# Distortion Analysis of MOS Track-and-Hold Sampling Mixers Using Time-Varying Volterra Series

Wei Yu, Subhajit Sen, and Bosco H. Leung, *Senior Member, IEEE*

*Abstract*—A time-varying theory of Volterra series is developed and applied in the sampled-data domain to solve for harmonic and intermodulation distortion of a MOS-based track-and-hold sampling mixer with a nonzero fall-time LO waveform. Distortion due to sampling error is also calculated. These results, when combined with the continuous-time solution, quantify harmonic and intermodulation distortion of a track-and-hold type mixer completely. Closed-form solutions are obtained. As a practical consequence, it is shown that for certain fall-time, the distortion of track-and-hold mixers can be better than what would be predicted by a simple application of time-invariant Volterra series theory.

*Index Terms*— Harmonic distortion, intermodulation distortion, intermediate frequency (IF), radio frequency (RF), time-varying Volterra series, track-and-hold sampling mixer.

## I. INTRODUCTION

INCREASING demand for digital wireless personal communication devices, and steady improvements in the MOS device performance of scaled VLSI technologies, have allowed the feasibility of performing A/D conversion on radio-frequency (RF) signals directly or on their mixed down (IF) versions [1], [2]. Early A/D conversion of the received signal in a receiver chain pushes front-end signal processing functions such as channel-selection filtering and automatic gain control (AGC) into the digital domain, and thus helps to improve tolerance against process variations inherent in analog dominated front-end architectures and therefore the overall performance of the receiver. The key element crucial for good distortion performance of an A/D converter that digitizes RF signals is the track-and-hold circuit at the input of the A/D converter. This is because unlike a conventional baseband A/D converter, the track-and-hold circuit in an RF A/D application is inherently a mixer that mixes the RF signal down to baseband or a lower IF frequency in the process of continuous-time to discrete-time conversion. Compared to the conversion of a band-limited signal in a conventional Nyquist-rate A/D converter, the mixing down process in a narrow-band VHF or UHF signal is much more nonlinear with nonlinear products increasing with input RF frequency. Linearity is crucial because the presence of adjacent channel interfering signals at the input generates spurious components in the desired channel band. The track-and-hold mixer also differs from a conventional mixer (such as a Gilbert multiplier) in that the frequency of the sampling clock (which serves as the equivalent of a local oscillator) may be a submultiple of the local oscillator (LO) frequency in a conventional mixer. In a subsampling track-and-hold mixer, the $m$th harmonic of a sampling-clock with fundamental frequency $f_s$ may be mixed down with a carrier-tone at frequency $f_o$ to generate a baseband tone at frequency $(f_o - mf_s)$.

In this paper, the harmonic and intermodulation distortion of MOS track-and-hold mixers is thoroughly analyzed and related with simulation results. It is shown that a complete picture of distortion can be obtained by applying Volterra series analysis, hitherto applied only to continuous-time time-invariant nonlinear systems, to time-varying circuits that are observed every clock period in the discrete-time domain such as in the case of a track-and-hold circuits in A/D converters. We have done this by developing a theory of Volterra analysis for time-varying systems valid under the assumption that the time-constant of the circuit is much smaller than the sample clock period. We have demonstrated that the incorporation of the time-varying analysis to Volterra series is essential for an accurate estimation of distortion for arbitrary fall-time in the sampling clock waveform. The closed-form formulas we obtained are useful in providing physical insights to the circuit designer. Successful applications which used these formulas in designing the front-end sampling mixer have been reported, for example, in an 150-MHz 13-bit 12-mW A/D converter in [3], and in a 400-MHz 12-bit 18-mW A/D converter in [4].

The rest of the paper is organized as follows. Section II explains the basic operation of the track-and-hold circuit as a mixer under the assumption of zero fall-time LO waveform. Expressions for harmonic distortion ($HD_2$ and $HD_3$) and intermodulation ($IM_2$ and $IM_3$) are derived using Volterra Series. The difficulty of extending this result to an arbitrary gate waveform is explained. Section III develops the theory of time-varying Volterra Series and shows that it is possible to extend the theory to time-varying systems by assuming that the output is observed in the sampled-date domain. Expressions for time-varying distortion are derived

W. Yu was with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., N2L 3G1, Canada. He is now with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: weiyu@stanford.edu).

S. Sen was with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., N2L 3G1, Canada. He is now with Arcus Technology Limited in Bangalore, India (e-mail: sen@arcustech.com).

B. H. Leung is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., N2L 3G1, Canada (e-mail: bleung@sun14.vlsi.uwaterloo.ca).

Fig. 1.  A track-and-hold mixer.



Fig. 2.  Mixing distortion using two tones.
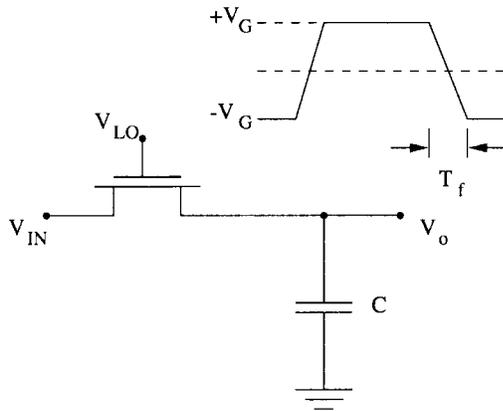
for the case of a simple track-and-hold mixer. In Section IV, sampling distortion is analyzed and expressions for harmonic and intermodulation distortion are derived. Section V describes SPICE simulation results for the simple track-and-hold mixer. These results are then explained using the theoretical results on the continuous-time, time-varying, and sampling distortion. Section VI analyzes the practical scheme of bottom-plate sampling using the theory developed. Conclusions are drawn in Section VII.

## II. TRACK-AND-HOLD SAMPLING MIXER

### A. Mixer Operation

A single switch track-and-hold sampling mixer consists of a MOS transistor followed by a sampling capacitor (Fig. 1). The input $V_{in}$ is at RF or IF frequency. As $V_{LO}$ applied at the gate goes high, the output $V_o$ tracks the input; as $V_{LO}$ goes low, the output $V_o$ is sampled and held on the capacitor. The output is considered in the sampled-data domain.

Before analyzing distortion, let us first define the terminology used in this paper. Suppose that a single pure sinusoidal input $\sin(\omega t)$ is applied at the input of the circuit, in this case to $V_{in}$, we define the harmonic distortion of $n$th order to be the magnitude of the $\sin(n\omega t)$ term at the output $V_o$. If two tones are applied at the input, frequencies at the sum and difference of input frequencies are present at the output. So, two different kinds of second order intermodulation (IM$_2$) are possible. For an input of $\sin(\omega_1 t) + \sin(\omega_2 t)$, IM$_2^+$ is defined as the magnitude of $\sin((\omega_1 + \omega_2)t)$ term at the output, and IM$_2^-$ that of $\sin((\omega_1 - \omega_2)t)$ term. Likewise, intermodulation of third order (or IM$_3$) is defined as the magnitude of $\sin((2\omega_1 - \omega_2)t)$ term when an input of $\sin(\omega_1 t) + \sin(\omega_2 t)$ is applied. The physical origin of HD$_2$ and IM$_2^+$ is the same, and HD$_2$ is related to IM$_2^+$ by a factor of 2. So, we are more interested in IM$_2^-$ than IM$_2^+$. In the following analysis, unless explicitly specified, IM$_2$ refers to IM$_2^-$.

For example, in the case that the RF input contains two tones at frequencies $f_{RF1}$ (say at 100.3 MHz) and $f_{RF2}$ (say at 100.4 MHz), respectively, with a sampling clock frequency $f_s$ at 100 MHz. The components visible in the baseband are shown in Fig. 2. The 100-kHz component is the IM$_2^-$ component at $f_{RF1} - f_{RF2}$. The 200-kHz component is due to IM$_3$ at
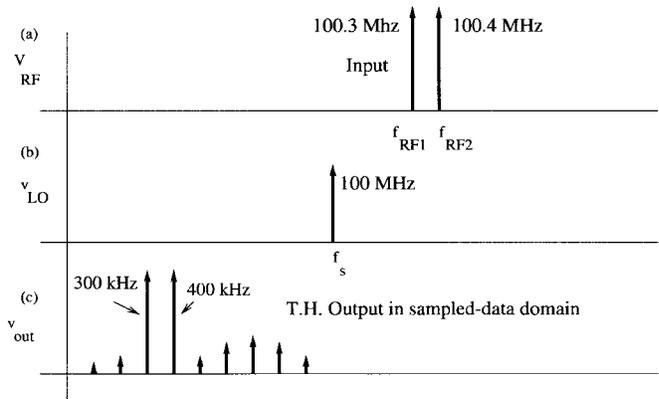
$2f_{RF2} - f_{RF2}$ (200.2 MHz) mixed down to baseband. The component at 700 kHz is the IM$_2^+$ component at $f_{RF1} + f_{RF2}$ (200.7 MHz) mixed down. We also observe HD$_2$ components at 600 and 800 kHz as well as HD$_3$ component at 900 kHz. We may note that by using a differential architecture, the even order distortion (e.g., HD$_2$, IM$_2$) may be made considerably smaller (by the mismatch factor) than in a single-ended architecture.

A rigorous distortion analysis of the simple track-and-hold circuit with arbitrary input frequency and arbitrary LO waveform is difficult for the following reasons.

- The relation between drain current and terminal voltages is nonlinear.
- The body effect and bias-dependent junction capacitors produce additional nonlinearities.
- The circuit is time varying due to the modulation of drain-source conductance by the applied gate LO waveform.
- The precise instant of sampling depends not only on the LO waveform but also on the input amplitude and frequency resulting in sampling distortion.

Previous work on track-and-hold distortion focused on its application in baseband A/D converters [5], [6]. These have mainly tackled the distortion problem of a CMOS switch by isolating the distortion into the "tracking error" and the "aperture error" components with low-frequency single-tone inputs. They used assumptions that are not valid for track-and-hold mixers applications, the most important of which is the circuit time-invariant assumption. In [7], track-and-hold type mixers for continuous-time output using Volterra Series approach were analyzed. Again, the analysis did not take into account the sampling and the time-varying nature of the circuit due to periodic but arbitrary gate LO waveforms.

In the following, we propose to analyze distortion using time-varying Volterra series. To simplify the problem, we will abstract the most essential features of MOS track-and-hold distortion by excluding the nonlinear bias-dependent components of junction capacitors and body effect from the MOS transistor model. Further analysis incorporating these second order effects results in a slight modification of only a few decibels. As a first step, we will solve for mixer distortion in the continuous-time mode, which corresponds to the case where LO waveform has a sharp cutoff. This also serves
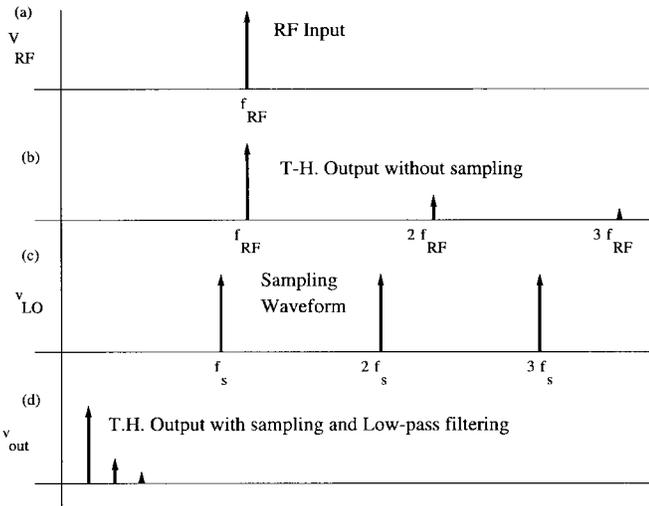
Fig. 3. Mixing using ideal sampling.

to illustrate the essential steps in applying Volterra series in distortion analysis.

### B. Mixer Using Ideal Sampling

The mixer operation under ideal sampling can be approximated by the following simplified model. We assume that a constant high voltage is applied at the gate, an RF signal at frequency $f_{RF}$ is applied at the input, and that the output is sampled using an ideal sampling process. Since the circuit is nonlinear, we expect the input signal to be attenuated by the gain of the network along with harmonic components at $nf_{RF}$ where $n = 2, 3, \cdots$, as shown in Fig. 3(b). The ideal sampling process followed by an ideal low-pass filter with a cutoff frequency at $f_s/2$ is then applied to the output. The effect is to replicate the output spectrum by shifting it by $nf_s$, where $n = 1, 2, \cdots$, so that a shifted component of output is mixed down to near DC. This is shown in Fig. 3(d). We note that harmonic components at $nf_{RF}$ now reappear near DC at $nf_{RF} - nf_s$ and close to $f_{RF} - f_s$ in the baseband. We may also note that it is possible to mix down the output to the same positions in the baseband by using a sampling clock with a frequency that is a subharmonic of $f_s$ (i.e., subsampling). The observation is that the harmonic distortion of the mixer without sampling [i.e., Fig. 3(b)] is the same as the harmonic distortion of the sampled signal in the baseband [i.e., Fig. 3(d)] if ideal sampling is assumed. Therefore, under the ideal sampling condition, harmonic distortion can be calculated as if sampling has not occurred, and the system is operating in the continuous-time mode.

The mixing process with ideal sampling just described may now be used to approximate a track-and-hold circuit at whose gate an ideal square-wave is applied. When the gate voltage is at $V_G$, the MOS transistor is in the triode region and the output tracks the input voltage. When the MOS turns OFF instantaneously, as in the case of an ideal square-wave gate voltage, the tracked output voltage is held onto the capacitor. Assuming that the time-constant of the circuit is much smaller than the gate ON period, the output voltage reaches a steady-state value within several time-constants after the MOS turns

ON. (This is true even for high-frequency RF if a subsampling technique is used with a sufficient subsampling factor.) Under this condition, the distortion components seen in the discrete-time frequency domain are the same as that for the case when the track-and-hold is working in the continuous-time mode with the gate voltage at $V_G$ and the MOS transistor stays ON. Therefore, to calculate distortion components for the ideal sampling case, we only need to analyze distortion for the continuous-time mode, which can be done using the time-invariant Volterra series theory.

### C. Volterra Series and Harmonic Distortion

Under some general continuity conditions, a time-invariant system may be expanded into the sum of a linear term, a second-order term, and a third-order term, etc. [8]. Mathematically, if the input is $x(t)$, the output $y(t)$ is represented by a series of the type,

$$
\begin{aligned}
y(t) = &\int_{-\infty}^{\infty} h_1(t - \tau_1)x(\tau_1)\, d\tau_1 \\
&+ \iint_{-\infty}^{+\infty} h_2(t - \tau_1, t - \tau_2)x(\tau_1)x(\tau_2)\, d\tau_1\, d\tau_2 \\
&+ \iiint_{-\infty}^{+\infty} h_3(t - \tau_1, t - \tau_2, t - \tau_3) \\
&\cdot x(\tau_1)x(\tau_2)x(\tau_3)\, d\tau_1\, d\tau_2\, d\tau_3 + \cdots
\end{aligned}
\tag{1}
$$

whose convergence is uniform. More succinctly, we write,

$$
y(t) = \boldsymbol{H_1}[x(t)] + \boldsymbol{H_2}[x(t)] + \boldsymbol{H_3}[x(t)] + \cdots
\tag{2}
$$

where $\boldsymbol{H_n}$ represents the $n$th order operator with kernel $h_n$. This series is called the Volterra series. It can be shown that Volterra kernels form a basis for sufficiently well-behaved systems, i.e., a mildly nonlinear system can be expanded in a Volterra series in one and only one way.

Expanding a system in Volterra series enables us to find its frequency response and hence its sinusoidal response, so Volterra series is directly related to harmonic distortion [9]. For example, the response to a sinusoidal input for a second-order system contains a term at twice the input frequency and a term at DC. The magnitude of each term is related to the corresponding terms in the Volterra series. More specifically, for an input of $A \cdot \cos(\omega_0 t)$, the output is

$$
\begin{aligned}
y(t) = &\left(\frac{A}{2}\right)^2 H_2(j\omega_0, j\omega_0)e^{2j\omega_0 t} \\
&+ \left(\frac{A}{2}\right)^2 H_2(-j\omega_0, -j\omega_0)e^{-2j\omega_0 t} \\
&+ \left(\frac{A}{2}\right)^2 H_2(j\omega_0, -j\omega_0) + \left(\frac{A}{2}\right)^2 H_2(-j\omega_0, j\omega_0).
\end{aligned}
$$

The third-order term can be derived in a similar way. Following the definition for harmonic distortion, harmonic distortion is related to Volterra kernels in the following way:

$$
\mathrm{HD}_2 = \frac{A}{2} \cdot \frac{H_2(\omega, \omega)}{H_1(\omega)}
\tag{3}
$$

$$
\mathrm{HD}_3 = \frac{A^2}{4} \cdot \frac{H_3(\omega, \omega, \omega)}{H_1(\omega)}.
\tag{4}
$$

Similarly, intermodulation can be quantified as well:

$$\text{IM}_2 = A \cdot \frac{H_2(\omega, -\omega)}{H_1(\omega)} \tag{5}$$

$$\text{IM}_3 = \frac{3A^2}{4} \cdot \frac{H_3(\omega, \omega, -\omega)}{H_1(\omega)}. \tag{6}$$

Therefore, obtaining the Volterra series coefficients is the key in calculating harmonic distortion and intermodulation.

### D. MOS Mixer Volterra Coefficients

We now analyze harmonic distortion of an ideal sampling mixer using the continuous-time model by calculating its Volterra coefficients. First, it can be shown by a Volterra analysis that at sufficiently low frequencies the distortion is mainly determined by the nonlinear source-to-drain $I$–$V$ characteristic of the MOS transistor and that distortion due to bias-dependent junction capacitance is quite small for typical values of MOS model parameters. We also choose to ignore body effect for simplicity in analysis, but note that this choice modifies neither our method of distortion analysis nor the significant conclusions regarding distortion in MOS track-and-hold mixers. Further, the only parasitic capacitance of importance is the junction capacitance at the output side whose effect is to increase the value of the sampling capacitor. Therefore, in deriving the Volterra series coefficient, we can assume the configuration in Fig. 1 and use the most basic MOS equation

$$I_D = K(V_{GS} - V_t)V_{DS} - \frac{K}{2} V_{DS}^2 \tag{7}$$

where $K = \mu C_{ox} W/L$, and $V_t$ is the threshold voltage. Either the input or the output side may be regarded as source. The system differential equation thus becomes

$$gV_o + C \frac{dV_o}{dt} = gV_{\text{in}} - \frac{K}{2}(V_{\text{in}}^2 - V_o^2) \tag{8}$$

where $g = K(V_G - V_t)$. Express $V_o(t)$ in its Volterra series

$$V_o(t) = v_1(t) + v_2(t) + v_3(t) + \cdots$$

where $v_n(t) = \boldsymbol{H_n}[V_{\text{in}}(t)]$, substitute the expansion into (8), and keep only the first three order terms, so we get

$$g(v_1 + v_2 + v_3) + C \frac{d}{dt}(v_1 + v_2 + v_3)$$
$$= gV_{\text{in}} - \frac{K}{2}(V_{\text{in}}^2 - v_1^2 - 2v_1 v_2).$$

Note, $V_{\text{in}}$ is first order since the input is a pure sinusoidal; the term $v_1^2$ is second order; and $v_1 v_2$ is third order. Now, equating the coefficients, we have

$$\begin{cases} gv_1 + C \dfrac{dv_1}{dt} = gV_{\text{in}} \\[2mm] gv_2 + C \dfrac{dv_2}{dt} = \dfrac{K}{2}(v_1^2 - V_{\text{in}}^2) \\[2mm] gv_3 + C \dfrac{dv_3}{dt} = K v_1 v_2. \end{cases} \tag{9}$$

The solutions to this set of differential equations are the Volterra coefficients $H_1$, $H_2$, and $H_3$. Since we assume a constant gate voltage, hence a constant $g$, we can show that

$$H_1(\omega_1) = \frac{g}{g + j\omega_1 C} = 1 \tag{10}$$

$$H_2(\omega_1, \omega_2) = \frac{\dfrac{K}{2}(H_1(\omega_1)H_1(\omega_2) - 1)}{g + j(\omega_1 + \omega_2)C}$$
$$= -\frac{j(\omega_1 + \omega_2)C}{2K(V_G - V_t)^2} \tag{11}$$

$$H_3(\omega_1, \omega_2, \omega_3) = \frac{K \cdot H_1(\omega_1) \cdot H_2(\omega_2, \omega_3)}{g + j(\omega_1 + \omega_2 + \omega_3)C}$$
$$= \frac{-j(\omega_1 + \omega_2 + \omega_3)C}{3K(V_G - V_t)^3} \tag{12}$$

where the assumption of small time-constant ($g \gg j\omega C$) is used, and also the expressions are symmetrized. Substituting (10)–(12) into (3)–(6) yields the harmonic distortion and intermodulation

$$\text{HD}_2 = \frac{A}{2} \cdot \frac{j\omega C}{K(V_G - V_t)^2} \tag{13}$$

$$\text{HD}_3 = \frac{A^2}{4} \cdot \frac{j\omega C}{K(V_G - V_t)^3} \tag{14}$$

$$\text{IM}_2 = 0 \tag{15}$$

$$\text{IM}_3 = \frac{A^2}{4} \cdot \frac{j\omega C}{K(V_G - V_t)^3}. \tag{16}$$

It must be emphasized that the above analysis assumes a constant gate voltage. This applies to the case of perfect square LO sampling if the system time-constant is much smaller than the sampling period. However, if the sampling-clock voltage deviates from the ideal square wave, two effects start to emerge which will invalidate the continuous-time model. If the gate voltage falling edge is not instantaneous, the approximate time-constant of the circuit, as given by

$$\tau = \frac{C}{g} = \frac{C}{K(V_{GS} - V_t)},$$

gradually increases since the MOS ON resistance increases as the MOS leaves the triode region. In addition, the precise instant of sampling depends not only on the gate falling edge slew-rate waveform but also on the input amplitude and frequency, which results in sampling distortion. This leads to a situation where the final sampled voltage on the sampling capacitor when the MOS transistor turns OFF deviates from the steady-state voltage on the capacitor predicted by Volterra series theory when the MOS transistor is ON. Therefore, the sampled voltage can no longer be predicted with a time-invariant theory. This leads us to develop the theory of the time-varying Volterra Series.

### III. TIME-VARYING VOLTERRA SERIES

As mentioned in the previous section, it is necessary to generalize the Volterra series to time-varying systems in order to analyze the track-and-hold mixer with an arbitrary LO waveform. In this section, we develop a time-varying Volterra

series, explore its frequency response, and show that the time-varying Volterra series can be applied in the sampled-data domain to solve for time-varying distortion exactly.

### A. Time-Varying Systems

The notion of impulse response can be generalized to time-varying systems by adding another time variable [10]. The impulse response of a linear time-varying system, $h_1(t, \tau)$, is a function of two time variables, where $t$ is the observation time and $\tau$ the launch time. $h_1(t, \tau)$ represents the output observed at time $t$ for an impulse input launched at time $\tau$. The system response for an arbitrary input can be found from $h_1(t, \tau)$ using the following:

$$y_1(t) = \int_{-\infty}^{\infty} h_1(t, \tau) x(\tau) \, d\tau. \tag{17}$$

Under general continuity conditions, a linear time-varying system is completely characterized by its two-dimensional impulse response, otherwise known as the kernel.

Higher order kernels are generalized in a similar way. For example, a second order kernel has two launch time variables and an observation time variable. $h_2(t, \tau_1, \tau_2)$ is the system response to two impulses launched at time instants $\tau_1$ and $\tau_2$. In general, a mildly nonlinear time-varying system has the following Volterra series expansion:

$$
\begin{aligned}
y(t) = & \int_{-\infty}^{\infty} h_1(t, \tau_1) x(\tau_1) \, d\tau_1 \\
& + \iint_{-\infty}^{\infty} h_2(t, \tau_1, \tau_2) x(\tau_1) x(\tau_2) \, d\tau_1 \, d\tau_2 \\
& + \iiint_{-\infty}^{\infty} h_3(t, \tau_1, \tau_2, \tau_3) \\
& \cdot x(\tau_1) x(\tau_2) x(\tau_3) \, d\tau_1 \, d\tau_2 \, d\tau_3 + \cdots. \tag{18}
\end{aligned}
$$

Recall that in the linear time-invariant case, the impulse response may be found in either time domain or frequency domain. The frequency domain solution is often much easier to obtain than the time domain impulse response. However, for time-varying systems, frequency response is no longer well defined and, hence, it is necessary to directly apply time domain impulses at the system input to find the output expression. This amounts to solving the differential equation directly, which is not trivial in general. However, for first-order differential equations such as the track-and-hold mixer equation, this method is feasible.

Consider the mixer equation (9), where $g$ is time varying. The zero state response of the first-order equation

$$gv_1 + C \frac{dv_1}{dt} = gV_{\text{in}}$$

can be obtained using an integrating factor

$$v_1(t) = \int_0^t \exp\left(-\int_\mu^t \frac{g(\xi)}{C} \, d\xi\right) \frac{g(\mu)}{C} V_{\text{in}}(\mu) \, d\mu. \tag{19}$$
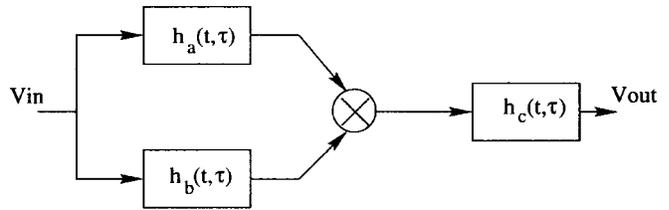


Fig. 4.   Composition of second-order system from first-order systems.

If we set $V_{\text{in}}(\mu) = \delta(\mu - \tau)$ in the above, we obtain the impulse response of the first-order time-varying system

$$
h_1(t, \tau) = \begin{cases} \exp\left(-\int_\tau^t \frac{g(\xi)}{C} \, d\xi\right) \dfrac{g(\tau)}{C} & \text{if } t \geq \tau \\ 0 & \text{if } t < \tau. \end{cases} \tag{20}
$$

We would also like to calculate the second-order kernel. In general, a second-order system can be thought of as the composition of linear systems as shown in Fig. 4. Suppose that the impulse responses of the linear systems are $h_a(t, \tau)$, $h_b(t, \tau)$, and $h_c(t, \tau)$, respectively, the second-order kernel can be computed as follows:

$$h_2(t, \tau_1, \tau_2) = \int_{-\infty}^{\infty} h_c(t, \tau) h_a(\tau, \tau_1) h_b(\tau, \tau_2) \, d\tau. \tag{21}$$

The expression for third order systems is similar. If in Fig. 4, $h_a$ is second order, and therefore the overall system is third order, the overall system kernel is then

$$h_3(t, \tau_1, \tau_2, \tau_3) = \int_{-\infty}^{\infty} h_c(t, \tau) h_a(\tau, \tau_1, \tau_2) h_b(\tau, \tau_3) \, d\tau. \tag{22}$$

Therefore, as was done for the time-invariant case, $h_2$ and $h_3$ can be solved consecutively. For example, the second-order kernel is found by substituting $v_1(\tau)$ from (19) and setting input as an impulse in the solution to the second-order equation in (9)

$$
\begin{aligned}
v_2(t) = & \int_0^t \exp\left(-\int_\tau^t \frac{g(\xi)}{C} \, d\xi\right) \frac{g(\tau)}{C} \left(\frac{K}{2}\right) \\
& \cdot (v_1^2(\tau) - V_{\text{in}}^2(\tau)) \, d\tau. \tag{23}
\end{aligned}
$$

Since the differential equations are linear and first order, there is no theoretical difficulty in obtaining the solution. Following this method, the complete time domain solution to the system can be obtained.

Now, in a track-and-hold mixer where the output points of interest are at the sampled points, we need to sample $v_2$ and apply Fourier analysis to the sampled points to obtain the second order distortion. This procedure, although straightforward in theory, is tedious even for a simple system such as the track-and-hold mixer. However, as will be shown in the next section, by exploring the frequency domain interpretation and by taking advantage of the fact that output is in the sampled-data domain, the problem can be simplified significantly.

### B. Sampling Process

Let sampling occur at time 0, let $h_1(t, \tau)$ be the time-varying kernel of the system with a nonzero fall-time LO

waveform, the linear term in the sampled output voltage is

$$y(0) = \int_{-T}^{0} h_1(0, \tau)x(\tau)\,d\tau + \text{ZIR} \qquad (24)$$

where $T$ is the sampling period, and ZIR is the zero input response due to the initial condition. Since we have assumed that the RC time-constant for the track-and-hold mixer is much smaller than the sampling period, ZIR is negligible. Also because of the small time-constant, we may replace the lower limit of integration by $-\infty$

$$y(0) = \int_{-\infty}^{0} h_1(0, \tau)x(\tau)\,d\tau. \qquad (25)$$

Further, since the output voltage of interest is at the sampling instant, the time-varying characteristic of importance occurs during the time within a few time-constants before the sampling instant. This means that the same Volterra kernel applies to every sampling instant, except the kernels are shifted by $nT$, integer multiples of the sampling period. To simplify computation, we want to keep the Volterra kernel identical for each sampling period. Therefore, instead of shifting the kernel, we shift the input signal backward in time by $nT$ and keep the sampling instant at time 0. This effectively keeps the functional form of the kernel identical for all samples. Further, there is nothing special about sampling points $nT$. So, if sampling occurs at an arbitrary time $t$, the sampled voltage can be represented as

$$y(t) = \int_{-\infty}^{0} h_1(0, \tau)x(\tau + t)\,d\tau. \qquad (26)$$

This $y(t)$ is a fictitious signal, whose value at $t$ represents the sampled value of the output if sampling is to occur at $t$. If sampling occurs at instants $nT$, the output voltage samples are just $y(T)$, $y(2T)$, $\cdots$, $y(nT)$. Hence, the nonideal sampling of the output signal is reduced to the ideal sampling of $y(t)$, where $y(t)$ is related to the input $x(t)$ by (26). Now, because the harmonic distortion of the ideal sampled $y(nT)$ is the same as the harmonic distortion of the continuous signal $y(t)$, the problem of calculating distortion in nonideal sampled output is reduced to the problem of calculating the distortion in the continuous signal $y(t)$. The transformation from (24)–(26) also reduces a time-varying system to a time-invariant system, as the relation between $x$ and $y$ in (26) is time invariant. This situation is analogous to the analysis of discrete control systems where, although a zero-order-hold is not a time-invariant operation in the continuous-time domain, the input–output relation is nevertheless time invariant in the sampled-data domain.

It remains to find the impulse response of the linear time-invariant system (26). Denote the impulse response in time and frequency domain by $\hat{h}_1(t)$ and $\hat{H}_1(\omega)$, respectively. To find $\hat{h}_1$, set $x(t) = \delta(t)$ in (26), it follows that

$$\hat{h}_1(t) = \begin{cases} h_1(0, -t) & \text{if } t \geq 0 \\ 0 & \text{if } t < 0. \end{cases} \qquad (27)$$

The frequency response is its Fourier transform

$$\hat{H}_1(\omega) = \int_{-\infty}^{0} h_1(0, \tau)e^{j\omega\tau}\,d\tau \qquad (28)$$

where, again, $h_1(t, \tau)$ is the time-varying kernel.

The same technique applies to higher order systems. For a second-order system, the sampled-data domain response is

$$y(t) = \int_{-\infty}^{0}\int_{-\infty}^{0} h_2(0, \tau_1, \tau_2)x(\tau_1 + t)x(\tau_2 + t)\,d\tau_1\,d\tau_2 \qquad (29)$$

and the frequency response is

$$\hat{H}_2(\omega_1, \omega_2) = \int_{-\infty}^{0}\int_{-\infty}^{0} h_2(0, \tau_1, \tau_2)e^{j\omega_1\tau_1}e^{j\omega_2\tau_2}\,d\tau_1\,d\tau_2. \qquad (30)$$

Again, the second-order kernel may be obtained from its composition from first-order systems as in (21).

The frequency domain expression is of the most interest. It is in fact possible to bypass the time domain expression and obtain frequency domain result directly. Assuming the composition form as in Fig. 4, substituting (21) into (30), then

$$\hat{H}_2(\omega_1, \omega_2) = \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{-\infty}^{0} h_c(0, \tau_3)h_a(\tau_3, \tau_1) \\ \cdot h_b(\tau_3, \tau_2)e^{j\omega_1\tau_1}e^{j\omega_2\tau_2}\,d\tau_3\,d\tau_1\,d\tau_2.$$

More succinctly, and for convenience only, define

$$\hat{H}_1(\omega, t) = \int_{-\infty}^{t} h_1(t, \tau)e^{j\omega\tau}\,d\tau \qquad (31)$$

and

$$\hat{H}_2(\omega_1, \omega_2, t) = \int_{-\infty}^{t}\int_{-\infty}^{t} h_2(t, \tau_1, \tau_2) \\ \cdot e^{j\omega_1\tau_1}e^{j\omega_2\tau_2}\,d\tau_1\,d\tau_2 \qquad (32)$$

we arrive at the formula

$$\hat{H}_2(\omega_1, \omega_2, t) = \int_{-\infty}^{t} h_c(t, \tau)H_a(\omega_1, \tau)H_b(\omega_2, \tau)\,d\tau. \qquad (33)$$

In particular, setting $t = 0$ gives the sampled frequency domain kernel for the system of Fig. 4

$$\hat{H}_2(\omega_1, \omega_2) = \int_{-\infty}^{0} h_c(0, \tau)H_a(\omega_1, \tau)H_b(\omega_2, \tau)\,d\tau. \qquad (34)$$

The third-order frequency domain representation is similarly derived. If $h_a$ in Fig. 4 is second order, then

$$\hat{H}_3(\omega_1, \omega_2, \omega_3, t) \\ = \int_{-\infty}^{t} h_c(t, \tau)H_a(\omega_1, \omega_2, \tau)H_b(\omega_3, \tau)\,d\tau \qquad (35)$$

and the overall frequency domain kernel is

$$\hat{H}_3(\omega_1, \omega_2, \omega_3) = \int_{-\infty}^{0} h_c(0, \tau)H_a(\omega_1, \omega_2, \tau)H_b(\omega_3, \tau)\,d\tau. \qquad (36)$$

### C. Time-Varying Mixer

We now analyze the track-and-hold mixer of Fig. 1 in the sampled-data domain. Assume the input side is source, the

output drain,[1] from the basic MOS equation, we have

$$K(V_g - V_s - V_t)(V_d - V_s) - \frac{K}{2}(V_d - V_s)^2 + C\frac{dV_d}{dt} = 0.$$

We define MOS conductance as $\hat{g} = K(V_{gs} - V_t)$. The MOS enters the cutoff region at $V_{gs} - V_t = 0$, so the value of $\hat{g}$ goes to zero at each sampling point. Now, since we have assumed that the RC time-constant is small, the region of most importance is the time right before the sampling instant. Therefore, we can approximate $\hat{g}$ by a linear function prior to each sampling point. This assumption is most valid if the system time-constant is smaller than the fall-time of the gate voltage. In practical cases where this is not true, the analysis yields a limiting case and provides a useful bound. Now, the slope of the linearly varying $g$ is just the slope of the cutting edge $V_g$ minus the slope of the input signal $V_s$ to a first-order approximation. So, if we define $\beta = (2KV_G)/T_f$, where $V_G$ and $T_f$ are as indicated in Fig. 1, and define $g = -\beta t$, the MOS equation becomes

$$\left(g - K\frac{dV_s}{dt} \cdot t\right)(V_d - V_s) - \frac{K}{2}(V_d - V_s)^2 + C\frac{dV_d}{dt} = 0.$$

Let $V_s = V_{in}$, $V_d = V_o$, expand $V_o$ into its Volterra series as before, and collect the first three order terms, we have

$$\begin{cases} gv_1 + C\dfrac{dv_1}{dt} = gV_{in} \\[2mm] gv_2 + C\dfrac{dv_2}{dt} = K\dfrac{dV_{in}}{dt}t(v_1 - V_{in}) + \dfrac{K}{2}(v_1 - V_{in})^2 \\[2mm] gv_3 + C\dfrac{dv_3}{dt} = K\dfrac{dV_{in}}{dt}tv_2 + K(v_1 - V_{in})v_2. \end{cases}$$
$$(37)$$

To calculate $H_1$, we use (31) where the time domain $h_1(t,\tau)$ is as found in (20).

$$\hat{H}_1(\omega, t) = \int_{-\infty}^{t} h_1(t,\tau)e^{j\omega\tau}\,d\tau$$
$$= \int_{-\infty}^{t} \exp\left(-\int_{\tau}^{t}\frac{g(\xi)}{C}\,d\xi\right)\frac{g(\tau)}{C}e^{j\omega\tau}\,d\tau$$
$$= \int_{-\infty}^{t} e^{-k^2(\tau^2-t^2)}(-2k^2\tau)e^{j\omega\tau}\,d\tau$$

where the substitution of $g = -\beta t$ and $k^2 = \beta/2C$ are made. This integral cannot be evaluated analytically. However, we are interested in the value of the integral when $t$ is close to 0, so the upper limit of the integral is close to zero. The integrand is a rapidly increasing function of $\tau$ as $\tau$ approaches 0; so, effectively, the only relevant portion of the integration is when $\tau$ is close to 0. Therefore, we can assume that $e^{j\omega\tau} \approx 1 + j\omega\tau$. In this case, the integral becomes:[2]

$$\hat{H}_1(\omega, t) = 1 + j\omega t - j\omega\sqrt{\frac{\pi}{4}}\cdot\frac{1}{k}\cdot e^{k^2t^2}(\text{erf}(kt)+1). \quad (38)$$

So, setting $t = 0$ gives

$$\hat{H}_1(\omega) = 1 - j\omega\sqrt{\frac{\pi}{4}}\cdot\frac{1}{k}. \quad (39)$$

[1] Assuming the output side being source will give a slightly different differential equation, but the form of the answer is the same.

[2] The error function is defined as $\text{erf}(t) = (2/\sqrt{\pi})\int_0^t e^{-x^2}\,dx$.

The second-order term is evaluated using (33), where the time-varying kernel is obtained from its composition from first-order terms. Mechanically, $v_1$ is substituted by $\hat{H}_1$, and $V_{in}$ is substituted by $e^{j\omega t}$

$$\hat{H}_2(\omega_1, \omega_2, t) = \int_{-\infty}^{t}\exp\left(-\int_{\tau}^{t}\frac{g(\xi)}{C}\,d\xi\right)\cdot\frac{K}{C}\cdot j\omega_1$$
$$\cdot e^{j\omega_1\tau}\cdot\tau\cdot(\hat{H}_1(\omega_2,\tau) - e^{j\omega_2\tau})\,d\tau$$
$$+ \int_{-\infty}^{t}\exp\left(-\int_{\tau}^{t}\frac{g(\xi)}{C}\,d\xi\right)\cdot\frac{K}{2C}$$
$$\cdot(\hat{H}_1(\omega_1,\tau) - e^{j\omega_1\tau})$$
$$\cdot(\hat{H}_1(\omega_2,\tau) - e^{j\omega_2\tau})\,d\tau. \quad (40)$$

Again, approximate $e^{j\omega t}$ by its Taylor expansion, substitute (38) to (40)

$$\hat{H}_2(\omega_1, \omega_2, t)$$
$$= \omega_1\omega_2\cdot\frac{K}{C}\cdot\sqrt{\frac{\pi}{4}}\cdot\frac{1}{k}\cdot e^{k^2t^2}\int_{-\infty}^{t}\tau(\text{erf}(k\tau)+1)\,d\tau$$
$$- \omega_1\omega_2\cdot\frac{K}{C}\cdot\frac{\pi}{8}\cdot\frac{1}{k^2}\cdot e^{k^2t^2}\int_{-\infty}^{t}e^{k^2\tau^2}(\text{erf}(k\tau)+1)^2\,d\tau.$$
$$(41)$$

Fortunately, at $t = 0$, the last two integrals may be evaluated numerically

$$\hat{H}_2(\omega_1, \omega_2) = -\omega_1\omega_2\left(\frac{K}{C}\right)\cdot\frac{b}{k^3} \quad (42)$$

where $b = 0.375$ comes from the evaluation of definite integrals. (The relative contributions from the two integrals are comparable.) Finally, $\hat{H}_3$ is also obtained by its composition from first- and second-order systems

$$\hat{H}_3(\omega_1, \omega_2, \omega_3, t)$$
$$= \int_{-\infty}^{t}\exp\left(-\int_{\tau}^{t}\frac{g(\xi)}{C}\,d\xi\right)\cdot\frac{K}{C}\cdot j\omega_1\cdot e^{j\omega_1\tau}$$
$$\cdot\tau\cdot\hat{H}_2(\omega_1,\omega_2,\tau)\,d\tau + \int_{-\infty}^{t}\exp\left(-\int_{\tau}^{t}\frac{g(\xi)}{C}\,d\xi\right)\cdot\frac{K}{C}$$
$$\cdot(\hat{H}_1(\omega_3,\tau) - e^{j\omega_3\tau})\cdot\hat{H}_2(\omega_1,\omega_2,\tau)\,d\tau. \quad (43)$$

We need to evaluate the above expression for $t = 0$. Because each $\hat{H}_2$ term contains two integrals, there is a total of four definite integrals. By substituting the expression for $\hat{H}_1$ from (38), and $\hat{H}_2$ from (41), we can evaluate the integral numerically

$$\hat{H}_3(\omega_1, \omega_2, \omega_3) = j\omega_1\omega_2\omega_3\cdot\left(\frac{K}{C}\right)^2\cdot\frac{c}{k^5} \quad (44)$$

where $c = 0.234$. Equations (39), (42), and (44) are the final solution to the sampled time-varying Volterra series. To summarize the results in terms of circuit parameters

$$\hat{H}_1(\omega) = 1 - j\omega\cdot\sqrt{\frac{C}{K}}\left(\frac{T_f}{V_G}\right)^{1/2}\cdot a \quad (45)$$

$$\hat{H}_2(\omega_1, \omega_2) = -\omega_1\omega_2\cdot\sqrt{\frac{C}{K}}\left(\frac{T_f}{V_G}\right)^{3/2}\cdot b \quad (46)$$

$$\hat{H}_3(\omega_1, \omega_2, \omega_3) = j\omega_1\omega_2\omega_3\cdot\sqrt{\frac{C}{K}}\left(\frac{T_f}{V_G}\right)^{5/2}\cdot c \quad (47)$$

where again $a = 0.886$, $b = 0.375$, $c = 0.234$ numerically. Recall that this result is obtained by assuming a linearly decreasing $g$. In reality, $g$ is nearly constant for a large part of the sampling period. Therefore, this solution is an asymptotic case.

Comparing the Volterra coefficients for the nonideal sampling MOS track-and-hold mixer with the ideal sampling case, at low frequency we find that nonlinearity due to the time-varying nature is much smaller than the ideal sampling case. However, time-varying distortion varies with the cube of input frequency; so, theoretically, at high enough frequency, the time-varying distortion will overtake the continuous-time distortion. This happens at an RF frequency around 2–3 GHz. The fact that time-varying distortion is small at low frequency can be intuitively explained. There are two sources of nonlinearity from the MOS governing equation

$$I_D = K(V_{GS} - V_t)V_{DS} - \frac{K}{2}V_{DS}^2.$$

An obvious source of nonlinearity is the square term $V_{DS}^2$, but this effect is secondary for the ideal sampling case, where the major contributing factor is the input signal dependent conductance $g = K(V_{GS} - V_t)$. However, in a nonideal sampling mixer, the first-order signal dependence in $g$ is eliminated because sampling always occurs when $V_{GS} - V_t = 0$, hence the local behavior of $g$ prior to cutoff is approximately same for each sampling point. Then, the only nonlinear factors left are the signal dependence in the derivative of the conductance and the $V_{DS}^2$ term, which are significantly smaller. However, we are not getting something for free here. Although in the case of a nonideally sampled mixer, the $K(V_{GS} - V_t)$ term causes much less input dependence in $g$, the $V_{GS}$ term manifests as an altogether different source of distortion. Because cutoff occurs when $V_{GS} - V_t = 0$, the cut-off time is now input signal dependent. In other words, the distortion in the amplitude domain is shifted to the time domain by nonideal sampling. The signal dependent cutoff time is called sampling error, which will be discussed in detail in the next section.

## IV. SAMPLING ERROR

In the development of time-varying distortion analysis in sampled-data domain, the output voltage is assumed to be sampled at equally spaced instants. This assumption is not true if the gate voltage has a nonzero fall-time. The instance of sampling is when $V_{GS} - V_t = 0$, so the time at which sampling occurs depends not only on gate voltage but also on the input voltage. The signal-dependent sampling time introduces additional distortion term, which can again be quantified using Volterra series.

Suppose the input signal is smooth—call it $f(t)$—and we intend to sample the input at time 0. Let the gate voltage have a finite-slope falling edge with slope $\alpha$, where $\alpha = 2V_G/T_f$ as shown in Fig. 5, i.e., the falling edge has the form $e(t) = V_G - \alpha t$. Since the MOS enters cutoff at $V_G - V_S = V_t$, and the input is applied at the source side, sampling occurs at the instant when $e(t) - V_t = f(t)$. Now, use the approximation
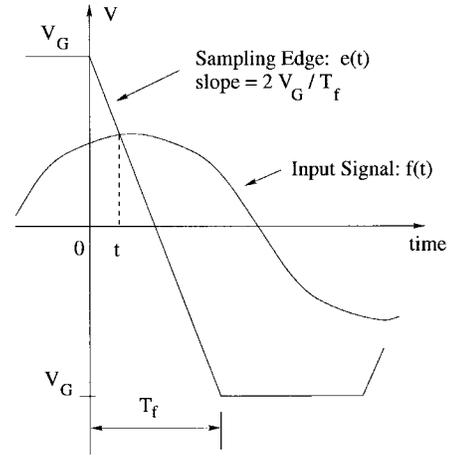


Fig. 5.   Sampling error.

$f(t) \approx f(0) + f'(0)t$, and solve for the cutoff time $t$

$$t = \frac{V_G - V_t - f(0)}{\alpha + f'(0)}$$

$$\approx \frac{V_G - V_t - f(0)}{\alpha} \cdot \left(1 - \frac{f'(0)}{\alpha} + \frac{f'(0)^2}{\alpha^2}\right) \quad (48)$$

where we assumed that $\alpha$ is large. Substitute the above expression for $t$ into the Taylor expansion for $f(t)$ around 0, and collect the first three order terms, we have

$$f(t) = f(0) - \frac{f'(0)f(0)}{\alpha} + \left(\frac{f''(0)f^2(0)}{2\alpha^2} + \frac{f'(0)^2 f(0)}{\alpha^2}\right) \quad (49)$$

where again the assumption of large $\alpha$ is used. So, due to the finite fall-time, the sampled value $f(t)$ differs from the desired value $f(0)$ by additional signal-dependent terms, which produce distortion. Since the sampling time is arbitrary, (49) is valid if time 0 is replaced by an arbitrary sampling time.

The signal dependence can be analyzed by Volterra series. The linear term in (49) is $f(0)$, so $H_1 = 1$. The second-order term is the product of derivative with the function itself, and its Volterra series coefficient can be found using the composition rule (21) followed by symmetrization. The second order coefficient is found to be

$$H_2(\omega_1, \omega_2) = -\frac{j(\omega_1 + \omega_2)}{2\alpha} \quad (50)$$

and for the third-order,

$$H_3(\omega_1, \omega_2, \omega_3) = -\frac{(\omega_1 + \omega_2 + \omega_3)^2}{6\alpha^2}. \quad (51)$$

Using these expressions in (3)–(6), harmonic distortion and intermodulation can be calculated exactly

$$\mathrm{HD}_2 = \frac{A}{4}\left(\frac{j\omega T_f}{V_G}\right) \quad (52)$$

$$\mathrm{HD}_3 = \frac{3A^2}{32}\left(\frac{\omega T_f}{V_G}\right)^2 \quad (53)$$

$$\mathrm{IM}_2 = 0 \quad (54)$$

$$\mathrm{IM}_3 = \frac{A^2}{32}\left(\frac{\omega T_f}{V_G}\right)^2. \quad (55)$$

Comparing the sampling distortion as given by (52)–(55) to the continuous-time distortion as given by (13)–(16), we can show that the sampling distortion becomes comparable to the continuous-time distortion when the fall-time is in the order of $\sqrt{\tau T}$, where $\tau$ is the RC time-constant formed by the MOS resistor and the load capacitor, and $T$ is the period of the input signal. Therefore, sampling error becomes a bigger problem at high frequency, where $\tau$ has to be small.

## V. SIMULATION RESULTS AND COMPARISON TO THEORY

SPICE simulations are performed to verify the theoretical results derived in the previous sections for the cases of continuous-time, time-varying, and sampling distortion. The three factors have to be combined in order to fully explain the distortion of a track-and-hold mixer whose gate voltage has a finite falling slope.

Although RF is normally carried in the 900-MHz range, we start off the simulation at 100 MHz assuming the mixer is used in a first-stage IF digitizer application. The track-and-hold mixer is fed with either a single tone or two tones near 100 MHz at its IF input. For example, to estimate intermodulation distortion, two tones at 100.3 and 100.4 MHz are applied at the IF input, and a 100-MHz square-wave sampling clock is applied at the gate. Consistent with our standing assumptions, the RC time-constant during ON period has been chosen to be approximately 20 ps, and therefore much smaller than the 5 ns ON period of the sampling clock. Finally, we note that the results presented in this section are applicable in the RF frequency range as well.

### A. Simulation Accuracy and Speedup

A number of precautions have to be taken to ensure numerical accuracy, convergence, and charge conservation in the simulations. The simulation accuracy has to be calibrated with ideal sine tones and simple linear models. The maximum time-step value and time-step update algorithm also have to be chosen consistent with the dynamic range to be expected from simulations. The time-step parameter (delmax) is crucial in determining simulation accuracy. Our experience shows that a delmax as low as 0.1 ps is necessary to achieve reasonably accurate results. Charge conservation is particularly difficult to achieve. It depends on the level of the MOS model chosen, the MOS device capacitance (capop) model, and the integration method. Our experience has been that the charge conservation models (capop = 4 and capop = 9 in HSPICE) either have poor convergence properties or produce numerical inaccuracy. We have bypassed the charge-injection and charge-conservation problems by calculating the device capacitances separately and inserting linear capacitances into the circuit and using a no-capacitance (capop = 5) model. Without these precautions, the simulation results can be rendered totally useless.

SPICE simulation for very small values of delmax can make the simulation time impractically long. A major factor in speeding up simulation has been the insight that the RC time-constant during the ON period of the MOS transistor is much smaller than the ON period. Thus, since the circuit reaches steady state within only a tiny fraction of the period, we may
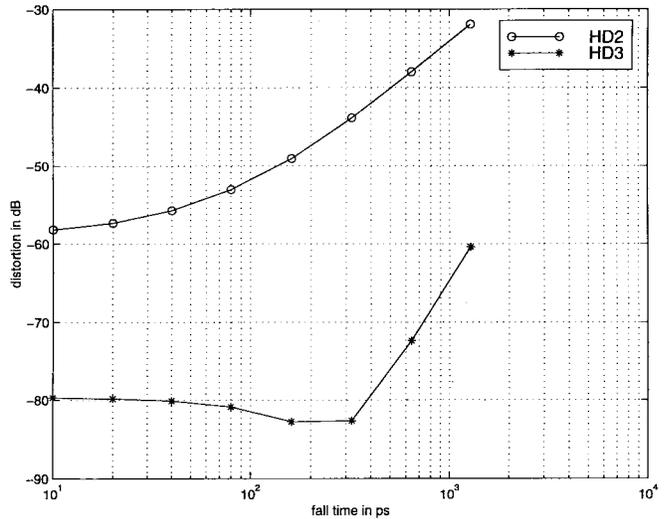


Fig. 6. Harmonic distortion ($HD_2$ and $HD_3$) of a track-and-hold mixer versus fall-time at 100 MHz (body effect excluded).

only need to simulate around the sampling edge. This has been accomplished by repeated simulations at uniform phase intervals of the input sine-wave for a total phase interval of 360°, which corresponds to exactly one period of the mixed down frequency.

### B. Harmonic Distortion

Fig. 6 plots the second- and third-order harmonic distortion of a track-and-hold mixer with a MOS switch $W/L$ of 102 $\mu$m/0.8 $\mu$m. The sampling capacitance is 0.2 pF. An additional capacitance of 0.152 pF is added to account for source/drain capacitance. A square-wave with fall-time varying between 10 ps and 1.28 ns is applied at the gate. In this simulation, the SPICE model parameter gamma has been set to 0 to exclude substrate bias modulation of the threshold voltage. The $HD_3$ curve can clearly be divided into three regions:

1) a relatively flat portion corresponding to small fall-times up to a few tens of picoseconds;
2) a region with distortion decreasing with fall-time and having a distinct minimum; and
3) a region with distortion increasing with fall-time occurring for large fall-times and with a distinct slope of 20 dB/decade for $HD_2$ and 40 dB/decade for $HD_3$.

Region 1 shows an asymptotic behavior toward 0 fall-time. This asymptotic distortion value agrees with the prediction from continuous-time (time-invariant) harmonic distortion model (13)–(16). This is to be expected since we have shown in Section II that for fall-time of 0, the harmonic distortion of the MOS track-and-hold distortion approaches the continuous-time harmonic distortion.

Region 3 displays asymptotic behavior toward a log-linear relation that increases at 20 dB/decade for $HD_2$ and 40 dB/decade for $HD_3$. This line agrees exactly with the distortion predicted by sampling error (52)–(55). Hence, for large fall-times, harmonic distortion is limited by sampling distortion.

Region 2 can be explained by time-varying distortion of a MOS track-and-hold mixer at whose gate a nonzero fall-time sampling signal is applied. In Section III, it was pointed out
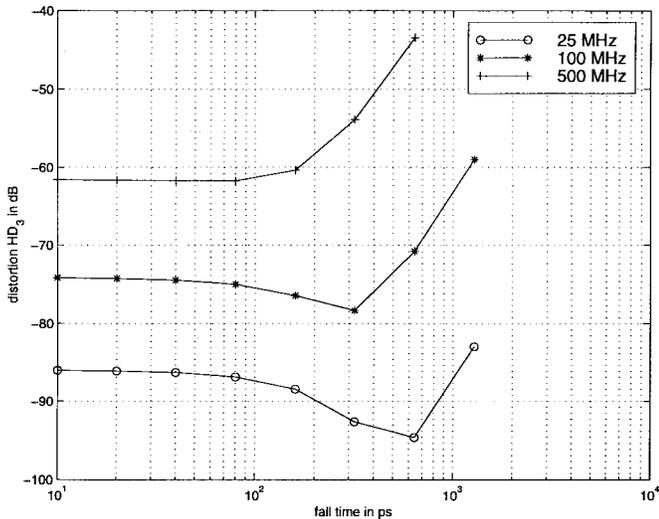
Fig. 7. Effect of input frequency on harmonic distortion HD$_3$ (body effect included).



Fig. 8. Intermodulation (IM$_2^+$ and IM$_3$) of a track-and-hold mixer versus fall-time at 100 MHz.

that for the case of nonideal sampling, if sampling error is neglected, harmonic distortion, now attributed to time-varying distortion, becomes very small. This is responsible for the dip in distortion curve in region 2 since in this intermediate region the fall-time is large enough so that contribution due to the continuous-time operation is small, while at the same time the fall-time is small enough that sampling distortion does not manifest itself. Overall, the simulation result shows quantitative agreement with the theory in regions 1 and 3, and qualitative agreement in region 2.

In order to highlight the role of time-varying distortion, we note that the third order sampling distortion is proportional to the square of the input frequency $\omega$. Thus, for large fall-times, by reducing the input frequency we may be able to reduce sampling distortion sufficiently enough for time-varying distortion to manifest itself over a wider window (i.e., region 2) of the sweep of fall-time. Conversely, by increasing the input frequency, sampling-distortion may increase to the point that it completely swamps out the time-varying distortion. This trend is confirmed in Fig. 7 which plots HD$_3$ versus fall-time for 25, 100, and 500 MHz. Body effect is included in these simulations.

Thus, we have shown both by simulation and by time-varying analysis that a region with a distortion smaller than the 0 fall-time asymptotic value predicted by continuous-time distortion exists, which explains the decreasing distortion and the minimum observed in this region. The practical significance of this result is that the distortion of a MOS track-and-hold mixer can actually be better than what would be predicted by time-invariant continuous-time Volterra series theory by a proper choice of fall-time. Note that this does not occur for HD$_2$ and IM$_2$ in this particular case because sampling distortion is bigger than the continuous-time distortion even for small fall-times.

### C. Intermodulation Distortion

Fig. 8 shows the intermodulation distortion of the MOS track-and-hold mixer obtained by applying two tones at 100.3
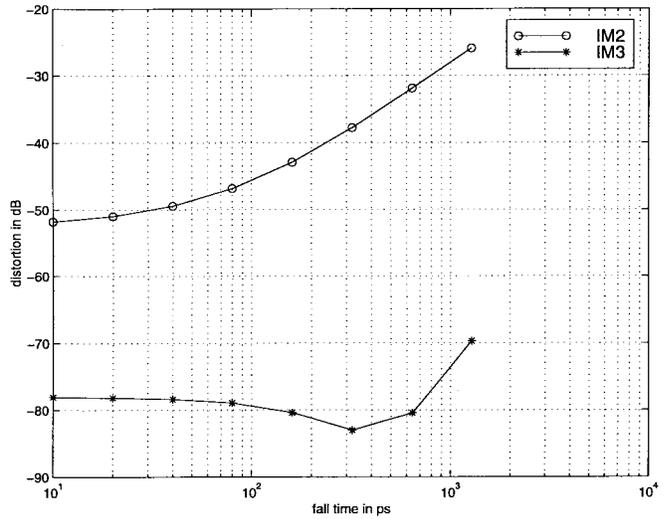
and 100.4 MHz at the IF input and a 100-MHz sampling clock at the gate. IM$_2^-$ resulting from the difference of the two input frequencies is very small. IM$_2^+$, which behaves in the same trend as HD$_2$, is shown. The distortion characteristics again match very well with theoretical predictions both for continuous-time formulas developed in Section II and for the sampling error distortion developed in Section IV. The initial decrease in IM$_3$ curve as the fall-time increases may be explained by the system time-varying nature. In general, intermodulation distortion is slightly worse than harmonic distortion, and they exhibit similar behavior with respect to fall-time.

## VI. BOTTOM-PLATE SAMPLING

In this last section, we solve for distortion in the bottom-plate sampling scheme as a practical application of the theory developed. The bottom-plate sampling scheme is shown in Fig. 9. The input is at $V_{in}$, and the output is taken as the differential voltage between $X$ and $Y$. During sampling, the bottom transistor opens first, which effectively fixes the amount of charge on the sampling capacitor $C_s$. Then, the top transistor opens, which fixes the voltage levels of $X$ and $Y$. In the figure, $C_j$ is included as the junction capacitor of the top switch, which can be fairly large, and $C_p$ is included as the junction capacitor of the bottom switch, which is typically much smaller.

The main advantage of bottom-plate sampling is that it eliminates the charge injection problem. When a MOS transistor turns OFF, the charges that comprise the inversion layer are injected to either the source or the drain side. Which side the charges go to depends on the relative impedance. In the track-and-hold mixer of Fig. 1, the impedance on the output side $(1/j\omega C_s)$ is small, so it attracts the extra charges, which disturb the output signal level. This does not happen in bottom-plate sampling. At the opening of the bottom transistor, charges are injected to the ground, while at the opening of the top plate, the charges are injected back to the input. In either case, little extra charges are dumped to the sampling capacitor.
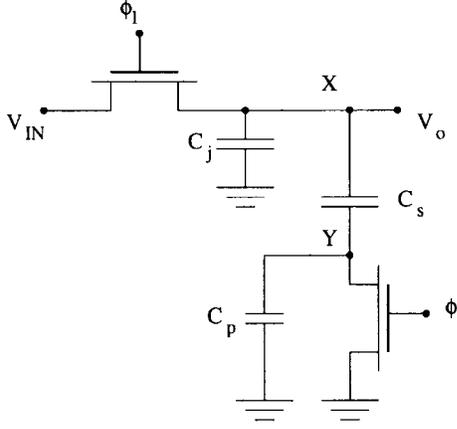
Fig. 9.   Bottom plate sampling.

To analyze this circuit, we first model the transistors as simple switches. It is easy to show that the output voltage can be expressed as

$$V_o = \left(\frac{C_s}{C_s + C_p}\right) \cdot V_{bottom} + \left(\frac{C_p}{C_s + C_p}\right) \cdot V_{top} \quad (56)$$

where $V_{bottom}$ and $V_{top}$ are the voltages at node $X$ when the bottom switch and top switch each opens, respectively. The total distortion in the output is, therefore, the sum of two factors, one due to the opening of the bottom switch, and the other due to the opening of the top switch. Distortions due to the two parts should be analyzed separately. Each part would consist of three factors: the continuous distortion, the time-varying distortion, and the sampling distortion.

### A. Distortion in Opening Bottom Switch

We first notice that the sampling error is small in opening the bottom switch. This is because the source and the drain of the bottom switch are kept at a constant voltage at the ground level. This eliminates signal-dependent sampling.

Next, the continuous-time distortion can be calculated from the coupled differential equations. Assume that both transistors are in the triode region with device constants $K_1$ and $K_2$, respectively. Define $g_1 = K_1(V_{G_1} - V_T)$ and $g_2 = K_2(V_{G_2} - V_T)$. The governing equations are

$$\begin{cases} g_1(V_{in} - X) - \dfrac{K_1}{2} V_{in}^2 + \dfrac{K_1}{2} X^2 = C_s \dfrac{dX}{dt} - C_s \dfrac{dY}{dt} \\ g_2 Y - \dfrac{K_2}{2} Y^2 = C_s \dfrac{dX}{dt} - (C_s + C_p) \dfrac{dY}{dt} \end{cases}$$
$$(57)$$

where $X$ and $Y$ are the voltages on the nodes $X$ and $Y$. We assume the Volterra series expansions

$$\begin{aligned} X(t) &= \boldsymbol{H_1}[V_{in}(t)] + \boldsymbol{H_2}[V_{in}(t)] + \boldsymbol{H_3}[V_{in}(t)] + \cdots \\ Y(t) &= \boldsymbol{G_1}[V_{in}(t)] + \boldsymbol{G_2}[V_{in}(t)] + \boldsymbol{G_3}[V_{in}(t)] + \cdots. \end{aligned} \quad (58)$$

If we assume that $g_1 \gg j\omega C_s$ and $g_2 \gg j\omega C_p$, the output of interest, $H_n - G_n$ is close to the case without the bottom transistor. This is not unexpected because $H_n$'s are almost independent of $g_2$, and $H_n$ is much larger than $G_n$. So, $H_n$ in this case is the same as $H_n$ in the case without the bottom
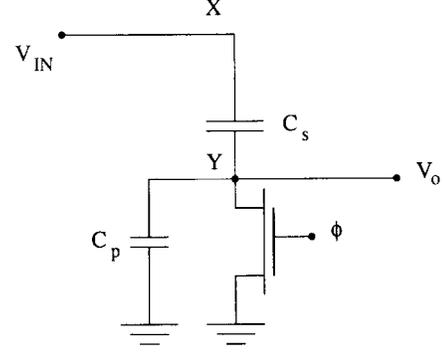


Fig. 10.   Simplified bottom plate sampling.

switch, as given by (14)–(16). In particular,

$$\text{HD}_3 = \frac{A^2}{4} \cdot \frac{j\omega(C_s + C_j)}{K_1(V_G - V_t)^3} \quad (59)$$

$$\text{IM}_3 = \frac{A^2}{4} \cdot \frac{j\omega(C_s + C_j)}{K_1(V_G - V_t)^3}. \quad (60)$$

Note that the continuous-time distortion is determined by the $(W/L)$ ratio of the top transistor.

Finally, we need to take the time-varying distortion into account. The difficulty is that we have a second-order differential equation with nonconstant coefficients for which solutions are not easily found. However, the situation may be simplified. When the bottom switch opens, the resistance between the source and the drain increases with time and eventually reaches infinite; but the top switch stays closed, hence its resistance is relatively small. So we can neglect the top switch altogether and simplify the problem to that in Fig. 10. With only one capacitive node, this circuit is modeled by a first order equation which we can solve.

We are ultimately interested in the differential voltage $(X–Y)$. But since node $X$ is purely sinusoidal, the distortion in $(X–Y)$ is same as the distortion in $Y$. Let $V_o$ be the voltage in node $Y$. The differential equation for the system is as follows:

$$\begin{cases} g v_1 + (C_p + C_s) \dfrac{dv_1}{dt} = C_s \dfrac{d}{dt} V_{in} \\ g v_2 + (C_p + C_s) \dfrac{dv_2}{dt} = \dfrac{K_2}{2} v_1^2 \\ g v_3 + (C_p + C_s) \dfrac{dv_3}{dt} = K_2 v_1 v_2 \end{cases} \quad (61)$$

where again $V_o = v_1 + v_2 + v_3 + \cdots$, and $g = K_2(V_G - V_t)$.

Using the methodology developed in Section IV, time-varying distortion may now be derived. Note here we have $d/dt$ at the right side of the first-order equation, which must be kept as an operator in the derivation, i.e.,

$$h(t, \tau) = \exp\left(-\int_\tau^t \frac{g(\xi)}{C_p + C_s} d\xi\right) \left(\frac{C_s}{C_s + C_p}\right) \frac{d}{d\tau}$$

and

$$\begin{aligned} H_1(\omega, t) &= \int_{-\infty}^t h(t, \tau) e^{j\omega\tau} d\tau \\ &= \int_{-\infty}^t \exp\left(-\int_\tau^t \frac{g(\xi)}{C_p + C_s} d\xi\right) \\ &\quad \cdot \left(\frac{C_s}{C_s + C_p}\right) \frac{d}{d\tau} e^{j\omega\tau} d\tau. \end{aligned} \quad (62)$$

After carrying out the necessary computation, the final expressions for time-varying distortion turn out to be

$$H_1(\omega) = j\omega \left( \frac{C_s}{C_s + C_p} \right) \cdot \frac{a}{k} \tag{63}$$

$$H_2(\omega_1, \omega_2) = -\omega_1 \omega_2 \cdot K_2 \cdot \frac{C_s^2}{(C_s + C_p)^3} \cdot \frac{b}{k^3} \tag{64}$$

$$H_3(\omega_1, \omega_2, \omega_3) = j\omega_1 \omega_2 \omega_3 \cdot K_2^2 \cdot \frac{C_s^3}{(C_s + C_p)^5} \cdot \frac{c}{k^5} \tag{65}$$

where $k$ is related to fall-time as defined before, and $a = 0.886$, $b = 0.375$, $c = 0.234$ numerically. Again, these distortion components are small.

Therefore, as the bottom switch opens, for a small fall-time, distortion is bounded by distortion produced by the continuous-time operation as expressed in (59) and (60). Since both the time-varying and the sampling distortion are small, distortion due to bottom switch is small at large fall-times.

### B. Distortion in Opening Top Switch

The distortion analysis for the top switch is considerably simpler. This is because when the bottom switch is open, the circuit is identical to the case in Fig. 1, except that the capacitor is primarily due to parasitics. The distortion again consists of three parts: continuous-time, time-varying, and sampling distortion.

The continuous-time distortion is given by (13)–(16)

$$HD_3 = \frac{A^2}{4} \cdot \frac{j\omega C}{K_1 (V_G - V_t)^3} \tag{66}$$

$$IM_3 = \frac{A^2}{4} \cdot \frac{j\omega C}{K_1 (V_G - V_t)^3} \tag{67}$$

where the capacitor $C$ is

$$C = \frac{C_s C_p}{C_s + C_p} + C_j. \tag{68}$$

Since the capacitor in this case is primarily due to the parasitic which is much smaller than in the single switch case, the continuous-time distortion is small. The sampling distortion is the same as before, as given by (52)–(55)

$$HD_3 = \frac{3A^2}{32} \left( \frac{\omega T_f}{V_G} \right)^2 \tag{69}$$

$$IM_3 = \frac{A^2}{32} \left( \frac{\omega T_f}{V_G} \right)^2. \tag{70}$$

Again, the time-varying distortion is very small.

Therefore, as the top switch opens, because both continuous-time and time-varying distortion are small, the total distortion is mainly due to sampling error, which is significant at large fall-times.

### C. Total Distortion

We see that distortion due to the opening of the bottom switch comes from continuous-time operation, and that due to the opening of the top switch comes from sampling, so the total distortion is a linear combination of the two as in (56).

$$HD_3 = \left( \frac{C_s}{C_s + C_p} \right) \cdot \frac{A^2}{4} \cdot \frac{j\omega(C_s + C_j)}{K_1 (V_G - V_t)^3}$$
$$+ \left( \frac{C_p}{C_s + C_p} \right) \cdot \frac{3A^2}{32} \left( \frac{\omega T_f}{V_G} \right)^2 \tag{71}$$

$$IM_3 = \left( \frac{C_s}{C_s + C_p} \right) \cdot \frac{A^2}{4} \cdot \frac{j\omega(C_s + C_j)}{K_1 (V_G - V_t)^3}$$
$$+ \left( \frac{C_p}{C_s + C_p} \right) \cdot \frac{A^2}{32} \left( \frac{\omega T_f}{V_G} \right)^2. \tag{72}$$

Therefore, distortion for a bottom-plate sampler at small fall-time is bounded by the continuous distortion, and distortion at large fall-time is bounded by sampling distortion. The major difference between this and the simple case of the track-and-hold mixer is that the distortion due to sampling (when the top switch opens) is lessened by a factor of $C_p/(C_s + C_p)$, which is significant. So, distortion due to nonzero fall-time is less a problem in bottom-plate sampling than in the simple case of Fig. 1. This is another advantage of bottom-plate sampling.

## VII. CONCLUSIONS

The operation of a MOS track-and-hold circuit as a mixer is explained. It is shown that conventional nonlinear time-invariant analysis using Volterra series, while useful for explaining the behavior of the circuit for ideal sampling clock waveform, fails to account for its behavior for nonzero fall-time of the sampling clock. The theory of time-varying Volterra series is developed and applied to the MOS track-and-hold circuit in sampled-data domain. Expressions for harmonic and intermodulation distortion for nonzero fall-time LO waveform are derived. Nonzero fall-time also introduces sampling distortion for which distortion formulas are derived. SPICE simulation on the track-and-hold sampling mixer is performed, and the results are successfully explained for all fall-times by combining the theories of continuous-time, time-varying, and sampling distortion. Finally, the practical scheme of bottom-plate sampling is analyzed.

## REFERENCES

[1] P. Y. Chen, A. Rofougaran, K. A. Ahmed, and A. A. Abidi, "A highly linear 1-GHz CMOS downconversion mixer," in *Proc. Euro. Solid-State Circuits Conf.,* Sevilla, Spain, 1993, pp. 226–227.
[2] A. Hairapetian, "An 81-MHz IF receiver in CMOS," *IEEE J. Solid-State Circuits,* vol. 31, pp. 1981–1986, Dec. 1996.
[3] S. Sen and B. Leung, "A 150MHz, 13-bit 12.5mW IF digitizer using sampling mixer," in *Proc. IEEE Custom Integrated Circuits Conf.,* 1998, p. 233.
[4] A. Namdar and B. Leung, "A 400MHz, 12-bit 18mW IF digitizer using mixer inside the sigma-delta modulator loop," in *Digest Int. Solid State Circuits Conf. (ISSCC),* San Francisco, CA, 1999.
[5] Y.-M. Lin, "Performance limitations on high-resolution video-rate analog-digital interfaces," Ph.D. thesis, Univ. Calif., Berkeley, 1990.
[6] D. Cline, private communication, 1995.
[7] C. D. Keys, "Low-distortion mixers for RF communications," Ph.D. thesis, Univ. Calif., Berkeley, 1994.

[8] M. Schtzen, *The Volterra and Wiener Theories of Nonlinear Systems.* New York: Wiley, 1980.
[9] R. Meyer, "Advanced integrated circuits for communications," Cal VIEW, 1995, UC Berkeley EECS 242 Course Notes.
[10] C. D. Hull, "Analysis and optimization of monolithic RF downconversion receivers," Ph.D. thesis, Univ. Calif., Berkeley, 1992.
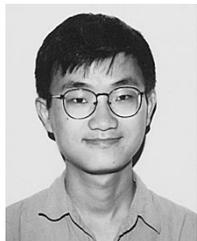
**Wei Yu** received the joint B.S. degree in computer engineering and mathematics from the University of Waterloo in 1997, and the M.S. degree in electrical engineering from Stanford University in 1998. He is currently a Ph.D. candidate in the information systems laboratory in the Electrical Engineering Department at Stanford University. His current research interest is in the transmission theory for digital communications.

**Subhajit Sen** was born in Varanasi, India, in 1961. He received the B.Tech. degree in electronics engineering from the Institute of Technology, Banaras Hindu University; the M.S. degree from the Louisiana State University, Baton Rouge; and the Ph.D. degree from the University of Waterloo, Ont., Canada, in 1984, 1991, and 1997, respectively. During his Ph.D. studies he investigated the problem of sampling and A/D conversion of RF/IF signals. Prior to this work he also worked on high-speed BiCMOS opamp circuit design for switched-capacitor and IF processing applications.

Between 1984 and 1988, he worked in Semiconductor Complex Limited, Chandigarh, India, on analog MOS LSI circuit design. He is presently employed as Project Manager at Arcus Technology Limited, Bangalore, India, a subsidiary of Avant! Corporation, Fremont, CA, in the areas of analog, mixed-signal, and high-frequency circuit design including a 1.2 Giga-samples/s A/D and D/A converter.

**Bosco H. Leung** (S'84–M'85–SM'92) received the Ph.D. degree from the University of California, Berkeley, in 1987, the M.Sc. degree from the California Institute of Technology in 1980, and the B.Sc. degree from the Rensselaer Polytechnic Institute in 1979, all in electrical engineering.

From 1980 to 1983, he was with Northern Telecom, Canada, working on analog circuit design. He is currently a Professor with the Electrical and Computer Engineering Department of the University of Waterloo, Canada. His main research interest is in CMOS and BiCMOS mixed analog/digital integrated circuits, in particular, on analog to digital converters. He has published more than 30 technical papers and has received patents in this area. He has recently developed ultralow-power low distortion mixer and A/D converters for direct conversion based wireless receivers and is also working on frequency synthesizers and other front ends design for wireless receiver.

Dr. Leung is currently an Associate Editor with the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II.