Analysis of a balanced analog multiplier for an arbitrary number of signed inputs

Brian Hong*[1] and Ali Hajimiri[1]

[1] *Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA*

Received 13 January 2016; Revised 17 May 2016; Accepted 21 June 2016; Published online 8 August 2016

SUMMARY

We present an extension of the double-balanced current-commutating analog multiplier (also known as the Gilbert cell) that enables the multiplication of an arbitrary number of signed differential input voltages. A general analysis of the circuit for an arbitrary device nonlinearity is provided, and simulations on a bulk CMOS process as well as measurement results of a discrete bipolar implementation are reported.

KEY WORDS: analog multiplier; Gilbert cell; nonlinear circuit; transistor; intermodulation; bandwidth

1 INTRODUCTION

The double-balanced four-quadrant analog multiplier, invented in the 1960s by Howard Jones [1] and then improved upon by Barrie Gilbert [2, 3], is ubiquitous in numerous modern electronic systems. Shown in Figure 1, this topology effectively multiplies its two input voltages V_{RF} and V_{LO}. Furthermore, the double-balanced nature of the circuit prevents non-product terms (i.e., terms involving V_{RF} or V_{LO} alone) from appearing at or 'feeding through' to the output V_{out}, making it useful in various settings. As a result of its many applications (e.g., amplitude modulation, phase detection, active frequency mixing), it has also been the subject of much research and academic investigation (e.g., [4–11]).

In this expository paper, we look at a natural generalization of this topology that allows for the multiplication of an arbitrary number of analog input voltages of any polarity. Through its analysis, we illustrate how the salient features of a seemingly complicated circuit can be intuitively understood by properly modularizing its topology and exploiting its inherent symmetry. In doing so, we will be able to isolate the system's bare-bone operational concepts from tedious technical details that clutter up the main results.

2 THE TOPOLOGY—A GENERAL DISCUSSION

The *stacked*, balanced, current-commutating analog multiplier is shown in Figure 2. Notice that this topology features n pairs of differential pairs (known as switching stages), whose inputs are cross-coupled and outputs are connected in parallel, 'stacked' on top of one another between the load and the bottom-most differential pair driven by $V_{\mathrm{in},0}$ (known as the transconductance stage). The topology reduces to the standard (dual-input) analog multiplier of Figure 1 for $n = 1$. Versions of this circuit with $n = 2$ have been reported in [12–17] without a detailed general analysis.

*Corresponding author's e-mail: bhong@caltech.edu

This is the peer reviewed version of the following article: *International Journal of Circuit Theory and Applications*, Volume 45, Issue 4, Pages 483–501, April 2017; which has been published in final form at DOI: 10.1002/cta.2243. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

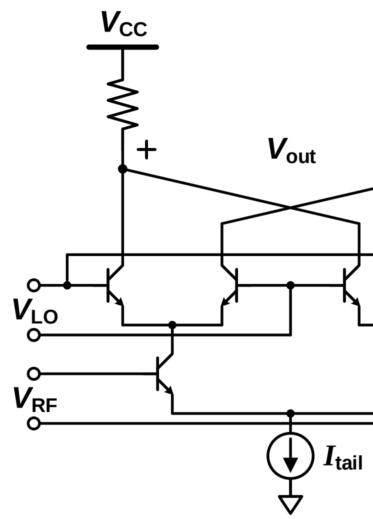


Figure 1: An NPN bipolar implementation of the double-balanced current-commutating analog multiplier (aka Gilbert cell).

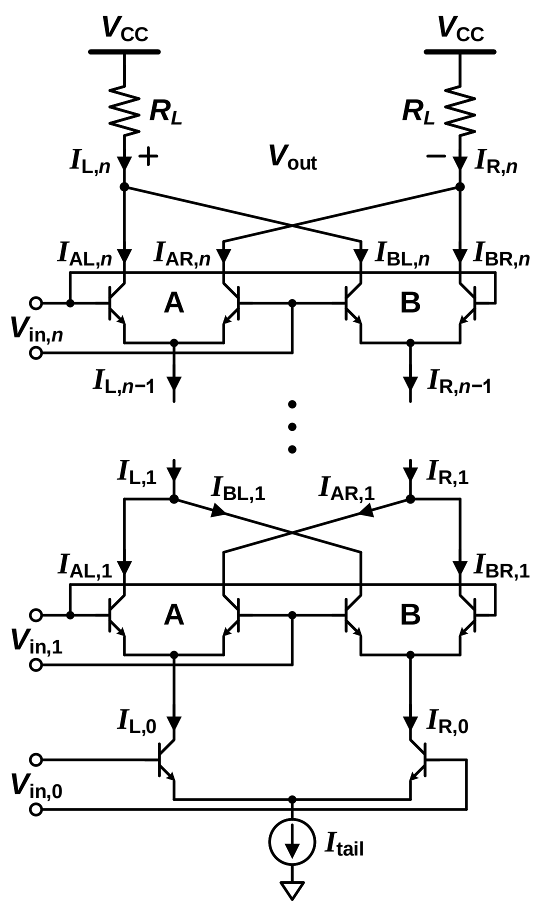


Figure 2: The stacked, balanced, current-commutating analog multiplier. The circuit can also be implemented using field-effect transistors.

To understand how this circuit works, let us make several observations. First, notice that the circuit is balanced with respect to each input. To see this, say one of the inputs, $V_{\mathrm{in},k}$, is zero. Then, the output current of that stage $\left(I_{\mathrm{L},k} - I_{\mathrm{R},k}\right)$ and therefore of any stage above it will be zero, resulting in zero output. Thus, terms that are not a product involving *every* single input cannot feedthrough to the output.

Furthermore, we argue that the output's polarity is reversed whenever the polarity of any one of the inputs is reversed. To see this, observe that switching the polarity of $V_{\mathrm{in},k}$ interchanges the currents $I_{\mathrm{L},k}$ and $I_{\mathrm{R},k}$. If $k = n$, the argument is finished. If $k < n$, notice that because $I_{\mathrm{L},k}$ and $I_{\mathrm{R},k}$ are the tail currents of the two differential pairs of stage $(k+1)$, we have *effectively* reversed the polarity of $V_{\mathrm{in},k+1}$ with respect to the circuit. Recursively propagating this reasoning up the stack of switching stages, we deduce that the polarity of the output is therefore ultimately reversed.

Also note that this topology requires a voltage headroom of $(n+1)V_{\mathrm{act}} + I_{\mathrm{tail}}R_L$ between the supply and the tail current, where V_{act} is the minimum voltage that must be dropped across a transistor to keep it in the proper 'active' region of operation (e.g., $V_{\mathrm{CE,sat}}$, $V_{\mathrm{DS,sat}}$) maximized over all possible operating conditions.

In the following sections, we develop a general analytical framework for computing the output voltage V_{out} in terms of the input voltages. We assume the transistor current I is a nonlinear, monotonically increasing function $f(V)$ of the transistor's control voltage V (e.g., $|V_{\mathrm{BE}}|$, $|V_{\mathrm{GS}}| - |V_{\mathrm{T}}|$). Note that second order effects (e.g., Early effect, channel length modulation, body effect) will be neglected.

3 THE DIFFERENTIAL PAIR

We begin by considering the differential pair, shown in Figure 3, as it is the fundamental building block of the balanced current-commutating analog multiplier. The differential input voltage V_{in} steers the tail current I_{T} between the two transistors, thereby controlling the differential output current $I_{\mathrm{out}} := I_{\mathrm{L}} - I_{\mathrm{R}}$. Throughout, we assume the transistors are matched—that is, they have identical properties.

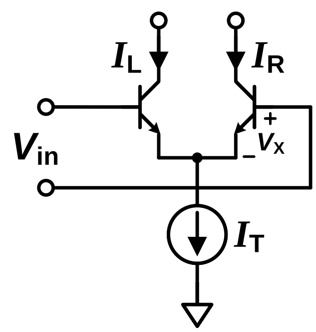


Figure 3: The differential pair. The circuit can also be implemented using field-effect transistors.

First, write the following relationship between the transistors' currents and the tail current:

$$I_{\mathrm{L}} + I_{\mathrm{R}} = f(V_{\mathrm{X}} + V_{\mathrm{in}}) + f(V_{\mathrm{X}}) = I_{\mathrm{T}} \tag{1}$$

where V_{X} is the control voltage of the rightmost transistor.[‡] We can solve for V_{X} in terms of V_{in} and I_{T}: $V_{\mathrm{X}} \equiv g\left(V_{\mathrm{in}}, I_{\mathrm{T}}\right)$; so, I_{L} and I_{R} can be written solely in terms of V_{in} and I_{T}:

$$I_{\mathrm{L}} = f[g(V_{\mathrm{in}}, I_{\mathrm{T}}) + V_{\mathrm{in}}] \tag{2a}$$
$$I_{\mathrm{R}} = f[g(V_{\mathrm{in}}, I_{\mathrm{T}})]\,. \tag{2b}$$

Then, we can write the differential output current as a function of the input voltage and the tail current:

$$I_{\mathrm{out}} := I_{\mathrm{L}} - I_{\mathrm{R}} \equiv h(V_{\mathrm{in}}, I_{\mathrm{T}})\,. \tag{3}$$

Here, we point out two important facts:

[‡]For bipolar transistors, we can modify the right-hand-side of Equation 1 to αI_{T} (where $\alpha \equiv I_{\mathrm{C}}/I_{\mathrm{E}}$) for a more accurate result.

1. h must be an odd function of V_{in} due to symmetry.

2. $h = 0$ if $I_T = 0$, assuming the reverse leakage current through the transistors is negligible.

The asymptotic behavior of h will be discussed in detail in Section 4.

We will now compute examples of h for several well-known transistors.

3.1 Bipolar Differential Pair

For bipolar junction transistors, we take V to be the base-emitter voltage $|V_{\text{BE}}|$. Assuming the forward active region of operation and neglecting the Early effect, the transistor current is

$$f(V) = I_S\left(e^{V/V_{\text{th}}} - 1\right) \tag{4}$$

where I_S is the transistor's saturation current and $V_{\text{th}} := kT/q$ is the thermal voltage. Then [21]

$$h(V_{\text{in}}, I_T) = (\alpha I_T + 2I_S)\tanh\left(\frac{V_{\text{in}}}{2V_{\text{th}}}\right) \tag{5}$$

where $\alpha \equiv I_C/I_E$. Because the reverse leakage current I_S is typically numerous orders of magnitude smaller than the tail current, we have

$$h(V_{\text{in}}, I_T) \cong \alpha I_T \tanh\left(\frac{V_{\text{in}}}{2V_{\text{th}}}\right). \tag{6}$$

3.2 Square-Law MOSFET Differential Pair

For metal-oxide-semiconductor field-effect transistors (MOSFETs), we take V to be the overdrive voltage $|V_{\text{GS}}| - |V_T|$. Assuming pinch-off (i.e., saturation) and neglecting channel length modulation,

$$f(V) = KV^2 \cdot \mathbb{1}\{V \geq 0\} \tag{7}$$

where $K := (\mu C_{\text{ox}}/2)(W/L)$, and the indicator function ensures that the transistor turns off when its overdrive voltage is negative. Then [21]

$$h(V_{\text{in}}, I_T) = \begin{cases} V_{\text{in}}\sqrt{2KI_T - (KV_{\text{in}})^2}, & |V_{\text{in}}| < \sqrt{\dfrac{I_T}{K}} \\ \text{sign}(V_{\text{in}}) \cdot I_T, & \text{otherwise} \end{cases}. \tag{8}$$

3.3 Short-Channel MOSFET Differential Pair

Accounting for velocity saturation, the drain current of Equation 7 can be modified as [21]

$$f(V) = \left(\frac{KV^2}{1 + \dfrac{V}{E_{\text{sat}}L}}\right) \cdot \mathbb{1}\{V \geq 0\} \tag{9}$$

where E_{sat} is the saturation electric field strength (i.e., the saturation velocity is μE_{sat}). Assuming the carriers in the channel are deeply velocity saturated (i.e., $E_{\text{sat}}L \ll V$), Equation 9 becomes

$$f(V) \cong KE_{\text{sat}}LV \cdot \mathbb{1}\{V \geq 0\}. \tag{10}$$

Then it is straightforward to see that

$$h(V_{\text{in}}, I_T) = \begin{cases} KE_{\text{sat}}LV_{\text{in}}, & |V_{\text{in}}| < \dfrac{I_T}{KE_{\text{sat}}L} \\ \text{sign}(V_{\text{in}}) \cdot I_T, & \text{otherwise} \end{cases}. \tag{11}$$

DOI: 10.1002/cta

3.4 Subthreshold Conduction

Due to the increasing prevalence of the use of metal-oxide-semiconductor (MOS) transistors under subthreshold conduction within low-power applications, it is prudent to quickly mention their operation at this point. If $|V_{\mathrm{GS}}| \leq |V_{\mathrm{T}}|$, then the MOSFET's *I-V* characteristic is given by [21]

$$I_{\mathrm{D}} \propto \exp\left(\frac{|V_{\mathrm{GS}}|}{nV_{\mathrm{th}}}\right) \cdot \left(1 - e^{-|V_{\mathrm{DS}}|/V_{\mathrm{th}}}\right), \tag{12}$$

where $n > 1$ is an ideality factor.[§] Reasonably assuming that the drain-source voltage is at least several thermal voltages (e.g., $|V_{\mathrm{DS}}| > 3V_{\mathrm{th}}$), the dependence of the drain current on the drain voltage vanishes and we can take the transistor's control voltage to be the gate-source voltage $V = |V_{\mathrm{GS}}|$. Then,

$$f(V) \propto \exp\left(\frac{V}{nV_{\mathrm{th}}}\right). \tag{13}$$

In other words, a MOS transistor under weak inversion behaves similarly to a bipolar transistor, except with a worse turn-on. Because the proportionality constant in Equation 13 is typically orders of magnitude smaller than the bias current, appealing to the derivation from Section 3.1, we can compute

$$h(V_{\mathrm{in}}, I_{\mathrm{T}}) \cong I_{\mathrm{T}} \tanh\left(\frac{V_{\mathrm{in}}}{2nV_{\mathrm{th}}}\right) \tag{14}$$

which is identical to Equation 6 except for the factor of n.

Note that subthreshold conduction becomes apparent in two different scenarios. Most importantly, if $|V_{\mathrm{GS}}| \leq |V_{\mathrm{T}}|$ at the differential pair's operating or bias point, then obviously, Equation 14 accurately characterizes the differential pair for essentially all input voltages V_{in} of interest. On the other hand, if the transistors are biased in strong inversion (and the square-law prevails, for example), for sufficiently large inputs $|V_{\mathrm{in}}|$ (e.g., $|V_{\mathrm{in}}| \geq \sqrt{I_{\mathrm{T}}/K}$ for square-law MOSFETs), due to subthreshold leakage, one of the transistors will *not* simply 'turn off' abruptly and conduct no current whatsoever as implied by Equations 8 and 11. Instead, there will be a *smooth*, 'exponential-decay like' transition that causes $h(V_{\mathrm{in}}, I_{\mathrm{T}})$ to asymptotically approach $\mathrm{sign}(V_{\mathrm{in}}) \cdot I_{\mathrm{T}}$ in accordance with Equation 14. For our purposes though, this latter scenario does not influence the overall behavior of the differential pair in a practically significant way.

Figure 4 depicts normalized theoretical plots of $h(V_{\mathrm{in}}, I_{\mathrm{T}})$ vs. V_{in} based on Equations 6, 8, and 11.

4 THE CRUCIAL ROLE OF NONLINEARITY

Before we proceed with an analysis of the stacked analog multiplier, we briefly discuss why $f(\cdot)$ must be nonlinear in order for multiplication to occur. To see this, assume f is linear and notice that the output current can be written as

$$\begin{aligned} h(V_{\mathrm{in}}, I_{\mathrm{T}}) &= f(V_X + V_{\mathrm{in}}) - f(V_X) \\ &= f(V_{\mathrm{in}}) \end{aligned} \tag{15}$$

which depends solely (and linearly) upon the differential input voltage V_{in} and not on the tail current I_{T}. In this scenario, the differential behavior of each switching stage becomes independent of and therefore isolated from the operation of the stage below it, preventing the input voltages from multiplying.[¶]

Of course, this is never truly an issue, at least in large-signal: Transistors, being unilateral devices, can appreciably conduct current in only one direction. Consequently, they *turn off* for control voltages that do not exceed some threshold, making f inherently nonlinear. We therefore intuit that differential input voltages above a certain magnitude (say V_{sw}, which depends on I_{T}) will turn off one of the pair's transistors, thereby completely switching the tail current

[§]To elaborate slightly further,

$$n = 1 + \frac{C_S}{C_{\mathrm{ox}}},$$

where C_S is the capacitance of the semiconductor bulk and C_{ox} is the oxide capacitance.

[¶]In fact, due to the balanced nature of the topology, one can easily deduce that the output will be identically zero if h is independent of I_{T}.

DOI: 10.1002/cta

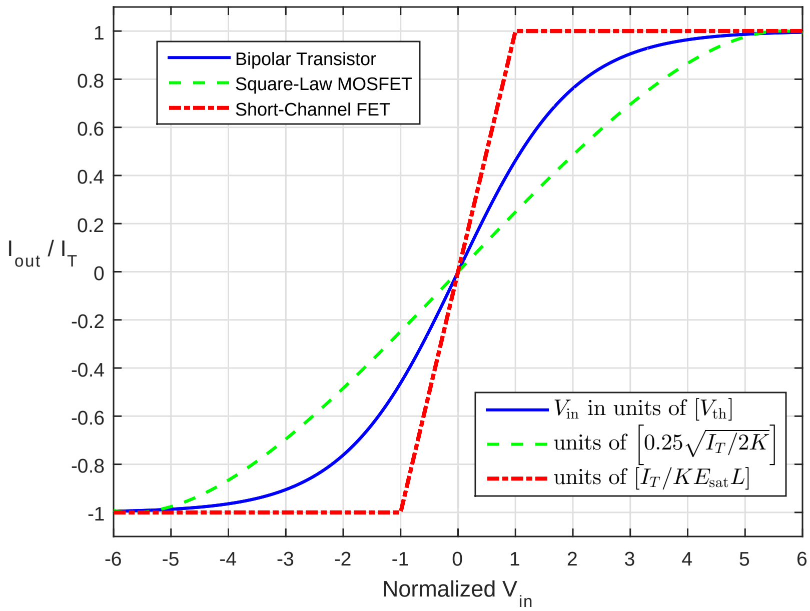


Figure 4: Current switching in a differential pair as a function of the input voltage. Base current and subthreshold conduction were ignored. Due to the chosen normalizations for the $V_{\rm in}$-axis, the plot scales may not be *quantitatively* comparable.

to the other side (Figure 4). That is,[‖]

$$h(V_{\rm in}, I_{\rm T}) = \operatorname{sign}(V_{\rm in}) \cdot I_{\rm T} \quad \text{for} \quad |V_{\rm in}| \geq V_{\rm sw}. \tag{16}$$

As a result, $|h|$ is fundamentally bounded by and therefore dependent on $I_{\rm T}$.

What if $f(\cdot)$ is locally linear in some neighborhood of the transistor's operating or bias point, resulting in a range of input voltages (centered around 0) for which $h(V_{\rm in}, I_{\rm T})$ is independent of the tail current $I_{\rm T}$ (refer to Section 3.3 for an example)? In this situation, if the input voltage to any switching stage remains poised in this range, multiplication cannot occur and the output will be zero. This claim will be proven analytically below in Section 5.2. Essentially, multiplication requires the differential pair's output current $h(V_{\rm in}, I_{\rm T})$ for *every* switching stage to depend on $I_{\rm T}$ at the point where the input voltage $V_{\rm in}$ is situated.

5 GENERAL ANALYSIS

We now proceed with an analysis of the output voltage of the stacked analog multiplier. Throughout, we will assume that the transistors in each *stage* are identical but may differ from stage to stage. Referring to Figure 2,

$$\begin{aligned} V_{\rm out} &= -R_L\left(I_{{\rm L},n} - I_{{\rm R},n}\right) \\ &= -R_L\left[\left(I_{{\rm AL},n} - I_{{\rm AR},n}\right) + \left(I_{{\rm BL},n} - I_{{\rm BR},n}\right)\right] \\ &= -R_L\left[h_n(V_{{\rm in},n}, I_{{\rm L},n-1}) + h_n(-V_{{\rm in},n}, I_{{\rm R},n-1})\right] \\ &= -R_L\left[h_n(V_{{\rm in},n}, I_{{\rm L},n-1}) - h_n(V_{{\rm in},n}, I_{{\rm R},n-1})\right]. \end{aligned} \tag{17}$$

[‖]In the presence of reverse bias or subthreshold leakage, $V_{\rm sw}$ is not well-defined. The more precise statement is

$$h(V_{\rm in}, I_{\rm T}) \to \operatorname{sign}(V_{\rm in}) \cdot I_{\rm T} \quad \text{as} \quad |V_{\rm in}| \to \infty,$$

assuming the leakage current is *small*. See Section 3.1.

DOI: 10.1002/cta

To proceed, the analysis depends on the specific form of h. However, the main idea is that for any $k \in \{1,\ldots,n\}$, we can decompose $I_{\mathrm{L},k} = I_{\mathrm{AL},k} + I_{\mathrm{BL},k}$ and $I_{\mathrm{R},k} = I_{\mathrm{AR},k} + I_{\mathrm{BR},k}$. Then, using $f(\cdot)$ and $g(\cdot)$ for stage k, we can write

$$\begin{aligned} I_{\mathrm{L},k} &= f\left[g\left(V_{\mathrm{in},k}, I_{\mathrm{L},k-1}\right) + V_{\mathrm{in},k}\right] + f\left[g\left(V_{\mathrm{in},k}, I_{\mathrm{R},k-1}\right)\right] \\ I_{\mathrm{R},k} &= f\left[g\left(V_{\mathrm{in},k}, I_{\mathrm{L},k-1}\right)\right] + f\left[g\left(V_{\mathrm{in},k}, I_{\mathrm{R},k-1}\right) + V_{\mathrm{in},k}\right]. \end{aligned} \tag{18}$$

For $k = 0$, we use Equation 2 with $V_{\mathrm{in}} = V_{\mathrm{in},0}$ and $I_{\mathrm{T}} = I_{\mathrm{tail}}$.

We will now explore several important specific cases.

5.1 Output Current Proportional to Tail Current

If $h(V_{\mathrm{in}}, I_{\mathrm{T}})$ is proportional to I_{T}, the analysis proceeds via induction rather quickly. Say

$$h(V_{\mathrm{in}}, I_{\mathrm{T}}) = \zeta(V_{\mathrm{in}})\, I_{\mathrm{T}} \tag{19}$$

where $\zeta(\cdot)$ is a unit-less, odd function bounded by ± 1. Then, from the last step of Equation 17, we have

$$V_{\mathrm{out}} = -R_L\, \zeta_n(V_{\mathrm{in},n})\, \left(I_{\mathrm{L},n-1} - I_{\mathrm{R},n-1}\right). \tag{20}$$

Comparison with the first step of Equation 17 allows us to use induction to obtain

$$V_{\mathrm{out}} = -R_L \left(I_{\mathrm{L},0} - I_{\mathrm{R},0}\right) \prod_{k=1}^{n} \zeta_k\left(V_{\mathrm{in},k}\right). \tag{21}$$

Therefore,

$$V_{\mathrm{out}} = -R_L I_{\mathrm{tail}} \prod_{k=0}^{n} \zeta_k\left(V_{\mathrm{in},k}\right). \tag{22}$$

Example with Measurements: Bipolar Multiplier

From Equation 6, the large-signal output current of a bipolar differential pair is

$$h(V_{\mathrm{in}}, I_{\mathrm{T}}) \cong \alpha I_{\mathrm{T}} \tanh\left(\frac{V_{\mathrm{in}}}{2V_{\mathrm{th}}}\right), \tag{23}$$

which is proportional to I_{T}. Therefore, comparing Equation 23 with Equation 19, we can easily compute the large-signal output voltage of a bipolar implementation of the stacked analog multiplier [13]:

$$\Longrightarrow V_{\mathrm{out}} = -R_L I_{\mathrm{tail}} \prod_{k=0}^{n} \alpha_k \tanh\left(\frac{V_{\mathrm{in},k}}{2V_{\mathrm{th}}}\right). \tag{24}$$

To verify this equation, a bipolar NPN implementation of the stacked analog multiplier with $n = 4$ was constructed using CA3083 transistor arrays. The tail current, $I_{\mathrm{tail}} = 1\,\mathrm{mA}$, was implemented using an emitter degenerated 2N3904 NPN transistor. A load resistance of $R_L = 1\,\mathrm{k\Omega}$ was used, and the supply was set to $V_{\mathrm{CC}} = 7\,\mathrm{V}$. The DC current gain of the transistors was experimentally estimated to be around $\beta = 100$. The common-mode voltage of the kth input was tuned to $(1.68 + k)\,[\mathrm{V}]$.

For each of a total of 100 experiments, 5 input voltages ranging from $\pm 25\,\mathrm{mV}$ to $\pm 130\,\mathrm{mV}$ were randomly generated using MATLAB and fed to the circuit using a USB-3106 DAQ board by Measurement Computing.** The output voltage was measured and compared against $\alpha^5 \prod_{k=0}^{4} \tanh\left(V_{\mathrm{in},k}/2V_{\mathrm{th}}\right)$ [V] from Equation 24. A histogram of the percent error is shown in Figure 5, demonstrating excellent agreement between theory and experiment.

** 'Small' input voltages were avoided because not all of the transistors in the array are matched to one another. It can easily be shown [21] that a (voltage-driven) unbalanced bipolar differential pair is equivalent to a balanced pair with a differential offset voltage equal to $V_{\mathrm{off}} = V_{\mathrm{th}} \ln(I_{\mathrm{SL}}/I_{\mathrm{SR}})$. Therefore, the impact of transistor mismatch on the error in the output current is more prominent for smaller input voltages. Experimentally, an output voltage of 3.3 mV was measured when all the inputs were nulled, which corresponds to a *mean* offset voltage of $V_{\mathrm{off}} = 17\,\mathrm{mV}$ for each pair.

DOI: 10.1002/cta

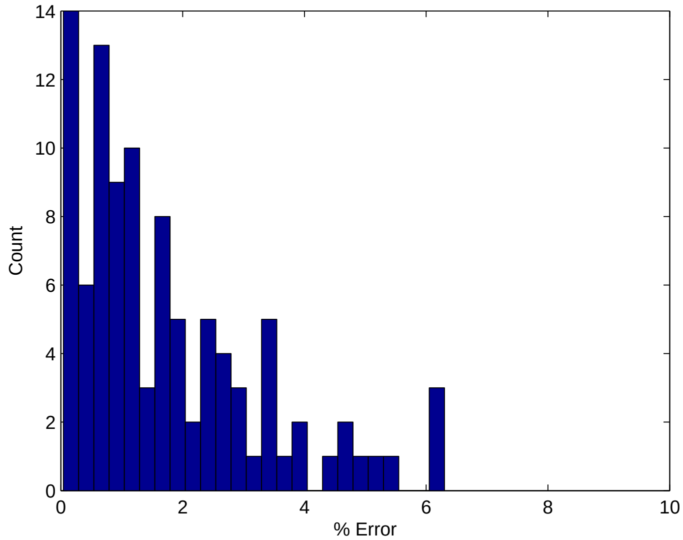


Figure 5: 25 bin histogram of the percent error over $N = 100$ trials. The mean of the percent error is 1.8%.

5.2 Small-Signal Multiplication

Here, we assume that the input voltages are sufficiently small such that any dependence on any particular input voltage is essentially linear. The transistor's transconductance g_m is defined as the rate of change of the output current with respect to the control voltage:

$$g_m := \frac{\partial I}{\partial V} = f'(V), \tag{25}$$

which can be expressed in terms of the transistor's bias current $I = I_{\text{bias}}$ by noting that $V = f^{-1}(I_{\text{bias}})$. So, we write $g_m \equiv g_m(I_{\text{bias}})$ to make this dependence explicit.

A simple small-signal analysis of the differential pair[††] reveals that its small-signal differential transconductance is

$$\left.\frac{\partial h(V_{\text{in}}, I_{\text{T}})}{\partial V_{\text{in}}}\right|_{V_{\text{in}}=0} \equiv \left.\left[\frac{\partial I_{\text{L}}}{\partial V_{\text{in}}} - \frac{\partial I_{\text{R}}}{\partial V_{\text{in}}}\right]\right|_{V_{\text{in}}=0} = g_m\left(\frac{I_{\text{T}}}{2}\right). \tag{26}$$

Using Equation 26 to linearize the last step of Equation 17 about $V_{\text{in},n} = 0$,

$$\begin{aligned} V_{\text{out}} &\approx -R_L V_{\text{in},n}\left[g_{m,n}\left(\frac{I_{\text{L},n-1}}{2}\right) - g_{m,n}\left(\frac{I_{\text{R},n-1}}{2}\right)\right] \\ &= -R_L V_{\text{in},n}\left[g_{m,n}\left(\frac{I_{\text{AL},n-1} + I_{\text{BL},n-1}}{2}\right) - g_{m,n}\left(\frac{I_{\text{AR},n-1} + I_{\text{BR},n-1}}{2}\right)\right]. \end{aligned} \tag{27}$$

Noting that if $V_{\text{in},k} = 0$ for any $k \in \{0, 1, \ldots, n\}$, then $I_{\text{L},k} = I_{\text{R},k} = I_{\text{tail}}/2$, we now linearize about $V_{\text{in},n-1} = 0$ to obtain

$$\begin{aligned} V_{\text{out}} &\approx -R_L V_{\text{in},n} V_{\text{in},n-1} \cdot \frac{1}{2} \left.\frac{\partial g_{m,n}}{\partial I_{\text{bias}}}\right|_{V_{\text{in},n-1}=0} \cdot \left.\left[\frac{\partial I_{\text{AL},n-1}}{\partial V_{\text{in},n-1}} - \frac{\partial I_{\text{AR},n-1}}{\partial V_{\text{in},n-1}} + \frac{\partial I_{\text{BL},n-1}}{\partial V_{\text{in},n-1}} - \frac{\partial I_{\text{BR},n-1}}{\partial V_{\text{in},n-1}}\right]\right|_{V_{\text{in},n-1}=0} \\ &= -R_L V_{\text{in},n} V_{\text{in},n-1} \cdot \frac{1}{2} \left.\frac{\partial g_{m,n}}{\partial I_{\text{bias}}}\right|_{I_{\text{bias}}=I_{\text{tail}}/4} \cdot \left[g_{m,n-1}\left(\frac{I_{\text{L},n-2}}{2}\right) - g_{m,n-1}\left(\frac{I_{\text{R},n-2}}{2}\right)\right]. \end{aligned} \tag{28}$$

By induction, we see that

$$V_{\text{out}} \approx -R_L V_{\text{in},1}\left[g_{m,1}\left(\frac{I_{\text{L},0}}{2}\right) - g_{m,1}\left(\frac{I_{\text{R},0}}{2}\right)\right] \cdot \prod_{k=2}^{n}\left(\frac{1}{2} V_{\text{in},k} \left.\frac{\partial g_{m,k}}{\partial I_{\text{bias}}}\right|_{I_{\text{bias}}=I_{\text{tail}}/4}\right). \tag{29}$$

[††] Rigorously, one can use Equations 1 and 2 to linearize Equation 3 about $V_{\text{in}} = 0$.

DOI: 10.1002/cta

Finally, we linearize about $V_{\text{in},0} = 0$ to obtain[‡‡]

$$v_{\text{out}} = -R_L\, g_{m,0}\left(\frac{I_{\text{tail}}}{2}\right) v_{\text{in},0} \cdot \prod_{k=1}^{n}\left(\frac{1}{2} v_{\text{in},k}\, \frac{\partial g_{m,k}}{\partial I_{\text{bias}}}\bigg|_{I_{\text{bias}}=I_{\text{tail}}/4}\right) \tag{30}$$

where a lowercase variable indicates a small-signal quantity.

Here, we briefly revisit the concept from Section 4 that multiplication is a strictly nonlinear phenomenon, even within the small-signal regime. Notice from Equation 30 that the key requisite for small-signal multiplication is that the transconductances of all the switching stage transistors ($g_{m,k}$, $k > 0$) change with the bias current at the operating point. This dependency is clearly non-existent if $f(\cdot)$ is locally linear in the vicinity of the bias point $I_{\text{bias}} = I_{\text{tail}}/4$, as this also implies that $h(V_{\text{in}}, I_{\text{T}})$ does not change with I_{T} in some neighborhood of $V_{\text{in}} = 0$. For this reason, multiplication of *small-signals* is not possible if any of the switching stages are constructed from *deeply* velocity-saturated short-channel MOS transistors, for example.

Example: Square-Law MOS Multiplier

Let us apply this result to an MOS implementation of the stacked analog multiplier. Recalling that the transconductance of a square-law MOSFET is $g_m(I_{\text{bias}}) = 2\sqrt{KI_{\text{bias}}}$, we get

$$\Longrightarrow v_{\text{out}} = -\sqrt{2} R_L I_{\text{tail}} \prod_{k=0}^{n}\left(\sqrt{\frac{K_k}{I_{\text{tail}}}}\, v_{\text{in},k}\right). \tag{31}$$

5.3 Hard-Switching Inputs

Finally, we explore the case where all the inputs are large enough such that for each differential pair, the current is essentially completely switched to one side (i.e., $|V_{\text{in}}| \geq V_{\text{sw}}$). Then, it is easily seen that in this scenario, the output current is proportional to the tail current with $\zeta(V_{\text{in}}) = \text{sign}(V_{\text{in}})$. Therefore,

$$V_{\text{out}} = -R_L I_{\text{tail}} \prod_{k=0}^{n} \text{sign}\big(V_{\text{in},k}\big) \tag{32}$$

which is positive if and only if an odd number of the inputs are negative. So, if 'positive' and 'negative' are interpreted in a binary fashion, we see that the stacked analog multiplier performs an 'exclusive or' (XOR) of all the inputs.

Notice that in the hard-switching situation, exactly one transistor from each stage is on, and therefore the entirety of the tail current will flow through a single path from the supply to ground (out of 2^{2n+1} possible paths).

5.4 Mixture of Small-Signal and Hard-Switching Inputs

Let $S \subset \{0, 1, \ldots, n\}$ be a (nonempty) index set. Suppose it is known that $V_{\text{in},k}$ is a small-signal if $k \in S$, whereas $V_{\text{in},k}$ is a large, hard-switching input if $k \notin S$. Denote $q := \min(S)$ as corresponding to the bottom-most stage driven by a small-signal input. Then, it can be shown that

$$v_{\text{out}} = -R_L\, g_{m,q}\left(\frac{I_{\text{tail}}}{2}\right) v_{\text{in},q} \cdot \prod_{\substack{k\in S\\ k\neq q}}\left(\frac{1}{2} v_{\text{in},k}\, \frac{\partial g_{m,k}}{\partial I_{\text{bias}}}\bigg|_{I_{\text{bias}}=I_{\text{tail}}/4}\right) \cdot \prod_{k\notin S} \text{sign}\big(V_{\text{in},k}\big). \tag{33}$$

6 SIMULATION RESULTS

Here, we used Spectre to run transient simulations on a stacked analog multiplier with $n = 3$ (4 inputs) implemented using identical NMOS transistors on a 55-nm bulk CMOS process. A load of $R_L = 1\,\text{k}\Omega$ was used, and the tail was implemented with an ideal current source.

[‡‡]As a computational note, in light of Equation 25,

$$\frac{\partial g_m}{\partial I_{\text{bias}}} = \frac{f''(V)}{f'(V)} = \frac{f''\big(f^{-1}(I_{\text{bias}})\big)}{g_m(I_{\text{bias}})}.$$

DOI: 10.1002/cta

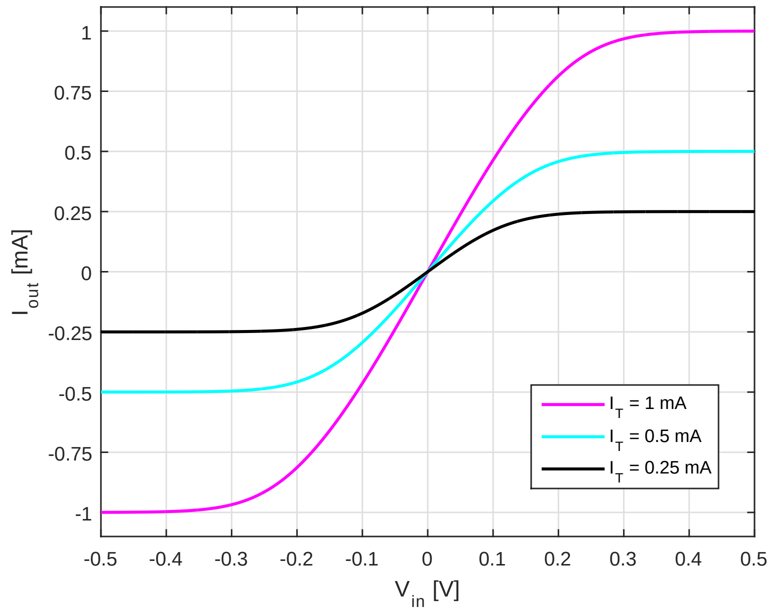


Figure 6: A simulated plot of the differential output current vs. the differential input voltage for various tail currents.

For a differential pair composed on this technology, the output current $h(V_{\text{in}}, I_{\text{T}})$ as a function of the input voltage V_{in} was simulated and is depicted in Figure 6 for various values of the tail current I_{T}.

6.1 Sinusoidal Inputs with Different Phases

Figure 7 shows the simulation result for sinusoidal input voltages, all of the same frequency, spaced at a phase of $\pi/2$ apart from one another:

$$V_{\text{in},k} = V_{\text{amp}} \sin\left(\omega_{\text{in}} t + \frac{\pi}{2} k\right) \tag{34}$$

where $V_{\text{amp}} = 100\,\text{mV}$, $f_{\text{in}} = 250\,\text{MHz}$, and $k = 0, 1, 2, 3$. Thus, in the small-signal limit, the output voltage is proportional to

$$-\prod_{k=0}^{3} \sin\left(\omega_{\text{in}} t + \frac{\pi}{2} k\right) \propto [\cos(4\omega_{\text{in}} t) - 1], \tag{35}$$

which is observed in the simulated output. Of course, the sine wave is not perfect, as Equation 35 is a small-signal limit.

6.2 Sinusoidal Inputs at Multiple Frequencies

Figure 8 shows the simulation result for sinusoidal input voltages at different frequencies:

$$V_{\text{in},k} = V_{\text{amp}} \sin[(k+1)\omega_0 t] \tag{36}$$

where $V_{\text{amp}} = 100\,\text{mV}$, $f_0 = 100\,\text{MHz}$, and $k = 0, 1, 2, 3$. Thus, in the small-signal limit, it can be shown that the output voltage consists of equally strong harmonics at DC, $6f_0$, $8f_0$, and $10f_0$:

$$\prod_{k=0}^{3} \sin[(k+1)\omega_0 t] \propto 1 - \cos(6\omega_0 t) - \cos(8\omega_0 t) + \cos(10\omega_0 t)\,. \tag{37}$$

Accounting for gain compression (Figure 6), we can use Equation 30 to roughly estimate the amplitude of the outputs seen in Figures 7 and 8. Simulations reveal $g_m(0.25\,\text{mA}) = 3.45\,\text{mS}$, $g_m{}'(0.125\,\text{mA}) = 12.9\,\text{V}^{-1}$, and an

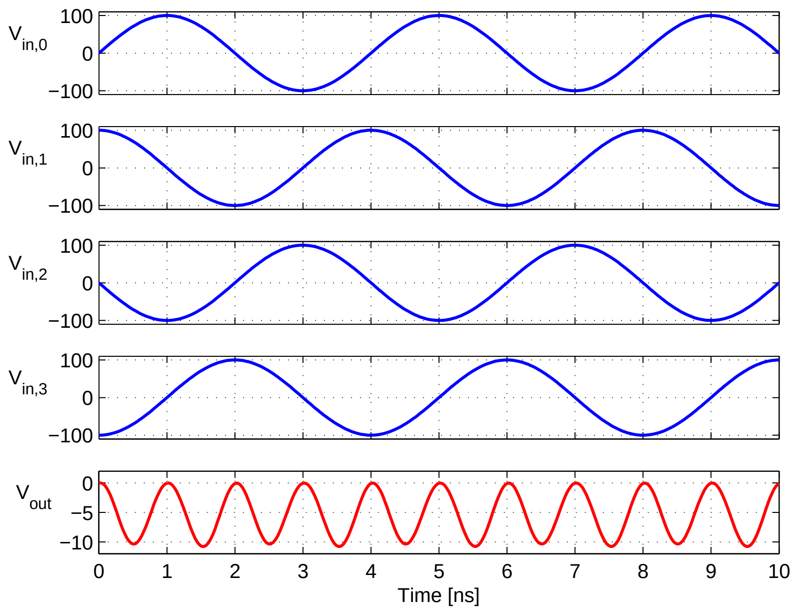


Figure 7: Sinusoidal inputs with different phases. Voltages are in units of [mV]. The tail current is 0.5 mA.

average per stage gain compression factor of 0.83 (defined as the factor by which the differential pair output current is reduced from $g_m V_{\text{in}}$). The constant of proportionality for Equations 35 and 37 is $1/8$. This results in an amplitude (Figure 8b) of

$$\frac{1}{8} \times (0.83)^4 \times 1\,\text{k}\Omega \times 3.45\,\text{mS} \times \left(\frac{12.9\,\text{V}^{-1}}{2}\right)^3 \times (0.1\,\text{V})^4 = 5.49\,\text{mV},$$

or a peak-to-peak amplitude (Figure 7) of 11 mV, which are reasonably close to the simulated amplitudes. Note also that our analysis does not account for the current consumed by the output resistance of the transistors, which further decreases the output amplitude.

6.3 Hard-Switching Square-Wave Inputs

Figure 9 shows the simulation result for ± 1 V amplitude, 125 MHz frequency, square-wave inputs spaced at a delay of 1 ns apart from one another. Equation 32 therefore predicts a 500 MHz square-wave output of amplitude $\pm I_{\text{tail}} R_L = \pm 1$ V, which is observed in the simulated output. Switching delays due to the transistors' capacitances are apparent.

7 HIGHER-ORDER INTERMODULATION PRODUCTS

To quantitatively characterize this circuit's nonlinearity, we can use a multi-dimensional Taylor series expansion of the output with respect to all of the input voltages[§§]:

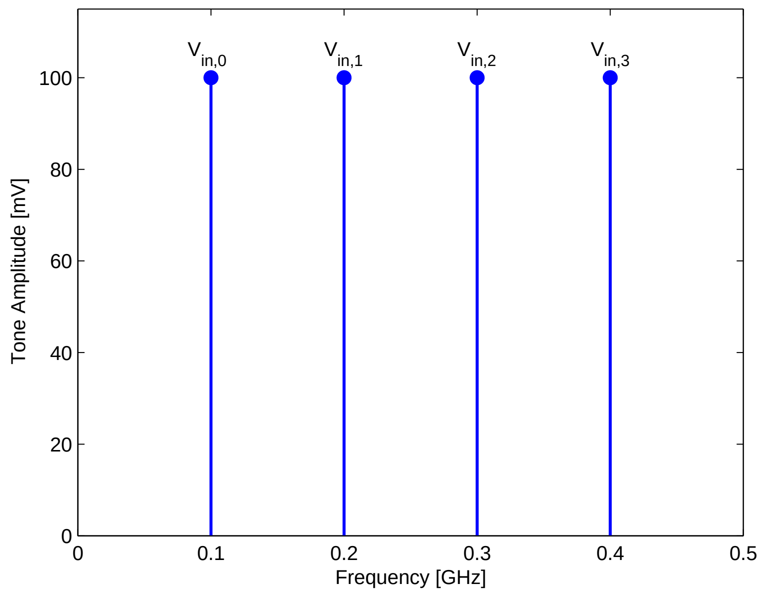
$$V_{\text{out}} = \sum_{p_0=1}^{\infty} \sum_{p_1=1}^{\infty} \cdots \sum_{p_n=1}^{\infty} G(p_0, \ldots, p_n) \prod_{k=0}^{n} V_{\text{in},k}{}^{p_k} \tag{38}$$

where the intermodulation product 'gain' is

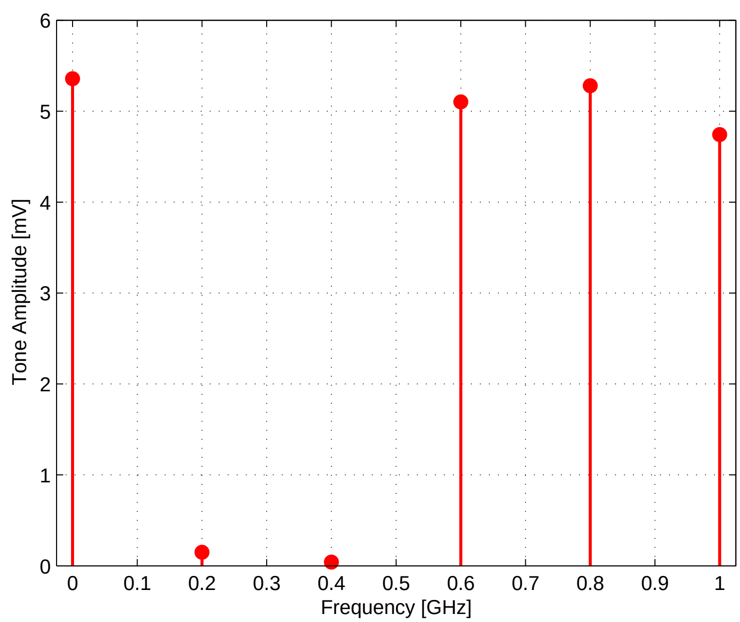
$$G(p_0, \ldots, p_n) = \frac{1}{p_0! \cdots p_n!} \cdot \left. \frac{\partial^{p_n + p_{n-1} + \cdots + p_0} V_{\text{out}}}{\partial V_{\text{in},n}{}^{p_n} \partial V_{\text{in},n-1}{}^{p_{n-1}} \cdots \partial V_{\text{in},0}{}^{p_0}} \right|_{V_{\text{in},k}=0\,\forall k}.$$

[§§]Note that the lower indices of summation are all 1 because the multiplier is balanced with respect to each input (refer to Section 2).

DOI: 10.1002/cta



(a) Inputs



(b) Output

Figure 8: Magnitude spectra of (a) the multi-frequency inputs and (b) the output. The spectrum was generated by computing a 64-point fast Fourier transform (FFT) of a 1μs transient simulation. The tail current is 0.5 mA. Roughly equally strong harmonics are seen in the output at 0 MHz (DC), 600 MHz, 800 MHz, and 1 GHz.

DOI: 10.1002/cta

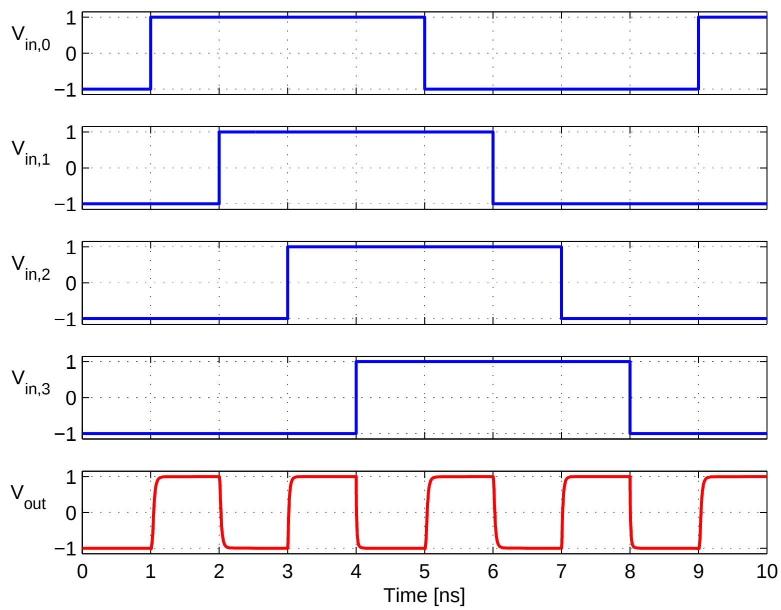


Figure 9: Square-wave inputs. Voltages are in units of [V]. The tail current is 1 mA.

It is apparent that the small-signal analysis of Section 5.2 deals with the 'first' term of this expansion (where $p_k = 1\,\forall k$). The purpose of this section is to derive the rest of the terms. The analysis is similar in spirit to that of Section 5.2, with some differences in the technical details. In order to strike a balance between accuracy and tractability, we make a crucial approximation in our analysis: Fix a stage $k \in \{1,\ldots,n\}$. Then, (1) *nonlinear* dependencies (i.e., beyond the first derivative) of the output V_{out} on the kth stage's input $V_{\text{in},k}$ and (2) *any* dependence of V_{out} on the subset of input voltages $V_{\text{in},0},\ldots,V_{\text{in},k-1}$ below the kth stage are effectively encapsulated by the kth stage's differential output current $\left(I_{\text{L},k} - I_{\text{R},k}\right)$. The precise meaning of this statement will become clear in the subsequent analysis. Physically, the intuition behind this statement follows from the circuit's balanced nature: The multiplier 'works' by generating differences between I_{L} and I_{R} at each stage; all other signals within the circuit (such as common-mode variations) should not appear at the output. Consequently, notice that this approximation is predicated on all the differential pairs within the multiplier being perfectly balanced.

First, we introduce the notation

$$M^{(p)}(I_{\text{T}}) \equiv \left.\frac{\partial^p h(V_{\text{in}}, I_{\text{T}})}{\partial V_{\text{in}}{}^p}\right|_{V_{\text{in}}=0}, \tag{39}$$

where h is defined in Equation 3. $M^{(p)}$ essentially represents pth order nonlinearities in the differential pair.

We start by differentiating the last step of Equation 17 with respect to $V_{\text{in},n}$ a total of p_n times:

$$\frac{\partial^{p_n} V_{\text{out}}}{\partial V_{\text{in},n}{}^{p_n}} = -R_L\left[\frac{\partial^{p_n} h_n(V_{\text{in},n}, I_{\text{L},n-1})}{\partial V_{\text{in},n}{}^{p_n}} - \frac{\partial^{p_n} h_n(V_{\text{in},n}, I_{\text{R},n-1})}{\partial V_{\text{in},n}{}^{p_n}}\right]. \tag{40}$$

Next, we use the chain rule to differentiate with respect to $V_{\text{in},n-1}$ *once*:

$$\frac{\partial^{1+p_n} V_{\text{out}}}{\partial V_{\text{in},n}{}^{p_n}\partial V_{\text{in},n-1}} = -R_L\left[\frac{\partial^{1+p_n} h_n(V_{\text{in},n}, I_{\text{L},n-1})}{\partial I_{\text{L},n-1}\partial V_{\text{in},n}{}^{p_n}}\frac{\partial I_{\text{L},n-1}}{\partial V_{\text{in},n-1}} - \frac{\partial^{1+p_n} h_n(V_{\text{in},n}, I_{\text{R},n-1})}{\partial I_{\text{R},n-1}\partial V_{\text{in},n}{}^{p_n}}\frac{\partial I_{\text{R},n-1}}{\partial V_{\text{in},n-1}}\right]. \tag{41}$$

This is where our above approximation comes into play: We assume the quantity $(I_{\text{L},n-1} - I_{\text{R},n-1})$ dominates all *higher-order* variations of the output with respect to $V_{\text{in},n-1}$ and *all* variations with respect to the lower input voltages $V_{\text{in},n-2},\ldots,V_{\text{in},0}$. Because the final answer is to be evaluated at the operating point where all the inputs are nulled, we

DOI: 10.1002/cta

decompose Equation 41 using

$$\frac{\partial^{1+p_n} h_n(V_{\text{in},n}, I_{\text{L},n-1})}{\partial I_{\text{L},n-1} \partial {V_{\text{in},n}}^{p_n}} \approx \frac{\partial^{1+p_n} h_n(V_{\text{in},n}, I_{\text{R},n-1})}{\partial I_{\text{R},n-1} \partial {V_{\text{in},n}}^{p_n}} \approx \left. \frac{\partial M_n^{(p_n)}(I_\text{T})}{\partial I_\text{T}} \right|_{I_\text{T} = {}^{I_\text{tail}}/_2} \tag{42}$$

which leads to

$$\Longrightarrow \left. \frac{\partial^{1+p_n} V_\text{out}}{\partial {V_{\text{in},n}}^{p_n} \partial V_{\text{in},n-1}} \right|_{V_{\text{in},n}=0} \approx -R_L \left[\left. \frac{\partial M_n^{(p_n)}(I_\text{T})}{\partial I_\text{T}} \right|_{I_\text{T} = {}^{I_\text{tail}}/_2} \cdot \frac{\partial (I_{\text{L},n-1} - I_{\text{R},n-1})}{\partial V_{\text{in},n-1}} \right]. \tag{43}$$

The first term in the brackets represents the output's dependence on the nth input; the second term captures the rest of the inputs. Noting that $I_{\text{L},k} - I_{\text{R},k} = h_k(V_{\text{in},k}, I_{\text{L},k-1}) - h_k(V_{\text{in},k}, I_{\text{R},k-1})\ \forall k = 1, \ldots, n$ and then differentiating the output voltage with respect to $V_{\text{in},n-1}$ another $p_{n-1} - 1$ times results in

$$\begin{aligned} \left. \frac{\partial^{p_n + p_{n-1}} V_\text{out}}{\partial {V_{\text{in},n}}^{p_n} \partial {V_{\text{in},n-1}}^{p_{n-1}}} \right|_{V_{\text{in},n}=0} &\approx -R_L \left. \frac{\partial M_n^{(p_n)}(I_\text{T})}{\partial I_\text{T}} \right|_{I_\text{T} = {}^{I_\text{tail}}/_2} \\ &\quad \cdot \left[\frac{\partial^{p_{n-1}} h_{n-1}(V_{\text{in},n-1}, I_{\text{L},n-2})}{\partial {V_{\text{in},n-1}}^{p_{n-1}}} - \frac{\partial^{p_{n-1}} h_{n-1}(V_{\text{in},n-1}, I_{\text{R},n-2})}{\partial {V_{\text{in},n-1}}^{p_{n-1}}} \right], \end{aligned} \tag{44}$$

Comparison with Equation 40 allows us to use induction to obtain

$$\begin{aligned} \left. \frac{\partial^{p_n + p_{n-1} + \cdots + p_1} V_\text{out}}{\partial {V_{\text{in},n}}^{p_n} \partial {V_{\text{in},n-1}}^{p_{n-1}} \cdots \partial {V_{\text{in},1}}^{p_1}} \right|_{V_{\text{in},n}, \ldots, V_{\text{in},2}=0} &\approx -R_L \prod_{k=2}^{n} \left. \frac{\partial M_k^{(p_k)}(I_\text{T})}{\partial I_\text{T}} \right|_{I_\text{T} = {}^{I_\text{tail}}/_2} \\ &\quad \cdot \left[\frac{\partial^{p_1} h_1(V_{\text{in},1}, I_{\text{L},0})}{\partial {V_{\text{in},1}}^{p_1}} - \frac{\partial^{p_1} h_1(V_{\text{in},1}, I_{\text{R},0})}{\partial {V_{\text{in},1}}^{p_1}} \right]. \end{aligned} \tag{45}$$

Finally, applying the same procedure and approximation to differentiate with respect to $V_{\text{in},0}$ a total of p_0 times, noting that $I_{\text{L},0} - I_{\text{R},0} = h(V_{\text{in},0}, I_\text{tail})$, and evaluating at $V_{\text{in},0} = 0$ gives us

$$\left. \frac{\partial^{p_n + p_{n-1} + \cdots + p_0} V_\text{out}}{\partial {V_{\text{in},n}}^{p_n} \partial {V_{\text{in},n-1}}^{p_{n-1}} \cdots \partial {V_{\text{in},0}}^{p_0}} \right|_{V_{\text{in},k}=0\, \forall k} \approx -R_L \prod_{k=1}^{n} \left. \frac{\partial M_k^{(p_k)}(I_\text{T})}{\partial I_\text{T}} \right|_{I_\text{T} = {}^{I_\text{tail}}/_2} \cdot M_0^{(p_0)}(I_\text{tail}). \tag{46}$$

Therefore, we have the following result[¶¶]:

$$G(p_0, \ldots, p_n) \approx -\frac{1}{p_0! \cdots p_n!} \cdot R_L \prod_{k=1}^{n} \left. \frac{\partial M_k^{(p_k)}(I_\text{T})}{\partial I_\text{T}} \right|_{I_\text{T} = {}^{I_\text{tail}}/_2} \cdot M_0^{(p_0)}(I_\text{tail}). \tag{47}$$

Based on Equation 47, we can infer a very important fact: *due to the balanced nature of the multiplier, only intermodulation products involving an odd power of every single input voltage will appear at the output.* This is because $h(V_\text{in}, I_\text{T})$ is an odd function of V_in, and the derivative of an odd function is an even function (and vice versa). As a result, only when p is odd will $M^{(p)}(\cdot)$ be nonzero.

Application: Gilbert cell RF input third-order IMP

To demonstrate the validity of the higher-order IMP analysis, we will use Equation 47 to examine the third-order intermodulation product of the radio frequency (RF) input for a standard down-conversion Gilbert cell mixer (Figure 1), and then compare the resulting calculation against simulation. Specifically, we will find the RF input voltage V_RF^* at which the mixing term $V_\text{RF} V_\text{LO}$ and the cubic nonlinearity ${V_\text{RF}}^3 V_\text{LO}$ have the same magnitude:

$$|G(1,1)|\, V_\text{RF}^* V_\text{LO} = |G(3,1)|\, {V_\text{RF}^*}^3\, V_\text{LO}. \tag{48}$$

¶¶Based on the fact that $M^{(1)}(I_\text{T}) = g_m(I_\text{T}/2)$, it is easy to check that Equation 47 reduces to Equation 30 when $p_k = 1\, \forall k = 0, \ldots, n$.

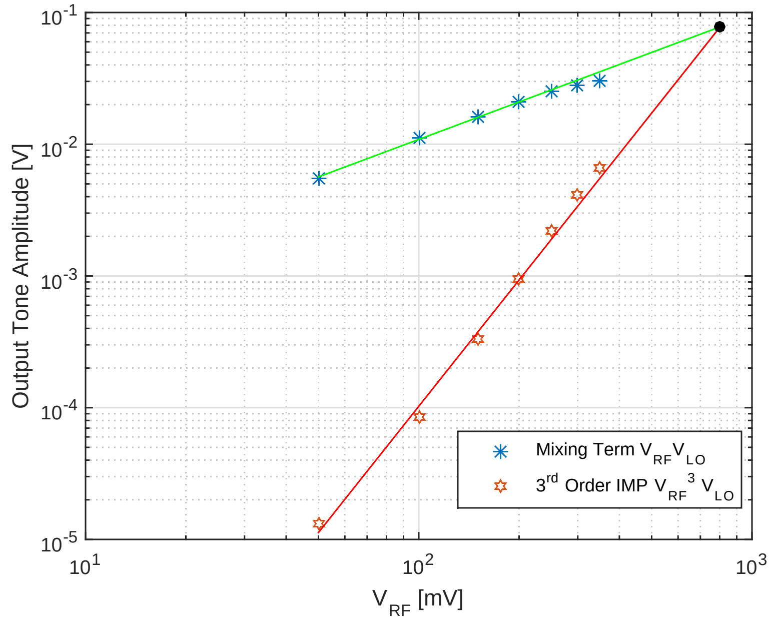


Figure 10: Comparison of fundamental mixing term with third-order IMP. Linear extrapolations on the data were used to obtain the intersection point, which occurs at $V^*_{\mathrm{RF}} = 803\,\mathrm{mV}$.

Solving for V^*_{RF} and using Equation 47 along with footnote ¶¶ gives us the simple expression

$$V^*_{\mathrm{RF}} = \sqrt{\left|\frac{G(1,1)}{G(3,1)}\right|} = \sqrt{\frac{6\,g_{m,\mathrm{RF}}}{\left|M^{(3)}_{\mathrm{RF}}\right|}}, \tag{49}$$

where the bias currents and voltages can be inferred from Equations 30 and 47.

Next, we used SpectreRF to run transient simulations of this circuit, which was implemented using identical NMOS transistors on a 55-nm bulk CMOS process. A load resistance of $R_L = 1\,\mathrm{k}\Omega$ was used, and the tail current $I_{\mathrm{tail}} = 1\,\mathrm{mA}$ was implemented with an ideal current source. To simulate V^*_{RF}, the input frequencies were set to $f_{\mathrm{LO}} = 190\,\mathrm{MHz}$ and $f_{\mathrm{RF}} = 200\,\mathrm{MHz}$, and the local oscillator (LO) amplitude was set to $V_{\mathrm{LO}} = 10\,\mathrm{mV}$. Figure 10 shows the appropriately scaled*** amplitudes of the output tones (obtained by computing a discrete Fourier transform (DFT) of the output waveform over 100 ns) at $f_{\mathrm{RF}} - f_{\mathrm{LO}} = 10\,\mathrm{MHz}$ (the mixing term) and at $3f_{\mathrm{RF}} - f_{\mathrm{LO}} = 410\,\mathrm{MHz}$ (the cubic nonlinearity) as a function of the RF amplitude.

Transistor simulations reveal $g_{m,\mathrm{RF}}(0.5\,\mathrm{mA}) = 5.588\,\mathrm{mS}$ and $\left|M^{(3)}_{\mathrm{RF}}(1\,\mathrm{mA})\right| = 55.63\,\mathrm{mA/V^3}$. This leads to a theoretically computed input intercept point of $V^*_{\mathrm{RF}} = 776\,\mathrm{mV}$, which is very close to the simulated value of 803 mV.

Finally, by replacing V_{LO} with $V_{\mathrm{LO}}\cos(\omega_{\mathrm{LO}}t)$ and V^*_{RF} with $V_{\mathrm{RF}}\cos(\omega_1 t) + V_{\mathrm{RF}}\cos(\omega_2 t)$, it is easily seen that the RF amplitude for which the output tones at $\omega_{1,2} - \omega_{\mathrm{LO}}$ and $(2\omega_{1,2} - \omega_{2,1}) - \omega_{\mathrm{LO}}$ have the same amplitude, which characterizes the RF input's two-tone third-order intercept point (IIP3), is given by $V_{\mathrm{RF}}|_{\mathrm{IIP3}} = \left(2/\sqrt{3}\right) V^*_{\mathrm{RF}}$.

***The amplitude of the tone at $3f_{\mathrm{RF}} - f_{\mathrm{LO}}$ was scaled by a factor of 4 because

$$\cos(x)\cos(y) = \frac{1}{2}\cos(x-y) + \text{other harmonics}$$

whereas

$$\cos^3(x)\cos(y) = \frac{1}{8}\cos(3x-y) + \text{other harmonics}.$$

8 FREQUENCY RESPONSE—BANDWIDTH ANALYSIS

While the previous section generalized the low-frequency, small-signal analysis of Section 5.2 by considering *higher-order* intermodulation products between the inputs, in this section, we extend the discussion of Section 5.2 along a different direction by looking at high-frequency behavior. Specifically, we will develop a simple framework for computing the small-signal 'bandwidth' of the stacked analog multiplier. To that end, let us assume small-signal sinusoidal inputs and look at how device parasitics influence the amplitude of the output. Although a bandwidth analysis of frequency mixers is typically somewhat complicated due to their time-varying nature [18, 19], we will introduce a reasonable approximation here that both significantly simplifies the analysis and imparts insight into some of the circuit's bandwidth limiting factors. Namely, in the presence of a sufficiently large input drive resistance R_S, the response of a differential pair to changes in its tail current (a common base/gate response) is significantly faster[†††] than to changes in the differential input voltage (a common emitter/source response). What this means for our multiplier is that the kth stage's dynamics can essentially be estimated by the action of the stage's (zero-value) time constant on its input frequency [20]; the time-varying tail currents $I_{\mathrm{L},k-1}$ and $I_{\mathrm{R},k-1}$ serve only to periodically but instantaneously switch, or modulate, the polarity of the output current $\left(I_{\mathrm{L},k} - I_{\mathrm{R},k}\right)$.

Under this framework, an input voltage $v_{\mathrm{in},k}$ to stage k at frequency $\omega_{\mathrm{in},k}$ is, to the first order, effectively altered by the following amplitude frequency response factor [8, 20]:

$$v_{\mathrm{in},k} \longrightarrow \frac{v_{\mathrm{in},k}}{\left|1 + j\omega_{\mathrm{in},k}\tau_k\right|} \tag{50}$$

where $1/\tau_k$ is the 3-dB bandwidth of stage k when it is driven 'in isolation'. This is defined as the bandwidth of the kth stage when it is removed from the multiplier and operated under the following conditions:

1. For $k > 0$, the tails $I_{\mathrm{L},k-1}$ and $I_{\mathrm{R},k-1}$ are connected to DC current sources with a small DC difference between them (otherwise the output would be zero).

2. The output port (i.e., where $I_{\mathrm{L},k}$ and $I_{\mathrm{R},k}$ flow into) is loaded with the equivalent 'load' resistance seen by the kth stage within the multiplier.

Alternatively, this definition is approximately equivalent to the bandwidth of the multiplier with respect to $v_{\mathrm{in},k}$ when all other inputs are excited by small DC voltages. Theoretically, τ_k can be estimated by the stage's zero-value time constant sum [20]

$$\tau_k = 2R_{S,k}\left[C_{\mathrm{gs},k} + \left(1 + g_{m,k}R_{L,k}\right)C_{\mathrm{gd},k}\right] + 2R_{L,k}C_{\mathrm{gd},k} \tag{51}$$

where R_S is the differential input voltage source resistance and $R_{L,k}$ is the equivalent single-ended load resistance seen by the kth stage. Clearly $R_{L,n} = R_L$; for $k < n$, $R_{L,k} \approx 1/2g_{m,k+1}$. Also note that for $k = 0$, the above expression should be halved because there are half as many parasitics.

Next, any load capacitor (effective or explicit) that appears in parallel with the load resistance R_L will be encountered by the output currents $I_{\mathrm{L},n}$ and $I_{\mathrm{R},n}$. Therefore, assuming a single-ended load capacitance C_L, we also have the following transformation [8]:

$$R_L \longrightarrow \left|R_L \middle\| \frac{1}{j\omega_{\mathrm{out}}C_L}\right| = \frac{R_L}{\left|1 + j\omega_{\mathrm{out}}R_LC_L\right|} \tag{52}$$

where ω_{out} is the frequency of the output tone of interest.

Note that these expressions are not meant to provide the complete amplitude response as a function of frequency; we seek only to use them to provide a rough estimate of the multiplier's small-signal 'bandwidth' with respect to each input. Also, we make no attempt to look at the phase response of the multiplier.

To test our framework, we used SpectreRF to run transient simulations on the multiplier with various input configurations, the results of which are summarized in Table I. The multipliers were implemented using NMOS transistors on a 55-nm bulk CMOS process with $R_L = 1\,\mathrm{k\Omega}$ and an ideal $I_{\mathrm{tail}} = 0.5\,\mathrm{mA}$. All the input voltages were sinusoidal with an amplitude of $50\,\mathrm{mV}$, and each input featured an internal source resistance of $R_S = 2\,\mathrm{k\Omega}$. Each stages' input frequency $f_{\mathrm{in},k}$ was either held constant or depended on an 'input frequency' f_{in} that was varied; the bandwidth for

[†††] Within the domain of lumped element circuits that are driven sinusoidally, speed is spoken of in a *phase-shift* (as opposed to *delay*) sense. The 'faster' a system responds to an input, the less phase shift there is between that input and the output.

Table I: Bandwidth Simulation Results

Scenario	Theoretical BW	Simulated BW	% Deviation
2 inputs, $f_{\text{in},0} = 5\,\text{GHz}$, $f_{\text{in},1} = f_{\text{in}}$ varied	5.55 GHz	5.05 GHz	9.4%
2 inputs, $f_{\text{in},0} = 100\,\text{MHz}$, $f_{\text{in},1} = f_{\text{in}}$ varied	6.53 GHz	6.55 GHz	0.3%
2 inputs, $f_{\text{in},0} = f_{\text{in}}$ varied, $f_{\text{in},1} = 5\,\text{GHz}$	5.55 GHz	5.55 GHz	0%
2 inputs, $f_{\text{in},0} = f_{\text{in}}$ varied, $f_{\text{in},1} = 100\,\text{MHz}$	6.53 GHz	7.40 GHz	12%
2 inputs, $f_{\text{in},k} = (k+1)f_{\text{in}}\,\forall k$	2.59 GHz	2.55 GHz	1.6%
3 inputs, $f_{\text{in},k} = (k+1)f_{\text{in}}\,\forall k$	1.43 GHz	1.25 GHz	13%
4 inputs, $f_{\text{in},k} = f_{\text{in}}\,\forall k$	2.43 GHz	2.05 GHz	17%

each scenario was then defined with respect to f_{in}. The 3-dB bandwidth of the multiplier was simulated by locating the value of f_{in} for which the magnitude of the output's mixing tone at $f_{\text{out}} = \sum_{k=0}^{n} f_{\text{in},k}$ (obtained by computing a DFT of the output waveform over 20 ns) decreases by 3 dB from what it is when f_{in} is low (50 MHz). The bandwidth was theoretically estimated by assuming the amplitude of the small-signal output voltage varies with frequency according to Equations 50 and 52 as

$$v_{\text{out}} \propto \frac{1}{\sqrt{1+(\omega_{\text{out}}R_L C_L)^2}} \prod_{k=0}^{n} \frac{1}{\sqrt{1+\left(\omega_{\text{in},k}\tau_k\right)^2}}, \tag{53}$$

where each stage's time constant was obtained via AC simulation of the multiplier with the stage of interest being driven in isolation. Each stage was designed to have a 3-dB bandwidth of roughly 7.9 GHz, leading to a time constant of $\tau_k = 20\,\text{ps}\ \forall k$. The load capacitance C_L was estimated via Miller's approximation:

$$C_L = C_{\text{gd},n}\left(1 + g_{m,n}R_L\right). \tag{54}$$

From transistor simulations, we found $g_{m,n}(0.125\,\text{mA}) = 2.09\,\text{mS}$ and $C_{\text{gd},n} = 3.44\,\text{fF}$.

Looking at the simulation results, we note several shortcomings of our relatively simple theoretical framework. Our model appears to be less accurate when there are more inputs, likely due to the fact that the 'phase shift' from the tail currents of the bottommost stages to the output, which our approximation neglects by construction, starts becoming appreciable as the number of intermediate stages between them increases. Also, notice that we fail to account for the fact that when there are 2 inputs, the LO stage seems to have an inferior bandwidth compared to the RF stage even though both stages in isolation have the same bandwidth.

On a final note, it appears based upon comparison with simulation that the dynamics due to the time-varying tail currents *cannot* be simply modeled as that of an output pole acting on the intermediate output frequencies of each stage.

9 CONCLUSION

By stacking switching stages within a double-balanced current-commutating analog multiplier, a generalization of the topology that allows for the multiplication of an arbitrary number of input voltages was realized. A general framework for analyzing the circuit given any nonlinear transistor *I-V* characteristic was formulated. Throughout, we demonstrated that by employing a modular and intuitive, as opposed to brute-force and exact, analysis ideology which (1) views the system in terms of its appropriate building blocks for a given investigative context or (2) exploits the system's inherent physical or mathematical structure; we were able to temper rigor with insight and efficaciously extract the system's most important attributes and dominant characteristics. In particular, the multiplier's small-signal characteristics, higher-order intermodulation products, and frequency response were looked at. Specifically, we observed that under the right conditions, the time-varying nature of the mixing process could be abstracted away, allowing us to effectively decouple the dynamics of the individual stages from one another. Simulations and measurements that confirmed the analysis were also reported.

In passing, we would like to mention two potential advantages of this circuit that were not discussed in detail. First, the stacked nature of the multiplier's topology makes it feasible within high-voltage processes (such as those used for RF power amplifier design). Also, the low output impedance seen by each stage (except possibly the last) may result in superior bandwidth properties when compared against a mere cascade of multiple two-input analog multipliers (e.g., Gilbert cells).

Acknowledgments

The authors would like to thank A. Safaripour and R. Fatemi of Caltech for technical disucssions and also A. Taeb of Caltech for proofreading this manuscript.

References

[1] Jones HE. Dual output synchronous detector utilizing transistorized differential amplifiers. U.S. Patent 3241078 A, March 15, 1966.

[2] Gilbert B. A new wide-band amplifier technique. *IEEE Journal of Solid-State Circuits* 1968; **SC-3**(4): 353–365.

[3] Gilbert B. A precise four-quadrant multiplier with subnanosecond response. *IEEE Journal of Solid-State Circuits* 1968; **SC-3**(4):365–373.

[4] Babanezhad JN, Temes GC. A 20-V four-quadrant CMOS analog multiplier. *IEEE Journal of Solid-State Circuits* 1985; **SC-20**(6):1158–1168.

[5] Sullivan PJ, Xavier BA, Ku WH. Low voltage performance of a microwave CMOS Gilbert cell mixer. *IEEE Journal of Solid-State Circuits* 1997; **32**(7):1151–1155.

[6] MacEachern LA, Manku T. A charge-injection method for Gilbert cell biasing. *Conference Proceedings. IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No.98TH8341)* 1998; **1**:364–368.

[7] Han G, Sánchez-Sinencio E. CMOS transconductance multipliers: a tutorial, *IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing* 1998; **45**(12):1550–1563.

[8] Kuntman H, Toker A. Novel non-linear macromodel suitable for SPICE simulation of analogue multipliers realized with bipolar and CMOS technologies. *International Journal of Circuit Theory and Applications* 1999; **27**:485-495.

[9] Terrovitis MT, Meyer RG. Noise in current-commutating CMOS mixers. *IEEE Journal of Solid-State Circuits* 1999; **34**(6):772–783.

[10] Darabi H, Abidi AA. Noise in RF-CMOS mixers: a simple physical model. *IEEE Transactions on Solid-State Circuits* 2000; **35**(1):15–25.

[11] Darabi H, Chiu J. A noise cancellation technique in active RF-CMOS mixers. *IEEE Journal of Solid-State Circuits* 2005; **40**(12):2628–2632.

[12] Choma J Jr. A three-level broad-banded monolithic analog multiplier. *IEEE Journal of Solid-State Circuits* 1981; **SC-16**(4):392–399.

[13] Kimura K. Some circuit design techniques using two cross-coupled, emitter-coupled pairs. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications* 1994; **41**(5):411–423.

[14] Johnson DA, Raman S. A packaged SiGe x2 sub-harmonic mixer for U-NII band applications, *IEEE Bipolar/BiCMOS Circuits Technolology Meeting* 2001; 159–162.

[15] Chien G. Gilbert cell and method thereof. U.S. Patent 7088982 B1, August 8, 2006.

[16] Wu TH, Tseng SC, Meng CC, Huang GW. GaInP/GaAs HBT sub-harmonic Gilbert mixers using stacked-LO and leveled-LO topologies. *IEEE Transactions on Microwave Theory and Techniques* 2007; **55**(5):880–889.

[17] Wu TH, Meng CC, Huang GW. High-gain high-isolation CMFB stacked-LO subharmonic Gilbert mixer using SiGe BiCMOS technology. *Microwave and Optical Technology Letters* 2007; **49**(5):1214–1216.

[18] Johansen TK, Vidkjær J, Krozer V. Analysis and design of wide-band SiGe HBT active mixers. *IEEE Transactions on Microwave Theory and Techniques* 2005; **53**(7):2389–2397.

[19] El-Nozahi M, Sánchez-Sinencio E, Entesari K. A 20–32-GHz wideband mixer with 12-GHz IF bandwidth in 0.18-μm SiGe process. *IEEE Transactions on Microwave Theory and Techniques* 2010; **58**(11):2731–2740.

[20] Hajimiri A. Generalized time- and transfer-constant circuit analysis. *IEEE Transactions on Circuits and Systems—I: Regular Papers* 2010; **57**(6):1105–1121.

[21] Gray PR, Hurst PJ, Lewis SH, Meyer RG. Analysis and Design of Analog Integrated Circuits (5th edn). Wiley: New York, 2010.

DOI: 10.1002/cta