlation between Y and Y_{REF} can be implemented by the random sampling process shown in Fig. 2(b), the next step is to prepare the "known" random variable, Y_{REF}, by using the asynchronous RSA process to measure the duty-cycle of $CK\tau_{REF}$, τ_{REF}/T, so that the Monte Carlo estimate of Y_{REF} with a relatively high accuracy, i.e., $\hat{E}[\bar{Y}_{REF}]$ with $Var[\bar{Y}_{REF}]$, becomes ready and "known" before measuring the "unknown" random variable, Y, with a targeted accuracy, i.e., $\bar{Y}$ with $Var[\bar{Y}]$. Since the pre-characterization of $\hat{E}[\bar{Y}_{REF}]$ with $Var[\bar{Y}_{REF}]$ must be done before the regular operation can be executed, the time consumption due to this pre-characterization is under the cost of the calibration procedure or instrument initialization, which is not counted toward the cost of conversion-rate under the RSA-based TDC specification. Note that the Monte Carlo estimate of Y_{REF} is expressed by $\hat{E}[\bar{Y}_{REF}]$ instead of $\bar{Y}_{REF}$ to represent its extremely high accuracy, obtained from the pre-characterization process. In addition, the non-idealities of the analog circuits, including TAC, VGA, and VCDL, due to process, voltage, and temperature (PVT) variations, can be also calibrated by this pre-characterization process described in Appendix.

With the "known" information, $\hat{E}[\bar{Y}_{REF}]$, ready and the cross-correlation between Y and Y_{REF} shown in (13), the samples of a new random variable, Y_{CV}, possessing the same expectation

of Y but a lower variance due to CVVR, can be formulated based on the per sample relation between $Y_{REF,n}$ and Y_n [46]:

$$Y_{CV,n} = Y_n - \mu_{CV} \cdot \left(Y_{REF,n} - \hat{E}[\bar{Y}_{REF}]\right) \tag{14}$$

$$E[Y_{CV}] = E[Y] - \mu_{CV} \cdot E[Y_{REF}] + \mu_{CV} \cdot E\left[\hat{E}[\bar{Y}_{REF}]\right]$$

$$= E[Y] = \frac{\tau}{T} = P_1 \tag{15}$$

$$Var[Y_{CV}] = Var\left[Y - \mu_{CV} \cdot \left(Y_{REF} - \hat{E}[\bar{Y}_{REF}]\right)\right]$$

$$\approx Var\left[Y - \mu_{CV} \cdot Y_{REF} + \mu_{CV} \cdot E[Y_{REF}]\right]$$

$$= Var[Y] + \mu_{CV}^2 \cdot Var[Y_{REF}] - 2\mu_{CV} \cdot Cov[Y, Y_{REF}] \tag{16}$$

where $Y_{CV,n}$ is the variance-reduced version of Y_n per sample; the error term, $(Y_{REF,n} - \hat{E}[\bar{Y}_{REF}])$, serves as a "control" per sample to produce $Y_{CV,n}$ in estimating $E[Y] = E[Y_{CV}]$ as shown in (15); μ_{CV} is the VR coefficient. To intuitively understand the reason for effective VR achieved by (14), the extreme case of the fully correlated condition can be quickly examined, i.e., $Y_n = Y_{REF,n}$ for all "n". With $\mu_{CV} = 1$, Equation (14) leads to the result of $Y_{CV,n} = \hat{E}[\bar{Y}_{REF}]$ for all "n"; this means that even a single sampled outcome, $Y_{CV,1}$, can achieve the accuracy of $\hat{E}[\bar{Y}_{REF}]$ without needing a large number of samples. The degree of correlation between Y and Y_{REF} can significantly affect the efficiency of CVVR and correspondingly reduce the required number of total samples, N_{DCO}, to reach the target accuracy, which is dominated by the accuracy of $\hat{E}[\bar{Y}_{REF}]$, $Var[\bar{Y}_{REF}]$, during the pre-characterization process. In other words, $Var[Y_{CV}]$ is less than $Var[Y]$ since the combination of the 2nd and 3rd terms in the 3rd line of (16) is negative and a function of $Cov[Y, Y_{REF}]$. Note that the approximation in the 2nd line of (16) is due to ignoring the finite accuracy of $\hat{E}[\bar{Y}_{REF}]$ by replacing it with $E[Y_{REF}]$ for the sake of simplicity.

In short, the key idea of CVVR is established on taking advantage of the pre-characterization process in two aspects: first, the more time or accuracy contributed by the pre-characterization, the more VR or conversion-rate can be gained; second, the higher correlation between "known" and "unknown" quantities, the more VR or conversion-rate can be enhanced as well. However, two major concerns can be raised in the circuit implementation: extra circuit power/area consumption and achievable degree of correlation, which are successively discussed in the following sub-sections.

A. Practical Realization of Control-Variate VR

The realization of $Y_{CV,n}$ in (14) actually requires a high-resolution and high-speed digital operation per sample, which can induce a certain amount of power/area overhead because of two facts: first, the pre-characterized, $\hat{E}[\bar{Y}_{REF}]$, and VR coefficient, μ_{CV}, possess high resolutions to meet the targeted accuracy per RSA process; second, although $\hat{E}[\bar{Y}_{REF}]$ and μ_{CV} are constant, the digital operation of (14) is executed as fast as the DCO frequency to generate the high-resolution $Y_{CV,n}$ per sample. Fortunately, the primary outcome per RSA process is the mean value of the sampled random variable, $\bar{Y}_{CV}$, so the per sample format in (14) can incorporate the averaging process to perform the RSA with CVVR technique:

$$\bar{Y}_{CV} = \frac{\sum_{n=1}^{N_{DCO}}\left[Y_n - \mu_{CV} \cdot \left(Y_{REF,n} - \hat{E}[\bar{Y}_{REF}]\right)\right]}{N_{DCO}}$$

$$= \bar{Y} - \mu_{CV} \cdot \left(\bar{Y}_{REF} - \hat{E}[\bar{Y}_{REF}]\right)$$

$$= \frac{N_Y}{N_{DCO}} - \mu_{CV} \cdot \left(\frac{N_{Y_{REF}}}{N_{DCO}} - \hat{E}[\bar{Y}_{REF}]\right) \tag{17}$$

$$E[\bar{Y}_{CV}] = E[\bar{Y}] - \mu_{CV} \cdot E[\bar{Y}_{REF}] + \mu_{CV} \cdot E\left[\hat{E}[\bar{Y}_{REF}]\right]$$

$$= E[\bar{Y}] = E[Y] = \frac{\tau}{T} = P_1 \tag{18}$$

$$Var[\bar{Y}_{CV}] = Var\left[\bar{Y} - \mu_{CV} \cdot \left(\bar{Y}_{REF} - \hat{E}[\bar{Y}_{REF}]\right)\right]$$

$$\approx Var[\bar{Y}] + \mu_{CV}^2 \cdot Var[\bar{Y}_{REF}] - 2\mu_{CV} \cdot Cov[\bar{Y}, \bar{Y}_{REF}]$$

$$= \frac{Var[Y]}{N_{DCO}} + \mu_{CV}^2 \cdot \frac{Var[Y_{REF}]}{N_{DCO}} - 2\mu_{CV} \cdot \frac{Cov[Y, Y_{REF}]}{N_{DCO}} \tag{19}$$

Equation (17), (18), and (19) seem similar to (14), (15), and (16), respectively, but they actually represent a power/area efficient realization of the RSA with CVVR technique. First, instead of calculating every single high-resolution $Y_{CV,n}$ in (14) and then taking the average over N_{DCO} samples, the realization method in (17) shows that $\bar{Y}_{CV}$ can be actually obtained by a one-time calculation in high-resolution after the parallel RSA processes of $\bar{Y}$ ($= N_Y/N_{DCO}$) and $\bar{Y}_{REF}$ ($= N_{Y_{REF}}/N_{DCO}$) are completed, so the high-speed random sampling-and-averaging circuits all stay as the 1-bit digital operations as shown in Fig. 1(a). Second, the expectation shown in (18) proves that the convergency of $\bar{Y}_{CV}$ is not affected by the simplified CVVR realization method from (14) to (17). Third, the theoretical variance of the RSA with CVVR technique in (19) still follows the weak law of large numbers.

With the low power/area realization method in (17) and pre-characterized $\hat{E}[\bar{Y}_{REF}]$, the last item for the RSA with CVVR technique is the optimal value of the VR coefficient, μ_{CV}, which can minimize the variance of $\bar{Y}_{CV}$ shown in (19):

$$\frac{\partial(Var[\bar{Y}_{CV}])}{\partial \mu_{CV}} = 2 \cdot \mu_{CV} \cdot Var[\bar{Y}_{REF}] - 2 \cdot Cov[\bar{Y}, \bar{Y}_{REF}] = 0$$

$$\mu_{CV} = \frac{Cov[\bar{Y}, \bar{Y}_{REF}]}{Var[\bar{Y}_{REF}]} = \frac{\sigma_{\bar{Y}}}{\sigma_{\bar{Y}_{REF}}} \cdot \rho_{\bar{Y}, \bar{Y}_{REF}} \tag{20}$$

By plugging (20) into (19), the minimum variance of the RSA with CVVR technique is shown as follows:

$$Var[\bar{Y}_{CV}] \approx Var[\bar{Y}] - \frac{Cov^2[\bar{Y}, \bar{Y}_{REF}]}{Var[\bar{Y}_{REF}]} \tag{21}$$

Clearly, the amount of variance reduction (the 2nd term of (21)) from the variance of an I.I.D. case (the 1st term of (21)) is mainly determined by the cross-correlation between $\bar{Y}$ and $\bar{Y}_{REF}$; this is the reason for finding V_{REF} as the maximum likelihood estimate of V_{VGA} under the available resolution of the ADC-DAC pair. Moreover, the conclusion shown in (21) can be further quantified by the gain of CVVR or conversion-rate, G_{CV}, of this technique, which is represented by the ratio between the RSA variances without and with CVVR:

$$G_{CV} = \frac{Var[\overline{Y}]}{Var[\overline{Y}_{CV}]} = \frac{Var[\overline{Y}] \cdot Var[\overline{Y}_{REF}]}{Var[\overline{Y}] \cdot Var[\overline{Y}_{REF}] - Cov^2[\overline{Y}, \overline{Y}_{REF}]}$$
$$= \frac{\sigma_{\overline{Y}}^2 \cdot \sigma_{\overline{Y}_{REF}}^2}{\sigma_{\overline{Y}}^2 \cdot \sigma_{\overline{Y}_{REF}}^2 - \sigma_{\overline{Y}}^2 \cdot \sigma_{\overline{Y}_{REF}}^2 \cdot \rho_{\overline{Y}, \overline{Y}_{REF}}^2} = \frac{1}{1 - \rho_{\overline{Y}, \overline{Y}_{REF}}^2} \quad (22)$$

Based on (20), (21), and (22), multiple important attributes and implementation methodologies of CVVR are discussed as follows. First, the gain of CVVR, G_{CV}, is purely determined by the degree of cross-correlation between the outcomes of the parallel RSA processes, $\overline{Y}$ and $\overline{Y}_{REF}$. For example, when they are uncorrelated, $\rho_{\overline{Y}, \overline{Y}_{REF}} = 0 = \mu_{CV}$, $G_{CV} = 1$, and then $Var[\overline{Y}_{CV}]$ = $Var[\overline{Y}]$; when they are fully correlated, $\rho_{\overline{Y}, \overline{Y}_{REF}} = 1 = \mu_{CV}$, G_{CV} = ∞, and then $Var[\overline{Y}_{CV}] \approx 0$ based on (21). Second, unlike the variance of SAVR in (8), which requires the auto-covariance sum of all $Cov[Y_n, Y_k]$, $n \neq k$ to be negative, the cross-covariance, $Cov[\overline{Y}, \overline{Y}_{REF}]$, of CVVR can be either positive or negative; both polarities can only create effective VR instead of VA because $Var[\overline{Y}_{CV}]$ in (21) is always less than or equal to $Var[\overline{Y}]$ regardless of the polarity of $Cov[\overline{Y}, \overline{Y}_{REF}]$, which has been cancelled by the same cross-covariance embedded in μ_{CV} as shown in (20). This attribute reflects the superior stability of CVVR than that of SAVR. Third, CVVR scales the variance in a reciprocal manner of G_{CV} based on (22). Therefore, if a certain amount of VR can be obtained by increasing N_{DCO}, theoretically the same amount of VR can be achieved by CVVR without increasing N_{DCO}. This is the reason for using G_{CV} to represent the gain of conversion-rate enhancement or measurement-time ($\propto N_{DCO} \cdot T_{DCO,AVG}$) reduction offered by CVVR to reach a certain requirement of variance or ENOB.

The three attributes discussed so far are all under the assumption of knowing the value of μ_{CV}, which indicates the significance of obtaining this optimal VR coefficient in this technique. However, based on μ_{CV} and $Cov[\overline{Y}, \overline{Y}_{REF}]$ in (20) and (11), respectively, if $E[Y]$ is unknown, which is the ideal RSA-based TDC result, then it is unlikely to find this optimal VR coefficient. Fortunately, the concept of the Monte Carlo method is also applicable to finding an estimate of μ_{CV} by replacing the expectations with their estimates from the RSA processes under a finite sampling number, N_{DCO}. Therefore, since Y and Y_{REF} are individual I.I.D. random variables, the estimate VR coefficient, $\hat{\mu}_{CV}$, can be derived by:

$$\mu_{CV} = \frac{Cov[\overline{Y}, \overline{Y}_{REF}]}{Var[\overline{Y}_{REF}]} = \frac{Cov[Y, Y_{REF}]}{Var[Y_{REF}]}$$
$$= \frac{E[(Y - E[Y]) \cdot (Y_{REF} - E[Y_{REF}])]}{E[(Y_{REF} - E[Y_{REF}])^2]}$$
$$\approx \frac{\sum_{n=1}^{N_{DCO}}[(Y_n - \overline{Y}) \cdot (Y_{REF,n} - \overline{Y}_{REF})]}{\sum_{n=1}^{N_{DCO}}(Y_{REF,n} - \overline{Y}_{REF})^2} = \hat{\mu}_{CV} \quad (23)$$

Though (23) eliminates the necessity of ideal expectations, it requires hardware to store the entire data sequences of Y_n and $Y_{REF,n}$ with N_{DCO} samples until the parallel RSA process outcomes, $\overline{Y}$ and $\overline{Y}_{REF}$, are ready for the calculations of $\hat{\mu}_{CV}$ in (23) and then $\overline{Y}_{CV}$ in (17). To avoid an explosion of circuit power/area overhead, $\hat{\mu}_{CV}$ in (23) can be reformulated by (24) which demonstrates significant hardware simplification to obtain $\hat{\mu}_{CV}$; it is basically a one-time multi-bit operation with N_Y, $N_{Y_{REF}}$, $N_{Y \cdot Y_{REF}}$ and N_{DCO} from the 1-bit accumulators after the completion of each RSA-based TDC process.

B. Efficiency of Control-Variate VR

Based on (22), the gain of CVVR, G_{CV}, substantially increases as $|\rho_{\overline{Y}, \overline{Y}_{REF}}|$ approaches "1" and decreases as $|\rho_{\overline{Y}, \overline{Y}_{REF}}|$ moves away from "1". To maintain the efficiency of CVVR, the cross-correlation between $\overline{Y}$ and $\overline{Y}_{REF}$ has to be confined up to a certain degree. According to (12) and (13) with the assumption that Y and Y_{REF} are individual I.I.D. random variables declared at the beginning of Section V, the cross-correlation between Y and Y_{REF} can be simply set or controlled by the overlap between τ and τ_{REF} of the $CK\tau$ and $CK\tau_{REF}$ waveforms, respectively, as shown in Fig. 2(b). Therefore, to maintain a high correlation between τ and τ_{REF} across the entire DR of $\tau \in [0, T]$, multiple pre-characterized $\hat{E}[\overline{Y}_{REF}]$ (i.e., $\hat{E}[\overline{Y}_{REF1}]$, $\hat{E}[\overline{Y}_{REF2}]$, $\hat{E}[\overline{Y}_{REF3}]$, ...) generated by their corresponding τ_{REF}-options (i.e., τ_{REF1}, τ_{REF2}, τ_{REF3}, ...) can be deployed. With a pre-set number of τ_{REF}-options, any τ under the RSA with CVVR technique needs to be coarsely quantized to find one of the τ_{REF}-options having the maximum overlap or correlation with τ. In other words, the chosen τ_{REF}-option is the maximum likelihood estimate of τ, so the most efficient CVVR can be performed under the available resource of the pre-characterized τ_{REF}-options with their $\hat{E}[\overline{Y}_{REF}]$. That is, if a finer step size between the adjacent τ_{REF}-options has been offered within $[0, T]$, which corresponds to more pre-characterized $\hat{E}[\overline{Y}_{REF}]$, then a larger correlation coefficient, $\rho_{\overline{Y}, \overline{Y}_{REF}}$ can be obtained for a higher amount of CVVR. The implementation of finding the best τ_{REF}-option is a part of the $CK\tau_{REF}$ generation as described in Section II. When V_{VGA} is converted from Δt to set the delay of CKF, it is also fed into a pair of coarse ADC and DAC to generate V_{REF} and then to set the delay of CKF_{REF} through the VCDL plotted by the blue solid-lines in Fig. 1(a). Therefore, the duty-cycle of $CK\tau_{REF}$, τ_{REF}/T, can be the one closet to τ/T. The resolution of generating τ_{REF}-options is set by the resolutions of the ADC-DAC pair and pre-characterization process. A higher resolution or number of τ_{REF}-options (or V_{REF}) can provide better CVVR, however, with higher ADC-DAC circuit overhead and pre-characterization effort.

$$\hat{\mu}_{CV} = \frac{\sum_{n=1}^{N_{DCO}}(Y_n \cdot Y_{REF,n} - Y_n \cdot \overline{Y}_{REF} - \overline{Y} \cdot Y_{REF,n} + \overline{Y} \cdot \overline{Y}_{REF})}{\sum_{n=1}^{N_{DCO}}(Y_{REF,n}^2 - 2 \cdot Y_{REF,n} \cdot \overline{Y}_{REF} + \overline{Y}_{REF}^2)} = \frac{\sum_{n=1}^{N_{DCO}}(Y_n \cdot Y_{REF,n}) + N_{DCO} \cdot \overline{Y} \cdot \overline{Y}_{REF}}{\sum_{n=1}^{N_{DCO}}(Y_{REF,n}^2) - 2 \cdot \overline{Y}_{REF} \cdot N_{DCO} \cdot \overline{Y}_{REF} + N_{DCO} \cdot \overline{Y}_{REF}^2}$$
$$= \frac{\frac{\sum_{n=1}^{N_{DCO}}(Y_n \cdot Y_{REF,n})}{N_{DCO}} - \overline{Y} \cdot \overline{Y}_{REF}}{\overline{Y}_{REF} - \overline{Y}_{REF}^2} = \frac{N_{DCO} \cdot N_{Y \cdot Y_{REF}} - N_Y \cdot N_{Y_{REF}}}{(N_{DCO} - N_{Y_{REF}}) \cdot N_{Y_{REF}}} \quad (24)$$
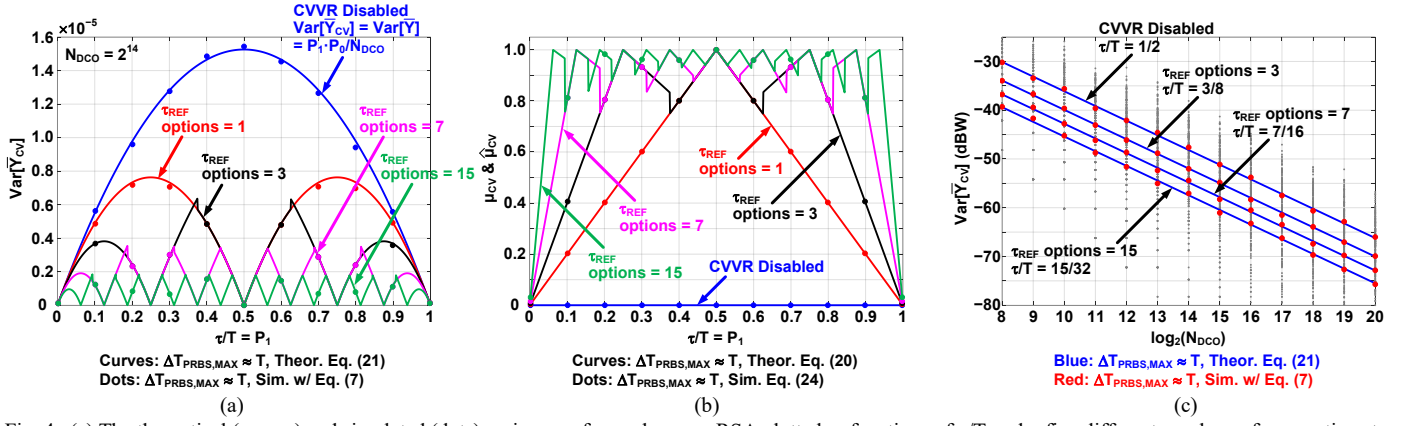
Fig. 4. (a) The theoretical (curves) and simulated (dots) variances of asynchronous RSA plotted as functions of τ/T under five different numbers of τ_{REF}-options to enable CVVR with $N_{DCO} = 2^{14}$ and $N_{EXP} = 2^{11}$. (b) The theoretical (curves) and simulated (dots) VR coefficients plotted as functions of τ/T according to the settings of CVVR in (a). (c) The worst-case theoretical (curves) and simulated (dots) variances of asynchronous RSA plotted as functions of N_{DCO} under four different numbers of τ_{REF}-options to enable CVVR with $N_{EXP} = 2^8$. Note that τ_{REF}/T step-size (LSB) = 1/2, 1/4, 1/8, and 1/16 for τ_{REF}-options = 1, 3, 7, and 15, respectively.

The variances of the RSA with CVVR technique vs. normalized DR are plotted in Fig. 4(a) under different numbers of τ_{REF}-options with $N_{DCO} = 2^{14}$ and $N_{EXP} = 2^{11}$, where $Mod[\Delta T_{PRBS,MAX}, T] \approx 0$ to ensure that Y and Y_{REF} are individual I.I.D. random variables, i.e., SAVR is disabled in these examples. The color coding is used to represent different numbers of τ_{REF}-options across the DR of τ. Multiple important conclusions have been demonstrated. First, the theoretical variances (solid curves) based on (21) match well with the experimental simulation results (solid dots) statistically processed by (7). Second, more τ_{REF}-options offered by the pre-characterized $\hat{E}[\bar{Y}_{REF}]$ and ADC-DAC hardware lead to lower quantization-noise power from the coarse ADC, which is equivalent to higher correlations between τ_{REF} and τ, i.e., $Cov[\bar{Y}, \bar{Y}_{REF}]$ in (21), across the entire DR, more variance lobes, more variance notches due to fully coincident correlations, and eventually more VR can be performed in average. Third, the discontinuities at the peaks of the variance lobes are due to the linear quantization process of the coarse ADC which sets the thresholds at the middle of two τ_{REF}-options, so two well-apart values of τ_{REF}/T (= $\bar{Y}_{REF}$) could be used in (21) for two close values of τ/T (= $\bar{Y}$). For example, in the case of τ_{REF}-options = 3, $\tau_{REF}/T = 0.25$ is chosen when τ/T is a little bit less than 0.375, otherwise $\tau_{REF}/T = 0.5$ when τ/T is just barely larger than 0.375. In any case, these discontinuities only create negligible increases of the variance peaks and render no reason for extra hardware to perform a non-linear quantization in the coarse ADC. Fourth, the theoretical VR coefficients (solid curves) based on (20) and its approximation (solid dots) for hardware-friendly implementation based on (24) are shown in Fig. 4(b) according to all parameter settings for the results shown in Fig. 4(a). As expected, when the correlation between τ and τ_{REF} increases in average due to more τ_{REF}-options, the values of μ_{CV} and $\hat{\mu}_{CV}$ tend to concentrate on the highly correlated regions, which are close to "1". In Fig. 4(c), the variances are plotted as functions of N_{DCO} on the dBW scale with different CVVR settings at the worst-cases quantization-noise condition in Fig. 4(a). Both theoretical (blue curves) and simulation (red dots) results are well aligned and follow the

weak law of large numbers with variance degradations at −3 dBW per octave of N_{DCO} or equivalently −6 dBW per octave of ENOB [53]. More importantly, the worst-case variances exhibit roughly −3 dBW per octave of the number of τ_{REF}-options, i.e., 3, 7, 15, and so forth, for all N_{DCO}. About the power consumption with vs. without CVVR, the DCO power stays the same due to its zero configuration-changes while the CKτ_{REF} data path plotted by the blue solid-lines in Fig. 1(a) is enabled for CVVR and occupies the primary power overhead. Thus, the TDC power numbers (TAC + VGA + coarse ADC-DAC + VCDL + edge-combiner + DFF + AND + clock buffer) with and without CVVR are 1.9 mW and 1.3 mW, respectively. Overall, the total RSA-based TDC power (DCO + TDC) is roughly increased by 14%, but the payback is more than 8× conversion-rate enhancement when τ_{REF}-options = 15.

VI. COLLABORATIVE VARIANCE REDUCTION

A. Practical Realization of Collaborative VR

The RSA with CoVR technique can take the complementary benefits from SAVR and CVVR without further power/area overhead since the required hardware for CVVR, including the pre-characterization of $\hat{E}[\bar{Y}_{REF}]$ and the parallel RSA process of $\bar{Y}_{REF}$, is basically isolated from the hardware for SAVR. In other words, the RSA processes of $\bar{Y}$ and $\bar{Y}_{REF}$ can perform their own SAVR not only simultaneously due to their common sampling clock (CK$_{DCO}$) but also individually due to their separate data paths as shown in Fig. 1(a); any VR in the RSA process of $\bar{Y}$ is only contributed by SAVR regardless of enabling CVVR or not, i.e., $\bar{Y} = \bar{Y}_{SA}$ and $Var[\bar{Y}] = Var[\bar{Y}_{SA}]$ when SAVR is enabled as shown in (8). To perform CoVR, the final one-time high-resolution calculation for incorporating CVVR into SAVR is still required as shown in (25) followed by the verification of convergence in (26).

$$\bar{Y}_{SACV} = \bar{Y}_{SA} - \hat{\mu}_{SACV} \cdot (\bar{Y}_{REF} - \hat{E}[\bar{Y}_{REF}])$$
$$\approx \frac{N_Y}{N_{DCO}} - \hat{\mu}_{CV} \cdot \left(\frac{N_{Y_{REF}}}{N_{DCO}} - \hat{E}[\bar{Y}_{REF}]\right) \quad (25)$$
$$E[\bar{Y}_{SACV}] = E[\bar{Y}_{SA}] - \hat{\mu}_{SACV} \cdot E[\bar{Y}_{REF}] + \hat{\mu}_{SACV} \cdot E[\hat{E}[\bar{Y}_{REF}]]$$
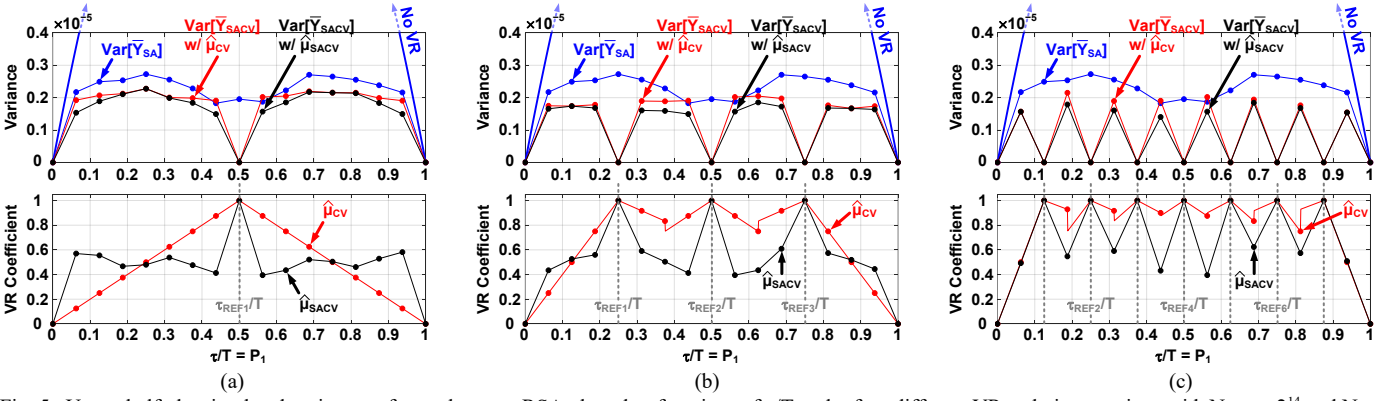
Fig. 5. Upper-half: the simulated variances of asynchronous RSA plotted as functions of τ/T under four different VR technique settings with $N_{DCO} = 2^{14}$ and $N_{EXP} = 2^{11}$, including no VR (blue curves), SAVR (blue dots), CoVR with suboptimal VR coefficients (red dots), and CoVR with optimal VR coefficients (black dots). Lower-half: the simulated suboptimal (red dots) and optimal (black dots) VR coefficients plotted as functions of τ/T for CoVR with $N_{DCO} = 2^{14}$ and $N_{EXP} = 2^{11}$. Note that (a), (b), and (c) have τ_{REF}-options = 1, 3, and 7, respectively, across the normalized DR of $\tau/T \in [0, 1)$.

$$= E[\bar{Y}_{SA}] = E[\bar{Y}] = E[Y] = \frac{\tau}{T} = P_1 \qquad (26)$$

$$Var[\bar{Y}_{SACV}] \approx Var[\bar{Y}_{SA}] + \hat{\mu}_{SACV}^2 \cdot Var[\bar{Y}_{REF}] \\ -2 \cdot \hat{\mu}_{SACV} \cdot Cov[\bar{Y}_{SA}, \bar{Y}_{REF}] \qquad (27)$$

$\bar{Y}_{SACV}$ and its variance, $Var[\bar{Y}_{SACV}]$, represent the RSA-based TDC result and quantization-noise power, respectively, when SAVR and CVVR are both enabled, i.e., CoVR.

However, the cross-correlation between $\bar{Y}_{SA}$ and $\bar{Y}_{REF}$ embedded in $\hat{\mu}_{SACV}$ is different from that in (23) since Y and Y_{REF} are no longer I.I.D. random variables whenever SAVR creates the auto-correlations among the samples of Y (as well as Y_{REF}) as discussed in Section III. In other words, the optimal VR coefficient, $\hat{\mu}_{SACV}$, for CoVR cannot be implemented by (24) because (12) and (13) are not valid anymore, i.e., $Cov[\bar{Y}, \bar{Y}_{REF}] \neq Cov[Y, Y_{REF}]/N_{DCO}$. Under this situation, the optimal VR coefficient for CoVR shown in (28) has to be derived from the original definition in (20) with the cross-covariance in (10). Obviously, the complexity of (28) is mainly caused by the auto-correlations within Y_{SA} and Y_{REF} due to SAVR, so all $E[Y_{SA,n} \cdot Y_{REF,k}]$ terms cannot be simplified as $E[Y_{SA,n}] \cdot E[Y_{REF,k}]$ like (11). Therefore, although the approximate CoVR coefficient, $\hat{\mu}_{SACV}$, only needs a one-time high-resolution calculation as $\hat{\mu}_{CV}$ does in (24), it requires $2 \cdot (N_{DCO} - 1)$ of 1-bit AND-gates and $N_{DCO} \cdot (N_{DCO} - 1)$ of 1-bit DFFs to obtain all products of $(Y_{SA,n} \cdot Y_{REF,n-k})$ and $(Y_{REF,n} \cdot Y_{REF,n-k})$ in the DCO sampling rate. More dramatically, all 1-bit products need to be individually accumulated across the maximum N_{DCO} cycles through $2 \cdot (N_{DCO} - 1)$ counters. These unreasonable hardware requirements are all reflected by the 2nd terms of the numerator and denominator in (28).

To resolve this practical issue, the simulation results shown in Fig. 5 demonstrate the comparison between two different VR coefficient scenarios for the RSA with CoVR technique: the

first scenario is to ignore the auto-correlation induced by SAVR and use the hardware-friendly but suboptimal coefficient, $\hat{\mu}_{CV}$, in (24) anyway; the second is to use the practically impossible but optimal coefficient, $\hat{\mu}_{SACV}$, in (28) for comparison purposes. In the upper-half of Fig. 5, SAVR is enabled to initially suppress the variance by around 9 dB (i.e., $Var[\bar{Y}_{SA}]$ in blue dots with $Mod[T_{DCO,MIN}, T] \approx T/2$, and $\Delta T_{PRBS,MAX} \approx T/8$) from the I.I.D. RSA process (i.e., $Var[\bar{Y}]$ in blue curve with $\Delta T_{PRBS,MAX} \approx T$) at $\tau/T = 0.5$ and $N_{DCO} = 2^{14}$. Then, CVVR further suppresses the variance with τ_{REF}-options = 1, 3, and 7 in Fig. 5(a), 5(b), and 5(c), respectively. Each figure contains the two CoVR coefficient scenarios: the suboptimal coefficient, $\hat{\mu}_{CV}$ (red dots), and the optimal coefficient, $\hat{\mu}_{SACV}$ (black dots). The corresponding values of the VR coefficients shown in the lower-half of Fig. 5(a), 5(b), and 5(c) all have a certain amount of discrepancy between $\hat{\mu}_{CV}$ and $\hat{\mu}_{SACV}$ especially when τ moves away from τ_{REF}-options. All simulated variances in the upper-half of Fig. 5 show that the optimal VR coefficients (black dots) do not offer significant variance reductions compared to those offered by suboptimal VR coefficients (red dots). The reason is because CVVR is effective only when Y (= Y_{SA}) and Y_{REF} (or $\bar{Y}_{SA}$ and $\bar{Y}_{REF}$) have high cross-correlations regardless of whether Y itself is an I.I.D. random variable or not. In other words, the CVVR technique provides pronounced VR only when τ is approaching any one of the τ_{REF}-options as shown in Fig. 4(a) and the upper-half of Fig. 5; meanwhile, both $\hat{\mu}_{CV}$ and $\hat{\mu}_{SACV}$ are almost identical and converging to "1" as shown in Fig. 4(b) and the lower-half of Fig. 5. On the other hand, when τ is at the middle of two τ_{REF}-options, the efficiency of CVVR is very low anyway due to low cross-correlations between Y and Y_{REF}, so the peaks of the variance lobes in Fig. 5 offered by the optimal VR coefficients are very close (< 1 dB difference) to those offered by the suboptimal VR coefficients even though there are some deltas between $\hat{\mu}_{CV}$ and $\hat{\mu}_{SACV}$ when

$$\mu_{SACV} = \frac{Cov[\bar{Y}_{SA}, \bar{Y}_{REF}]}{Var[\bar{Y}_{REF}]} = \frac{\sum_{n=1}^{N_{DCO}} E[Y_{SA,n} \cdot Y_{REF,n}] + 2 \cdot \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} E[Y_{SA,n} \cdot Y_{REF,k}] - N_{DCO}^2 \cdot E[Y_{SA}] \cdot E[Y_{REF}]}{\sum_{n=1}^{N_{DCO}} E[Y_{REF,n}^2] + 2 \cdot \sum_{n=2}^{N_{DCO}} \sum_{k=1}^{n-1} E[Y_{REF,n} \cdot Y_{REF,k}] - N_{DCO}^2 \cdot E[Y_{REF}] \cdot E[Y_{REF}]}$$

$$\approx \frac{\sum_{n=1}^{N_{DCO}} Y_{SA,n} \cdot Y_{REF,n} + 2 \cdot \sum_{k=1}^{N_{DCO}-1} \sum_{n=k+1}^{N_{DCO}} Y_{SA,n} \cdot Y_{REF,n-k} - N_{DCO}^2 \cdot \bar{Y}_{SA} \cdot \bar{Y}_{REF}}{\sum_{n=1}^{N_{DCO}} Y_{REF,n}^2 + 2 \cdot \sum_{k=1}^{N_{DCO}-1} \sum_{n=k+1}^{N_{DCO}} Y_{REF,n} \cdot Y_{REF,n-k} - N_{DCO}^2 \cdot \bar{Y}_{REF}^2} = \hat{\mu}_{SACV} \qquad (28)$$

τ stays away from any of the τ_{REF}-options.

In sum, to perform the RSA with CoVR technique, the suboptimal VR coefficient, $\hat{\mu}_{CV}$ in (24), is sufficient and practical from both performance and implementation standpoints. Also, the numbers of τ_{REF}-options can be minimized in the CoVR mode because the majority of quantization-noise power is suppressed by SAVR across the DR of τ, and the peak of the variance lobes, i.e., the worst-case variances of the red and black dots, almost stay at the same level across all scenarios in Fig. 5. This means the coarse ADC-DAC hardware and pre-characterization costs for CVVR can be minimized by enabling SAVR.

B. Efficiency of Collaborative VR

Besides creating local variance notches at $\tau = \tau_{REF}$-options, there are at least three more benefits offered by CVVR. First, it allows the RSA-based TDC to maintain high measurement accuracy for a variety of time-interval DRs without heavily scaling the conversion gains of the analog circuits. For the examples in Fig. 6, if an RSA-based TDC is designed to maintain the same resolutions for different applications with the DR from Δt_{MAX} down to $\Delta t_{MAX}/8$, the total Δt-to-τ (or Δt_{MAX}-to-T) conversion gain offered by the analog circuits, including K_{TAC}, K_{VGA}, and K_{DL}, is required to be scalable from 1$\times$ to 8$\times$, so both Δt_{MAX} and $\Delta t_{MAX}/8$ can be mapped to the single period of CKτ, T, as shown in the lower-half and upper-left of Fig. 6, respectively. On the other hand, with the existing variance notches offered by CVVR around any one of the τ_{REF}-options, the total conversion gain of the analog circuit can stay the same as ($K_{TAC} \cdot K_{VGA} \cdot K_{DL}$) and map $\Delta t_{MAX}/8$ to T/8 instead of T as shown in the upper-right of Fig. 6. This T/8 duration can be shifted to around one of the three variance notches for example, τ_{REF1} = T/4, by simply providing a time-domain offset, τ_{OS} = 3$\cdot$T/16, when generating the falling edge of CKτ, so the DR of $\Delta t_{MAX}/8$ is converted into the duration between τ_{OS} and τ_{MAX} (= τ_{OS} + T/8 = 5$\cdot$T/16). Though τ_{OS} causes an offset, which can be pre-calibrated easily, the result, $\overline{Y}_{SACV}$, has extremely low quantization-noise power, Var[$\overline{Y}_{SACV}$], because of the local variance notch offered by CVVR.

Second, the primary downside of SAVR is the sensitivity to the circuit/signal variations, especially the values of Mod[$T_{DCO,MIN}$, T] and $\Delta T_{PRBS,MAX}$ as discussed in Section III. Though SAVR can maintain some tolerances of these values, adding CVVR can definitely offer further improvement in suppressing the sensitivity. For the example in Fig. 7(a), $T_{DCO,MIN}$, $\Delta T_{PRBS,MAX}$, and T are intentionally set as Mod[$T_{DCO,MIN}$, T] $\approx$ 7$\cdot$T/16 and $\Delta T_{PRBS,MAX} \approx$ T/4, where the SAVR technique can still perform VR but with inconsistent quantization-noise power, Var[$\overline{Y}_{SA}$] (blue dots), across the DR of τ because Mod[$T_{DCO,MIN}$, T] is now about 6.25% (T/16) lower than the range requirement (0.5$\cdot$T $\leq$ Mod[$T_{DCO,MIN}$, T] $\leq$ 0.75$\cdot$T) for achieving the maximum flattened variance distribution. By enabling CoVR with a simple 2-bit ADC-DAC circuit, the variance notches offered by CVVR at τ_{REF1} (= 0.25$\cdot$T) and τ_{REF3} (= 0.75$\cdot$T) greatly bring down the overall variance, Var[$\overline{Y}_{SACV}$] (red dots), and suppresses the variance peaks due to the improper setup (Mod[$T_{DCO,MIN}$, T] $\approx$ 7$\cdot$T/16) or variation

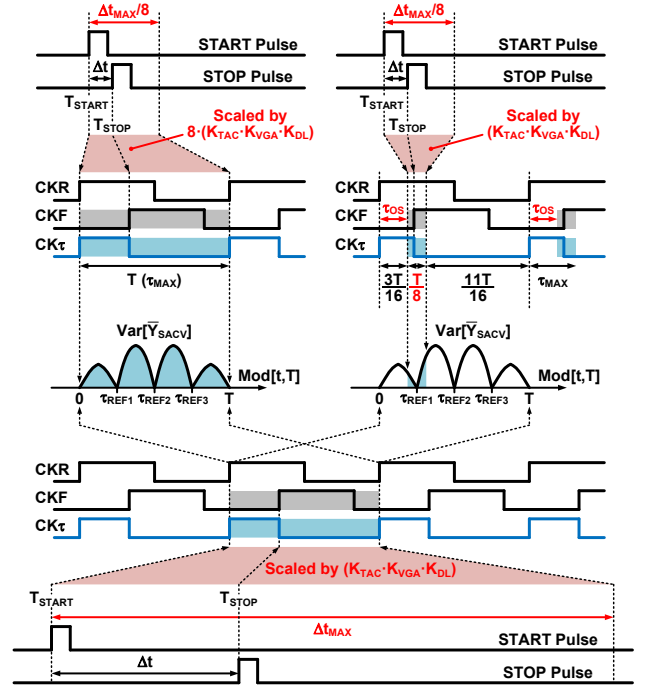Fig. 6. Lower-half: a regular conversion gain is designed to map Δt_{MAX} to T and to achieve an average variance requirement. Upper-left: an 8$\times$ amount of conversion gain is required to map $\Delta t_{MAX}/8$ to T and to maintain an average variance requirement. Upper-right: a regular conversion gain is sufficient to map $\Delta t_{MAX}/8$ to T/8, which can achieve even lower variance.

(6.25%) of $T_{DCO,MIN}$.

Third, in the case of maximizing the conversion-rate enhancement, $\Delta T_{PRBS,MAX}$ in SAVR has to be pushed down to the limit of the capacitor-size in CMOS process technology, such as T/128 or even smaller, while the SAVR rate of being a function of $|\tau/T - 0.5|$ gets more pronounced. As shown Fig. 7(b), when $\Delta T_{PRBS,MAX}$ is decreasing, the variance across $\tau/T \in$ [0, 1) is reduced in different rates, and the maximum rate is at τ/T = 0.5. Also, the variance curve by default (i.e., no VR) has the peak at τ/T = 0.5, so the variance at τ/T = 0.5 stays larger than the rest at $\tau/T \neq$ 0.5 when $\Delta T_{PRBS,MAX} >$ T/8; then the variance at τ/T = 0.5 becomes the local minimum when $\Delta T_{PRBS,MAX} <$ T/8 due to its fastest SAVR rate among all τ/T with the decrease of $\Delta T_{PRBS,MAX}$. Though the causes are different, the behavior of the variance lobes in Fig. 7(c) is similar to the case in Fig. 7(a), so this non-uniform quantization-noise power distribution can be again rectified by adding CVVR as shown in Fig. 7(c), where the average Var[$\overline{Y}_{SACV}$] across the entire DR of τ is about 3.5$\times$ lower than the average of Var[$\overline{Y}_{SA}$] when Mod[$T_{DCO,MIN}$, T] $\approx$ T/2, $\Delta T_{PRBS,MAX} \approx$ T/128, and τ_{REF}-options = 7.

VII. SUMMARY AND FUTURE WORK

The performance metrics of various RSA-based TDC techniques in [53] and this paper are summarized in Table I based on the behavioral-model and circuit simulations. Conventionally, the RSA technique possesses multiple advantages in high accuracy, calibratable linearity, scalable DR, high noise-immunity through 1-bit accumulators, low power/area, and simple read-out circuits, however, with the
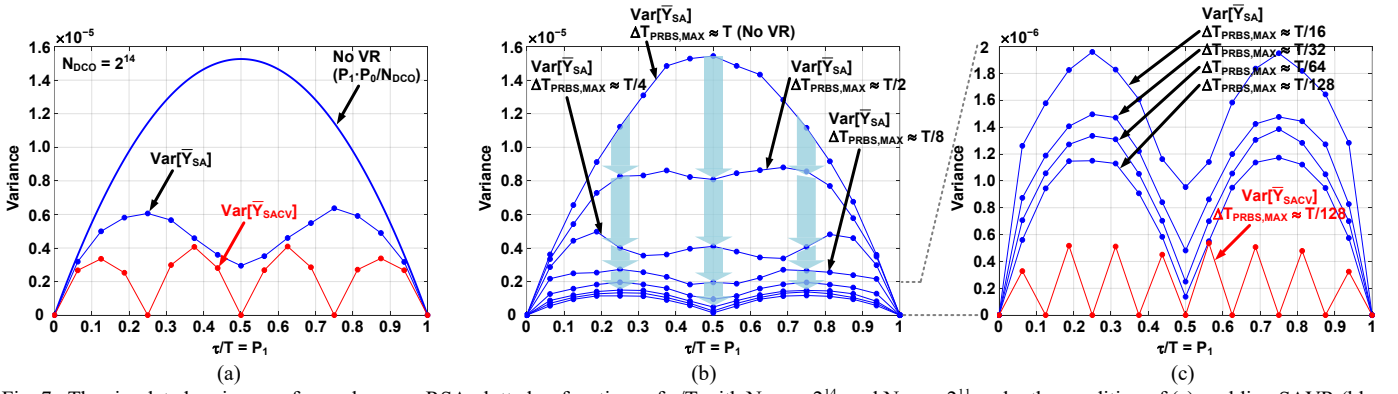
Fig. 7. The simulated variances of asynchronous RSA plotted as functions of τ/T with $N_{DCO} = 2^{14}$, and $N_{EXP} = 2^{11}$ under the condition of (a) enabling SAVR (blue dots) and CoVR (red dots) with $Mod[T_{DCO,MIN}, T] \approx 7 \cdot T/16$, $\Delta T_{PRBS,MAX} \approx T/4$, and τ_{REF}-options = 3; (b) enabling SAVR (blue dots) with $Mod[T_{DCO,MIN}, T] \approx T/2$, $\Delta T_{PRBS,MAX} \approx T$, T/2, T/4, T/8, T/16, T/64, and T/128; (c) enabling CoVR (red dots) with $Mod[T_{DCO,MIN}, T] \approx T/2$, $\Delta T_{PRBS,MAX} \approx T/128$, and τ_{REF}-options = 7.

downside of slow conversion-rates, which can be significantly mitigated by the VR techniques analyzed in this paper. The SAVR technique induces negative auto-covariance sums among the samples to improve the conversion-rate but causes sensitivity to the circuit/signal variations and non-uniform quantization-noise power across the DR. The CVVR technique requires the coarse quantization process in advance and then accomplishes the fine quantization process in real-time by paying the pre-characterization cost. Finally, the CoVR technique takes the complementary benefits from SAVR and CVVR to enhance the conversion-rate up to the range of 1.5 MHz with 12-ENOB accuracy and 0.36-pJ/step energy efficiency. Meanwhile, the total RSA-based TDC power (DCO + TDC) is roughly increased by 23% (> 7% + 14%) due to the additive and isolated hardware implementations of SAVR and

CVVR with extra clock distributions as shown in Fig. 1(a). In sum, the theoretical analyses, digitally implementable signal processing, and statistical verifications developed in [53] and this paper establish a concrete foundation for the forthcoming silicon-photonics RSA-based TDC realizations, which can open up a new path for both high-speed and high-resolution TCSCP applications in the emerging quantum technology.

APPENDIX

The calibration procedure of the RSA-based TDC contains two steps. In the pre-characterization process (1st step), the overall non-ideality is captured by injecting "known and individual" input signals (Δt) across the targeted DR from external pulse generators and collecting the corresponding digital outputs to form the input-output transfer curve of the

TABLE I
RSA-BASED TIME-INTERVAL DETECTION TECHNIQUE COMPARISON & SUMMARY

Work	TCAS-I'22 [53]	This Work	This Work	This Work
Technology for Simulations	22 nm	22 nm	22 nm	22 nm
Technique	Asyn. RSA	Asyn. RSA w/ SAVR	Asyn. RSA w/ CVVR	Asyn. RSA w/ CoVR
Sampling Clock Generator (DCO) Power	3 mW	3.1 mW	3 mW	3.1 mW
Sampling Frequency = $1/T_{DCO,AVG}$	4 GS/s	7.8 GS/s	4 GS/s	7.9 GS/s
Number of Sampling Phases = N_{PH} [53]	8	8	8	8
Dynamic Range (DR) scaled to T by TAC	10 ns < DR < 1 μs	10 ns < DR < 1 μs	10 ns < DR < 1 μs	10 ns < DR < 1 μs
* ENOB = $10 \cdot \log_{10}\left(\frac{P_1^2}{Var}\right)/6.02$ [53]	12 ENOB @ $N_{DCO} \approx 2^{24}$ 14 ENOB @ $N_{DCO} \approx 2^{28}$	12 ENOB @ $N_{DCO} \approx 2^{19}$ 14 ENOB @ $N_{DCO} \approx 2^{23}$	12 ENOB @ $N_{DCO} \approx 2^{20.9}$ 14 ENOB @ $N_{DCO} \approx 2^{24.9}$	12 ENOB @ $N_{DCO} \approx 2^{15.3}$ 14 ENOB @ $N_{DCO} \approx 2^{19.3}$
Effective Resolution (ER) = DR/2^{ENOB} [53]	0.61 ps < ER < 61 ps @ 14 ENOB	0.61 ps < ER < 61 ps @ 14 ENOB	0.61 ps < ER < 61 ps @ 14 ENOB	0.61 ps < ER < 61 ps @ 14 ENOB
Conversion-Rate (CVR) = $N_{PH}/(N_{DCO} \cdot T_{DCO,AVG})$ [53]	2 kHz @ 12 ENOB	120 kHz @ 12-ENOB	16 kHz @ 12-ENOB	1.5 MHz @ 12-ENOB
TDC Power	1.3 mW	1.5 mW	1.9 mW	2.2 mW
TDC FoM = (TDC Power)/($2 \cdot BW \cdot 2^{ENOB}$) = (TDC Power)/(CVR$\cdot 2^{ENOB}$)	159 pJ/step @ 12 ENOB	3.1 pJ/step @ 12 ENOB	29.0 pJ/step @ 12 ENOB	0.36 pJ/step @ 12 ENOB
DCO + TDC Power Ratio	1×	1.07×	1.14×	1.23×
CVR Ratio	1×	60×	8×	750×
TDC Area	0.01 mm²	0.01 mm²	0.018 mm²	0.018 mm²
Digital Filter (1-bit Acc.) Power	0.45 mW	0.90 mW	0.91 mW	1.81 mW
Digital Filter (1-bit Acc.) Outputs per RSA-based TDC process	N_Y, N_{DCO}	N_Y, N_{DCO}	N_Y, $N_{Y_{REF}}$, $N_{Y \cdot Y_{REF}}$, N_{DCO}	N_Y, $N_{Y_{REF}}$, $N_{Y \cdot Y_{REF}}$, N_{DCO}
One-time Multi-bit Digital Operations per RSA-based TDC Process	Eq. (6) in [53]	Eq. (3)	Eq. (17) & Eq. (24)	Eq. (25) & Eq. (24)
Theoretical Expectation	Eq. (15) in [53]	Eq. (3)	Eq. (18)	Eq. (26)
Theoretical Variance	Eq. (7) in [53]	Eq. (8)	Eq. (21)	Eq. (27)
Circuit Parameters @ T = 250 ps, $T_{DCO,MIN} \approx T/2$	$\Delta T_{PRBS,MAX} \approx T$	$\Delta T_{PRBS,MAX} \approx T/32$	$\Delta T_{PRBS,MAX} \approx T$, τ_{REF} LSB = T/16	$\Delta T_{PRBS,MAX} \approx T/128$, τ_{REF} LSB = T/8

* $P_1 = \tau/T \approx 0.5$, Var = Var[$\bar{Y}$] for ordinary RSA, Var = Var[$\bar{Y}_{SA}$] for RSA w/ SAVR, Var = Var[$\bar{Y}_{CV}$] for RSA w/ CVVR, Var = Var[$\bar{Y}_{SACV}$] for RSA w/ CoVR. When $P_1 \approx 0.5$, Var reaches the worst case under the I.I.D. random sampling condition [53].
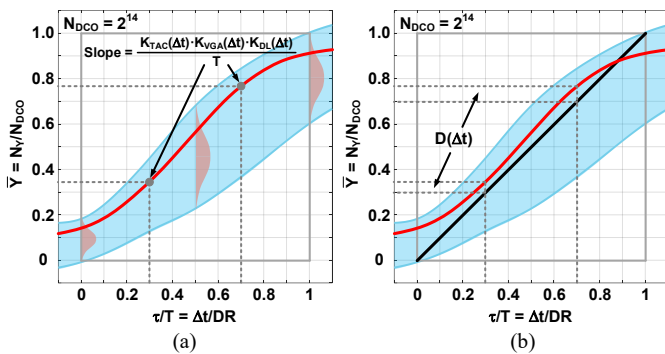
Fig. A1. 1000 input-output transfer curves of the RSA-based TDC distributed in the light-blue band obtained from Monte-Carlo simulations with 1000 time-invariant PVT conditions: (a) each time-invariant PVT condition possesses a unique transfer curve; (b) the input-output pairs on a certain transfer curve can be independently calibrated by their own $D(\Delta t)$ due to the random events of Δt.

entire TDC signal path. In Fig. A1, the light-blue band contains 1000 transfer curves obtained from the transistor-level Monte-Carlo simulations across the process-corners and random variations in a 22-nm CMOS process technology with constant temperature and analog/digital supplies. These 1000 transfer curves can be used to represent 1000 input-output relations of 1000 fabricated RSA-based TDC chips; i.e., each RSA-based TDC operating under a certain time-invariant PVT condition can possess a unique transfer curve within this light-blue band, e.g., the red curve in Fig. A1, which represents the time-invariant relation or mapping between an input Δt and its digital output ($\overline{Y} = N_Y/N_{DCO}$) of a specific RSA-based TDC. Meanwhile, the non-linearity (input-dependent gain error) of each transfer curve is mainly contributed by the analog circuits of TAC, VGA, and VCDL as shown in (6) and input-dependent slope equation in Fig. A1. More importantly, this non-linearity does not generate any harmonics and degrade the measured ENOB at all in the regular RSA-based TDC operation for TCSPC because the Δt detections from single photons are random events [1]. Thus, any input-output pair is independent from any other input-output pair even though they are on the same transfer curve. In sum, the 1st step is to obtain the transfer curve under a certain PVT condition, and then the non-linearity of this transfer curve can be corrected in the 2nd step.

In the regular TDC operation (2nd step), each digital output of a specific RSA-based TDC can be calibrated by its own transfer curve obtained from the 1st step under the same PVT condition. For the example in Fig. A1(b), the delta, $D(\Delta t)$, between the pre-characterized (red) and ideal (black) transfer curves at any certain Δt simply offers the information to simultaneously correct conversion offset and analog non-linearity. As mentioned, because of the independency among all received Δt and the individual RSA-based TDC process for each received Δt, the calibration by subtracting the $D(\Delta t)$ from the measured digital output is like the DC offset cancellation and can be done individually for each received Δt. Note that the accuracy requirement in the 1st step for generating the transfer curve or $D(\Delta t)$ must be higher than that in the 2nd step, which can be done by changing the setting of N_{DCO} in the RSA-based TDC for these two steps separately.

Overall, this calibration procedure only works effectively for the TDCs using 1-bit quantization, like RSA-based TDCs and 1-bit delta-sigma TDCs, because this approach can only correct the DC offset and analog non-linearity as described but not multi-level quantization mismatches, i.e., differential and integral non-linearities, which are avoided by the 1-bit quantizer in RSA-based TDCs.

REFERENCES

[1] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, Berlin, Germany: Springer, 2005.

[2] W. Becker, *The bh TCSPC Handbook*, 7th Ed., Berlin, Germany: Becker & Hickl GmbH, 2017.

[3] M. Wahl, "Time-correlated single photon counting," PicoQuant GmbH, Berlin, Germany, 2014. [Online]. Available: https://www.picoquant.com/images/uploads/page/files/7253/technote_tcspc.pdf

[4] N. A. W. Dutton *et al.*, "A time-correlated single-photon-counting sensor with 14GS/s histogramming time-to-digital converter," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, 2015, pp. 204–205.

[5] 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 in Quantum Electronics*, vol. 20, no. 6, pp. 364–373, Dec. 2014.

[6] C. Niclass *et al.*, "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.

[7] D. Tamborini *et al.*, "Compact, low-power and fully reconfigurable 10 ps resolution, 160 µs range, time-resolved single-photon counting system," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3827–3833, May 2016.

[8] M. Unternährer *et al.*, "Coincidence detection of spatially correlated photon pairs with a monolithic time-resolving detector array," *Opt. Express*, vol. 24, no. 25, pp. 28829–28841, Dec. 2016.

[9] 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.

[10] C. Veerappan *et al.*, "A 160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, 2011, pp. 312–313.

[11] R. M. Field *et al.*, "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, April 2014.

[12] C. Niclass *et al.*, "A 100-m range 10-frame/s 340 × 96-pixel time-of-flight depth sensor in 0.18-µm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2013.

[13] S. W. Hutchings *et al.*, "A reconfigurable 3-D-stacked SPAD imager with in-pixel histogramming for flash LIDAR or high-speed time-of-flight imaging," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 2947–2956, Nov. 2019.

[14] H. Seo *et al.*, "A 36-channel SPAD-integrated scanning LiDAR sensor with multi-event histogramming TDC and embedded interference filter," in *IEEE Symp. on VLSI Circuits*, Honolulu, HI, USA, 2020, pp. 1–2.

[15] P. Padmanabhan *et al.*, "A 256 × 128 3D-stacked (45nm) SPAD FLASH LiDAR with 7-Level coincidence detection and progressive gating for 100m Range and 10klux background light," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, 2021, pp. 112–113.

[16] 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 Tran. Circuits and Syst. I: Reg. Papers*, vol. 60, no. 3, pp. 557–569, March 2013.

[17] K. Karadamoglou *et al.*, "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.

[18] Y. Liu *et al.*, "Multi-stage pulse shrinking time-to-digital converter for time interval measurements," in *Proc. Eur. Microwave Integrated Circuits Conf.*, Munich, Germany, 2007, pp. 267–270.

[19] Y. Arai and M. Ikeno, "A time digitizer CMOS gate-array with a 250 ps time resolution," *IEEE J. Solid-State Circuits*, vol. 31, no. 2, pp. 212–220, Feb. 1996.

[20] L. Vercesi *et al.*, "Two-dimensions Vernier time-to-digital converter," *IEEE J. Solid-State Circuits*, vol. 45, no. 8, pp. 1504–1512, Aug. 2010.

[21] P. Dudek, S. Szczepanski, and J. V. Haltfield, "A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 240–247, Feb. 2000.

[22] M. Zanuso *et al.*, "Time-to-digital converter for frequency synthesis based on a digital bang-bang DLL," *IEEE Trans. Circuits and Syst. I: Reg. Papers*, vol. 57, no. 3, pp. 548–555, Mar. 2010.

[23] A. Mantyniemi *et al.*, "A CMOS time-to-digital converter (TDC) based on a cyclic time domain successive approximation interpolation method," *IEEE J. Solid-State Circuits*, vol. 44, no. 11, pp. 3067–3078, Nov. 2009.

[24] J. P. Jansson, A. Mäntyniemi, and J. Kostamovaara, "A CMOS time-to-digital converter with better than 10 ps single-shot precision," *IEEE J. Solid-State Circuits*, vol. 41, no. 6, pp. 1286–1296, June 2006.

[25] B. K. Swann *et al.*, "A 100-ps time-resolution CMOS time-to-digital converter for positron emission tomography imaging applications," *IEEE J. Solid-State Circuits*, vol. 39, no. 11, pp. 1839–1852, Nov. 2004.

[26] E. Temporiti *et al.*, "A 3.5 GHz wideband ADPLL with fractional spur suppression through TDC dithering and feedforward compensation," *IEEE J. Solid-State Circuits*, vol. 45, no. 12, pp. 2723–2736, Dec. 2010.

[27] M. Beck, "Comparing measurements of $g^{(2)}(0)$ performed with different coincidence detection techniques," *J. Optical Society of America B*, vol. 24, no. 12, pp. 2972–2978, Dec. 2007.

[28] M. Beck, *Quantum Mechanics – Theory and Experiment*, New York, NY, USA: Oxford University Press, 2012.

[29] U. Leonhard, *Measuring the Quantum State of Light*, Cambridge, UK: Cambridge University Press, 1997.

[30] C. Spee, J. I. de Vicente, B. Kraus, "Remote entanglement preparation," *Phys. Rev. A*, vol. 88, 010305, July 2013.

[31] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states," *Phys. Rev. Lett.*, vol. 68, pp. 3121–3124, May 1992.

[32] B. Slutsky *et al.*, "Security of quantum cryptography against individual attacks," *Phys. Rev. A*, vol. 57, pp. 2383–2398, April 1998.

[33] M. Hillery, V. Bužek, and André Berthiaume, "Quantum secret sharing," *Phys. Rev. A*, vol. 59, no. 3, pp. 1829–1834, March 1999.

[34] D. Kepshire *et al.*, "A microcomputed tomography guided fluorescence tomography system for small animal molecular imaging," *Rev. Sci. Instrum.*, vol. 80, no. 4, 043701, April 2009.

[35] S. Coda *et al.*, "Fluorescence lifetime spectroscopy of tissue autofluorescence in normal and diseased colon measured ex vivo using a fiber-optic probe," *Biomed. Opt. Express*, vol. 5, no. 2, pp. 515–538, Feb. 2014.

[36] L. Marcu, P.M.W. French, and D.S. Elson, *Fluorescence Lifetime Spectroscopy and Imaging: Principles and Applications in Biomedical Diagnostics*, Boca Raton, FL, USA: CRC Press, 2015.

[37] M. A. Abas, G. Russell, and D. J. Kinniment, "Embedded high-resolution delay measurement system using time amplification," *IET Computers & Digital Techniques*, vol. 1, no. 2, pp. 77–86, Mar. 2007.

[38] M. Lee and A. A. Abidi, "A 9 b, 1.25 ps resolution coarse-fine time-to-digital converter in 90 nm CMOS that amplifies a time residue," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 769–777, Apr. 2008.

[39] J. Kalisz, "Review of methods for time interval measurements with picosecond resolution," *Metrologia*, vol. 41, no. 1, pp. 17–32, Feb. 2004.

[40] S.-K. Shin *et al.*, "A 12 bit 200 MS/s zero-crossing-based pipelined ADC with early sub-ADC decision and output residue background calibration," *IEEE J. Solid-State Circuits*, vol. 49, no. 6, pp. 1366–1382, June 2014.

[41] S.-H. W. Chiang *et al.*, "A 10-bit 800-MHz 19-mW CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 935–949, April 2014.

[42] Y. Wu *et al.*, "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 and Syst. I: Reg. Papers*, vol. 67, no. 8, pp. 2532–2545, Aug. 2020.

[43] 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, June 2018.

[44] 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.

[45] Y. Cao, W. De Cock, M. Steyaert, and P. Leroux, "1-1-1 MASH ΔΣ time-to-digital converters with 6 ps resolution and third-order noise-shaping," *IEEE J. Solid-State Circuits*, vol. 47, no. 9, pp. 2093–2106, Sept. 2012.

[46] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, New York, NY, USA: Springer, 2003.

[47] Z. Botev *et al.*, *Variance Reduction*, New York, NY, USA: Wiley, 2017.

[48] E. Alon, V. Stojanovic, and M. Horowitz, "Circuits and techniques for high-resolution measurement of on-chip power supply noise," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 820–828, Apr. 2005.

[49] E. Alon *et al.*, "On-die power supply noise measurement techniques," *IEEE Trans. Advanced Packaging*, vol. 32, no. 2, pp. 248–259, May 2009.

[50] M. Mansuri, B. K. Casper, F. O'Mahony, "An on-die all-digital delay measurement circuit with 250fs accuracy," in *IEEE Symp. on VLSI Circuits*, Honolulu, HI, USA, 2012, pp. 98–99.

[51] T.-C. Hsueh *et al.*, "A 25.6Gb/s differential and DDR4/GDDR5 dual-mode transmitter with digital clock calibration in 22nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, 2014, pp. 444–445.

[52] T.-C. Hsueh *et al.*, "An on-die all-digital power supply noise analyzer with enhanced spectrum measurements," *IEEE J. Solid-State Circuits*, vol. 50, no. 7, pp. 1711–1721, July 2015.

[53] T. Wu *et al.*, "Random sampling-and-averaging techniques for single-photon arrival-time detections in quantum applications: theoretical analysis and realization methodology," *IEEE Trans. Circuits and Syst. I: Reg. Papers*, vol. 69, no. 4, pp. 1452–1465, April 2022.

[54] E. Säckinger, *Analysis and Design of Transimpedance Amplifiers for Optical Receivers*, Hoboken, NJ, USA: Wiley, 2018.

[55] S. Haykin, *Communication Systems*, New York, NY, USA: Wiley, 2001.

[56] B. Razavi, *Design of CMOS Phase-Locked Loops from Circuit Level to Architecture Level*, New York, NY, USA: Cambridge, 2020.

[57] S. Ghahramani, *Fundamentals of Probability*, Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[58] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[59] G. Upton and I. Cook, *A Dictionary of Statistics*, 3rd ed., Oxford, UK: Oxford, 2014.
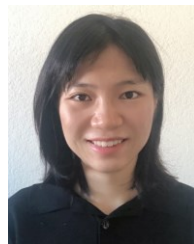
Ruoman Yang (Graduate Student Member, IEEE) received the B.S. degree in microelectronics from the University of Electronic Science and Technology of China, Chengdu, China, in 2017 and the M.S. degree in electrical and computer engineering from the University of California, San Diego, CA, in 2019, where she is currently pursuing the Ph.D. degree.

Her research interests include analog/mixed-signal integrated circuits, especially electrical/optical wireline transceivers.

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

His research interests 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, 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 in Intel Lab Signaling Research and an Analog Engineer in Intel I/O Circuit Technology, Hillsboro, Oregon. Since 2018, he has been an Assistant Professor in electrical and computer engineering at the University of California, San Diego. 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 (JSSC) Best Paper Award, and the 2020 NSF CAREER Award. He served on the Patent Committee for Intel Intellectual Property (Intel IP) and the Technical Committee for Intel Design & Test Technology Conference (DTTC) from 2016 to 2018. Since 2018, he has served on the Technical Program Committee for IEEE Custom Integrated Circuits Conference (CICC) and the Guest Associate Editor for IEEE Solid-State Circuits Letters (SSC-L).