Review Article

A Review of 5G Power Amplifier Design at cm-Wave and mm-Wave Frequencies

D. Y. C. Lie,1 J. C. Mayeda,1 Y. Li,2 and J. Lopez1,3

1Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX, USA
2Qorvo Phoenix Design Center, Phoenix, AZ, USA
3NoiseFigure Research LLC, Lubbock, TX, USA

Correspondence should be addressed to D. Y. C. Lie; donald.lie@ttu.edu

Received 12 November 2017; Accepted 28 January 2018; Published 4 July 2018

Academic Editor: Jesus Fontecha

Copyright © 2018 D. Y. C. Lie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The 5G wireless revolution presents some dramatic challenges to the design of handsets and communication infrastructures, as 5G targets higher than 10 Gbps download speed using millimeter-wave (mm-Wave) spectrum with multiple-input multiple-output (MIMO) antennas, connecting densely deployed wireless devices for Internet-of-Everything (IoE), and very small latency time for ultrareliable machine type communication, etc. The broadband modulation bandwidth for 5G RF transmitters (i.e., maximum possibly even above 1 GHz) demands high-power efficiency and stringent linearity from its power amplifier (PA). Additionally, the phased-array MIMO antennas with numerous RF front-ends (RFFEs) will require unprecedented high integration level with low cost, making the design of 5G PA one of the most challenging tasks. As the centimeter-wave (cm-Wave) 5G systems will probably be deployed on the market earlier than their mm-Wave counterparts, we will review in this paper the latest development on 15 GHz and 28 GHz 5G cm-Wave PAs extensively, while also covering some key mm-Wave PAs in the literature. Our review will focus on the available options of device technologies, novel circuit and system architectures, and efficiency enhancement techniques at power back-off for 5G PA design.

1. Introduction

The Fifth-Generation (5G) mobile networks is bringing in the latest wireless revolution, enabling wireless download speed exceeding 10 Gbps for eMBB (enhanced Mobile Broadband) applications, with 100x more wireless connected devices than 4G for mMTC (massive machine type communication) to enable IoE (Internet-of-Everything), and sub-1 ms latency for instant actions with URLL, mMTC (ultrareliable machine type communication) [1]. It will be extremely challenging to achieve those aggressive 5G performance metrics all at once, and thus the 5G revolution is expected to be happening in stages. In the prestandard 5G era at 2014, a benchmark 5 Gbps speed was already achieved in a live over-the-air test network from Ericsson using an innovative new radio interface concept in combination with advanced MIMO technology with wider bandwidths and shorter transmission time intervals at 15 GHz. For the higher frequency cm-Wave/mm-Wave 5G to take place, it will probably start from fixed wireless deployment, as Verizon has proposed its own 5G specification as “Verizon 5G wireless technology” or “V5G”. On the other hand, in March 2017, 3GPP published its first study item reports on the 5G New Radio (NR), the next generation 5G cellular network standard, and the likely global 5G standard for a new OFDM-based air interface designed to support the wide variation of 5G device-types, services, deployments, and spectrum. The biggest difference in V5G and 5G NR is the application focus: V5G is limited to fixed wireless access at 28 GHz, but the 5G NR is targeting all wireless communications applications (fixed and mobile) for all frequencies. V5G intended to deploy a high density of cm-Wave/mm-Wave small cells (i.e., base stations) that will communicate with commercial box set UEs, such as a wireless MODEM or a cable box. With the billions of wirelessly connected devices available for 5G, it becomes particularly critical that one must minimize the power consumption of individual wireless devices and back station/base station (BST) as well as the overall 5G system power consumption.
to achieve the critical reduction in energy usage spec by almost 90% over existing 4G networks [2]. Instead of only using the sub-6 GHz spectra like the 2G/3G/4G cellular networks have done in the past, at least some of the 5G devices and networks will also operate at the higher cm-Wave and mm-Wave frequencies to benefit from larger available spectrum bandwidth, smaller-sized massive MIMO phased-array antennas for 3-Dimensional Beamforming (3DBF).

It is well known the performance of a radio-frequency power amplifier (RF PA) can often dominate the overall transmitter (TX) performance, as its power-added efficiency (PAE) dictates the power and heat dissipation for the entire TX. For enhanced user experience and massive MIMO antennas at cm-Wave/mm-Wave frequencies, the 5G system will require more PAs to be integrated in the RF front-end modules (FEMs), making the design of a 5G PA more critical than that of a 4G PA. To any successful commercial 5G application, the output power ($P_{OUT}$), linearity, reliability, cost, and form factors of a PA are all very important.

Figure 1 illustrates an example of attractive 5G FEM IC array design in cm-Wave/mm-Wave for phased-array MIMO antennas. The 5G PA, low noise amplifier (LNA), T/R switches, phase shifter, and various passives are all integrated into the FEM IC as shown in Figure 1, whose architecture is rather different from their 3G/4G counterparts and also with a much higher level of IC integration. In some cases the antennas may be directly packaged on top of the FEM IC on the wafer-scale to achieve even higher integration with reasonable performance [3, 4]. The high integration requirement of FEM ICs and massive antenna systems may favor silicon-based technologies for 5G mobile products, even though GaAs or GaN FEMs usually have better performances than their silicon counterparts [3–7]. In addition to the high integration requirement, as the TX operation frequency moves to cm-Wave/mm-Wave frequencies, it has been well recognized as a very difficult task to design a high efficiency linear PA to overcome the overheating issue for successful massive MIMO realization. Note that we consider the 15 GHz and 28 GHz devices as operating at cm-Wave but not mm-Wave frequencies in this review, even though the industry often “mis-labels” those devices as “mm-Wave devices” for marketing purpose. For example, Qualcomm’s 5G New Radio (NR) demonstrated an impressive “5G mm-Wave” prototype phone in A.D. 2017 operating at 28 GHz, but it should really be called as cm-Wave prototype as it never operates above 30 GHz. In this paper, we will focus on surveying the latest key development and design examples on the cm-Wave 5G PA design (i.e., at 15 and 28 GHz), since the cm-Wave 5G devices and networks will most likely be deployed earlier than their mm-Wave counterparts. In a particular case, Qualcomm is accelerating mobile deployments for smartphones based on 5G NR Release-15 specification, where the cm-Wave RF front-end design is optimized in a smartphone form factor, with multi-MIMO, adaptive beamforming, and beam tracking, supporting 5G NR interoperability testing and over-the-air trials [8]. Many other companies are devoting lots of resources in realizing the 5G revolution as well, where we believe the sub-6 GHz bands and the cm-Wave frequencies will be utilized first. Note the 15-GHz band was expected to be allocated for 5G, but in WRC15 it was not assigned as a candidate band.

The outline of this paper is as follows. Section 2 discusses the device technology choices for 5G PAs. Sections 3 and 4 present some latest design examples of 5G PAs at 15 GHz and 28 GHz, respectively. Section 5 presents several efficiency enhancement techniques at power back-off for both cm-Wave and mm-Wave PAs, including dynamic supply modulations, all-digital architecture, and broadband Doherty PA with digital tuning.

2. The Choice of Device Technologies 5G PA

Today, the majority of handset PAs are still designed in III–V semiconductor devices technologies because of their superior frequency responses, breakdown robustness, and faster time-to-market than silicon-based counterparts [9, 10]. Since today’s base station PAs require rather high $P_{OUT}$, they are largely designed in low-cost silicon LDMS (Laterally Diffused MOSFETs) for sub-3.5 GHz bands, and in GaAs or GaN at higher frequencies, depending on the exact $P_{OUT}$ requirements [11, 12]. For example, GaN devices are capable of operating at a RF power density of 6–8 W/mm of gate periphery at 4G cellular bands and can deliver an impressive power density of 3.6 W/mm at 86 GHz in continuous-wave (CW) operation [11]. In a separate work, $P_{OUT}$ of 3.6 Watt at 83 GHz was achieved in pulse mode [11] that silicon-based PA technologies (LDMS, SiGe, and CMOS) simply cannot match. However, silicon-based RF PAs do have the advantages in offering higher monolithic integration with added functionalities (e.g., on-chip digital control/selection.
Wave/mm-Wave carrier frequency and the massive MIMO Wave/mm-Wave PAs in Figure 2 for various PA technologies. For applications where at power defense or aerospace applications, but here we expect can. Notemost reported GaN/Ku-bands PAs were for high-atcm-Wave/mm-Wave frequencies while the GaAs/GaN PAs counterparts, being notable to exceed a few Watts RF output silicon PAs still have difficulties competing with their III–V at 15GHz and above, Figure 2 indicates that state-of-the-art As mentioned above, the estimated $P_{\text{OUT}}$ requirements for 5G small cells (femtocells and picocells) are fairly low per PA (i.e., $<20$ dBm), meaning that they could be realizable by silicon-based PAs according to Figure 2. A 5G macrocell, on the other hand, may need to be GaN or GaAs PAs due to their larger $P_{\text{OUT}}$ requirements than those currently used in 4G LTE applications. As an example, Table 1 shows our estimated $P_{\text{OUT}}$ requirements for 5G small cells and large cells applications. One can see the 5G PAs used in femtocells and picocells both have fairly low $P_{\text{OUT}}$ requirements per PA (i.e., $<20$ dBm), which means they could be realizable by silicon-based PAs. A 5G macrocell, on the other hand, will probably need to utilize GaN or GaAs PAs due to their larger $P_{\text{OUT}}$ requirements. Power efficiency, robustness, and cost will eventually determine the preferred device technology for a given 5G PA application.

To see which PA device technologies can be the optimal one for a given 5G applications, we have plotted quite a few data points corresponding to the literature's latest cm-Wave/mm-Wave PAs in Figure 2 for various PA technologies. We have shown several best silicon-based PAs, where CMOS SOI stacked PAs with power combing and stacked SiGe PA demonstrated the best $P_{\text{OUT}}$ for mm-Wave silicon PAs [5–22]. For applications where $P_{\text{OUT}}$ needs to be above ~3 to 10 Watts at 15 GHz and above, Figure 2 indicates that state-of-the-art silicon PAs still have difficulties competing with their III–V counterparts, being not able to exceed a few Watts RF output at cm-Wave/mm-Wave frequencies while the GaAs/GaN PAs can. Note most reported GaN X/Ku-bands PAs were for high-power defense or aerospace applications, but here we expect GaN PA can be also be very attractive for the 5G PA market at $P_{\text{OUT}} \sim 3$–10 Watts level.

### 3. 5G PA Designed at 15 GHz in Several Device Technologies

As mentioned above, the estimated $P_{\text{OUT}}$ requirements for 5G small cells (femtocells and picocells) are fairly low per PA (i.e., $<20$ dBm), meaning that they could be realizable by silicon-based PAs according to Figure 2. A 5G macrocell, on the other hand, may need to be GaN or GaAs PAs due to their higher $P_{\text{OUT}}$ and PAE requirements. Cost and integration level will also be critical factors in deciding the preferred technology for a given 5G PA implementation. Nakatani et al. recently reported an impressive 15 GHz $5 \times 5 \text{mm}^2$ FEM IC that integrates a three-stage PA, a two-stage LNA, and a T/R switch in a 0.15 μm GaAs technology for 5G wideband massive MIMO [23]. The final stage of the PA is designed with a novel Doherty configuration using an output parasitic capacitance ($C_{d_s}$) compensation method as shown in Figure 3. Doherty topology was chosen to mitigate the PAs efficiency degradation at large power back-off; however, it is difficult to design a Doherty PA with high PAE over broadband due to the frequency-dependent $\lambda/4$ inverter required for output load modulation. For a conventional Doherty PA, it is well-known that the main amplifier (main) and auxiliary amplifier (Aux) matching circuits are directly connected at the extrinsic reference plane of the transistors (containing output parasitic capacit $C_{d_s}$). The matching networks transform the transistor’s output impedance to 50 Ω, and a $\lambda/4$ inverted transmission line (TL) is connected after main, which can vary the load at intrinsic nodes depending on $P_{\text{OUT}}$ levels. A conventional Doherty PA thus can realize high PAE at large $P_{\text{OUT}}$ back-off for narrow-band signals. However, for the proposed wideband 15 GHz Doherty PA in Figure 3, the frequency-dependent matching is removed with $C_{d_s}$ neutralization using the shunt inductors ($L_{\text{res}}$) through resonance. After the resonator tank of the main amplifier

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>RF $P_{\text{OUT}}$ (dBm)</th>
<th>Number of Users</th>
<th>RF $P_{\text{OUT}}$ Per PA (dBm)</th>
<th>Potential PA Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Femtocell</td>
<td>0–24</td>
<td>1 to 20</td>
<td>&lt;20</td>
<td>CMOS/SOI, SiGe, GaAs</td>
</tr>
<tr>
<td>Picocell</td>
<td>24–30</td>
<td>20 to 100</td>
<td>&lt;20</td>
<td>CMOS/SOI, SiGe, GaAs</td>
</tr>
<tr>
<td>Microcell</td>
<td>30–40</td>
<td>100 to 1000</td>
<td>&lt;27</td>
<td>GaAs, GaN, CMOS/SOI, SiGe</td>
</tr>
<tr>
<td>Macrocell</td>
<td>40–47</td>
<td>1000+</td>
<td>&gt;27</td>
<td>GaN, GaAs</td>
</tr>
</tbody>
</table>

![Figure 2: Recently published RF PA performances in silicon and III–V technologies in the literature with $P_{\text{OUT}}$ versus frequency [6]. Note even the state-of-the-art silicon PAs (in the “dashed rectangular box”) have difficulties reaching Watt level RF output at mm-Wave frequencies.](image-url)
two-stage cascode pre-drivers achieve more than 30 dB linear gate width. The high-power output stage, together with the characteristic impedance $R_{GaAs}$ technology [23, 24], © 2017 IEEE. A prototype 64-element array panel (80 × 80 mm$^2$) with its RF FEM (5 × 5 mm$^2$) for 5G cm-Wave applications in a 0.15 μm GaAs technology [23, 24], © 2017 IEEE.

in Figure 3, a $\lambda/4$ inverted TL is connected for the load modulation at the intrinsic node depending on its $P_{OUT}$. The characteristic impedance $R_{opt}$ in Figure 3 is the optimum impedance for high $P_{OUT}$ performance. Simulations suggest the bandwidth for effective load modulation for high PAE at the intrinsic node can be +/-10% at 15 GHz. Measured data shows the final stage Doherty PA and the three-stage PA achieved a drain efficiency (DE) at 8 dB back-off of 22% and 12%, respectively, at $V_{DD} = 4$ V at 14.5–15.0 GHz.

A picture of the 15 GHz RF FEM packaged in a 32-lead QFN using the above 3-stage GaAs PA is shown in Figure 4, together with the prototyped 64 elements array panel for 5G massive MIMO BST/small-cell applications [23, 24]. The FEM includes all RF active components (LNA, PA, phase shifter, switch, etc.), RF passive components (antenna, filters, combiner/divider, etc.), power management, and control ICs. The RF substrate is a multilayer PCB with RF passives in its internal layer. Each IC and external interface are mounted on its surface, while on the back the corresponding antenna elements are separated by a distance of $\lambda/2$.

For silicon-based cm-Wave 5G PAs, as indicated in Figure 2, recent literature shows CMOS SOI stacked PAs with/without power combining and stacked SiGe PA have demonstrated the best $P_{OUT}$ [15–22]. An example is given in Figure 5, where a three-stage, 4-stack FET CMOS 15 GHz 5G PA is reported [25]. The design of the PA uses 512 multigate-cells to form a 4-stack output device with 614 μm effective gate width. The high-power output stage, together with the two-stage cascode pre-drivers achieve more than 30 dB linear gain centered around 13.5 GHz. The PA achieves $P_{OUT} > 25.1$ dBm and a peak PAE of 32.4% at 13.5 GHz. The small-signal –3 dB bandwidth is 2.6 GHz (~20%). The PA provides the best combination of saturated $P_{OUT}$ and PAE reported to date for 15 GHz CMOS PAs.

Note most reported GaN X/Ku-bands PAs were for high-power defense or aerospace applications, but we expect GaN PA can be very attractive for the 5G PA market at $P_{OUT} \sim 3–10$ Watts level per PA. A high-efficiency two-stage fully integrated 15 GHz PA designed in a low-cost 0.25 μm GaN/SiC process was reported recently and shown in Figure 6 [26]. Postlayout simulation shows that the PA has a high small-signal gain $S_{21}$ of 24 dB, $S_{12} = -41.2$ dB, $S_{22} = -10.0$ dB, and $S_{11} = -12.3$ dB at 15 GHz. Its peak PAE reaches 36.6% at $P_{OUT} = 34$ dBm for CW input. Postlayout simulations also suggest that dynamic supply modulation from 28 V to 10 V may modestly improve the PA's efficiency at power back-off. When the PA is driven with 5/10/20 MHz LTE 16 QAM modulated signals, the simulated output spectra and adjacent channel leakage ratio (ACLR) at $P_{OUT, Linear} = 29.6$ dBm passed the LTE spectrum emission mask (SEM) without any predistortion, and the wider modulated signal bandwidth from 5 MHz to 20 MHz does not noticeably worsen the PA linearity in the RF + digital cosimulations. State-of-the-art literature survey suggests this highly efficient and linear 15 GHz GaN PA may be quite attractive for 5G PA applications at $P_{OUT} \sim 1$ Watt level, as the GaN PA has the smallest die size (1.78 × 0.78 mm$^2$), the 2nd highest peak PAE (i.e., 36.6%), and is the only one reported passing the output SEM specs with LTE 16 QAM input ($PAE_{Linear} = 15.0\%$), even though it is designed with the lowest device $f_T$ of ~27 GHz. However, measurement data is required to corroborate the postlayout simulation results.

4. 5G PA Designed at 28 GHz in Several Technologies

The work in [27] is one of the earliest examples from the industry that reports the design and its measured results of a fully integrated 28 GHz FEM IC, which is designed in a 0.15 μm GaAs pHEMT process (nominal $f_T/f_{max} = 65/95$ GHz, drain-to-gate breakdown voltage = 12 V). The die picture of the FEM IC is shown in Figure 7, which includes a three-stage PA, a three-stage LNA, and a single-pole, double-throw T/R switch. The PA consists of three common-source stages with gate peripheries of 160 μm, 300 μm, and 750 μm. Measured device data suggested a maximum power density of 630 mW/mm, and thus the output stage gate periphery was calculated to achieve 26 dBm $P_{OUT,1dB}$ on die, which can achieve 30 dBm linear EIRP (equivalent isotropically radiated power) in uplink, assuming these FEM ICs are used in a four elements array. The gate peripheries of the driver stages were selected to maintain high PAE with sufficient gain to drive the last stage into compression. All stages were biased in Class AB mode to optimize gain/efficiency trade-off; however, no measured linearity data was reported in this study. The output matching network was designed from nonlinear load-pull simulation for optimum power and PAE, while the interstage matching networks were designed to conjugately match two devices for maximum gain, using a combination of shunt
Figure 5: The circuit schematic of the three-stage multigate-cell stacked 15 GHz PA in 45 nm CMOS-SOI [25], © 2016 IEEE.

Figure 6: (a) Schematic for the two-stage 15 GHz fully integrated GaN PA MMIC design ([26]); (b) Postlayout simulations on the PA’s $S$-parameters; and (c) postlayout RF/digital cosimulated PA spectrum output with LTE 16 QAM 20 MHz input at 15 GHz. $V_{DD,1} = V_{DD,2} = 28$ V.

MIM (metal-insulator-metal) capacitors, shunt microstrip transmission lines (TLs), and series microstrip TLs to create bandpass filters and enhance frequency stability. The DC bias network also included a resistive, multiband reject filter to help stabilize the PA.

A first linear bulk CMOS PA targeting the low-power 28 GHz 5G mobile user equipment (UE) integrated phased-array transceivers application is reported in [28–30]. The output stage of the PA is first optimized for PAE at a desired error vector magnitude (EVM) and range. Next, inductive
source degeneration in the optimized output stage is used in a two-stage transformer-coupled PA shown in Figure 8. The design purposely broadens the interstage impedance matching bandwidth to help reducing distortion. A small 14 pH inductive degeneration is selected in this PA design, which lowered device power gain at $P_{\text{OUT}} = 12$ dBm from 10 to 8 dB and reduced its PAE at the same $P_{\text{OUT}}$ from 48% to 44% in measured test devices. It is possible that unwanted loss resistance in series with the 14 pH $L_{\text{deg}}$ also contributed to the PAE degradation. The chosen $L_{\text{deg}}$, however, did not degrade PA’s $P_{\text{OUT}}$ significantly. The authors had also minimized the ground path impedance for the test device of by using a wide and stacked metal mesh surrounding the device for grounding. The die picture of this 28 GHz PA is shown in Figure 9. The 28 nm bulk CMOS PA achieves $+4.2$ dBm linear $P_{\text{OUT}}$ with 9% PAE (at EVM = $-25$ dBc) using a 250 MHz 64 QAM OFDM input (PAPR ~ 9.6 dB). The PA also achieves 35.5%/10% PAE for CW input at $P_{\text{OUT}}$, sat/9.6 dB back-off. These are among the highest measured PAE values reported for K- and Ka-band CMOS PAs.

Next, a state-of-the-art 28 GHz 32-element phased-array transceiver IC designed in a $0.13 \mu m$ SiGe BiCMOS process ($f_T/f_{\text{max}} = 200/280$ GHz) with concurrent dual-polarized beams and 1.4° beam steering resolution for 5G communication was reported recently by Sadhu et al. [31] (Figure 9). The reported RFIC supports concurrent and independent dual-polarized 16-element beams (H and V) operation in either TX or RX mode and can be integrated in a volume-efficient antenna-in-package array. A new T/R switch at the shared antenna interface enables high $P_{\text{OUT}}$ without much PAE degradation of the PA, and a TL-based phase shifter achieves $<1^\circ$ RMS error and $<5^\circ$ phase steps for precise beam control and also minimizes the number of circuit components. As shown in Figure 10, each TX/RX signal path shares an antenna, a passive phase shifter, and a passive combiner/splitter between the TX and RX in TDD (time division duplex) operation using 3 TX/RX switches. Without getting into the design details of the entire phased-array SoC, we will only focus on the discussion on the RF front-end T/R IC portion of the SoC design, which contains the PA,
VGAs (variable gain amplifiers), LNA, phase shifter, and T/R switches.

Each RFFE IC as shown in Figure 10 includes a T/R switch that minimizes the insertion loss of the TX mode. In a traditional T/R switch, each λ/4 TL-based switch is in series with the PA or the LNA, resulting in similar insertion losses in TX versus RX modes. In this proposed design, the λ/4 switch at the output of the cascode PA is eliminated to avoid the TX-mode insertion loss to improve its PAE (Figure 11). Additionally, in the RX mode, the output admittance of the off-state PA presents a low conductance real part in parallel with a high susceptance inductive part. This proposed design, however, utilizes two switched capacitors to resonate out the inductive part, achieving a high real TX input impedance to maximize the RX signal flow to the LNA. The simulated TX impedances for different states of the 2-bit switched capacitor
are compared to a traditional switch impedance on the Smith Chart shown in Figure 11. The removal of the $\lambda/4$ switch in this TX design improves $P_{\text{OUT}}$ by 1.2 dB, while it only degrades the RX NF (noise figure) by 0.6 dB. This measured data indicates $P_{\text{OUT,Sat}} > 16$ dBm per signal path and PA + switch peak efficiency > 20%, while still maintaining a 6 dB NF for the LNA + switch block. The authors calculated that an additional 1.2 dB TX loss per path of the traditional T/R switch approach would have consumed 2.35 W (or 23%) more power in the FEIC than the proposed design to achieve the same $P_{\text{OUT,Sat}}$.

So far we have discussed various circuit architectures for 5G cm-Wave PAs at different integration levels, where GaN, GaAs, CMOS, and SiGe BiCMOS technologies have all been used. It is expected that the $P_{\text{OUT}}$ requirement, as well as the cost and robustness of the 5G FEM, will all be important in deciding the optimal device technology and RFFE IC architecture for a given 5G cm-Wave/mm-Wave PA design. Since it is the average PAE but not the peak PAE of the PA that determines the power and heat dissipation, efficiency enhancement techniques at power back-off shall be discussed next.

5. 5G PA Efficiency Enhancement Techniques at Power Back-Off

5.1. With Dynamic Supply Modulation. The 5G waveforms are with high PAPR and similar to 4G/WLAN, these waveforms will inevitably degrade PAs efficiency at power back-off and considerably worsen the average PAE of a PA. Therefore, both Doherty PA and supply-modulated PA are regarded very attractive for efficiency enhancement for 5G PA design. We will focus on the supply modulation techniques in this section, while discussing some more broadband Doherty 5G PA techniques in the next section to conclude.

Figure 12 shows a simplified transmitter block diagram using the envelope-tracking (ET) technique for supply modulation to improve PAE at power back-off [32]. The design uses a linear-assisted envelope modulator (EM, also known as envelope amplifier), where the RF input signal into the PA contains both AM (amplitude modulation) and PM (phase modulation) components. Li et al. presented a detailed design example by applying the ET techniques to have realized the
world-first fully monolithic single-chip silicon ET-PA with high-efficiency and broadband performance, covering several cellular frequency bands [33]. An envelope-tracked SiGe cascode PA with an integrated CMOS envelope modulator for 3 GPP LTE transmitters is shown in Figure 13 for this work. The entire single-chip ET-PA delivers the linear output power of 24.3 dBm with the overall system PAE of 42% at 2.4 GHz for the 3 GPP LTE 16 QAM modulation. Additionally, it exhibits a highly efficient broadband characteristic for multiband applications. Compared to the fixed-supply cascode PA, this single-chip ET-cascode PA meets the LTE spectral mask and error vector magnitude (EVM) spec at close to its $P_{1\text{dB}}$ compression without predistortion. The SiGe PA and the CMOS envelope modulator are both designed and fabricated in the TSMC 0.35 $\mu$m SiGe BiCMOS process on the same die. This work represents an essential integration step towards achieving a fully monolithic ET-PA for wideband wireless applications, paving the way for a potential single-chip 5G cm-Wave ET-PAs in the future. Note, besides DPD, one can apply envelope-shaping such as DC shifting and/or AC filtering to achieve optimal high-efficiency ET-PA design for broadband wireless applications [38].

Another example to illustrate supply-modulated PA can improve its efficiency of cm-Wave 5G PA applications is shown in Figure 14, where a highly efficient and linear two-stage fully integrated GaN PA operating at 15 GHz (the same design shown in Figure 6) is used to illustrate the point; however for brevity we will not include here the on-going work on optimizing the entire ET-PA system design, which requires the codesign of the EM with the cm-Wave GaN PA. SPICE simulations show that this 2-stage PA achieves an output 1 dB compression $P_{\text{OUT,1dB}} = 32.2$ dBm with a 28.2 dB gain and 30.0% PAE for CW operation at 15 GHz. Its PAE reaches 38.7% at $P_{\text{OUT}} = 34.0$ dBm with a gain of 22.0 dB. Postlayout SPICE simulations also suggest that dynamic supply modulation from 28 V to 10 V as shown in Figure 14 can significantly improve the PA’s efficiency at power back-off, where the PAE improvement at 6 dB back-off is around 10%. When the PA is driven with 5/10/20 MHz LTE 16 QAM modulated signals, the wider modulated signal

Figure 13: (a) Simplified circuitry schematic of the differential cascode SiGe PA used in [33]; (b) simplified block diagram of our CMOS envelope modulator with its connection to the differential cascode SiGe PA on the same die; (c) chip micrograph of the world-first single-chip ET-PA ($1.1 \times 1.5$ mm$^2$); (d) measured overall PAE and EVM (at $P_{1\text{dB}}$) of the single-chip ET-PA for the LTE 16 QAM 5 MHz signal at 0.8/1.75/2.4 GHz.
bandwidth from 5 MHz to 20 MHz does not worsen the PA linearity considerably according to the RF/analog/digital cosimulations (data not shown) [26]. Note as the EM in an ET-PA will consume significant power especially for high bandwidth input signal to satisfy PA’s linearity requirements, the actual PAE enhancement for this GaN ET-PA needs to be measured against the exact linearity specs for a given 5G waveform in the future as the enhancement is dependent on PAPR and signal bandwidth.

Not only can GaN devices offer the highest $P_{OUT}$ for cm-Wave PA design as suggested in Figure 2, they can also often deliver the highest peak PAE compared to GaAs and silicon devices. In order to improve on the envelope-tracking bandwidth of an ET-PA above 100–500 MHz (note the envelope bandwidth is at least twice greater than the RF signal bandwidth due to the nonlinear $I$-$Q$ to polar transformation [32, 33, 38]), we can also take advantage of the faster GaN devices for the broadband supply modulator design. For example, a recent study reports that a buck converter (i.e., a “switcher”) can be integrated with gate drivers in a Qorvo 0.15 $\mu$m RF GaN-on silicon carbide (SiC) technology, which have demonstrated over 90% peak efficiency at up to 200 MHz switching frequency and supply up to 15 Watt peak power [39]. Furthermore, to show the potential ET-PA integration required for some 5G PA applications, a GaN MMIC that integrated a fast dynamic supply and RF PA is shown in Figure 15, where a 10 W PA operating at 10 GHz is combined on the same die with a 100 MHz buck converter with drive circuitry and also with a UHF cascode PA. This ET-PA is packaged with connectors for all the input and control signals, and the only off-chip component is a filter that determines the bandwidth division between the buck and cascode circuits [34].

Another supply modulation technique simpler than ET is average power tracking (APT), in which the supply modulator only responds to the peak of the output at each average $P_{OUT}$ level. As reported in [40] for a multiband LTE handset SiGe PA, the APT-PA with optimal waveform engineering could still achieve similar efficiency as some of the ET-PAs without DPD. As a matter of fact, in today’s commercial market most of the low-/mid-tier handsets use APT while the high-tier handsets use the full-blown ET to track the instantaneous power supply voltage dynamically [39]. As signal PARs and bandwidths continue to raise in 5G, challenges remain on how to achieve a high-efficiency ET-PA, as the efficiency of its supply modulator degrades significantly with wider bandwidth. On the other hand, the APT-PA may become more realistic to provide power saving for 5G applications.

An interesting approach for improving the envelope-tracking speed of supply modulator is to use multiple converters in parallel, with time offsets between them. By coordinating the inputs of $N$ switchers, the output bandwidth can ideally be increased by a factor of $N$. For example, Florian et al. have reported an ET transmitter architecture based on the combination of a novel 3-bit supply modulator and digital predistortion (DPD) [35]. The proposed power converter is shown in Figure 16, which is based on a direct digital-to-analog conversion (DAC) architecture that implements the binary-coded sum of isolated DC voltages, allowing the synthesis of an output envelope waveform with voltage levels in a binary distribution. This design provides a better voltage resolution with respect to typical multilevel switched-sources topologies as there are 8 levels, which enables the correction of the residual discretization error in the ET transmitter by using DPD of the RF signal. This design has demonstrated that one can successfully remove the low-efficiency linear Op-Amp shown in Figure 13(b), potentially improving the overall efficiency of the ET-PA significantly. The proposed ET-PA has been tested with an L-band 30-W lateral-diffused MOS RF high-power amplifier (RF HPA) with 1.4 and 10 MHz LTE signals, where the EM achieved 92% and 83% efficiency, respectively, and the overall efficiencies of the transmitter system are 38.3% and 23.9% at 5.5 and 1.9 W of average RF output power, respectively. Figure 16(b) shows the measured normalized amplitude of the time-domain signal envelope.
at the HPA input and output under ET operation with and without DPD; one can see that, in order to keep the fidelity of the tracked envelope waveform, DPD is highly desired and this will consume more power at the baseband, adding more overhead for developing the power-DAC based ET-PA. Note the proposed power converter is designed using very fast discrete GaN-based power switches with an aggressively compact circuit layout to minimize losses at high operating frequencies, and the proposed direct power-DAC uses a binary asymmetric cascaded multilevel structure to perform ET on the PA with a fairly high supply voltage of 48 V (plus a DC offset voltage around 6 V). Therefore, the proposed technique in Figure 16 can be rather useful to defense/aerospace and 5G base station applications, but it is not clear whether it will be attractive for 5G handset applications yet.

5.2. With Digital-Intensive Design Methods

5.2.1. Digitally Controlled Power-DAC Architecture: Digital PA (DPA) with Power Combining. Digitally assisted and digital-intensive RF TX can benefit from the fast nm-CMOS/BiCMOS devices to provide functional flexibility with high integration and power combining and tuning done digitally, making them very attractive for multimode broadband applications with on-chip digital predistortion (DPD) [41, 42]. For sub-6 GHz band applications, recent all-digital RF transmitters using CMOS digital PA (DPA) and a direct quadrature architecture that can meet both cellular and WLAN specs have been reported [36]. The I/Q digital bit streams directly feed to the input of DPA with no CORDIC transformation; a 2D DPD look-up table (LUT) is developed in the digital front-end (DFE) to improve linearity. The predistorted baseband signal is upsampled to about 800 MHz before it reaches the DPA to suppress far-out noise. The sampling clock of the DFE is derived directly from RF carrier frequency \( f_c \) through a configurable frequency divider instead of using a PLL (phase-locked loop), and a DIV-2 circuit is used to generate the I/Q LO signals. The critical building block, DPA, is shown in Figure 17. Its RF output is taken by combining the currents from \( I\)-PA and \( Q\)-PA, and each \( I\)-PA/\( Q\)-PA is split into multiple power cells, controlled by 13-bit baseband signals BB to select the proper number of power cells. Since the RF output contains quantized baseband info, the DPA is not just a PA, as its output combines modulation and digital-to-analog conversion (DAC) as well. Current-mode Class-D topology is used in this design of switching pairs, where a transformer-compatible matching network is used to approximate zero-voltage switching (ZVS). For the DPD, a training sequence is applied as DPA input to characterize its quasistatic profile, and a 2D-LUT can be iteratively built-up through the loop-back path using gradient search. In normal transmission mode, the \( 32 \times 32\)-point 2D-LUT is used to map the baseband I/Q input into DPA I/Q control bits. The 2D-DPD can correct nonlinearity, I/Q mismatch, and carrier leakage. The entire TX IC is only 0.7 mm\(^2\) in a QFN package, with \( P_{\text{OUTSat}}\) of 24.7 dBm and peak DPA drain efficiency of 37%. For an 802.11 g 54 Mb/s
signal, the maximum $P_{\text{OUT,Linear}}$ is 18.8 dBm for the TX IC with $-25$ dB EVM and 17% DPA drain efficiency. The measured spurious emission is $-133$ dBc/Hz at WCDMA band 1 at 2.17 GHz with no external filtering. When tested with 80 MHz 256-QAM OFDM input, the all-digital TX IC achieves $P_{\text{OUT,Linear}}$ of 15.7 dBm with $-33$ dB EVM, showing great promise for broadband wireless standards, such as LTE and IEEE 802.11ac, and possibly for the cm-Wave/mm-Wave 5G applications.

To validate this kind of digital-intensive DPA/power-DAC design methodology can also work at mm-Wave frequencies, a novel highly linear direct digital-to-mm-Wave DACs PA architecture that can simultaneously achieve high $P_{\text{OUT,Sat}}$ (through large-scale on-chip power combining), good linearization (through dynamic load modulation), and improved back-off efficiency (through supply-switching and load modulation) has been recently reported [19]. This digitally controlled, supply-switched, and load modulated switching PA architecture shown in Figure 18 uses several switching mm-Wave PA unit-cells that can be individually turned ON or OFF with a digital control bit. These PAs are power-combined using a nonisolating power combiner to make an overall linear mm-Wave DAC with high back-off efficiency, through the load modulation of the combiner and turning OFF some of the PAs. A lumped $\lambda/4$ combiner that enables eight-way power combining with a high 75% measured efficiency at 45 GHz is designed using lumped spiral inductor with higher characteristic impedance to enable one-step, low-loss, eight-way power combining. Using the power combiner codesigned with stacked 45 nm SOI CMOS PAs results in an eight-way combined PA array with $P_{\text{OUT,Sat}} = 27$ dBm and a broadband 1 dB bandwidth from 33 to 46 GHz.

Another 45 nm SOI CMOS PA prototype, a 42.5 GHz 3-bit digital to mm-Wave PA array using the above-described linearization architecture achieves $P_{\text{OUT,Sat}} = 23.3$ dBm at 42.5 GHz, a highly linear digital control word (DCW) to output amplitude profile (Differential Nonlinearity (DNL) = 0.5 LSB; INL 1 LSB using end-point fit) and low AM-PM distortion [19]. It was also able to improve the PAs PAE at 6 dB back-off considerably by reaching an excellent ratio of $\text{PAE}_{-6 \text{dB}} / \text{PAE}_{\text{Peak}} = 67.7\%$.

5.2.2. Broadband Doherty 5G PA Using Digital Tuning. As mentioned in Section 3 earlier, Doherty PA is a highly promising design topology for enhancing the 5G PAs PAE at power back-off, even though it has been used mainly as a narrow-band PA efficiency enhancement technique in the past. However, due to its relative design simplicity and good performance, most cellular base stations today use the Doherty PA architecture instead of the slightly superior ET-PA architecture, except for the cases where the PA output power must change over a large dynamic range (say, $>10$ dB), which Doherty PA typically cannot handle as well as ET-PA [39]. The existing silicon mm-wave Doherty PAs in the literature mostly have very limited PAE enhancement at power back-off, mainly due to lossy Doherty power combiners, and/or nonoptimized main/auxiliary PA design operation [37, 43]. As described in Section 3, [19] reported an efficient 15 GHz PA operation over wideband frequency with a Doherty configuration using a parasitic output capacitance neutralization technique for the final stage PA. In addition, Hu et al. [37] have recently reported a fully integrated 28/37/39 GHz multiband Doherty PA for 5G massive MIMO applications. The PAE for the 5G Doherty PA at the power back-off levels are improved, and its bandwidth is also broadened by the low-loss transformer-based Doherty parallel power combiner. This new transformer-based combiner is based on the concept reported by [34] where three $\lambda/4$
lines are used to provide optimum impedance seen by the device output at the fundamental frequency across the wide frequency range of +/-23% at 2.2 GHz. In the design of Hu et al., however, they took it to the next level of integration by using two transformers absorbing these three λ/4 lines to achieve compact and true Doherty load modulation. What is particularly interesting in that design is the adoption of a digital-intensive tuning scheme, where a power-dependent uneven-feeding scheme is used to adaptively provide optimum main/auxiliary PA operation as shown in Figure 19. At the input, an on-chip transformer-based differential quadrature hybrid was used, while nine-section varactor-loaded transmission lines are deployed both for the main and for the auxiliary paths. Unlike the conventional Doherty PA where one relies on an adaptive biasing circuit on the auxiliary peaking amplifier to turn it on for higher \( P_{\text{in}} \) levels, the input conductance of this Class-C auxiliary PA can be tuned to increase significantly at higher \( P_{\text{in}} \), while that of the Class AB main PA remains almost identical. In addition, one also performs dynamically modulation of the auxiliary driver load to achieve larger power gain when \( P_{\text{in}} \) increases. Therefore, these tunings enable the rapid increase of auxiliary PA output current to obtain an optimum Doherty operation with little limitation on the modulation rate. A prototype is implemented in 0.13 μm SiGe BiCMOS and it achieved +16.8/+17.1/+17 dBm peak \( P_{\text{OUT}} \), with 20.3/22.6/21.4% peak PAE at 28/37/39 GHz. It also amplifies 3 Gb/s 64-QAM input with high-efficiency and reasonable linearity in all these three 5G bands (Figure 19).

For 5G systems, the use of wide bandwidth (>$100$ MHz) and massive MIMO are key technologies. Therefore, we expect that the efficiency enhancement technique of Doherty PA design may have good adaptability of digital-intensive PA to wide bandwidth and massive MIMO.

In the end, we would like to mention that it is not entirely clear if 5G PAs in the handsets would be required to go to 28 GHz cm-Wave range in the next couple of years as the standard is still evolving. As an example, one 5G startup PHAZR has developed a solution known as Quadplex, whose technology uses mm-Wave frequencies (24–40 GHz) for the downlink and the sub-6 GHz frequencies (3.5 GHz or 5 GHz) for the uplink, which the company and their collaborators believe this technology can "uniquely enable high-performance, cost-effective, and power-efficient 5G systems" [44]. Additionally, some latest 5G modulation methods, such as the advanced nonorthogonal multiple access

---

**Figure 18:** (a) Digitally controlled, load modulated power-DAC architecture for linear and highly efficiency mm-Wave PA design [19]; (b) schematic of the 33–46 GHz watt-class PA array prototype; (c) schematic of the two-stage 45 nm SOI CMOS stacked PA unit-cell used in the watt-class PA array prototype; (d) chip microphotograph of the 33–46 GHz watt-class PA array prototype. Chip dimensions are 3.2 mm × 1.3 mm without pads © 2013 IEEE.
(NOMA) scheme that serve multiple users at the same time/frequency/code but with different power levels, will complicate the MIMO 5G PA design optimization to achieve the overall highest TX power efficiency under various output power levels and users clustering scenarios [45].

6. Conclusions

In this paper, we have briefly surveyed some recent advanced and promising design trends on cm-Wave and mm-Wave 5G PAs. These PAs are designed in various device technologies: GaAs, GaN, SiGe, bulk CMOS, and CMOS SOI. We covered wideband Doherty PA in GaAs and in SiGe; stacked PA on SOI CMOS; differential bulk CMOS PA with neutralization cap and transformers; CMOS DPA; fully monolithic GaN PA; highly integrated RFFE with LNA, PA, phase shifter, switches for phased-array MIMO, and so forth. These PA designs presented potential solutions for successful 5G cm-Wave front-end IC designs, where unprecedented high PAE at both peak $P_{\text{OUT}}$ and power back-off with good linearity.
are required, while broadband operation and dense low-cost integration suitable for massive MIMO are also critical for the market adoption. For handsets and small cells/BST where the maximum $P_{\text{OUT}}$, linear requirement is low, typically varying ~4 dBm to ~36 dBm, silicon-based PAs can be attractive for highly integrated solutions with digital-intensive operations. For applications requiring higher $P_{\text{OUT}}$, on the other hand, GaAs and GaN technologies may still be largely used to achieve higher PA efficiency. We also discussed some techniques specifically for back-off efficiency enhancement, namely, using the dynamic supply modulations, all-digital PA architecture, and Doherty PA with digital tuning. We believe these techniques will continue to be explored extensively to pave the road to the 5G deployment.

**Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

**Acknowledgments**

The authors wish to acknowledge the funding support of DoD (Department of Defense), including the DARPA Microscale Power Conversion Program and TTU Keh-Shew Lu Regents Chair Endowment. They also like to thank Ms. D. Wang, Dr. N. Cahoon, Dr. A. Joseph, and Dr. D. Haramé at GlobalFoundries for IC fabrication. Finally, they like to thank Professor C. Kuo and Mr. J.-Y. Lai at National Chiao-Tung University (NCTU), Taiwan, for the support on the GaN IC design kit and fabrication.

**References**


