Analysis and Design of Capacitive Voltage Distribution Stacked MOS Millimeter-Wave Power Amplifiers

Mohammad Hassan Montaseri, Janne P. Aikio, Timo Rahkonen, *Member, IEEE*, and Aarno Pärssinen, *Senior Member, IEEE*

***Abstract*— Stacked MOS power amplifiers (PA) are commonly used in SOI nodes but also have the potential to be realized in bulk CMOS nodes. In this paper they are analyzed in millimeter wave regimes. The study focuses on the key limiting factors and in particular the optimum number of transistors from which the key performance parameters such as maximum possible operating frequency, output power, and efficiency are achieved. Based on the analysis, design trade-offs of stacked MOS PAs are presented. The frequency dependency of the optimum load presented to each stack is analyzed to express the overall performance of the mentioned PA topologies as a new optimization method. Additionally, it is shown how the optimal load variations translate into amplitude-to-amplitude/phase (AM-AM/PM) conversion distortions. The validity of the analysis is examined against simulations. The simulations are performed based on 8M1P CMOS 28nm technology and electromagnetic simulations in ADS Momentum.**

***Index Terms*— CMOS integrated circuits, power amplifier, millimeter wave integrated circuits, stacked MOS power amplifier.**

I. Introduction

SHANNON'S channel capacity states the higher the bandwidth the higher the data rate. This is the main motivation towards higher operating frequency and emerging 5G and 6G systems, which could offer several advantages such as reduced system size, portability, and hence lower power consumption. Moving to higher frequencies poses several design challenges including modification to the technology node mostly in the form of scaling for covering higher f_t/f_{max} which yields reduced power density of the corresponding semiconductor components.

Manuscript received 15 December 2021; revised 30 April 2022; accepted 5 June 2022. Date of publication 21 July 2022; date of current version 30 August 2022. This work was supported in part by the InfoTech Oulu Doctoral Program, in part by the Finnish Funding Agency for Technology and Innovation (Tekes), in part by Nokia, in part by Esju, in part by CoreHW, in part by the Academy of Finland 6Genesis Flagship under Grant 318927, in part by the Nokia Foundation, and in part by Tauno Tönning Säätiö. This article was recommended by Associate Editor N. Krishnapura. *(Corresponding author: Mohammad Hassan Montaseri.)*

Mohammad Hassan Montaseri and Aarno Pärssinen are with the Center for Wireless Communications, University of Oulu, 90570 Oulu, Finland (e-mail: mohammad.montaseri@oulu.fi; aarno.parssinen@oulu.fi).

Janne P. Aikio and Timo Rahkonen are with the Electronics Circuits and Systems Research Unit, University of Oulu, 90570 Oulu, Finland (e-mail: janne.aikio@oulu.fi; timo.rahkonen@oulu.fi).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSI.2022.3185301.

Digital Object Identifier 10.1109/TCSI.2022.3185301

Thus, the output power density offered by a single transistor is quite limited in practice. A great demand for compact solutions for wireless communications applications has promoted CMOS integrated circuits (IC) design. However, designing an IC at such high frequencies is challenging as the key performance characteristics of the transistors including gain, linearity, signal-to-noise ratio (SNR), etc. are remarkably degraded [1].

On the other hand, the PAPR of higher-order modulation schemes, such as m-QAM, (O)QPSK, and OFDM deployed in the communications systems, e.g. LTE or 5G NR, postulate stringent circuit design considerations which add to the already existing challenges [2], [3]. Power amplifiers are considered one of the most important building blocks of a transmitter as they dominantly determine the ultimate performance of wireless communications systems. To achieve both in- and out-of-band signal integrity, the mentioned building blocks must fulfill the required performance. This concerns with the key characteristics such as AM-AM/PM conversion distortions, reduced desensitization at both in-band and adjacent channels, signal blockage, and bit-error-rate (BER) decrease. Due to amplitude variation of the modulated signals, the PA needs large back-off from saturation to attain sufficient linearity. However, linearity translates into poor efficiency obliged by the mentioned modulation schemes being on the order of 1% – 10% at the desired frequency range. Poor efficiency poses form factor drawbacks, thermal management issues, and reduced system lifetime expectancy issues [2]–[20].

Stacking the transistors have been widely adapted specially in CMOS SOI solutions [4]–[20] and recently brought to bulk CMOS domain [21]–[26]. As it has been utilized more extensively, investigation of the characteristics and performance of the mentioned topology is needed. Since its modern introduction by Ezzeddine [8], all the designs and their analysis have been based on frequency independent formulations [4]–[26]. They also lack proper formulation on the AM-AM/PM conversion distortions, besides the classical transconductance, i.e. g_m, compression and C_{gs} variation, will advance on understanding the design trade-offs. It is shown in this paper that the traditional frequency independent design, underestimates the variation of the gain of the stacks and hence performance reduction of the stacked MOS PAs at a wide frequency band. Furthermore, the impact of amplitude

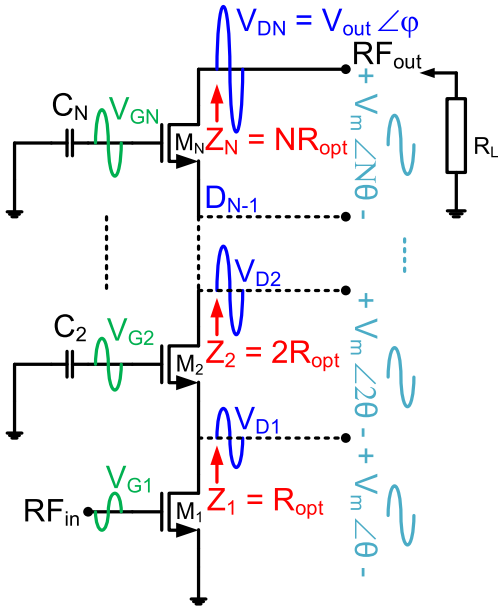


Fig. 1. A typical stacked MOS PA configuration.

variations on the gain variations, known as AM-AM conversion, as well as the phase variations, known as AM-PM conversion, is analyzed based on the projected optimal load variations, for the first time in this paper.

This paper is organized as follows. In section II the stacked MOS PA topology is reviewed briefly. A detailed analysis of the design dimensioning of the mentioned topology is presented in section III. Section IV discusses with the phase rotation compensation and AM-AM/PM is studied in section V. Finally the analysis is evaluated vs. simulation results in section VI.

II. Stacked MOS Topology

Scaling compels increased doping concentration to increase the operating frequency of MOSFETs. This has, in the meantime, resulted in decreased junction breakdown levels which binds the maximum possible voltage swing across the junctions. Excess increase in current density of a single MOS transistor, on the other hand, translates to reduced output impedance, which gives rise to the matching network transformation ratio, hence making it quite lossy. Accordingly, single-MOS-based power amplifier (PA) design is quite confined to a small available power density.

To alleviate the aforementioned problems, stacking transistors in a series connection on top of each other has been proposed [4]–[27] which is very well adapted specifically to SOI based technology nodes [4]–[20], to LDMOS [21] and MMIC [22], and recently to bulk CMOS technologies as well [23]–[27]. As can be seen in Fig. 1, this topology takes the advantages of cascading common source (CS) and common gate (CG) stages, which pretty much resembles a cascode amplifier topology, at the very first glance. There is a difference between the two. The gates in the stacked topology are not fully bypassed. The reason behind such technique is that in the cascode amplifier the fully bypassed gates result in the gain of each single inter-stage amplifier to remain ideally unity. And consequently, the overall signal swing will occur across the drain-source of the last stage making it more susceptible to break down. It should be noted that the focus of all the discussions is on the class A/AB operation.

Traditionally, the capacitors at the gates of the MOSFETs in the stacked PAs are dimensioned such that part of the signal swing is divided across them so that neither gate-source nor gate-drain junctions undergo breakdown levels. Besides equalizing the voltages across the transistors, the stacked circuit also makes impedance transformation from low at the bottom to high at the top: when the current in each stage is high, increasing voltage swing causes the apparent load impedance to increase stage by stage. The dimensioning of the mentioned capacitors was designed so that inter-stage matching, i.e. $R_{opt} = V_{DS_{DC}}/I_{D_{DC}}$ [28], was perfectly done [4]–[20] and [23]–[27].

With the increase in operating frequency, as specified in 5G systems, design of stacked MOS PAs has turned into a new challenge. This is not a straightforward procedure anymore, which is due to the fact that more and more high-order parasitics start to manifest with the increase of the operating frequency. Not only that but also scaling the transistors as well as the signal dependency of the parasitics add to the problem.

In the following sections we are investigating the stacked MOS amplifiers in different design aspects to clarify the design trade-offs and performance characteristics of the mentioned PA topologies.

III. Stacked MOS PA Analysis

A detailed analysis of the stacked MOS transistor PAs is discussed in this section. First a small-signal model is analyzed to be further referenced for silicon-on-insulator (SOI), and then a more general small-signal model is introduced to cover the issues corresponding to triple-well bulk CMOS technologies. After indicating the design parameters, tradeoff between output power, efficiency, operating frequency, number of stages, etc. are described.

A. Frequency Dependent Design Consideration

Stacking concept has been utilized to distribute the overall output signal equally among the stages so that none of the transistor junctions experience over-stressed conditions. Additionally, it constructs an internal impedance which ultimately performs load matching at the output. Without loss of generality, the conventional small signal model of Fig. 2(a) has extensively been utilized for the analysis purposes.

Ideally, the equations governing the circuit of Fig. 2 can be expressed by (1), as shown at the bottom of the next page. Due to the capacitive loading of the upper stages, (1) needs to be modified to include the reactive part of the load. In order for that Y_{n+1} is assumed to be comprised of an optimum conductive part as well as a non-desirable capacitive susceptance, i.e. $Y_{n+1} = G_{n+1} + jB_{n+1}$ (Fig. 2.b). Thus (1) can be rewritten in (2), as shown at the bottom of the next page. Solving (2) for the input admittance $Y_n = -V_{s_n}/I_{s_n}$

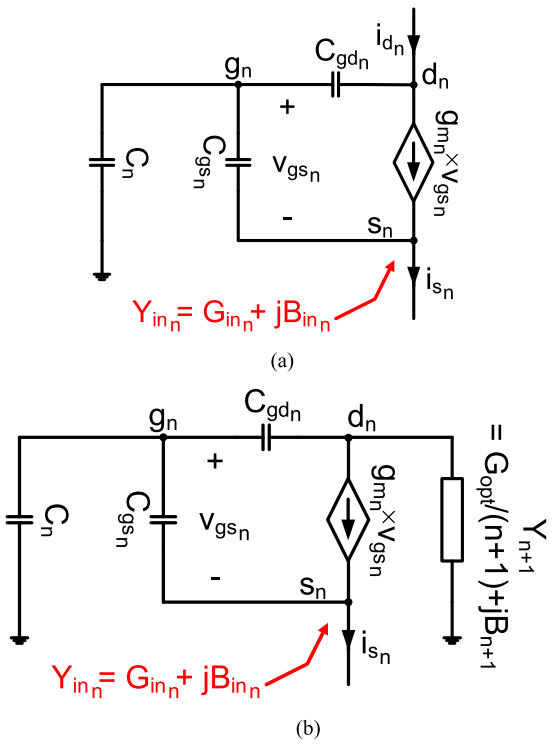


Fig. 2. Generic small-signal model of the MOS transistors in a stacked MOS PAs without (a) and with (b) the loading admittance of the succeeding stage.

yields (3).

$$Y_n = G_{opt_n} + jB_n = \frac{\left(g_{m_n} + j\omega C_{gs_n}\right)\left(\left(C_n + C_{gd_n}\right)Y_{n+1} + j\omega C_{gd_n}C_n\right)}{\left(C_{gs_n} + C_n + C_{gd_n}\right)Y_{n+1} + j\omega C_{gd_n}\left(C_{gs_n} + C_n\right) + g_{m_n}C_{gd_n}} \tag{3}$$

For the design purpose the real part of the input impedance Y_n needs to be equal to the desired optimum conductance, i.e. $\mathrm{Re}\{Y_n\} = G_{opt_n} = 1/\left(nR_{opt}\right)$. As will be explained later, the susceptance of Y_{n+1} can be compensated. Thus, substituting $G_{opt_{n+1}}$ with Y_{n+1}, for $1 < n \leq N-1$ we have,

$$\mathrm{Re}\{Y_n\} = G_{opt_n} = \frac{1}{nR_{opt}} = \frac{g_{m_n}(ac+bd) + \omega C_{gs_n}(ad-bc)}{c^2+d^2} \tag{4}$$

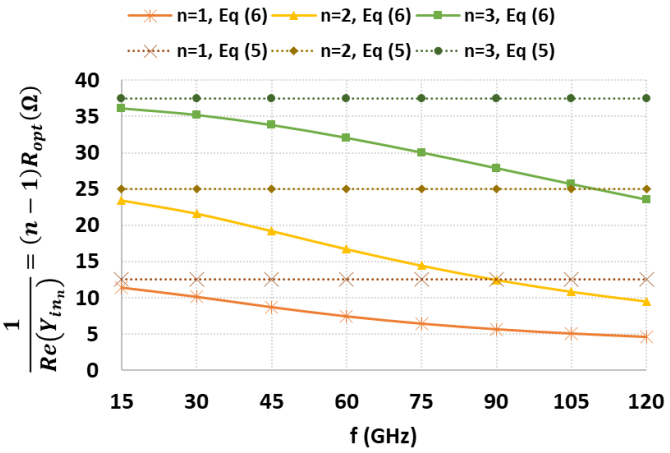


Fig. 3. The variation of the drain load as a result of frequency variation while using frequency independent design rule of (6) vs. frequency dependent rule (5). f_t of the transistors is 240GHz.

where

$$a = \left(C_n + C_{gd}\right)\left(\frac{G_{opt}}{n+1}\right) \tag{4.a}$$

$$b = \omega C_{gd_n} C_n \tag{4.b}$$

$$c = \left(C_{gs_n} + C_n + C_{gd_n}\right)\left(\frac{G_{opt}}{n+1}\right) + g_{m_n}C_{gd_n} \tag{4.c}$$

$$d = \omega C_{gd_n}\left(C_{gs_n} + C_n\right) \tag{4.d}$$

and $G_{opt} = 1/R_{opt}$. Equation (4) clearly illustrates the dependence of the $\mathrm{Re}\{Y_n\}$ on the operating frequency and in turn the frequency dependence of C_n dimensioning. In other words, the values of the gate capacitances C_n shown in Fig. 1 need to be designed depending on the operating frequency. Accounting for the frequency variations in the design of C_n's offers the opportunity to control the optimum stack loading. Therefore, the loading can be optimized versus the frequency band of interest. This is shown in Fig. 3 wherein the optimal loading is tuned at the desired frequency band. Solving (4) for C_n, yields (5), as shown at the bottom of the next page. If the operating frequency range is less than $f_t/10$, the gate capacitance C_{n+1} values asymptotically approach

$$C_n|_{f \leq \frac{f_t}{10} \ll f_t} = \frac{C_{gs_n} + C_{gd_n}\left(g_{m_n}R_{opt} + 1\right)}{g_{m_n}R_{opt}\,(n-1) - 1} \tag{6}$$

compliant with [15]. Although very short/simplistic and efficient for low frequencies, (6) is lacking the impact of the frequency on the C_n's dimensioning, which results in the

$$\begin{bmatrix} -j\omega C_{gd_n} & j\omega\left(C_{gs_n}+C_{gd_n}+C_n\right) & -j\omega C_{gs_n} & 0 & 0 \\ -\left(j\omega C_{ds_n}\right) & -\left(j\omega C_{gs_n}+g_{m_n}\right) & g_{m_n}+j\omega\left(C_{gs_n}+C_{ds_n}\right) & 0 & 1 \\ j\omega\left(C_{gd_n}+C_{ds_n}\right) & g_{m_n}-j\omega C_{gd_n} & -\left(g_{m_n}+j\omega C_{ds_n}\right) & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_d \\ V_g \\ V_s \\ I_d \\ I_s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ nV_m \\ (n+1)V_m \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 0 & 1 & -\frac{1}{g_{m_n}+j\omega C_{gs_n}} \\ -\frac{C_{gd_n}}{C_{gs_n}} & 1+\frac{C_{gd_n}+C_n}{C_{gs_n}} & 0 \\ \frac{j\omega C_{gd_n}+Y_{n+1}}{g_{m_n}} & 1-\frac{j\omega C_{gd_n}}{g_{m_n}} & 0 \end{bmatrix} \begin{bmatrix} V_{d_n} \\ V_{g_n} \\ I_{s_n} \end{bmatrix} = \begin{bmatrix} V_{s_n} \\ V_{s_n} \\ V_{s_n} \end{bmatrix} \tag{2}$$

loads to drop with frequency. Due to this, (6) is misleading as soon as the operating frequency surpasses $f_t/10$. Disregarding the higher frequency impact on the circuit performance, leads both optimal load reduction and impedance mismatching and hence gain drop at the mentioned bands. As will be shown later, this in turn translates to performance degradation. The optimal loading drop posed by (6) in comparison with that of (5) is graphically shown in Fig. 3. A 4-stacked MOS PA with $C_{gs} = 450fF$, $C_{gd} = 50fF$, $g_m = 600mS$ and swept operating frequency up to 120 GHz, i.e. $f_t/2$, has been utilized. As can be seen, the optimal value required for nR_{opt} at the drain of the nth MOS drops with frequency due to low frequency approximation of the gate capacitances, i.e. C_n predicted by (6), which yields drastic gain decrease at high frequencies and hence delivered power to the output of the stacked MOS PA topologies. Reflected by constant curves in Fig. 3, this is while the C_n's designed by (5) tend to tune and keep the load at its optimal value at the desired operating frequency. This is shown to reserve the performance of the PA, later in section VI. Illustrated in Fig. 4 are the frequency dependency of (5) – (7). As can be seen, the C_n's designed based on (5) take operating frequency into account thus need to be re-dimensioned accordingly, whilst those based on (6) are constant over all frequencies. Specifically, the optimal loading in accordance with (5) remains constant as the corresponding C_n's are adapted for the desired operating frequency hence the optimal loading is retained, compliant with Fig. 3. Equation (6) on the other hand, does not guarantee optimal loading at all frequencies as it is frequency independent. The simulations are conducted for an example of 4-stacked MOS PA with $C_{gs} = 450fF$, $C_{gd} = 50fF$, $g_m = 600mS$ and swept operating frequency up to 250GHz.

To get more insight into the matter, yet another important extreme value of the Re $\{Y_n\}$ at the infinite operating frequency is obtained and solved for gate capacitances C_n. This puts an upper bound on the C_n's. Accordingly, the required value for dimensioning the gate capacitances C_n are varying between (6) and (7) as a function of the frequency. Bearing in mind that (6) and (7) put a lower and an upper bound on the C_n's, correspondingly, one can exploit any interpolation between the two as an approximation of (5), based on the requirements, if minor errors are acceptable in the design and later in performance of the stacked MOS PAs.

$$C_n|_{f\gg\frac{f_t}{10}} = \frac{C_{gs_n}\left(\sqrt{\frac{(g_{m_n}R_{opt}+1)(n-1)}{n}}+1\right)}{\left(g_{m_n}R_{opt}\,(n-1)-1\right)}. \tag{7}$$

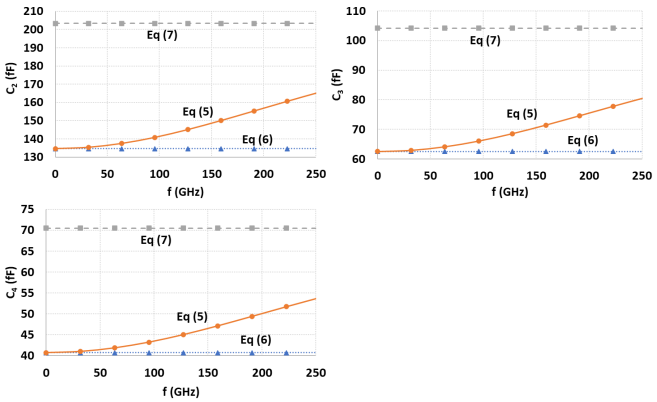


Fig. 4. The required gate capacitances values, i.e. C_n, of a 4-stack PA at different operating frequencies, based on calculations in (5).

If the frequency dependent part of the design is ignored, i.e. only (6) used as a design dimensioning rule, the optimal load presented to a stack starts to dramatically degrade from the desired value after a certain frequency, as in Fig. 3. Thus, the gain provided by each stacked transistor and hence the power delivered to the load is reduced. So, the gate capacitances C_n need to be dimensioned in accordance with the operating frequency, as shown in Fig. 4, as proposed by either (5) or any estimation/interpolation using/between (6) and (7).

Considering the small signal model of the MOS transistor at a certain operating frequency and disregarding the impact of the feedback capacitance of C_{gd}, a simplified qualitative gain analysis yield $Av_n \cong g_{m_n}(n+1)\,R_{opt}\times C_n\big/\left(C_n+C_{gs_n}\right)$ which affirms a homographic behavior with an upper bound asymptote of $g_m(n+1)\,R_{opt}$. (6) and (7) define a fixed lower and upper bound values for the gain, respectively. The problem with the designs based on (7) manifests at lower frequency bands where the PA experiences breakdown due to bypass property of the gate capacitances. In this mode the PA behaves more like a (small signal) cascode amplifier with its topmost stacked transistor being under the maximum junction stress. On the other hand, at higher frequency bands, design with (6) ends in lack of gain, hence transduced power and PAE. This is shown in Fig. 3 wherein the optimal load starts to decrease from its desired value at higher frequencies. Therefore, designs with (6) is appropriate only at lower frequencies whilst designs with (7) is appropriate only at higher bands. On the contrary, (5) predicts the design requirements and optimal load versus the desired frequency band and thus an optimum performance can be achieved. Moreover, the impact of the proposed design

$$C_{n+1} = \frac{-\beta+\sqrt{\beta^2-4\alpha\gamma}}{2\alpha} \tag{5}$$

$$\alpha = \left(g_m-\left(G_{opt}/n\right)\right)\left(\omega^2C_{gd}^2R_{opt}^2\,(n+1)^2+1\right) \tag{5.a}$$

$$\beta = \left(g_m-\left(2G_{opt}/n\right)\right)\left(g_mR_{opt}C_{gd}\,(n+1)+C_{gd}+C_{gs}\right)-\frac{2\,(n+1)^2\,\omega^2C_{gd}^2C_{gs}\,R_{opt}}{n}+g_mC_{gd} \tag{5.b}$$

$$\gamma = -\left(C_{gd}+C_{gs}\right)\left(g_mC_{gd}\,(n+2)+g_{opt}\left(C_{gd}+C_{gs}\right)\right)/n-C_{gd}^2\,(n+1)\,R_{opt}\left(\omega^2C_{gs}^2+g_m^2\right)/n \tag{5.c}$$

method, i.e. using (5), on the key performance characteristics of the PA compared to that of earlier approach, i.e. (6), is later shown in section VI Fig. 17.

B. Phase Shift/Rotation Impact on the Performance of Stacked MOS PA

The real part of the Y_n, i.e. Re $\{Y_n\}$ in (3), is the desired term which was discussed in the previous section. The imaginary part of it, i.e. Im $\{Y_n\}$, however is the unwanted term which directly impacts the performance of the stacked MOS PA topology. This aspect has been mostly ignored in previous analyses.

In fact, the presence of parasitic elements leads to the angles, θ, of drain-source voltage vectors to gradually rotate per stacked MOS transistor along the PA. From (3) we have $\text{Im}\{Y_n\} = \left(\omega C_{gs}\left(ac+bd\right) - g_m\left(ad-bc\right)\right) / \left(C^2+d^2\right)$. One can realize the dependence of (9) on the device dimensioning, biasing, transistor parasitics, operating frequency, and C_n. In other words, θ being defined as $tan^{-1}\left(Im\left(Y_n\right)/Re\left(Y_n\right)\right)$, follows a complex functionality of all the mentioned parameters of the form $\theta = f\left(W/L, I_{ds}, V_{gs}, g_m, C_{gs}, C_{gd}, C_n, \omega\right)$. With the assumption of uniform phase rotation, superposition of all the drain-source voltages yields the maximum amplitude of the output of interest (V_{DN} in Fig. 1) to be

$$|V_{DN}(N)| = \left|\sum_{n=1}^{N} V_{DS_n}\right|, \tag{8}$$

where V_{DS_n} is $V_m e^{jn\theta}$; wherein V_m and θ express the maximum voltage swing and the uniform phase rotation across the drain-source junctions. Thus,

$$|V_{DN}(N)| = \left|\sum_{n=1}^{N} V_m e^{jn\theta}\right| = V_m \times \frac{sin(N\theta/2)}{sin(\theta/2)}. \tag{9}$$

The output power can then be expressed by

$$P_{out}(N) = \frac{V_{DN}^2}{2R_L} = \frac{V_m^2}{2R_L} \times \frac{sin^2(N\theta/2)}{sin^2(\theta/2)}. \tag{10}$$

As the load R_L is distributed along with the N number of stages and each stage is designed to match its optimal load, i.e. R_{opt}, (10) must be modified by $R_L = NR_{opt}$. Thus, we have

$$P_{out}(N) = \underbrace{\frac{V_m^2}{2R_{opt}}}_{I} \times \underbrace{\frac{sin^2(N\theta/2)}{N\ sin^2(\theta/2)}}. \tag{11}$$

The first term in (11), i.e. I, is the maximum power that can be obtained from a non-stacked, i.e. single device, MOS transistor PA. We call the second term, i.e. II, as *Stacking Factor* (SF), which shows the dependence of the SF parameter and hence output power of the mentioned PA topologies to the number of stages as well as phase variation across each of them. As explained in the remaining part of this section, SF is used to define the maximum number of stages in a stacked MOS PA topology.

Plotting the SF versus the number of stacked transistors based on analytical equations, i.e. the second term in (11),

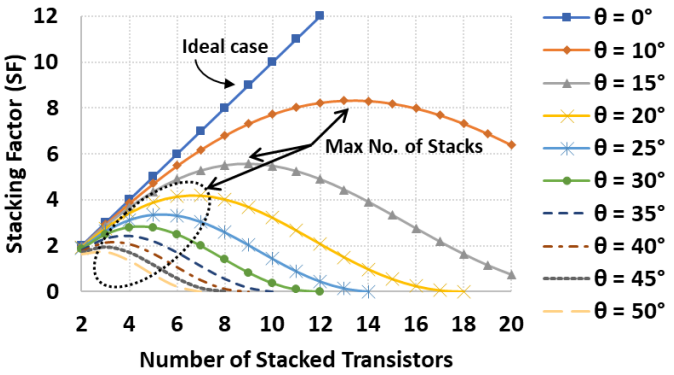


Fig. 5. Power multiplication by stacking factor as a function of the number of stacked MOS transistors for different phase variations, starting from a phase rotation of 10° per stage.

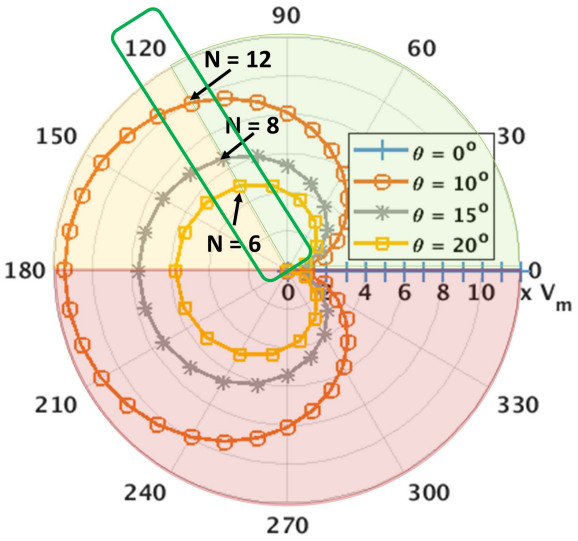


Fig. 6. Voltage gain curvature as a function of phase rotation per stage. The green area is the optimal region where voltage/power increases with gradual efficiency decrease, the yellow area is the non-optimal region where in voltage increases, however efficiency decreases drastically, and the red area is the inoperable region where both power and efficiency decrease.

for different phase variations proves highly informative as it quantifies the relationship between the undesired phase rotation and output power (Fig. 5). In the absence of phase rotation, i.e. $\theta = 0°$, which is a representative of the ideal condition, the SF increases in direct proportion to the number of transistors, i.e. $SF = N$, whilst this is not the case as phase rotation manifests along the stack. Moreover, the more the phase rotation, the more drastically degrades the output power from its ideal case.

Deduced from Fig. 5, it is of great importance to define the optimum number of stacked transistors. The reason can be understood from Fig. 5 and that is, after some point, adding more stages to the PA will start to deteriorate the performance of the PA.

There exist several different approaches to figure out the maximum number of stacked transistors. The first one is to read the optimum number of stacked transistors intuitively from the SF plot of Fig. 5. As shown in Fig. 6, the second

one is to plot the voltage signal swing of (9). As another approach, one can maximize (11) with respect to the number of stacked MOS transistors, i.e. N, for a given phase rotation per stack/stage, i.e. θ. Last but not least is to derive a formula that illustrates to what extent the power can be amplified, as will be shown later. All the above approaches are discussed in the following.

The impact of phase rotation along the transistors has already been expressed in Fig. 5. As mentioned earlier, it is not allowed to increase the number of stacks unboundedly. Conversely, after some point the SF starts to decrease as the number of transistors increases. Hence adding more stages after the mentioned point will not improve performance anymore. It can already be seen that the mentioned point, which is the maximum output power, is a function of the phase rotation. For example, in case the phase shift along each stage is 10°, the point of maximum number of stacked transistors i.e. maximum output power is somewhere around 13; and adding the next stages only results in performance degradation. This corresponds to a total of 130° phase shift approximately.

By the same token, for single stage phase rotation of 15° and 20°, the maximum number of transistors is 9 and 7, which corresponds to total phase rotation of 135° and 140°, respectively. The mentioned numbers are all around a unique optimal total phase rotation boundary which can be described in the following.

To further analyze the previous statements, (9) is plotted in the polar form of Fig. 6 to show both the phase rotation and voltage increase per stack. In accordance with the concept of Fig. 5 is the increase of voltage amplitude up to some maximum level and after that, the signal starts to degrade when more stages are added. To complete the foregone discussion, the maximum voltage amplitude happens at the phase of 180° after which the amplitude starts to degrade. Exploring Fig. 6 gives more insight into the supply requirement of a stacked MOS PA design. Let us first start with the example of phase rotation of 10° per stage. Adding first 12 stages will increase the output voltage amplitude to a level of $10 \times V_m$, approximately, where V_m is the maximum tolerable drain-source voltage of each single stacked MOS transistor. In order to get the mentioned amount of signal amplitude at the output it is required to have $12 \times V_m$ as for DC biasing of the overall structure. That means the maximum efficiency of the PA reduces to 83 percent of the maximum theoretical efficiency. This corresponds to a total phase rotation of 120° over all the stages. Although addition of more stages will increase the maximum amplitude, the output amplitude varies only marginally in a way that adding for example the next 8 stages will not even add $2 \times V_m$ more signal swing. And for this to happen the PA requires $8 \times V_m$ more DC power supply. Hence, simply 75 percent of the DC power is lost in such circumstances.

Exploring other amounts of phase rotation per stage converge to the same 120° boundary. The cases of 10°, 15° and 20° are shown in Fig. 6. This is still an intuitive method of estimating the optimum number of transistors in a stacked MOS PA. In the following, we will provide analytical approaches to characterize theoretically the boundary. It should be noted

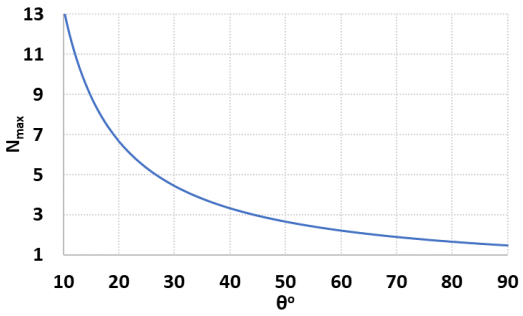


Fig. 7. Solution of (11) maximized for the optimum number of the transistors vs. phase rotation per transistor.

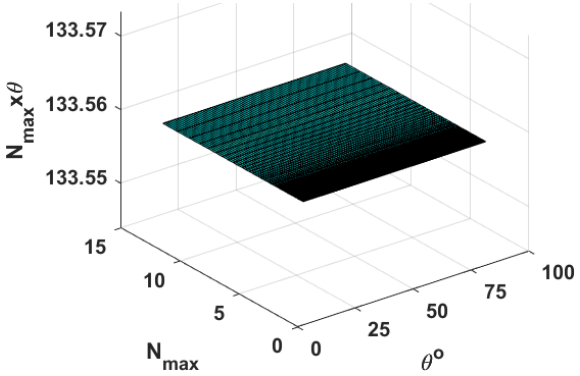


Fig. 8. The constant $N.\theta$ product, showing the optimum number of stacks follows the product rule.

that the stacked MOS PAs may be utilized in this respect as long as no feedback is applied to the PA circuit, which is the case in many applications.

$$\frac{\partial}{\partial N}\left(\frac{sin^2(N\theta/2)}{Nsin^2(\theta/2)}\right) = \frac{N\theta \sin\left(\frac{N\theta}{2}\right)\cos\left(\frac{N\theta}{2}\right) - \sin^2\left(\frac{N\theta}{2}\right)}{N^2 \sin^2\left(\frac{\theta}{2}\right)} = 0. \tag{12}$$

In order to maximize (11) with respect to the number of stacked transistors, i.e. N, for a given phase rotation per stage, i.e. θ, the derivative of the term SF in (11) must be calculated. Thus, we have (12). Solving (12) for the maximum number of transistors, i.e. N_{max}, results in

$$N_{max} = \frac{1}{\theta} \times \tan\left(\frac{N_{max}\theta}{2}\right). \tag{13}$$

Equation (13) is nonlinear which requires numerical methods to be solved. One can use any nonlinear method to solve (13) without loss of generality. Here Newton's method has been used to solve (13). The maximum number of transistors (N_{max}) in the stack to optimize the performance is shown in Fig. 7 as a function of phase rotation per stage.

Interestingly the product $N_{max} \times \theta$ is always constant and equal to 133.6°; this is shown in Fig. 8. This is a gain-bandwidth counterpart which can be used as a *rule-of-thumb* in the design of stacked MOS PAs. Given θ using either simulation or transistor parameters, the maximum and/or optimum number of transistors, which can be stacked, can

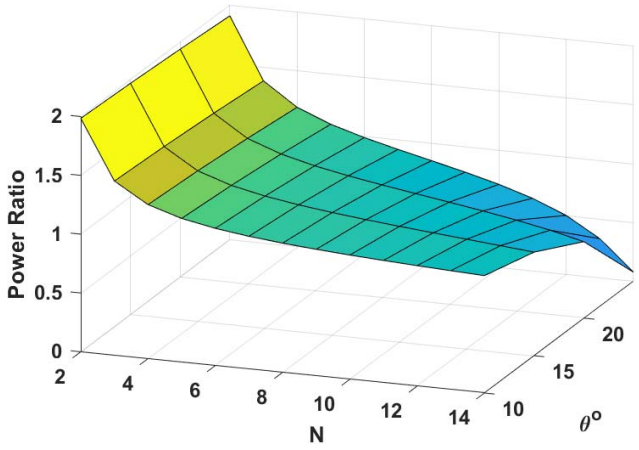


Fig. 9. Power ratio gain vs. number of transistors for different phase variations along each single stack MOS transistor.

be obtained. Conversely, for a required number of stacked transistors, the θ should be kept under the product number for optimal performance.

Finally, calculating $P_{out}(N)$ from (11) for N and $N-1$ number of transistors and dividing them to find the optimum number of transistors that still offers additive power gain, is also informative. In other words, solving the power ratio $P(N)/P(N-1)$ for N to find the maximum number of transistors when the power ratio gain is still greater than unity. Thus, we have

$$\frac{P(N)}{P(N-1)} = \frac{(N-1)\sin^2(N\theta/2)}{N\sin^2((N-1)\theta/2)} > 1. \tag{14}$$

Plotted (14) against the number of transistors for different values of phase rotation per stage is illustrated in Fig. 9. Consider the curve corresponding to the 10° phase rotation per stack; for the first 13 stages the power ratio gain is still above unity, i.e. 1. In other word if phase rotation per stage is 10°, by adding up to 13 transistors, the PA still offers power gain. From the 14th stage onwards, the additional stages will only act as attenuators. A similar statement can be made for other phase rotation values as well. This is plotted in Fig. 9 for phase rotations of 10° – 40° per stack, with 5° steps.

All the phase rotation calculations discussed tend to converge to the same total phase ∼130° which was already introduced in the previous paragraphs.

IV. Phase Rotation Compensation

Initially the capacitances C_n at the gate of each stacked MOS transistor (Fig. 1) have been designed to fulfill two functions: 1) capacitive voltage divider to limit the voltage swing across the junctions [7]–[11], [13], [9], [16]–[21], and 2) tune the real part of the input impedance to the optimal load for the preceding stage, i.e. R_{opt_n}, proposed by (5). However as pointed out in (1) – (5), the admittance seen by each transistor essentially conveys an imaginary part B_n, which needs to be compensated for. The impact of the B_n part was also discussed in previous section. As was shown through (9) – (14), the B_n part originates the phase rotation per stack and hence yields performance degradation.

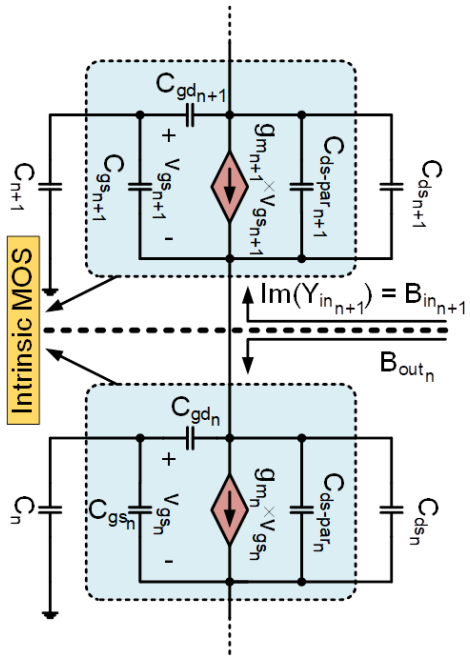


Fig. 10. Negative capacitance compensation method [11].

There exists also a mismatch between the susceptances looking upwards and downwards. In other words, $B_{in_{n+1}}$ is not necessarily equal to B_{out_n}. This is clearly explained through calculating imaginary parts of the admittances, which is the very first origin of discrepancy.

Thus, using the high frequency transistor model of Fig. 2 and two stacked MOS transistors in the middle of the structure (Fig. 10), in the presence of the drain-source compensation capacitances $C_{ds_{ii}}$ and neglecting the channel length modulation effect, we have

$$B_{in_{n+1}} \approx \frac{\omega}{n}\left(\frac{C_{gs_{n+1}}}{g_{m_{n+1}}R_{opt}} - C_{ds-par_{n+1}} - C_{ds_{n+1}}\right), \tag{15}$$

and

$$B_{out_n} \approx \frac{\omega}{n}\left(\frac{C_{gd_n}}{g_{m_n}R_{opt}} + C_{ds-par_n} + C_{ds_n} + C_{gd_n}\right). \tag{16}$$

The term $C_{ds-par_{ii}}$ in (15) and (16) are MOS transistor drain-source parasitic capacitances. It is seen that (15) differs from (16) which originates another design inaccuracy due to the discrepancy between the admittances seen towards different directions, which must be compensated for. The issues above were neglected in [8]–[10], [12]–[15], when dimensioning the devices. Equating (16) to the conjugate of (15), we have

$$\frac{\omega}{n}\left(\frac{C_{gs_{n+1}}}{g_{m_{n+1}}R_{opt}} - C_{ds-par_{n+1}} - C_{ds_{n+1}}\right) = -\frac{\omega}{n}\left(\frac{C_{gd_n}}{g_{m_n}R_{opt}} + C_{ds-par_n} + C_{ds_n} + C_{gd_n}\right). \tag{17}$$

Solving (17) for $C_{ds_{n+1}}$ results in

$$C_{ds_{n+1}} = \frac{C_{gs_{n+1}}}{g_{m_{n+1}}R_{opt}} - C_{ds-par_{n+1}} + \frac{C_{gd_n}}{g_{m_n}R_{opt}} + C_{ds-par_n} + C_{ds_n} + C_{gd_n}. \tag{18}$$

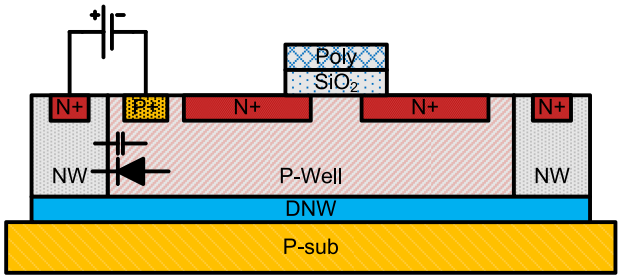


Fig. 11. Simplified cross-section of deep n-well (DNW) process and its most dominant parasitics in bulk CMOS technology.

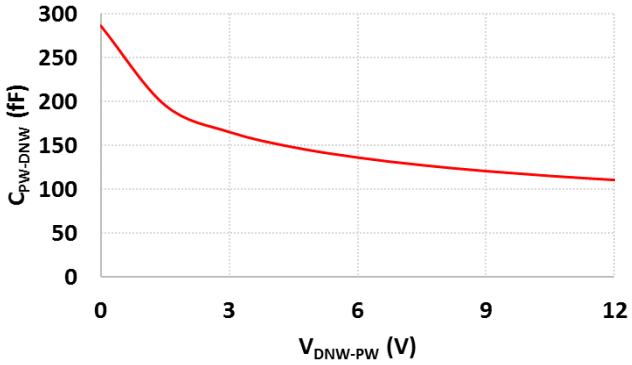


Fig. 12. Bias dependency of PW-DNW parasitic capacitance.

It should be noted that the first stage does not need to perform compensation, so we have forced C_{ds_1} to be 0. Thus, we have

$$C_{ds_{n+1}} = n \times \left(\frac{C_{gs} + C_{gd}}{g_m R_{opt}} + C_{gd} \right). \tag{19}$$

The dimensioning rule of (19) guarantees proper phase detuning along the stack. Equation (19) accounts for the Miller effect of C_{gd}, i.e. The term $C_{gd}/g_m R_{opt}$, in phase compensation and/or interstage matching. Simulations prove the importance of it in multi stack PA design at mm-wave regime. This is while, the mentioned term is missing from the compensation method proposed in [8] and [15]. Although the calculations were performed for the proposed negative capacitance compensation method, the approach can be applied to other detuning techniques without loss of generality. More importantly, $C_{ds_{n+1}}$ can be dimensioned to compensate for more parasitics as well.

A. Bulk CMOS Considerations

As explained in [21], to reduce the body effect on the AM-AM conversion it is recommended to utilize triple-well technique [7]. The body isolation based on the mentioned process technique however poses two issues: a diode and a parasitic capacitance are formed between the deep N-well (DNW) and P-well, which must be considered when designing the stacked MOS PA (Fig. 11).

The effect of the former can be simply minimized by reverse biasing the PW-DNW junction diode, however, to compensate the effect of the parasitic capacitance, formed by p-well and the DNW, one needs to consider the bias dependence of the mentioned parasitic capacitance (Fig. 12) [21].

Fig. 10 needs modification to take triple-well bulk MOS parasitics into account. This is shown in Fig. 13. Calculating the susceptances looking upwards and downwards, equating them, and solving for compensating drain-source capacitances, yields (20), [21].

$$C_{DS_{n+1}} = n \times \frac{C_{gs} + C_{gb} + C_{gd}\left(1 + g_m R_{opt}\right)}{g_m R_{opt}} + \sum_{i=2}^{n} i \times C_{PW-DNW_i}. \tag{20}$$

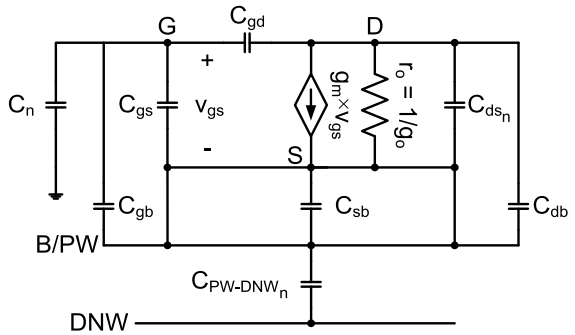


Fig. 13. Modified High-frequency small-signal model of a bulk CMOS transistor along with the effect of deep n-well parasitic and capacitive voltage divider (C_n) and negative capacitance compensator (C_{DS_n}) [21].

V. Amplitude to Amplitude/Phase Conversion Distortion

Up to this point all the analyses were based on the small signal domain approximation where the transistor parameters vary negligibly if at all. As soon as the input signal grows beyond such assumption, the PA manifests nonlinear behaviors. A direct consequence of which is gain compression known as amplitude to amplitude (AM-AM) conversion distortion. Also due to the presence of both intrinsic and extrinsic dynamic components, such as parasitic capacitances as well as gate capacitive voltage division network, the amplitude to phase (AM-PM) conversion distortion is inevitable [4]–[33].

A. AM-AM Conversion Distortion

After the gate capacitances, i.e. C_{n+1}, have been fixed in the PA design based on (5), the values of $R_{opt_n} = \mathrm{Re}\left\{Y_{in_n}\right\}$, expressed in (4), are ideally required to remain constant. However, this was shown not to be the case over the frequency variations in previous sections. Moreover, g_m and C_{gs} are also amplitude dependent parameters (Fig. 14) which definitely alter the value of $\mathrm{Re}\left\{Y_{in_n}\right\}$ departing it from the desired value of $R_{opt_n} = 1/G_{opt_n}$, Fig. 14(c). This in turn degrades the gain translating to AM-AM conversion distortion. It should be noted that C_{gd} also is an amplitude dependent parameter with a minor impact compared to the foregone parameters

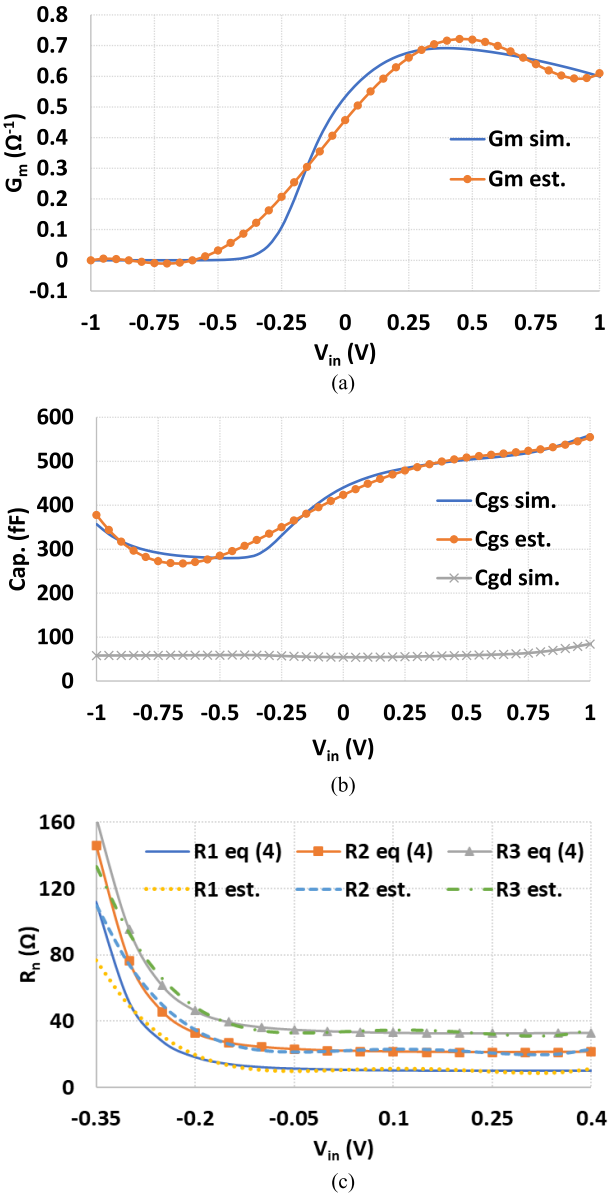


Fig. 14. Simulated and estimated variation of the (a) g_m, (b) C_{gs}, C_{gd}, and (c) calculated and estimated variation of the R_{opt_n} of an NMOS corresponding to the input drive with $C_{gs} = 450fF$, $C_{gd} = 50fF$, and $g_m = 600mS$, at $V_{gs} = 0.55V$ and an operating frequency of 28 GHz.

(Fig. 14 b). Conversely, the impact of C_{gd} on the $\mathrm{Re}\{Y_{in_n}\}$ manifests as a gain compression in the "Miller effect" which is already calculated in (4). Hence, taking the transconductance compression in the calculations should be fair enough.

It should be noted that, 4th order polynomial has been utilized for estimating R_{opt}'s in Fig. 14(c). This is chosen for the simplicity of hand calculations purposes. Although the simple low order polynomial is giving wrong estimate in negative signal swing side, increasing the order leads to calculations complexity. Based on simulation results presented in section VI, using such lower order polynomials shows very well matching with the final design, however it is evident one need to fine tune the values for better match. It is possible to express the input-output characteristics of the PA based on Volterra/power series [28]. However, to keep the analysis simple enough for hand calculation purposes, the effect of higher order nonlinearities on the first term of the Volterra series are considered in the following analysis, i.e. $V_{out} = \sum a_n V_{in}^n$, wherein a_n's are yet to be determined for the total amplitude dependent output signal as well as the gain of the PA. Given $V_{in} = A_c \cos(\omega t)$, the fundamental harmonic of the output signal hence the gain of the PA, using binomial formula, can be expressed as

$$V_{out_{fund}} = A_c \cos(\omega t) \times \sum_{n(=0)} \binom{2n}{n} \frac{a_{2n+1} \times A_c^{2n}}{2^{2n}}. \quad (21.a)$$

$$AV_{fund} = \sum_{n(=0)} \binom{2n}{n} \frac{a_{2n+1} \times A_c^{2n}}{2^{2n}}. \quad (21.b)$$

Based on the concept of Fig. 1, the overall output signal of a stacked MOS PA (or in general any technology) is the accumulation/summation of the signals across the drain-source of each single stage, i.e. v_{ds_n}. Hence the drain-source signal of each single stage of the PA can be expressed as:

$$v_{ds_n} = \int G_{m_n}(v_{gs_n})\, dv_{gs_n} \times \left(Z_n(v_{gs_n}) - Z_{n-1}(v_{gs_{n-1}})\right). \quad (21)$$

Thus, the overall gain can be expressed as an accumulation of the gains of the whole stages, i.e. $AV_{total} = \sum_n v_{ds_n}/V_{in}$. In case of identical dimensioning the total gain can be written as

$$AV_{total} \cong \frac{N \times \int G_m(V_{in})\, dV_{in} \times R_{opt}(V_{in})}{V_{in}}. \quad (22)$$

Since the parameter $G_m(V_{in})$ is technology dependent and the governing equation of the $R_{opt_n}(V_{in})$, i.e. (4), is quite complex, their values can be estimated with several different methods. Here, due to its widespread application, the power series approximation method has been used to estimate both (Fig. 14).

$$G_m(V_{in}) = g_{0_n} + g_{1_n} V_{in} + g_{2_n} V_{in}^2 + g_{3_n} V_{in}^3 + g_{4_n} V_{in}^4 + g_{5_n} V_{in}^5 \quad (23)$$

$$R_{opt_n}(V_{in}) = r_{0_n} + r_{1_n} V_{in} + r_{2_n} V_{in}^2 + r_{3_n} V_{in}^3 + r_{4_n} V_{in}^4 \quad (24)$$

where g_{i_n}'s and r_{i_n}'s are fitting parameters extracted for the technology node of interest and the optimum load of interest at the desired quiescent bias point, respectively. With a one tone sinusoidal continues wave (CW) of $V_{in} = A_c \cos(\omega t)$, and the fact that the higher order harmonics are filtered out and keeping odd harmonics up to 5th term, substituting (23) and (24) in (22), yields (25), as shown at the bottom of the next page, which can be simplified in terms of identical stacks.

Given the fitting parameters g_{i_n}'s and r_{i_n}'s, the output voltage and/or gain can be plotted vs input amplitude hence the AM – AM conversion distortion can be estimated. This is shown in the simulation results section.

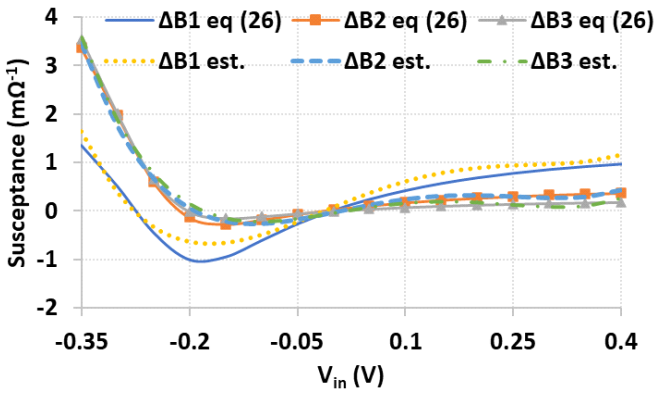


Fig. 15. The variation of the compensated susceptance imposed on each stage at the operating frequency of 28 GHz.

B. AM – PM Conversion Distortion

When calculating (4), $\mathrm{Im}\,(Y_n) = B_n$ was supposed to be fully compensated, i.e. it is required to be zero. As explained in previous subsection, the susceptance of the load projected to each stack varies with amplitude of the input signal. In other words, B_n does not remain zero for the whole input amplitude range. This in fact translates to phase variation of the load seen by each stack and must be quantified to envision the impact of the AM – PM on the modulation schemes.

To define the variation of the susceptance, called ΔB_n, from the ideal zero value, first the non-compensated value is extracted from (26), as shown at the bottom of the page, where b_m and d_m are defined in (27). b_m and d_m are complementary modifications to the definitions of coefficients "*b*" and "*d*" in (4.b) and (4.d), respectively.

$$b_m = B_{in_{n+1}}\left(C_n + C_{gd}\right) + b, \tag{27.a}$$

$$d_m = B_{in_{n+1}}\left(C_{gs} + C_n + C_{gd}\right) + d, \tag{27.b}$$

where "b" and "d" are defined in (4.b) and (4.d), respectively. Subtracting the compensated nominal value defined by (20) from (26) gives the ΔB_n. The susceptance variation of ΔB_n vs input signal is plotted in Fig. 15.

By the same token, ΔB_i can be approximated using third order polynomials as depicted in Fig. 15. To proceed with AM – PM conversion distortion quantification, the phase rotation per stage is calculated as

$$\Delta\theta_n = \tan^{-1}\left(\Delta B_{n_{eff}}\left(A_c\right) \times R_{opt_{n_{eff}}}\left(A_c\right)\right). \tag{28}$$

The "*eff*" subscription in (28) is the root mean square (RMS) calculated from polynomial estimation of the corresponding parameter. Still the AM – PM conversion distortion due to input matching is required to complete the analysis. In this respect, the approach proposed in [31] is followed to support the rest of the analysis in this section. To express the effective values of the terms in (28), i.e. $R_{opt_{n_{eff}}}$ and $\Delta B_{n_{eff}}$, their corresponding nominal values expressed in (24) and (29) are first plotted/extracted against input amplitude, shown in Figs 14(c) and 15, respectively.

$$\Delta B_n\left(V_{in}\right) = b_{1_n} V_{in} + b_{2_n} V_{in}^2 + b_{3_n} V_{in}^3, \tag{29}$$

The plots are then estimated based on power series curve-fitting to extract the coefficients (Figs 14(c) and 15). Substituting the $V_{in} = A_c \cos(\omega t)$ into the extracted polynomials and calculating the RMS values yields the $R_{opt_{n_{eff}}}$ and $\Delta B_{n_{eff}}$ reflected in (30) and (31), as shown at the bottom of the next page, respectively. By the same token, the variation of the effective value of the gate-source capacitance, i.e. $\Delta C_{gs_{eff}}$, is obtained from Fig 14(b) and expressed in (34), as shown at the bottom of the next page.

$$C_{gs}\left(V_{in}\right) = C_0 + C_1 V_{in} + C_2 V_{in}^2 + C_3 V_{in}^3 + C_4 V_{in}^4 \tag{32}$$

Accordingly, the overall phase rotation with respect to amplitude variations, i.e. the AM – PM conversion distortion,

TABLE I
TRANSISTORS' PARAMETERS

Parameter	Value	Parameter	Value
V_{DD}	3.6 V	g_m	650 mS
I_D	100 mA	C_{gs}	420 fF
R_1	11 kΩ	C_{gd}	50 fF
$R_2 - R_4$	18 kΩ	C_{bb}	12 fF
R_5	7 kΩ	R_L	50 Ω
		R_{opt}	12.5 Ω
C_2	182 fF	C_{ds2}	253.6 fF
C_3	71 fF	C_{ds3}	632 fF
C_4	45.5 fF	C_{ds4}	1.146 pF

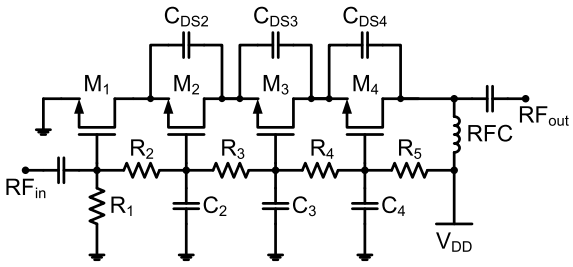


Fig. 16. Schematic of the designed 4-stacked mm-wave PA circuit.

$$a_1\left(A_c\right) \approx \sum_{n=1}^{N}\left(g_{0_n} r_{0_n} + A_c^2\left(\frac{g_{2_n} r_{0_n}}{4} + \frac{3g_{1_n} r_{1_n}}{8} + \frac{3g_{0_n} r_{2_n}}{4}\right) + A_c^4\left(\frac{g_{4_n} r_{0_n}}{8} + \frac{5g_{3_n} r_{1_n}}{32} + \frac{5g_{2_n} r_{2_n}}{24} + \frac{5g_{1_n} r_{3_n}}{16} + \frac{5g_{0_n} r_{4_n}}{8}\right)\right)$$
$$\cong N \times \left(g_0 r_0 + A_c^2\left(\frac{g_2 r_0}{4} + \frac{3g_1 r_1}{8} + \frac{3g_0 r_2}{4}\right) + A_c^4\left(\frac{g_4 r_0}{8} + \frac{5g_3 r_1}{32} + \frac{5g_2 r_2}{24} + \frac{5g_1 r_3}{16} + \frac{5g_0 r_4}{8}\right)\right). \tag{25}$$

$$Im\left\{Y_n\right\} = B_n = \left(\frac{\omega C_{gs}\left(ac + b_m d_m\right) - g_m\left(a d_m - b_m c\right)}{c^2 + d_m^2}\right) \tag{26}$$

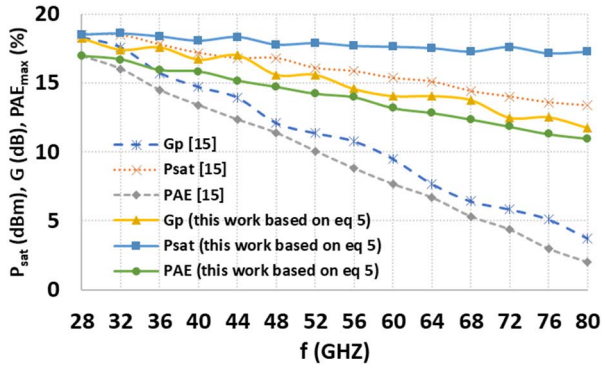


Fig. 17. Comparison between the simulated results of the proposed PA design dimensioning rule of (5) and that of conventional methods explained in [15], in both methods phase compensation method of (20) [21] has been applied. Simulations based on TSMC28nm bulk CMOS PDK has been exploited for both methods.

throughout the whole structure can be expressed as in (35), as shown at the bottom of the page.

VI. Simulation Results

In order to verify the proposed design methodology, first the impact of the proposed gate capacitance, C_n, dimensioning on the performance of the stacked MOS PA topology was simulated and compared to that of the conventional model presented in [15]. Then, a 28GHz four-stack CMOS PA was designed and simulated based on 28nm bulk 8M1P CMOS technology (Fig. 16). Table I summarizes the power supply rail and transistor's parameters. An R_L of 50Ω defines the optimum load to be 12.5Ω per stacked transistor, i.e. R_{opt}. Resistors R_1 – R_5 were used for DC biasing of transistors M_1 – M_4. Their resistance values were chosen to be much higher than the impedances of the gate capacitances C_2 – C_4 at the desired frequency band. Each single transistor device has 32 gate fingers with a total width of 600nm.

Fig. 17 shows simulated power gain, G_p, saturated output power, P_{sat}, and PAE_{max} at different frequencies for a 4-stack MOS PA based on the method introduced in [15] along with that of proposed in this work. Using (5) along with (20), which was applied in both cases, offers an approximately 8dB of power gain, 4dB of maximum output power, and 7% of maximum efficiency improvement at the maximum operating frequency of 80GHz, which corresponds to almost 200% gain, 30% output power and 450% efficiency increase, respectively. This performance increase is quite essential in mm-wave applications.

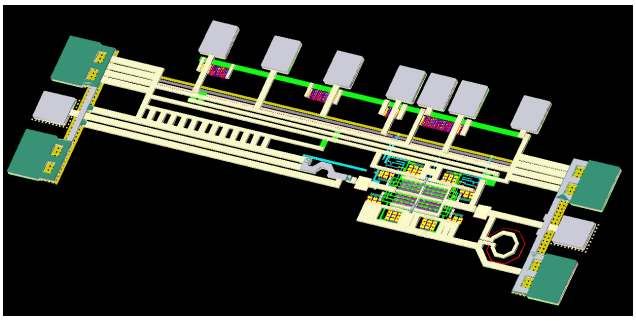


Fig. 18. 3D view of the EM simulated PA structure.

Fig. 18 displays the EM structure of the layout of a 28GHz 4-stacked MOS PA simulated in ADS Momentum. In order to be closer to realistic circuit behavior, the EM structure includes all the interconnects from signal pads to the matching networks and then to the input/output of the PA and biasing pads to the biasing nodes of the PA. In other words, any metallic interconnect starting from (and including) M1 to M8 is included in the EM structure. It is worth bearing in mind that ideal ground is considered only at the ground pad extent consistent with the probe tips and the ground what is laid out inside the chip does not include any ideal ground connection. For this reason, the stability simulations include the ground network effects. Simulation was configured to conduct adaptive frequency sampling (AFS) using microwave engine (μW-Eng) to account for more radiation losses, coupling in/between the metallic routings. It should be noted that active parts extracted based on provided PDK model along with the parasitic extraction routine.

As mentioned in section IV, due to mismatch between the susceptance of the consecutive stages as well as additional

$$R_{opt_{n_{eff}}}(A_c) = \sqrt{\frac{35}{128}A_c^8 r_{4_n}^2 + A_c^6\left(\frac{5}{8}r_{2_n}r_{4_n} + \frac{5}{16}A_c^6 r_{3_n}^2\right) + A_c^4\left(\frac{3}{4}r_{0_n}r_{4_n} + \frac{3}{4}r_{1_n}r_{3_n} + \frac{3}{8}r_{2_n}^2\right) + A_c^2\left(r_{0_n}r_{2_n} + \frac{1}{2}r_{1_n}^2\right) + r_{0\_n}^2} \tag{30}$$

$$\Delta X_{n_{eff}}(A_c) = \frac{A_c}{4} \times \sqrt{45\left(A_c^2 b_{3_n} + \frac{2}{5}x_{1_n}\right)^2 + \frac{4}{5}b_{1_n}^2 + 24A_c^2 b_{2_n}^2} \tag{31}$$

$$\Delta C_{gs}(V_{in}) \cong \frac{\partial C_{gs}}{\partial V_{in}} \times V_{in} = C_1 V_{in} + 2C_2 V_{in}^2 + 3C_3 V_{in}^3 + 4C_4 V_{in}^4 \tag{33}$$

$$\Delta C_{gs_{eff}}(A_c) = \frac{A_c}{4} \times \sqrt{A_c^2\left(70\left(A_c^2 C_4 + \frac{4}{7}C_2\right)^2 + \frac{8}{7}C_2^2\right) + 45\left(A_c^2 C_3 + \frac{2}{5}C_1\right)^2 + \frac{4}{5}C_1^2} \tag{34}$$

$$\Delta\theta_{total} = \tan^{-1}\left(\omega_o \Delta C_{gs_{eff}} R_s\right) + \sum_n \tan^{-1}\left(\Delta B_{n_{eff}}(A_c) \times R_{opt_{n_{eff}}}(A_c)\right) \tag{35}$$

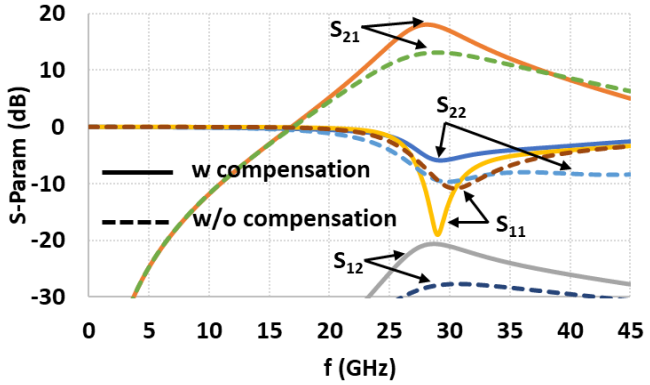


Fig. 19. S-parameters improvement of the designed 28GHz 4-stack MOS PA before and after the proposed phase compensation method (20) [21].

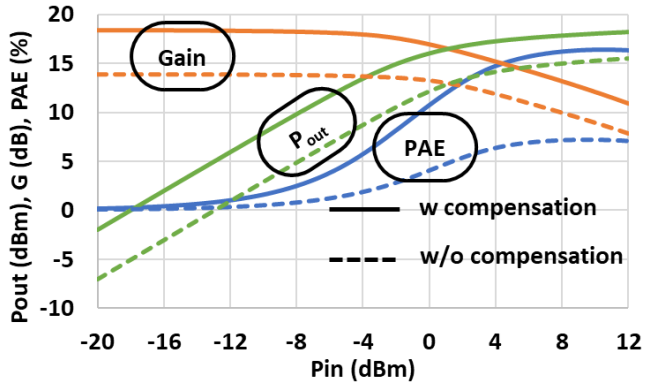


Fig. 20. Performance improvement offered by the proposed compensation method at 28GHz.

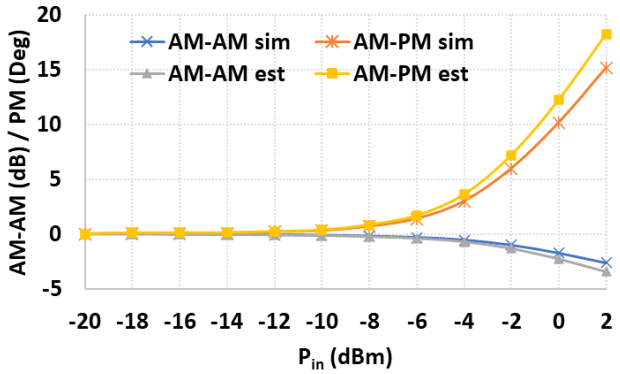


Fig. 21. Simulated vs. analytical AM-AM/PM conversion. Analytical results are calculated from (26) and (35).

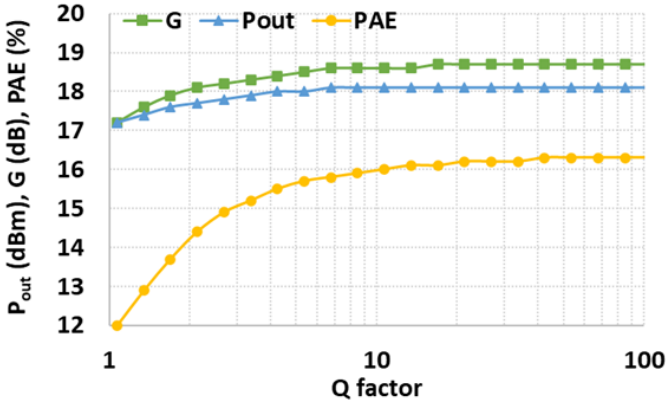


Fig. 22. The impact of the Q factor of the C_n's and C_{ds_n}'s on the performance of the designed stacked MOS PA circuit.

capacitance of pw-dnw, phase misalignment between the stacked transistors poses performance degradation. To compensate for such misalignment the method presented in [21] was deployed, the impact of which over different frequency bands is reflected in Fig. 19. The mentioned method offered a ~20% improvement in phase alignment between the transistors which yielded an improvement in gain and input match, but the reverse isolation and output match were somewhat degraded.

Large signal properties of the designed 28GHz 4-stack PA are plotted in Fig. 20. Based on the design guidelines in this work along with the phase compensation method of [21], the performance of the designed PA is improved by almost 30% in power gain, 200% in PAE, and it gives 5dB more output power.

The AM-AM, the orange dash dotted and yellow dotted lines, and AM-PM conversion, the solid blue and gray dashed lines, properties of the designed PA are illustrated in Fig. 21. Based on (21) – (35), the estimations were calculated for each single stage separately, then added up to form the final AM-AM/PM distortion. Fig. 20 shows a good agreement between the simulated results and theoretical analysis described in this work.

To get an insight into the impact of the quality factor, Q, of the gate capacitances C_n's and drain-source capacitances C_{ds_n}'s, the PDK capacitances were replaced by the ideal capacitances in series with ideal resistors and the value of the resistors swept corresponding to sweep Q factor in the simulation setup. The corresponding results are shown in Fig. 22.

As can be seen from Fig. 22, so long as the Q factor of the mentioned capacitances are above 5, the performance of the PA remains within 0.5dB and/or 0.5% discrepancy with respect to its infinite Q factor counterpart. The simulated Q factor of the capacitances laid out using the mentioned PDK are above 25 which satisfy the design requirements.

VII. Conclusion

High frequency impact on stacked MOS PAs was studied in this paper. Based on analysis, it was shown the traditional device dimensioning is only valid for frequencies up to $f_t/10$. After that frequency it was demonstrated that the optimal load degrades drastically necessitating a modification in device dimensioning, which was proposed in this paper. The impact of the phase rotation on the performance as well as the optimum number of the stacked transistors were studied, and it was shown that the product $N_{max} \times \theta$ is always constant and equal to 133°. After reviewing a negative capacitance compensation method, the AM-AM and AM-PM conversion distortion due to variation of the optimum load was studied based on sensitivity analysis. The theoretical expressions were

evaluated against the simulations. Finally, a 28GHz 4-stack MOS PA was designed and co-simulated using EM tools along with the passive structures of the circuit. The simulation results confirm the validity of the analysis expressed in this work. The results in the paper will help to minimize the inevitable performance degradation as a function of operating frequency in the PA design.

REFERENCES

[1] *Georgia Tech Electronics and Micro-System Lab*. Accessed: Oct. 2, 2021. [Online]. Available: https://gems.ece.gatech.edu/PA_survey.html

[2] T. Tuovinen, N. Tervo, and A. Parssinen, "Downlink output power requirements with an experimental-based indoor LOS/NLOS MIMO channel models at 10 GHz to provide 10 Gbit/s," in *Proc. 46th Eur. Microw. Conf. (EuMC)*, Oct. 2016, pp. 505–508.

[3] N. Tervo, J. Aikio, T. Tuovinen, T. Rahkonen, and A. Parssinen, "Effects of PA nonlinearity and dynamic range in spatially multiplexed precoded MIMO systems," in *Proc. 22th Eur. Wireless Conf.*, 2016, pp. 294–299.

[4] H.-F. Wu, Q.-F. Cheng, X.-G. Li, and H.-P. Fu, "Analysis and design of an ultrabroadband stacked power amplifier in CMOS technology," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 1, pp. 49–53, Jan. 2016.

[5] K. Cho and R. Gharpurey, "An efficient class-G stage for switching RF power amplifier applications," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 4, pp. 597–601, Apr. 2019.

[6] Y.-F. Tsao and H.-T. Hsu, "A 52–58 GHz power amplifier with 18.6-dBm saturated output power for space applications," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 6, pp. 1927–1931, Jun. 2021.

[7] J. Cui, S. Helmi, Y. Tang, and S. Mohammadi, "Stacking the deck for efficiency: RF-to millimeter-wave stacked CMOS SOI power amplifiers," *IEEE Microw. Mag.*, vol. 17, no. 12, pp. 55–69, Dec. 2016.

[8] A. K. Ezzeddine, H. C. Huang, and J. L. Singer, "UHiFET—A new high-frequency high-voltage device," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2011, pp. 1–4.

[9] J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Multigate-cell stacked FET design for millimeter-wave CMOS power amplifiers," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 2027–2039, Sep. 2016.

[10] S. R. Helmi, J.-H. Chen, and S. Mohammadi, "High-efficiency microwave and mm-wave stacked cell CMOS SOI power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2025–2038, Jul. 2016.

[11] K. Ning and J. F. Buckwalter, "An 18-dBm, 57 to 85-GHz, 4-stack FET power amplifier in 45-nm SOI CMOS," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2018, pp. 1453–1456, doi: 10.1109/MWSYM.2018.8439649.

[12] K. Ning and J. F. Buckwalter, "A 28-GHz, 18-dBm, 48% PAE stacked-FET power amplifier with coupled-inductor neutralization in 45-nm SOI CMOS," in *Proc. IEEE BiCMOS Compound Semiconductor Integr. Circuits Technol. Symp. (BCICTS)*, Oct. 2018, pp. 85–88, doi: 10.1109/BCICTS.2018.8550832.

[13] Z. Zong *et al.*, "A 28 GHz two-way current combining stacked-FET power amplifier in 22 nm FD-SOI," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Mar. 2020, pp. 1–4, doi: 10.1109/CICC48029.2020.9075906.

[14] O. El-Aassar and G. M. Rebeiz, "A cascaded multi-drive stacked-SOI distributed power amplifier with 23.5 dBm peak output power and over 4.5-THz GBW," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 3111–3119, Jul. 2020.

[15] H.-T. Dabag, B. Hanafi, F. Golcuk, A. Agah, J. F. Buckwalter, and P. M. Asbeck, "Analysis and design of stacked-FET millimeter-wave power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 4, pp. 1543–1556, Apr. 2013.

[16] J. P. Aikio, M. Hietanen, N. Tervo, T. Rahkonen, and A. Parssinen, "Ka-band 3-stack power amplifier with 18.8 dBm p_{sat} and 23.4% PAE using 22 nm CMOS FDSOI technology," in *Proc. IEEE Topical Conf. RF/Microwave Power Model. Radio Wireless Appl. (PAWR)*, Jan. 2019, pp. 1–3.

[17] M. H. Montaseri, T. Rahkonen, J. Aikio, and A. Parssinen, "Performance improvement of multi-stacked CMOS mm-wave power amplifiers based on negative capacitance phase compensation," in *Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits (INMMIC)*, Jul. 2018, pp. 1–3.

[18] M. H. Montaseri, J. Aikio, T. Rahkonen, and A. Parssinen, "Optimum number of transistors in stacked CMOS millimeter-wave power amplifiers," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–4.

[19] M. H. Montaseri, J. Aikio, T. Rahkonen, and A. Parssinen, "Design of stacked-MOS transistor mm-wave class C amplifiers for Doherty power amplifiers," in *Proc. IEEE Nordic Circuits Syst. Conf. (NORCAS), NORCHIP Int. Symp. System Chip (SoC)*, Oct. 2018, pp. 1–5.

[20] S. Pornpromlikit, J. Jeong, C. D. Presti, A. Scuderi, and P. M. Asbeck, "A watt-level stacked-FET linear power amplifier in silicon-on-insulator CMOS," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 57–64, Jan. 2010.

[21] J. P. Aikio and T. Rahkonen, "A comprehensive analysis of AM–AM and AM–PM conversion in an LDMOS RF power amplifier," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 2, pp. 262–270, Feb. 2009.

[22] A. B. Amado-Rey *et al.*, "Analysis and development of submillimeter-wave stacked-FET power amplifier MMICs in 35-nm mHEMT technology," *IEEE Trans. THz Sci. Technol.*, vol. 8, no. 3, pp. 357–364, May 2018.

[23] M. H. Montaseri, R. Vuohtoniemi, J. Aikio, T. Rahkonen, and A. Parssinen, "Design of multi-stacked CMOS mm-wave power amplifiers for phased array applications using triple-well process," in *Proc. IEEE Nordic Circuits Syst. Conf. (NORCAS), NORCHIP Int. Symp. System Chip (SoC)*, Oct. 2018, pp. 1–5.

[24] C.-W. Wu, Y.-H. Lin, Y.-H. Hsiao, C.-F. Chou, Y.-C. Wu, and H. Wang, "Design of a 60-GHz high-output power stacked-FET power amplifier using transformer-based voltage-type power combining in 65-nm CMOS," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4595–4607, Oct. 2018.

[25] Y. Chang, B.-Z. Lu, Y. Wang, and H. Wang, "A Ka-band stacked power amplifier with 24.8-dBm output power and 24.3% PAE in 65-nm CMOS technology," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2019, pp. 316–319.

[26] D. Manente, F. Padovan, D. Seebacher, M. Bassi, and A. Bevilacqua, "A 28-GHz stacked power amplifier with 20.7-dBm output p_{1dB} in 28-nm bulk CMOS," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 170–173, 2020.

[27] K. Dasgupta, S. Daneshgar, C. Thakkar, J. Jaussi, and B. Casper, "A 26 dBm 39 GHz power amplifier with 26.6% PAE for 5G applications in 28 nm bulk CMOS," in *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, Boston, MA, USA, Jun. 2019, pp. 235–238.

[28] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA, USA: Artech House, 2006.

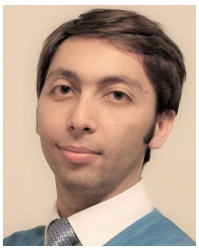
[29] S. Golara, S. Moloudi, and A. A. Abidi, "Processes of AM–PM distortion in large-signal single-FET amplifiers," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 2, pp. 245–260, Feb. 2017.

[30] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. C-29, no. 11, pp. 1715–1720, Nov. 1981.

[31] S. Kulkarni and P. Reynaert, "A 60-GHz power amplifier with AM–PM distortion cancellation in 40-nm CMOS," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2284–2291, Jul. 2016.

[32] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadband system," in *Proc. 2nd Eur. Conf. Satell. Commun. (ECSC)*, Oct. 1991, pp. 179–184.

[33] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.



Mohammad Hassan Montaseri received the M.Sc. degree in electrical engineering from the University of Mazandaran, Babol, Iran, in 2010.

He is currently with the University of Oulu, Oulu, Finland. His research interest includes RF front-end design for mm-wave/(sub) THz ICs for wireless communications applications.

Janne P. Aikio was born in Haukipudas, Finland, in 1977. He received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Oulu, Oulu, Finland, in 2002 and 2007, respectively. He is currently a Post-Doctoral Researcher with the University of Oulu. His technical interests lie in the field of millimeter-wave integrated power amplifiers and modeling and analysis of distortion of RF power amplifiers.

Timo Rahkonen (Member, IEEE) was born in Jyvaskylä, Finland, in 1962. He received the Diploma Engineer, Licentiate, and Doctor of Technology degrees from the University of Oulu, Oulu, Finland, in 1986, 1991, and 1994, respectively. Currently, he is a Professor in circuit theory and circuit design with the University of Oulu, conducting research on linearization and error correction techniques for RF power amplifiers and A/D and D/A converters.

Aarno Pärssinen (Senior Member, IEEE) received the M.Sc., Licentiate in Technology, and Doctor of Science degrees in electrical engineering from the Helsinki University of Technology, Finland, in 1995, 1997, and 2000, respectively.

From 1994 to 2000, he was with the Electronic Circuit Design Laboratory, Helsinki University of Technology, working on direct conversion receivers and subsampling mixers for wireless communications. In 1996, he was a Research Visitor with the University of California at Santa Barbara. From 2000 to 2011, he was with Nokia Research Center, Helsinki, Finland. During 2009 to 2011, he served as a member for the Nokia CEO Technology Council. From 2011 to 2013, he was with Renesas Mobile Corporation, Helsinki, working as a Distinguished Researcher and the RF Research Manager. From October 2013 to September 2014, he was an Associate Technical Director of Broadcom, Helsinki. Since September 2014, he has been with the Centre for Wireless Communications, University of Oulu, Oulu, Finland, where he is currently a Professor. He leads the devices and circuits research area in 6G flagship program financed by the Academy of Finland. He has authored and coauthored one book, two book chapters, more than 180 international journals and conference papers, and holds several patents. He is also one of the original contributors to Bluetooth low energy extension, now called as BT LE. His research interests include wireless systems and transceiver architectures for wireless communications with special emphasis on the RF and analog integrated circuit and systems design.

Dr. Pärssinen was a recipient of European Microwave Price on the best paper of the European Microwave conference 2020. He served as a member for the Technical Program Committee of International Solid-State Circuits Conference from 2007 to 2017, where he was the Chair of European Regional Committee from 2012 to 2013; and the Chair of the Wireless Sub-Committee from 2014 to 2017. He has been serving as a Solid-State Circuits Society Representative for IEEE 5G Initiative from 2015 to 2019.