

A Wideband IM3 Cancellation Technique for CMOS Π - and T-attenuators

Wei Cheng, Mark S. Oude Alink, *Student Member, IEEE*, Anne Johan Annema, *Member, IEEE*, Gerard J.M. Wienk, Bram Nauta, *Fellow, IEEE*

Abstract—A wideband IM3 cancellation technique for CMOS attenuators is presented. With proper transistor width ratios, the dominant distortion currents of transistor switches cancel each other. As a result, a high IIP3 robust to PVT variations can be achieved without using large transistors. Two prototypes in a 0.16 μm standard bulk CMOS process are presented: a Π -attenuator with four discrete settings obtains +26 dBm IIP3 and +3 dBm 1dB-compression point (CP) for 50 MHz to 5 GHz with only 0.0054 mm^2 active area, and a similar T-attenuator system which obtains +27 dBm IIP3 and +13 dBm CP for 50 MHz to 5.6 GHz with only 0.0067 mm^2 active area.

Index Terms—attenuator, CMOS, IM3 cancellation, intermodulation distortion, linearity

I. INTRODUCTION

IN receiver paths and in spectrum analyzers, gain control blocks are typically used to limit the incident power to a level that the receiver circuitry can handle without degrading the linearity [1]. In transmitter paths, stringent power control is also desirable in a pre-distortion or correction loop before a power amplifier [2]. Gain control is often implemented with variable-gain amplifiers, but when only attenuation is required, attenuators based on FET transistors show superior linearity, power handling capability and power consumption [2]–[5].

The Π -attenuator and T-attenuator shown in Fig. 1 are widely-used gain-control elements [4]–[6]. For linear-in-dB controllability, the three transistors in Fig. 1(a) and (b) are used as voltage-controlled resistors. By properly changing the gate voltages of the transistors between V_{SS} and V_{DD} , a continuously-controlled signal attenuation level can be achieved while maintaining input/output matching [4]. In these attenuators, the main source of distortion is the voltage swing across the transistors [6]. Alternatively, for better linearity, the transistors can be replaced by passive resistors (which are usually much more linear) together with transistor switches as shown in Fig. 1(c) and (d). Several of these branches in parallel can then provide discrete-step attenuation [5].

A lot of effort has been devoted to improving the linearity and power handling capability of continuously-tunable attenuators [2]–[4] and discrete-step attenuators [5]. Adaptive bootstrapped body biasing [2] is used in a cascaded Π -attenuator to suppress the body-related parasitic effects and to improve 1dB-compression point (CP). The stacked-FET technique used in [3] reduces the IM3 distortion by distributing the voltage swing

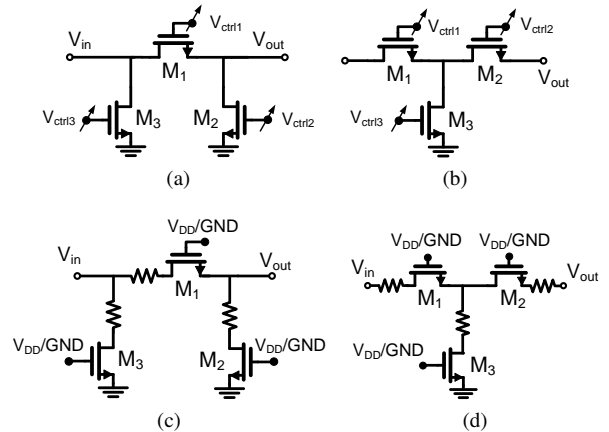


Fig. 1. Schematic of (a), (c) Π -attenuator and (b), (d) T-attenuator. (a) and (b) are continuously-tunable attenuators, (c) and (d) are discrete-step attenuators.

over many FETs in series to reduce the drain-source voltage swing per FET. However, the large parasitic capacitances of the large transistors required by this technique lower the bandwidth and increase the minimum insertion loss (IL) at high frequencies. Moreover, the capacitive nonlinearities will limit the highest achievable IIP3. Therefore, this technique is mainly effective in SOI CMOS [3]. In [6], T-attenuators are shown to be more linear than Π -attenuators, especially at higher attenuation settings. Their prototype two-stage cascaded T-attenuator obtains an IIP3 of +20 dBm in [4]. For a discrete-step implementation with low switch-on resistance to minimize distortion, the switches have to be large, resulting in less bandwidth and larger active area. In [5], a Π -attenuator with parallel branches is designed with discrete attenuation steps, which achieves +23 dBm IIP3 in the TV band.

In [7], we presented a wideband IM3 cancellation technique for discrete-step Π -attenuators in bulk CMOS that alleviates the tradeoff between IIP3 and transistor size (and thus bandwidth). This technique relies on canceling the distortion currents of series and shunt transistor switches, enabling highly linear attenuators without large transistor switches. In [8], similar IM3 cancellation was shown by simulations in a voltage divider. In this paper, we elaborate on this IM3 cancellation technique and show that it can be applied to both Π - and T-attenuators. We first analyze the proposed technique for both attenuator types in section II, and then discuss the effect of parasitic capacitances, nonlinear capacitances and PVT variations in section III. We verify the technique by measurements on two prototypes in section V, and we end

Manuscript received ...; revised ...

The authors are with the IC-Design Group, Centre for Telematics and Information Technology (CTIT), University of Twente, 7500 AE Enschede, The Netherlands (e-mail: w.cheng@utwente.nl).

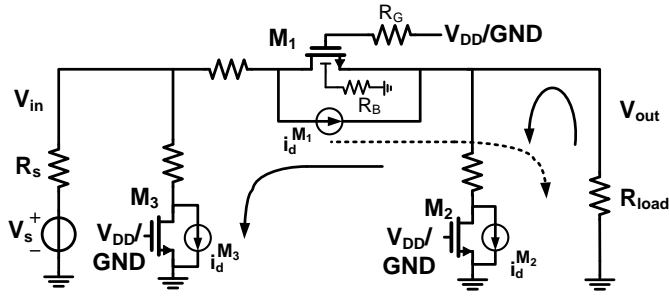


Fig. 2. Illustration of IM3 cancellation principle in a Π -attenuator.

with conclusions in section VI.

II. ATTENUATOR DISTORTION ANALYSIS

For the Π -attenuator with continuous attenuation settings, shown in Fig. 1a, higher attenuation is achieved mainly by increasing the resistance of the series device M_1 . Simultaneously, the control voltage of the shunt devices M_2 and M_3 adjusts their channel resistance for input/output matching. At high attenuation settings, the channel resistance of M_1 is large, so that a large part of the input signal drops across M_1 . As a result, the nonlinear channel resistance of M_1 generates relatively high levels of distortion.

Similarly, for the T-attenuator, shown in Fig. 1b, higher attenuation is achieved mainly by decreasing the resistance of the shunt device M_3 , while series devices M_1 and M_2 together with M_3 provide input/output matching. The channel resistance of M_3 is small to short the signal to ground, resulting in less distortion by M_3 . Consequently, T-attenuators in general are more linear than the Π -attenuators, especially at higher attenuation settings [6]. Nevertheless, for discrete-step attenuators, we will show that by properly sizing the switches, the linearity of both attenuator types can be improved to a similar level.

A. Π -attenuator

Fig. 2 shows a signal source, a Π -attenuator and its load, where the input power source is modeled as a voltage source $v_s = 2v_{in}$ (V_{IN} is the magnitude of v_{in}) with source impedance R_s . Assuming perfect matching, the input voltage for the attenuator is v_{in} and the gain is defined by $A = v_{out}/v_{in}$ (thus the attenuation is $1/A$). Large resistors in series with gate and bulk of M_1 force the gate and bulk voltages to follow the average of source and drain voltage: with sufficiently large resistors these voltages are purely AC-coupled via the parasitic capacitors of the transistors [5], resulting in $v_{gs,M_1} \approx v_{bs,M_1} \approx v_{ds,M_1}/2$. These bootstrapping resistors extend the bandwidth of the attenuator, and minimize the distortion caused by all nonlinearities related to v_{gs,M_1} and v_{bs,M_1} [5]. For M_2 and M_3 , these bootstrapping resistors are not used since they have negligible impact on all relevant performance parameters (e.g. linearity and bandwidth), as indicated in simulations. All transistors are assumed to have minimum length for maximum bandwidth.

As a first-order approximation, we assume that the distortion current between drain and source is dominant (its direction

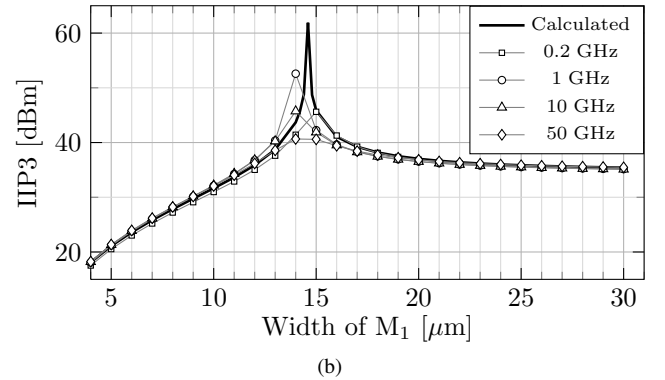
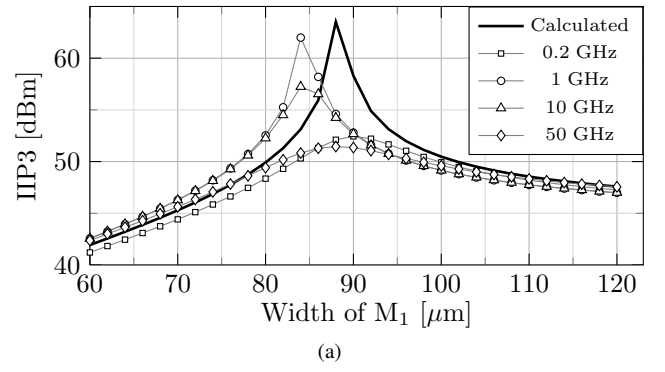


Fig. 3. Simulated (marker) and calculated (line) IIP3 as a function of W_{M_1} for two Π -attenuators ((a) for $A = -6$ dB and (b) for $A = -18$ dB) at different signal frequencies with $W_{M_2} = 20 \mu\text{m}$ and $W_{M_3} = 40 \mu\text{m}$.

is defined from drain to source) [6]. Applying the general nonlinearity model given in [8] to the Π -attenuator shown in Fig. 2 and only including the third-order nonlinearity, the voltage IM3 output is given by (see Appendix A for the derivation):

$$v_{out}^{\omega_{IM3}} \approx \frac{3V_{IN}^3}{8R_s^3(1+A)^4} \times (A^4 r_{on,M_1}^4 [16Y_{030,M_1} + 8Y_{120,M_1} + 4Y_{210,M_1} + 2Y_{300,M_1}] - A[1-A]^4 [A^2 r_{on,M_2}^4 Y_{030,M_2} + r_{on,M_3}^4 Y_{030,M_3}]) \quad (1)$$

where Y_{nml} is the nonlinear admittance between drain and source [8], defined as

$$Y_{nml} = G_{nml} + j\omega_{IM3} C_{nml}$$

with

$$G_{nml} = \frac{1}{n!m!!} \left. \frac{\partial^{n+m+l} I_{ds}}{\partial v_{gs}^n \partial v_{ds}^m \partial v_{bs}^l} \right|_{\substack{v_{gs}=v_{GS}, \\ v_{ds}=v_{DS}, \\ v_{bs}=v_{BS}}} \quad (2)$$

$$C_{nml} = \frac{1}{n!m!!} \left. \frac{\partial^{n+m+l} Q_d}{\partial v_{gs}^n \partial v_{ds}^m \partial v_{bs}^l} \right|_{\substack{v_{gs}=v_{GS}, \\ v_{ds}=v_{DS}, \\ v_{bs}=v_{BS}}} \quad (3)$$

The term ω_{IM3} is the frequency where the IM3 component locates, $r_{on} = 1/G_{010}$ is the small-signal on-resistance, G_{030} is the third-order output conductance nonlinearity, G_{300} is the third-order transconductance nonlinearity, and G_{210} and G_{120} are the cross-modulation nonlinearities. C_{030} , C_{300} , C_{210}

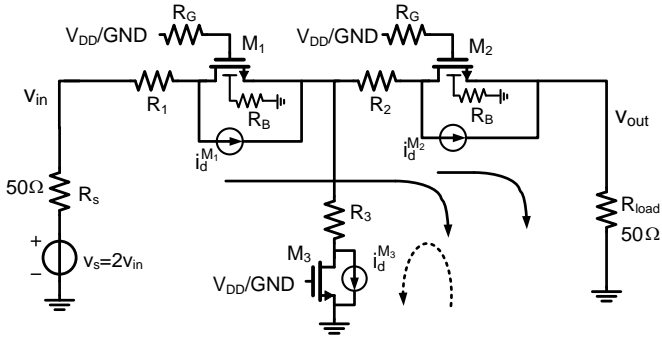


Fig. 4. Illustration of IM3 cancellation within a T-attenuator.

and C_{120} are their capacitive counterparts. As the switched-on transistors stay in the deep triode region, the third-order output admittance nonlinearity is dominant [8]. This allows for simplification of (1) into

$$v_{\text{out}}^{\omega_{\text{IM3}}} \approx \frac{3V_{\text{IN}}^3}{8R_s^3(1+A)^4} \left(16A^4 r_{\text{on},M_1}^4 Y_{030,M_1} - A[1-A]^4 [A^2 r_{\text{on},M_2}^4 Y_{030,M_2} + r_{\text{on},M_3}^4 Y_{030,M_3}] \right). \quad (4)$$

To a first-order approximation, $1/r_{\text{on}}$, G_{030} and C_{030} are proportional to transistor width W (assuming fixed channel length). Defining $r_{\text{on}} = K_r/W$, $G_{030} = K_G W$ and $C_{030} = K_C W$ we find

$$v_{\text{out}}^{\omega_{\text{IM3}}} \approx \frac{3V_{\text{IN}}^3 K_r^4}{8R_s^3(1+A)^4} (K_G + j\omega_{\text{IM3}} K_C) \times \left(\frac{16A^4}{W_{M_1}^3} - \frac{(1-A)^4 A^3}{W_{M_2}^3} - \frac{(1-A)^4 A}{W_{M_3}^3} \right) \quad (5)$$

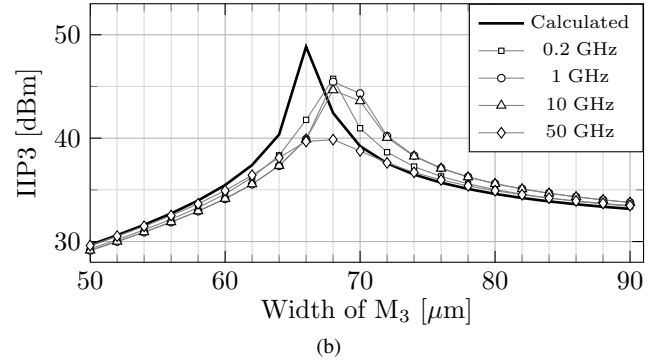
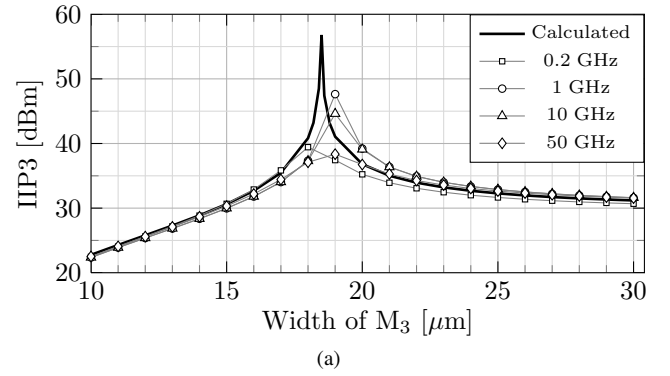
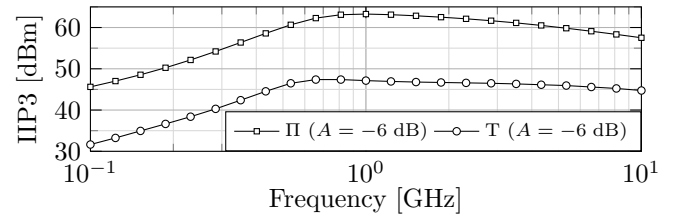
which indicates that for a certain A (note: $0 < A < 1$), the distortion current of M_1 can cancel the distortion from M_2 and M_3 for specific combinations of W_{M_1} , W_{M_2} , and W_{M_3} . This is illustrated in Fig. 2, where the distortion current of M_1 flows into the load R_{load} , while the distortion currents of M_2 and M_3 flow out of R_{load} in accordance with the sign in (4).

Solving for W_{M_1} , the switch width that yields zero IM3 when only output admittance nonlinearity is taken into account is:

$$W_{M_1,\text{opt}} \approx \frac{2^{\frac{4}{3}} A}{1-A} \left(\frac{(1-A)A^2}{W_{M_2}^3} + \frac{1-A}{W_{M_3}^3} \right)^{-\frac{1}{3}}. \quad (6)$$

Equation (4) shows that, in general, the IM3-distortion decreases with larger transistors (since the voltage swings across the transistors are smaller); equation (4) also shows that the IM3 cancellation can be achieved without requiring wide transistors: only a certain transistor dimension ratio needs to be satisfied. This latter point breaks the tradeoff between linearity and bandwidth. However, the combination of W_{M_1} , W_{M_2} , and W_{M_3} to obtain IM3 cancellation depends on the attenuation setting, so it mandates the use of a discrete-step attenuator.

Fig. 3 shows circuit simulation results for two Π -attenuators ($A = -6$ dB and $A = -18$ dB) for signal frequencies up to 50 GHz by sweeping W_{M_1} for fixed $W_{M_2} = 20 \mu\text{m}$ and


 Fig. 5. Simulated (marker) and calculated (line) IIP3 as a function of W_{M_3} for two T-attenuators (a) for $A = -6$ dB and (b) for $A = -18$ dB at different signal frequencies with $W_{M_1} = 40 \mu\text{m}$ and $W_{M_2} = 40 \mu\text{m}$.

 Fig. 6. Simulated IIP3 as a function of frequency for a Π -attenuator with 6 dB attenuation ($W_{M_1} = 88 \mu\text{m}$, $W_{M_2} = 20 \mu\text{m}$ and $W_{M_3} = 40 \mu\text{m}$) and a T-attenuator with 6 dB attenuation ($W_{M_1} = 40 \mu\text{m}$, $W_{M_2} = 40 \mu\text{m}$ and $W_{M_3} = 19 \mu\text{m}$).

$W_{M_3} = 40 \mu\text{m}$, which are rather small values.¹ W_{M_2} is chosen smaller than W_{M_3} because M_2 contributes less distortion than M_3 (see (4)) at the output of the attenuator. To keep 50Ω impedance matching and the desired attenuation, R_1 , R_2 and R_3 are set accordingly (thus R_1 is swept along with W_{M_1}). The IIP3 is extrapolated for an input power of -10 dBm with tones at $f_{\text{RF}} \pm 1.6$ MHz (thus 3.2 MHz spacing). For center frequencies up to 50 GHz, the simulation results agree very well with the simple model of (4), which validates our assumption for deriving (4) that the drain-source nonlinearity is dominant. The optimum width is well predicted, although the calculated IIP3 near the optimum is too optimistic as it neglects other nonlinearities than G_{030} and C_{030} . For

¹All simulations are performed in Spectre, using the PSP compact MOSFET model [9] fitted to our $0.16 \mu\text{m}$ CMOS process. The PSP model is known to correctly fit derivatives up to the third order [10], [11] and to satisfy the so-called Gummel symmetry test (details for this test can be found in [12], [13]), which is essential for accurate simulation of distortion.

$A = -6$ dB, the optimum W_{M_1} is $88 \mu\text{m}$ according to (5) while the optimum W_{M_1} according to simulations is between $84 \mu\text{m}$ and $90 \mu\text{m}$, depending on frequency. For $A = -18$ dB, the optimum W_{M_1} according to (5) is $15 \mu\text{m}$ while the optimum W_{M_1} is between $14 \mu\text{m}$ and $15 \mu\text{m}$ according to simulations.

For small W_{M_1} , M_1 is dominant for the IM3 output. As W_{M_1} increases, its distortion decreases and hence IIP3 increases until the IIP3 is dominated by M_2 and M_3 , yielding a saturated sub-optimum value because W_{M_2} and W_{M_3} are fixed. The IIP3 peaking area for $A = -6$ dB is less sensitive to absolute width variations than for $A = -18$ dB since larger W_{M_1} is used for $A = -6$ dB. As suggested by (4) and elaborated in section III-C, the IM3 cancellation is robust against process spread since it only relies on the ratio of transistor widths, assuming that the relative spread of passive resistors in the attenuators is small. As the switches operate in very deep triode with their gate connected to V_{DD} , the threshold voltage mismatches hardly play a role. The effect of device mismatch around the IIP3 peaking region can be reduced by increasing the width of all the switches with the same factor, at the cost of reduced bandwidth.

A similar analysis can be performed for IM2-distortion, which will yield a similar equation, but with different optimum width ratios. Because in this paper we focus on improving IIP3, it will not be further discussed here.

B. T-attenuator

A similar analysis can be performed for the T-attenuator shown in Fig. 4, yielding (see Appendix A for the derivation)

$$v_{\text{out}}^{\omega_{\text{IM3}}} \approx \frac{3V_{\text{IN}}^3}{64R_s^3} \times (-8(1-A)^4 r_{\text{on},M_3}^4 Y_{030} M_3 + Ar_{\text{on},M_1}^4 [8Y_{030,M_1} + 4Y_{120,M_1} + 2Y_{210,M_1} + Y_{300,M_1}] + A^3 r_{\text{on},M_2}^4 [8Y_{030,M_2} + 4Y_{120,M_2} + 2Y_{210,M_2} + Y_{300,M_2}]) \quad (7)$$

Since for our $0.16 \mu\text{m}$ CMOS process G_{030} contributes dominantly to the attenuator output distortion for $f_{\text{RF}} < 10$ GHz, (6) can be simplified to

$$v_{\text{out}}^{\omega_{\text{IM3}}} \approx \frac{3V_{\text{IN}}^3 K_r^4 K_G}{8R_s^3} \left(\frac{A}{W_{M_1}^3} + \frac{A^3}{W_{M_2}^3} - \frac{(1-A)^4}{W_{M_3}^3} \right) \quad (8)$$

As illustrated in Fig. 4, the distortion currents of M_1 and M_2 flow into R_{load} while the distortion current of M_3 flow out of R_{load} . The optimum width for M_3 that leads to IM3 cancellation thus is

$$W_{M_3,\text{opt}} \approx (1-A)^{\frac{4}{3}} \left(\frac{A}{W_{M_1}^3} + \frac{A^3}{W_{M_2}^3} \right)^{-\frac{1}{3}}. \quad (9)$$

A similar simulation as for the Π -attenuator is performed to illustrate the IM3-canceling behavior; the results are shown in Fig. 5. Here we swept W_{M_3} for a -6 dB and for a -18 dB setting, with $W_{M_1} = 40 \mu\text{m}$ and $W_{M_2} = 40 \mu\text{m}$, and the resistors again set to achieve a 50Ω input/output impedance matching and the targeted attenuation. The optimum width is again well predicted by the analytical expression (8). For $A = -6$ dB, $W_{M_3,\text{opt}}$ is $19 \mu\text{m}$ as predicted by (8) for the

IM3 cancellation while $W_{M_3,\text{opt}}$ in accordance with simulations is between $17 \mu\text{m}$ and $19 \mu\text{m}$, depending on frequency. For $A = -18$ dB, the calculated $W_{M_3,\text{opt}} = 66 \mu\text{m}$ and the simulated $W_{M_3,\text{opt}}$ is at $69 \mu\text{m}$.

III. LIMITING FACTORS FOR IM3 CANCELLATION

A. Parasitic capacitance

In the previous analysis, it was assumed that the distortion currents of the transistors have either 0° or 180° phase shift with respect to each other. This is a valid assumption at low frequencies, but at higher frequencies parasitic capacitances introduce a different phase shift for each distortion current, which leads to degraded distortion cancellation. This is illustrated in Fig. 6 where the simulated IIP3 of the optimized $A = -6$ dB attenuators of sections II-A and II-B is shown as a function of frequency. Clearly the IIP3 degrades for high frequencies due to the phase shift caused by parasitics. The degraded IIP3 towards low frequencies is because the AC-bootstrapping for the series devices becomes less effective. The resulting less ideal v_{gs} -variations at lower frequencies result in more distortion via the v_{gs} -related nonlinearity terms (e.g. Y_{300} , Y_{210} and Y_{120}) and therefore cause lower IIP3 at lower frequencies. Using more delicate bootstrapping circuit can make AC-bootstrapping effective at lower frequencies, but that comes at the cost of more area or power.

B. Effect of nonlinear capacitance

Each attenuator branch is optimized for IM3 cancellation at a specific attenuation value. During operation only one branch is enabled. The nonlinear parasitic capacitances of off-state transistor switches now set an upper bound on the maximum IIP3 that can be achieved by the enabled attenuator branch. As high linearity is desirable for high attenuation settings, to demonstrate the effect of nonlinear capacitances, we simulate the -24 dB attenuator branch in Fig. 7a and Fig. 7b as a function of f_{RF} , with and without the other switched-off branches connected to the system. As shown in Fig. 8, the nonlinear capacitances of the off-state switches in the disabled attenuator branches reduce the IIP3 of the enabled attenuator, especially for high frequencies.

C. Effect of Process, Voltage and Temperature Variations

As indicated by (5) and (8), the proposed IM3 cancellation technique relies on transistor width ratios, which makes it inherently fairly robust against PVT variations. For the -12 dB Π -attenuator in Fig. 7a (M_{1d} , M_{2d} , M_{3d}) and the -6 dB T-attenuator in Fig. 7b (M_{1d} , M_{2d} , M_{3d}), we ran 200 Monte Carlo simulations using a realistic production variation model at different temperatures and different V_{DD} to check the effect of PVT-variations. Fig. 9 shows the simulated range of IIP3 at 1 GHz as a function of temperature for $V_{DD} = 1.8$ V (nominal supply) and $V_{DD} = 1.5$ V. It shows, for a wide temperature range (-50°C to 100°C), that both attenuators always achieve > 30 dBm IIP3, even when the supply voltage drops to 1.5 V.

Overall, the proposed IM3 cancellation is relatively robust for PVT-variations. Wide transistor switches reduce the

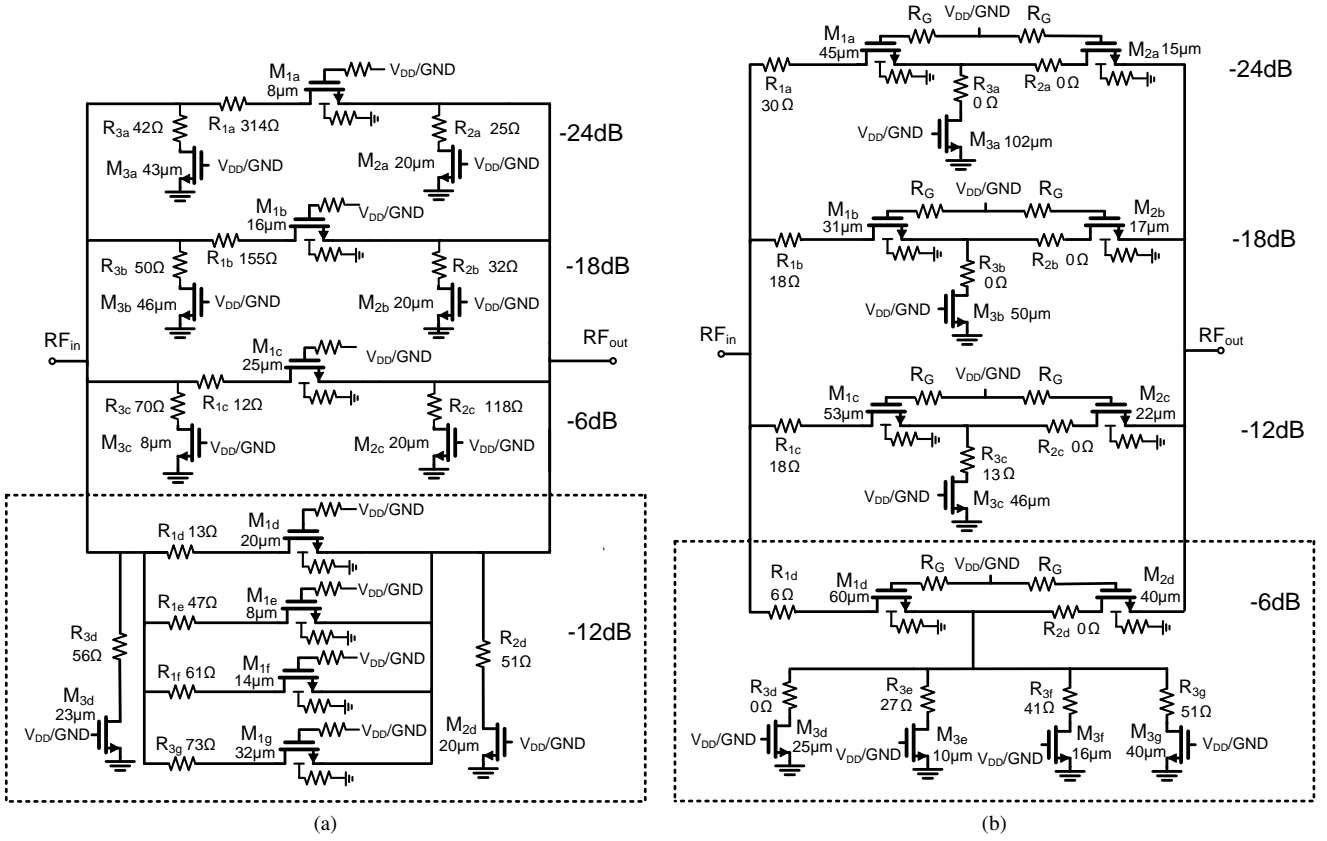


Fig. 7. The schematic of attenuator networks implemented in a 0.16 μm bulk CMOS process. (a) Π -attenuator network and (b) T-attenuator network.

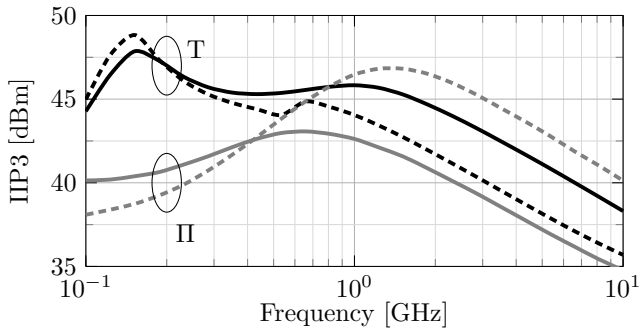


Fig. 8. Simulated IIP3 without (solid) and with (dashed) nonlinear capacitances of off-state switches taken into account for the -24 dB attenuation branch in both Π -attenuator network and T-attenuator network shown in Fig. 7.

sensitivity of IM3 cancellation over mismatch, but limit the bandwidth and introduce two factors that limit maximum achievable IIP3: capacitance nonlinearity and phase shift due the parasitic capacitances. As a result, careful optimization is necessary.

IV. DESIGN

To verify the proposed concept, the Π - and the T-attenuators of Fig. 7 are implemented in a standard 0.16 μm bulk CMOS process. Both attenuator systems contain two blocks for two different measurement purposes: 1) an attenuator block ($A = -12$ dB for the Π -attenuator, $A = -6$ dB for the T-attenuator, encircled in Fig. 7) for demonstrating the validity of (5) and (8)

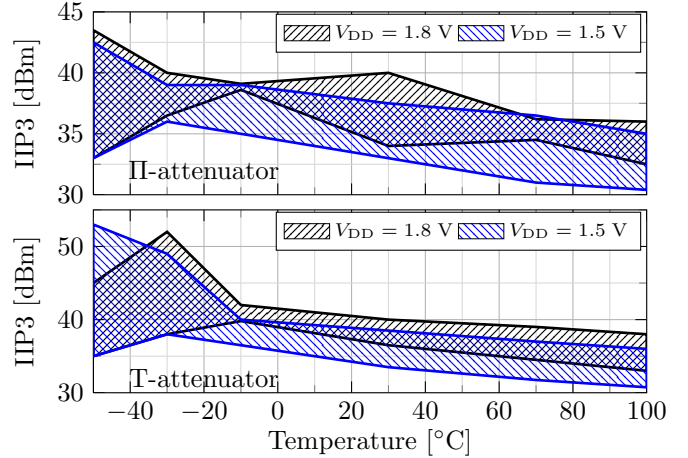


Fig. 9. The simulated IIP3 range (maximum and minimum) of 200 Monte Carlo simulations for mismatch and process spread at 1 GHz as a function of temperature when $V_{DD} = 1.8$ V (full supply) and $V_{DD} = 1.5$ V. Top for -12 dB Π -attenuator and bottom for -6 dB T-attenuator.

and 2) a four-step attenuator with 6 dB, 12 dB, 18 dB and 24 dB attenuation.

In the 12 dB attenuation block of the Π -attenuator, each of the four branches is designed for 12 dB attenuation, but has different width for M_1 to vary and verify the amount of IM3 cancellation. Therefore, this mimics a Π -attenuator with selectable W_{M_1} for fixed W_{M_2} (20 μm) and W_{M_3} (23 μm).

The four-step Π -attenuator system contains the upper three

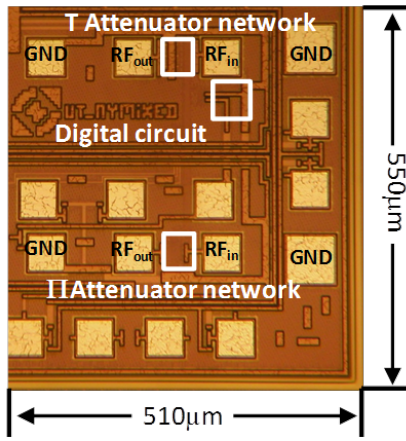


Fig. 10. The chip micrograph for both attenuators fabricated in a standard $0.16\ \mu\text{m}$ CMOS process. The IC is shared with another design, hence the additional lines and bondpads.

attenuator branches in Fig. 7a (all optimized for IM3 cancellation) and the attenuator branch in the $-12\ \text{dB}$ block that is optimized for IM3 cancellation (M_{1d} , M_{2d} , M_{3d}). During operation, only one branch is enabled. For isolation and bootstrapping purposes, the gate and bulk of M_1 are connected to the controlling voltage via $40\ \text{k}\Omega$ resistors; the gates and bulks of the shunt devices are connected directly to the controlling voltages to save area. For minimum signal attenuation, the transistors M_{1a} , M_{1b} , M_{1c} , M_{1d} are enabled, and the shunt transistors are disabled, yielding an additional $-1.8\ \text{dB}$ setting that sets the minimum IL of this system. Poly resistors are used for the series and shunt resistance in the attenuator because of their high linearity (IIP3 around $+50\ \text{dBm}$ according to simulations).

The T-attenuator is designed in a similar way, with the minimum attenuation equal to $-1.2\ \text{dB}$. A digital decoder provides the controlling voltage ($V_{\text{DD}} = 1.8\ \text{V}$ for enabling and $V_{\text{SS}} = 0\ \text{V}$ for disabling), and is shared by the attenuators. Simulated nominal IIP2 for the II-attenuator (T-attenuator) is $+55\ \text{dBm}$ ($+45\ \text{dBm}$) for all settings at $f_{\text{RF}} = 2.5\ \text{GHz}$ with a two-tone spacing up to $1\ \text{GHz}$, and can be improved by increasing the switch size.

The chip micrograph is shown in Fig. 10. The active areas of the digital decoder (not optimized for minimum area), the II-attenuator system and the T-attenuator system are $60 \times 65\ \mu\text{m}^2$, $50 \times 30\ \mu\text{m}^2$ and $54 \times 53\ \mu\text{m}^2$ respectively.

V. MEASUREMENTS

The measurements are performed by on-wafer probing. The CP is extrapolated from an input power of $-20\ \text{dBm}$, and IIP3 from an input power of $-15\ \text{dBm}$ with $3.2\ \text{MHz}$ two-tone spacing. All simulations to compare with measurements include the estimated bondpad capacitances ($100\ \text{fF}$) at the input/output of the attenuator system.

A. Verification of IM3 cancellation

We use the $-12\ \text{dB}$ branch of the II-attenuator ($-6\ \text{dB}$ for the T-attenuator) to demonstrate our IM3 cancellation theory.

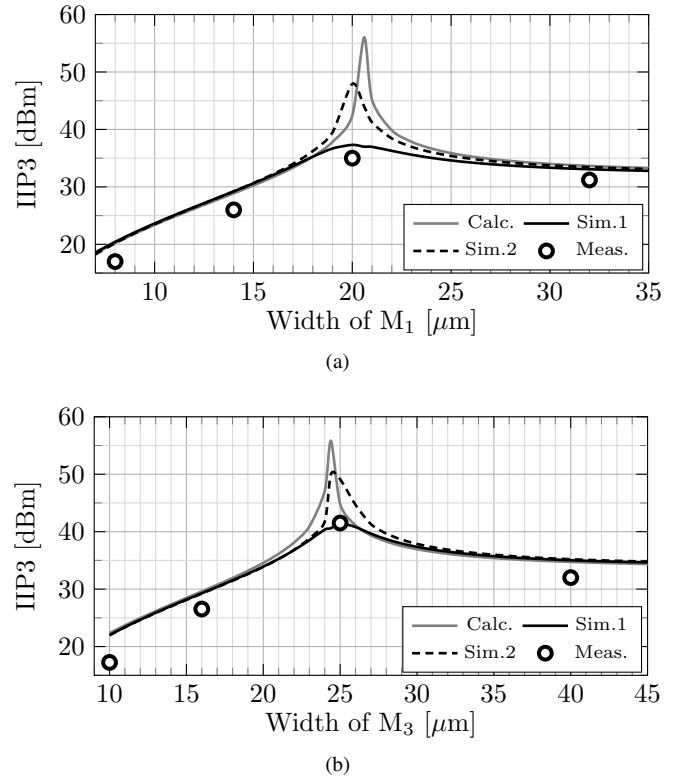


Fig. 11. Measured and simulated IIP3 for $f_{\text{RF}} = 1\ \text{GHz}$ for the (a) II-attenuator as a function of W_{M_1} (b) T-attenuator as a function of W_{M_3} . Sim.1 refers to simulation with the nonlinear capacitance of the off-state switches taken into account, Sim.2 to simulation where it is not taken into account.

The comparison between the II and T attenuator topologies is discussed in section V-B. The measured, simulated and calculated IIP3 using (4) ((7)) as a function of W_{M_1} (W_{M_3}) at $1\ \text{GHz}$ is shown in Fig. 11. The trend for the IIP3-peaking in the measurements, simulations and calculations by using our model are in good agreement. Note that the IIP3 peak for $W_{M_1} = 20\ \mu\text{m}$ ($W_{M_3} = 25\ \mu\text{m}$) is a reliable data point since the same result within $1\ \text{dB}$ variation is measured in ten randomly-chosen samples, which will be shown in section V-B. The small difference between measured and simulated IIP3 may be due to limited accuracy of transistor modeling and bondpad parasitics. For the optimum W_{M_1} (W_{M_3}), simulations show that the achievable maximum IIP3 is limited by the nonlinear capacitances of the off-state switches by more than $10\ \text{dB}$ in this particular case.

IIP3 as a function of f_{RF} is shown in Fig. 12. The measurement shows that the optimum width provides highest IIP3 up to $3\ \text{GHz}$. The difference between measurements and simulations are probably due to unaccounted parasitics. As frequency increases, simulation shows that the IIP3 improvement using optimal device size is limited by the nonlinear capacitances of the off-state switches in the attenuator networks, which is shown in Fig. 12.

Fig. 13a shows the IIP3 curves for the different W_{M_1} (W_{M_3}) at $1\ \text{GHz}$. For the particular device sizing in our implementation, it shows that the IM3 improvement becomes less effective for input powers above approximately $-8\ \text{dBm}$ due to higher-order nonlinearities. Nevertheless, the IM3-products at the

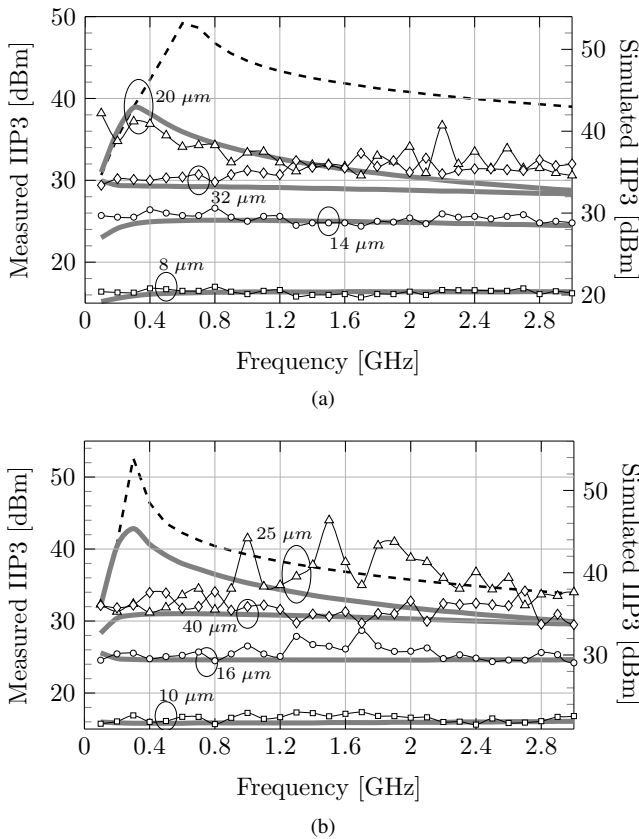


Fig. 12. Measured (markers) and simulated (lines) IIP3 vs f_{RF} for the (a) Π -attenuator for different W_{M1} (b) T-attenuator for different W_{M3} . Solid lines for simulation results with the effect of nonlinear capacitances of the off-state switches, dashed lines for simulation results without the effect of nonlinear capacitances of the off-state switches, symbols for measurement results.

TABLE I
CP-MEASUREMENTS FOR DIFFERENT TRANSISTOR WIDTHS. OPTIMUM FOR IIP3 IN BOLD.

Π -attenuator		T-attenuator	
W_{M1} [μm]	CP [dBm]	W_{M3} [μm]	CP [dBm]
8	5.6	10.0	16.5
14	12	16.0	15.8
20	15.7	25	14.5
32	16.0	40	12.5

optimum width remain the lowest up to even higher input powers. By scaling up the transistors, the voltage swing across them will be less, and thus the IM3 curve will follow the small-signal third-order behavior up to higher input powers. Table I shows that the optimum sizing affects CP by less than 2 dB.

B. Discrete-step attenuator

The measured and simulated S_{11} and S_{21} (50 Ω reference) for the different settings of the Π -attenuator (T-attenuator) are shown in Fig. 14. Due to a mistake in the decoder design, the minimum attenuation setting of -1.8 dB (-1.2 dB) cannot be enabled, so we only show their *simulated* values. Due to unaccounted parasitics, the measured S_{21} deviates > 1.6 dB for $f_{RF} > 5$ GHz ($f_{RF} > 5.6$ GHz) for the -24 dB setting. The T-attenuator requires wider shunt devices for the high attenuation settings, and thus wider series devices for IM3

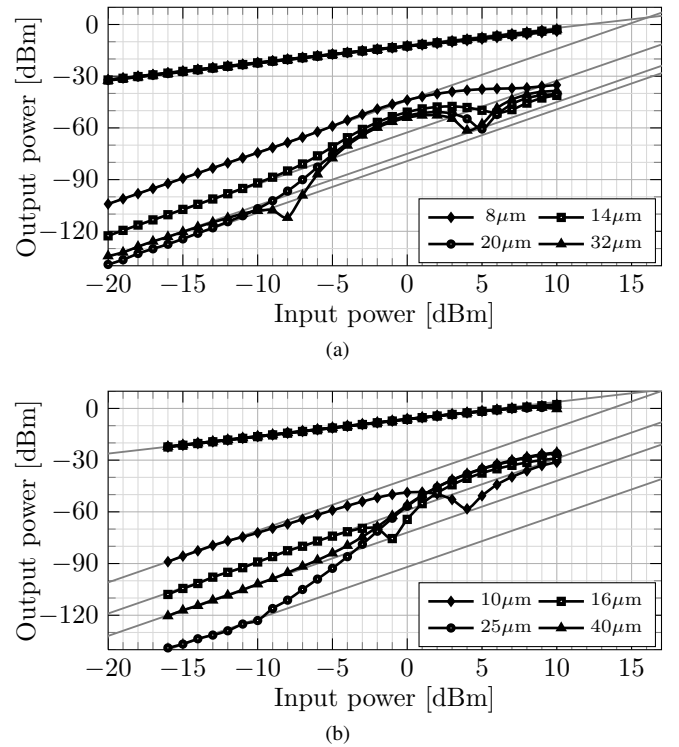


Fig. 13. Measured IIP3 curves vs input power ($f_{RF} = 1$ GHz) for the (a) Π -attenuator for the different W_{M1} and (b) T-attenuator for the different W_{M3} .

cancellation. This results in less bandwidth ($S_{11} < -10$ dB) and more area compared to the Π -attenuator system.

The measured IIP3 curves as a function of input power at $f_{RF} = 1$ GHz are shown in Fig. 15. The IIP3 for the Π -attenuator (T-attenuator) are respectively 31 dBm (40 dBm), 33 dBm (34 dBm), 38 dBm (30 dBm) and 36 dBm (35 dBm) for attenuation settings -6 dB, -12 dB, -18 dB and -24 dB. Again, for high input powers higher-order nonlinearities kick in.

Fig. 16 summarizes the measured IIP3 for various f_{RF} . Measurement and simulation results show a similar trend as frequency increases. The discrepancy may be caused by unaccounted (nonlinear) capacitances. Due to bandwidth limitations of our measurement setup, IIP3 cannot be measured below $f_{RF} = 50$ MHz. For both attenuators, IIP3 is above $+30$ dBm in the TV bands (0.05–1 GHz). The T-attenuator obtains an IIP3 $> +25$ dBm for the whole range 0.05–10 GHz, while the Π -attenuator obtains an IIP3 $> +26$ dBm for 0.05–5 GHz and IIP3 $> +24$ dBm for 0.05–10 GHz. At higher f_{RF} , extra phase shifts caused by the parasitic capacitances degrades the IM3 cancellation.

The measured IIP3 of ten dies in one wafer for $f_{RF} = 1$ GHz shows $< \pm 1.5$ dB IIP3 variation, as shown in Fig. 17. In a 200-run Monte Carlo simulation for mismatch and process spread, the difference between mean IIP3 and minimum IIP3 is smaller than 6 dB and IIP3 is higher than 32 dBm for all samples, which shows the robustness of this IM3 cancellation technique.

At $f_{RF} = 2.5$ GHz, measured curves for a two-tone spacing of 100 kHz and 30 MHz show negligible difference compared to the 3.2 MHz spacing, which is confirmed by simulations for

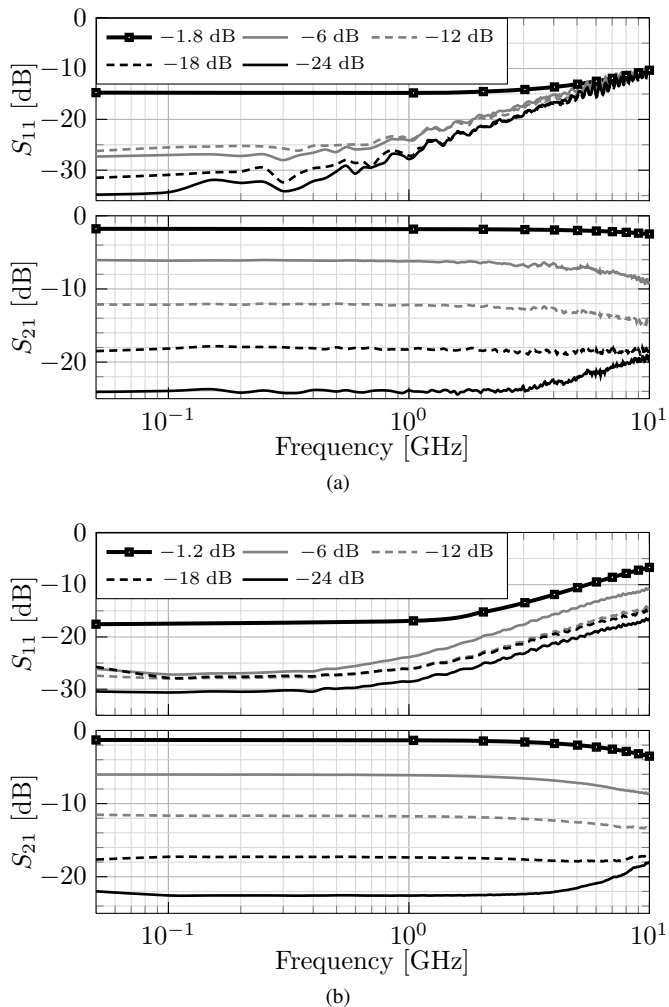


Fig. 14. Measured S_{11} and S_{21} of the four-step (a) Π -attenuator system and (b) T-attenuator system. For the minimum-attenuation settings, simulation results are shown.

a spacing from 100 kHz to 1 GHz at the same f_{RF} .

The CP of the Π -attenuator is larger than +3 dBm for the frequency range from 0.05–10 GHz, see Fig. 18. For attenuation branches $A = -18$ dB and $A = -24$ dB, M_1 is quite small and hence experiences a relatively large voltage swing, causing $CP < 10$ dBm for $f_{RF} < 1$ GHz. Using wider M_1 in these settings can increase CP. The T-attenuator has a higher CP of 11 dBm, because the two devices in series divide the voltage between v_{in} and v_{out} . Moreover, each device is wider, thus takes less voltage swing and generates less distortion. At lower frequencies the AC-bootstrapping becomes less effective, increasing v_{gs} of the series transistors, thus generating more distortion and decreasing CP. Increasing the gate series resistor can alleviate this problem, but making it too large may increase the noise and/or slow down the transition when changing the attenuation setting.

C. Benchmarking

In Table II we compare the two optimized designs with state-of-the-art attenuators. Both the Π - and T-attenuator system using the proposed IM3 cancellation technique achieve very high linearity, and simultaneously, high bandwidth for

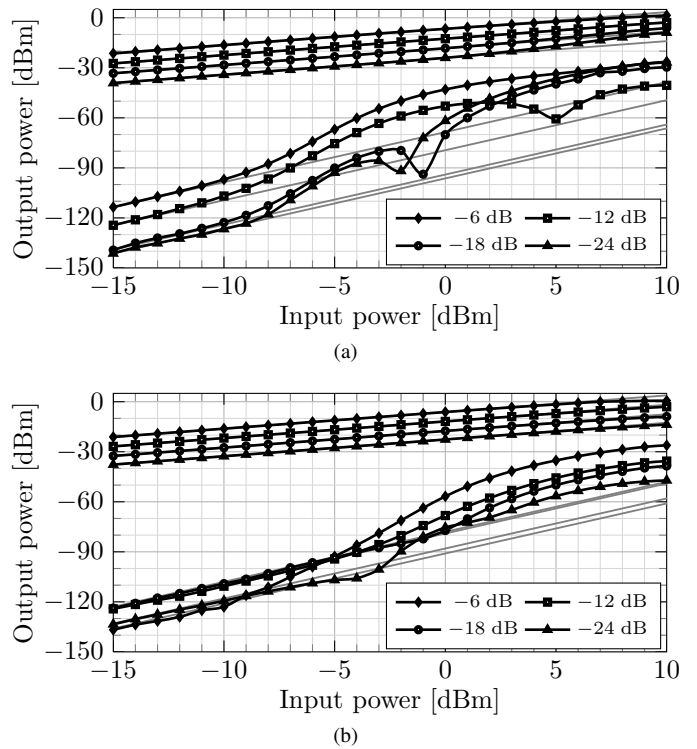


Fig. 15. Measured IIP3 curves vs input power at 1 GHz for different settings of the (a) Π -attenuator and (b) T-attenuator.

a very low active area in standard bulk CMOS. By using the proposed IM3 cancellation technique, the linearity of the Π -attenuator can be improved to the same level of the T-attenuator for similar transistor sizes.

VI. CONCLUSIONS

A wideband IM3 cancellation technique is introduced for CMOS Π -attenuators and T-attenuators. For specific transistor width ratios, the dominant distortion currents cancel at the load, which results in a high IIP3, even for relatively small transistors. Simple, yet accurate equations for transistor width dimensioning are introduced. This technique alleviates the trade-off between bandwidth and area on the one hand and high linearity on the other hand, without introducing extra devices. It thus enables highly linear wideband CMOS attenuators with small active area. A four-step Π -attenuator system designed in $0.16 \mu\text{m}$ CMOS using this IM3 cancellation achieves > 30 dBm IIP3 for the TV bands (0.05–1 GHz), > 26 dBm IIP3 for 0.05–5 GHz and > 3 dBm CP for 0.05–10 GHz, with only 0.0054 mm^2 of active area. A four-step T-attenuator system design achieves similar performance. Both measurement and simulation results show good robustness of this IM3 cancellation technique.

ACKNOWLEDGEMENTS

We thank NXP for chip fabrication donation, and G. van der Weide, E.A.M. Klumperink, M.C.M. Soer and H. de Vries for help.

TABLE II
 COMPARISON WITH STATE-OF-THE-ART ATTENUATORS.

	Huang [2]	Youssef [5]	Dogan [6]	Granger-Jones [3]	Ku [14]	This work (Π)	This work (T)
CMOS	0.18 μm	65 nm	0.13 μm	SOI	0.18 μm	0.16 μm	0.16 μm
V_{DD} [V]	1.8	1.2	1.2	5	N/A	1.8	1.8
Chip area [mm^2]	0.28	0.05	0.7	N/A	0.5	0.0054	0.0067
Bandwidth [GHz]	0.4–3.7	0.4–0.8	0.0–2.5	0.05–4.0	0.0–14	0.05–5.0	0.05–5.6
IIP3 [dBm]	+15	+23	+10	+47	+29	+30 / +27	+30 / +27
at RF [GHz]	0.7	0.4–0.8	10		10	0.05–1 / 5	0.05–1 / 5
CP ($P_{1\text{dB}}$) [dBm]	+6 (0.7 GHz)	N/A	+2.5	+30	+15 (10 GHz)	+3 (0.05–1 GHz) +10 (1.0–10 GHz)	+11 (0.05–10 GHz)
Att. flatness [dB]	2.6	N/A	2.6	3.0	0.7	1.6	1.6
Max. attenuation [dB]	33	48	42	40	31.5	24	24
Min. attenuation [dB]	0.96–2.9	N/A	0.9–3.5	2.4–4.0	3.7–10	1.8–2.4 (simulation)	1.3–2.2 (simulation)
Return loss [dB]	> 9	> 12	> 8.2	> 14	> 9	> 14	> 10
Control mode	linear-in-dB	discrete step	linear-in-dB	linear-in-dB	discrete step	discrete step	discrete step

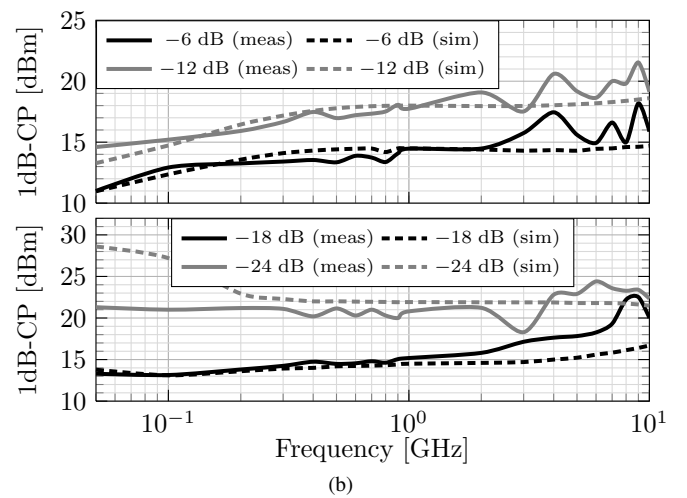
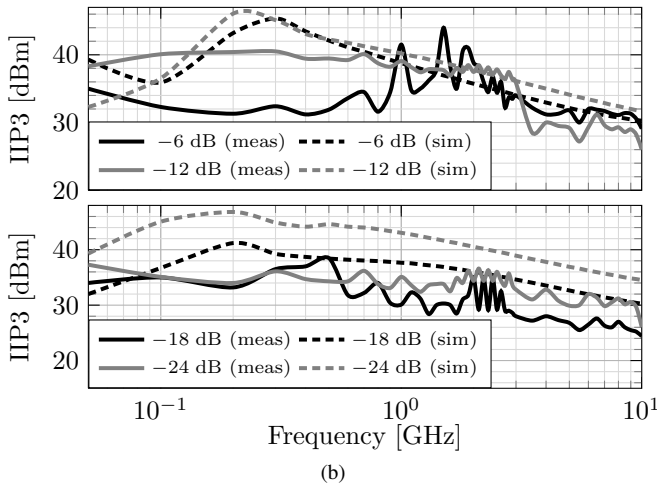
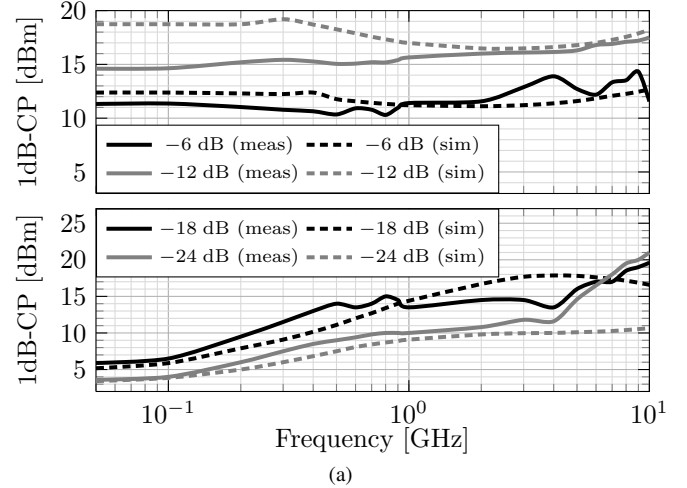
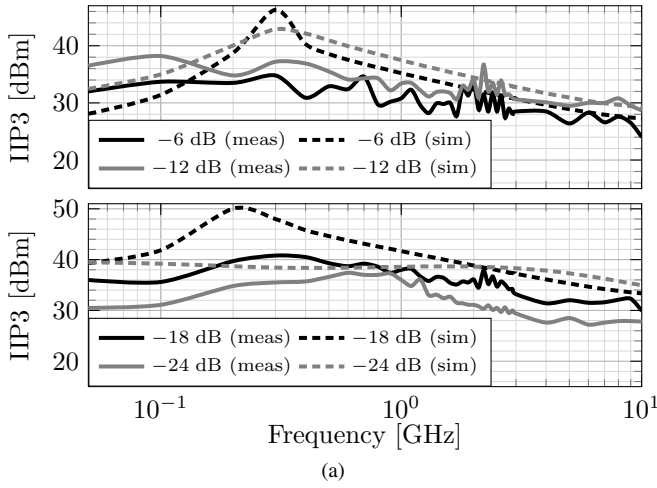

 Fig. 16. Measured and simulated IIP3 vs f_{RF} for different settings of the (a) Π -attenuator and (c) T-attenuator.

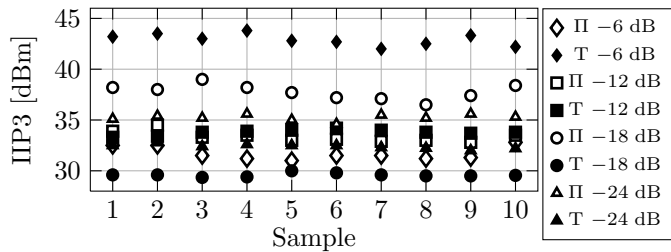
 Fig. 18. Measured CP ($P_{1\text{dB}}$) vs f_{RF} for different settings of the (a) Π -attenuator and (b) T-attenuator.


Fig. 17. Measured IIP3 for ten samples at 1 GHz for both attenuators.

REFERENCES

- [1] M. S. Oude Alink, E. A. M. Klumperink, A. B. J. Kokkeler, M. C. M. Soer, G. J. M. Smit, and B. Nauta, "A CMOS-compatible spectrum analyzer for cognitive radio exploiting crosscorrelation to improve linearity and noise performance," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 3, pp. 479–492, Mar. 2012.
- [2] Y.-Y. Huang, W. Woo, Y. Yoon, and C.-H. Lee, "Highly linear RF CMOS variable attenuators with adaptive body biasing," *IEEE J. Solid-State Circuits*, vol. 46, no. 5, pp. 1023–1033, 2011.

- [3] M. Granger-Jones, B. Nelson, and E. Franzwa, "A broadband high dynamic range voltage controlled attenuator MMIC with IIP3 > +47dBm over entire 30dB analog control range," in *Proc. IEEE MTT-S Int. Microwave Symp. Digest (MTT)*, 2011, pp. 1–4.
- [4] H. Dogan, R. G. Meyer, and A. M. Niknejad, "Analysis and design of RF CMOS attenuators," *IEEE J. Solid-State Circuits*, vol. 43, no. 10, pp. 2269–2283, 2008.
- [5] A. Youssef, J. Haslett, and E. Youssoufian, "Digitally-controlled RF passive attenuator in 65 nm CMOS for mobile TV tuner ICs," in *Proc. IEEE Int. Circuits and Systems (ISCAS) Symp.*, 2010, pp. 1999–2002.
- [6] H. Dogan and R. G. Meyer, "Intermodulation distortion in CMOS attenuators and switches," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 529–539, 2007.
- [7] W. Cheng, M. S. Oude Alink, A. J. Annema, G. J. M. Wienk, and B. Nauta, "A wideband IM3 cancellation technique for CMOS attenuators," in *Proc. IEEE Int. Solid-State Circuits Conf. - Dig. Tech. Papers*, Feb. 2012, pp. 78–79.
- [8] W. Cheng, M. S. Oude Alink, A. J. Annema, J. A. Croon, and B. Nauta, "RF circuit linearity optimization using a general weak nonlinearity model," *IEEE Trans. Circuits Syst. I*, vol. early access, no. 99, 2012.
- [9] [Online]. Available: http://www.nxp.com/models/mos_models/psp/
- [10] A. J. Scholten, G. D. J. Smit, B. A. De Vries, L. F. Tiemeijer, J. A. Croon, D. B. M. Klaassen, R. van Langevelde, X. Li, W. Wu, and G. Gildenblat, "The new CMC standard compact MOS model PSP: Advantages for RF applications," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1415–1424, 2009.
- [11] X. Li, W. Wu, A. Jha, G. Gildenblat, R. van Langevelde, G. D. J. Smit, A. J. Scholten, D. B. M. Klaassen, C. C. McAndrew, J. Watts, C. M. Olsen, G. J. Coram, S. Chaudhry, and J. Victory, "Benchmark tests for MOSFET compact models with application to the PSP model," *IEEE Trans. Electron Devices*, vol. 56, no. 2, pp. 243–251, 2009.
- [12] P. Bendix, P. Rakers, P. Wagh, L. Lemaitre, W. Grabinski, C. C. McAndrew, X. Gu, and G. Gildenblat, "RF distortion analysis with compact MOSFET models," in *Proc. Custom Integrated Circuits Conf the IEEE 2004*, 2004, pp. 9–12.
- [13] C. C. McAndrew, "Validation of MOSFET model source–drain symmetry," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2202–2206, 2006.
- [14] B.-H. Ku and S. Hong, "6-bit CMOS digital attenuators with low phase variations for X-band phased-array systems," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 7, pp. 1651–1663, 2010.

APPENDIX A

ATTENUATOR DISTORTION DERIVATION

In this section we neglect the linear capacitances of the transistors under the assumption that the attenuator is operating in the frequency far below f_t . Nevertheless, the capacitive nonlinearity of the transistor switches is included during derivation. Applying the general nonlinearity model given in [8] to both the Π -attenuator of Fig. 2 and the T-attenuator of Fig. 4, and only including the third-order nonlinearity yields

$$\begin{aligned} v_{\text{out}}^{\omega_{\text{IM}3}} &\approx H_{M_1} i_d^{M_1} + H_{M_2} i_d^{M_2} + H_{M_3} i_d^{M_3} \\ &= \frac{3}{4} V_{\text{IN}}^3 (H_{M_1} [\zeta_{030,M_1} + \zeta_{300,M_1} + \zeta_{210,M_1} + \zeta_{120,M_1}] \\ &\quad + H_{M_2} \zeta_{030,M_2} + H_{M_3} \zeta_{030,M_3}) \end{aligned} \quad (10)$$

where $H_{M_k} = v_{\text{out}}/i_d^{M_k}$ and $\zeta_{nml,M_k} = \beta_{nml,M_k} Y_{nml,M_k}$, with $\beta_{nml,M_k} = v_{\text{gs},M_k}^n v_{\text{ds},M_k}^m v_{\text{bs},M_k}^l / v_{\text{in}}^3$ for $n, m, l \in \mathbb{N}$ with $n+m+l \in (1, 2, 3)$ and Y_{nml} defined as in (2). For the Π -attenuator (T-attenuator), H and β are calculated using the model shown in Fig. 19 (Fig. 20).

For the Π -attenuator, let $R_x = R_1 + r_{\text{on},M_1}$, $R_y = R_2 + r_{\text{on},M_2} = R_3 + r_{\text{on},M_3}$, and $R_{\text{load}} = R_s$. For input matching to R_s and output matching to R_{load} , we find $R_x = R_s(1-A)/(1+A)/2A$ and $R_y = R_s(1+A)/(1-A)$. As a result, H

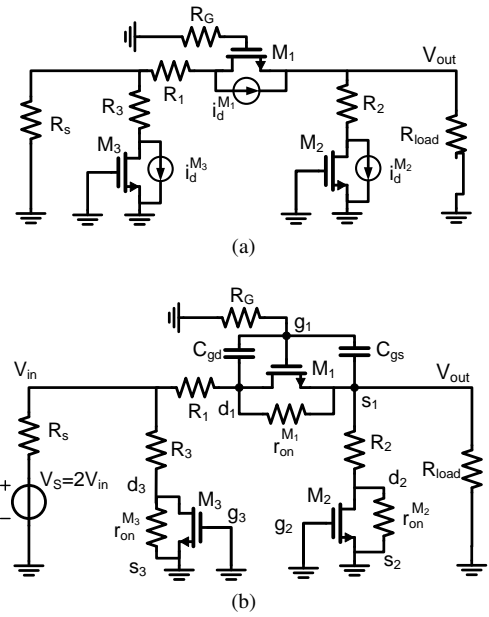


Fig. 19. Equivalent models for the Π -attenuator (a) for calculating the H function and (b) for calculating the β function.

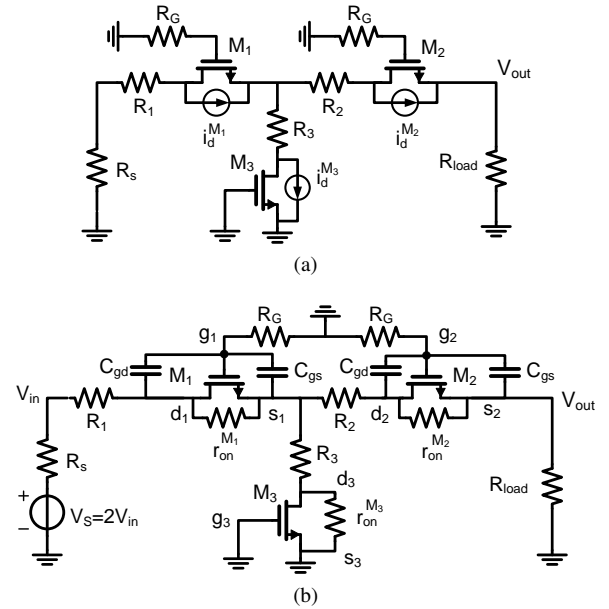


Fig. 20. Equivalent models for the T-attenuator (a) for calculating the H function and (b) for calculating the β function.

and β are found to be

$$H_{M_1} = \frac{A r_{\text{on},M_1}}{1+A} \quad H_{M_2} = -\frac{(1-A) r_{\text{on},M_2}}{2(1+A)} \quad (11)$$

$$H_{M_3} = -\frac{A(1-A) r_{\text{on},M_3}}{2(1+A)}$$

$$\begin{aligned} \beta_{030,M_1} &= \left(\frac{R_{\text{on},M_1}}{R_x + R_y \parallel R_{\text{load}}} \right)^3 & \beta_{300,M_1} &= \frac{\beta_{030,M_1}}{8} \\ \beta_{210,M_1} &= \frac{\beta_{030,M_1}}{4} & \beta_{120,M_1} &= \frac{\beta_{030,M_1}}{2} \\ \beta_{030,M_2} &= \left(\frac{R_y \parallel R_{\text{load}}}{R_x + R_y \parallel R_{\text{load}}} \right)^3 & \beta_{030,M_3} &= \left(\frac{r_{\text{on},M_3}}{R_y} \right)^3 \end{aligned} \quad (12)$$

Substituting (10) and (12) in (9) results in (1).

For the T-attenuator let $R_x = R_1 + r_{\text{on},M_1}$, $R_y = R_2 + r_{\text{on},M_2} = R_3 + r_{\text{on},M_3}$ and $R_{\text{load}} = R_s$. For input/output matching, we have $R_x = R_s(1 - A)/(1 + A)$ and $R_y = 2AR_s/(1 - A)(1 + A)$. Thus we find

$$H_{M_1} = \frac{Ar_{\text{on},M_1}}{2} \quad H_{M_2} = \frac{r_{\text{on},M_2}}{2} \quad H_{M_3} = \frac{(A - 1)r_{\text{on},M_3}}{2} \quad (13)$$

$$\begin{aligned} \beta_{030,M_1} &= \left(\frac{r_{\text{on},M_1}}{R_s}\right)^3 & \beta_{300,M_1} &= \frac{\beta_{030,M_1}}{8} \\ \beta_{210,M_1} &= \frac{\beta_{030,M_1}}{4} & \beta_{120,M_1} &= \frac{\beta_{030,M_1}}{2} \\ \beta_{030,M_2} &= \left(\frac{Ar_{\text{on},M_2}}{R_s}\right)^3 & \beta_{300,M_2} &= \frac{\beta_{030,M_2}}{8} \\ \beta_{210,M_2} &= \frac{\beta_{030,M_2}}{4} & \beta_{120,M_2} &= \frac{\beta_{030,M_2}}{2} \\ \beta_{030,M_3} &= \left(\frac{(1 - A)r_{\text{on},M_3}}{R_s}\right)^3 \end{aligned} \quad (14)$$

Substituting (12) and (14) in (9) results in (6).



analog/RF IC design.

Wei Cheng was born in Wuhan, Hubei, China, in 1980. He received the B.E. degree in measuring technology and instrument from Xiamen University, Xiamen, China, in 2002, the M.Sc. and Ph.D. degree in electrical engineering from the University of Twente (UT), Enschede, The Netherlands, in 2006 and 2012 respectively. From 2002 to 2003, he was an academic employee with the department of mechanical and electrical engineer of Xiamen University, Xiamen, China. He is now with Qualcomm in San Diego, CA. His research interest includes CMOS



Mark S. Oude Alink (S'09) was born on June 20, 1984, in Hengelo, the Netherlands. He received the B.Sc. degree in computer science in 2004 and the M.Sc. degrees in electrical engineering and computer science (both cum laude) in 2008, all from the University of Twente (UT), Enschede, the Netherlands. He is currently working towards the Ph.D. degree at the IC-Design group and the Computer Architecture for Embedded Systems group at the UT. His research includes system design, spectrum sensing and modulation for cognitive radio.



Anne Johan Annema (M'00) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Twente (UT), Enschede, the Netherlands. In 1995 he joined Philips Research, Eindhoven, The Netherlands. In 2000 he returned to the UT where he is associate professor in the IC-Design group. He is also part-time consultant in industry and in 2001 he co-founded ChipDesignWorks.



Gerard J.M. Wienk was born on October 11, 1958, in Hengelo, The Netherlands. From 1982 till 2001 he was with different companies working in the field of computer hardware and operating systems. In 1992 he received the B.Sc. degree in electrical engineering from the Hogeschool Enschede, The Netherlands. In July 2001 he joined the IC-Design group of the CTIT institute, University of Twente, Enschede, The Netherlands as a CAD support engineer.



Bram Nauta (M'91-SM'03-F'08) was born in Hengelo, the Netherlands, in 1964. In 1987, he received the M.Sc. degree (cum laude) in electrical engineering from the University of Twente, Enschede, the Netherlands. In 1991, he received the Ph.D. degree from the same university on the subject of analog CMOS filters for very high frequencies.

In 1991, he joined the Mixed-Signal Circuits and Systems Department of Philips Research, Eindhoven, the Netherlands, where he worked on high-speed A/D converters and analog key modules. In 1998, he returned to the University of Twente as a full Professor heading the IC Design group, which is part of the CTIT Research Institute. His current research interest is high-speed analog CMOS circuits.

He served as Associate Editor of IEEE Transactions on Circuits and Systems-II (1997–1999). He was Associate Editor (2001–2006) and later the Editor-in-Chief (2007–2010) of IEEE Journal of Solid-State Circuits. He is member of the technical program committees of the IEEE International Solid-State Circuits Conference (ISSCC), where he is the 2012 Program Committee vice-Chair, the European Solid-State Circuit Conference (ESSCIRC), and the Symposium on VLSI Circuits. He is co-recipient of the ISSCC 2002 and 2009 “Van Vessel Outstanding Paper Award”, is distinguished lecturer of the IEEE, elected member of IEEE-SSCS AdCom and is IEEE fellow.