Spectral Efficiency and Energy Efficiency of OFDM Systems: Impact of Power Amplifiers and Countermeasures

Jingon Joung, Chin Keong Ho, and Sumei Sun

***Abstract*—In wireless communication systems, the nonlinear effect and inefficiency of power amplifier (PA) have posed practical challenges for system designs to achieve high spectral efficiency (SE) and energy efficiency (EE). In this paper, we analyze the impact of PA on the SE-EE tradeoff of orthogonal frequency division multiplex (OFDM) systems. An ideal PA that is always linear and incurs no additional power consumption can be shown to yield a decreasing convex function in the SE-EE tradeoff. In contrast, we show that a practical PA has an SE-EE tradeoff that has a turning point and decreases sharply after its maximum EE point. In other words, the Pareto-optimal tradeoff boundary of the SE-EE curve is very narrow. A wide range of SE-EE tradeoff, however, is desired for future wireless communications that have dynamic demand depending on the traffic loads, channel conditions, and system applications, e.g., high-SE-with-low-EE for rate-limited systems and high-EE-with-low-SE for energy-limited systems. For the SE-EE tradeoff improvement, we propose a PA switching (PAS) technique. In a PAS transmitter, one or more PAs are switched on intermittently to maximize the EE and deliver an overall required SE. As a consequence, a high EE over a wide range SE can be achieved, which is verified by numerical evaluations: with 15% SE reduction for low SE demand, the PAS between a low power PA and a high power PA can improve EE by 323%, while a single high power PA transmitter improves EE by only 68%.**

***Index Terms*—Energy efficiency, spectral efficiency, power amplifier, power amplifier switching, OFDM.**

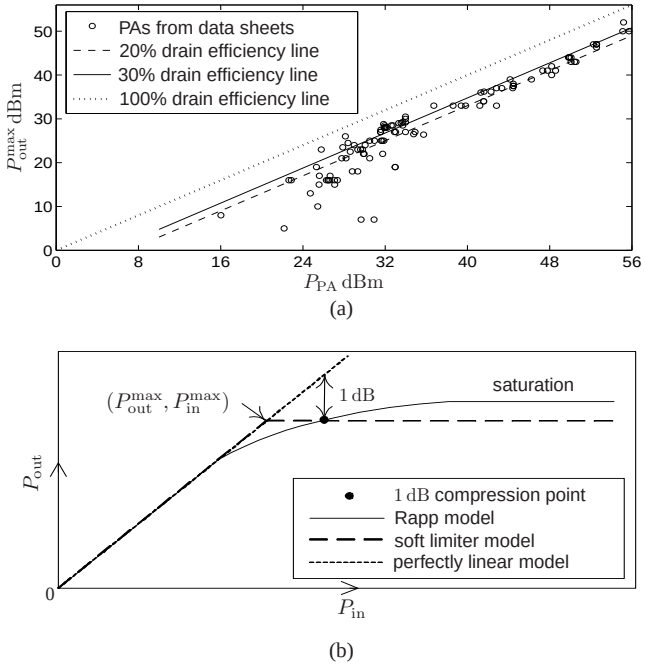

Fig. 1. Two fundamental characteristics of a practical PA. (a) Efficiency: maximum output power $P_{\rm out}^{\rm max}$ (at the linear region) versus consumed power $P_{\rm PA}$. (b) Nonlinearity: PA output power $P_{\rm out}$ versus PA input power $P_{\rm in}$.

I. Introduction

WIRELESS access communication networks consume significant amount of energy to overcome fading and interference, compared to fixed line communication networks [3], [4]. In wireless networks, energy is mostly consumed at the base station (BS) [3], of which a substantial fraction of 50%–80% of overall power is consumed at power amplifiers (PAs) [5]. A measure of the PA efficiency is given by the drain efficiency η that is the ratio of PA output power $P_{\rm out}$ to PA power consumption $P_{\rm PA}$, i.e., $\eta = P_{\rm out}/P_{\rm PA}$. Fig. 1(a) plots PA maximum output power $P_{\rm out}^{\rm max}$ versus $P_{\rm PA}$, based on our survey of commercially available PAs for which we give a summary of the key parameters in Table II in Appendix A. From Fig. 1(a), we see that η at $P_{\rm out}^{\rm max}$ is typically between 20% and 30%, which confirms that the overhead incurred at PA is substantial. To ensure high energy efficiency (EE), the PA characteristics have to be carefully considered in system designs.

On the other hand, high spectral efficiency (SE) is needed to support the growing demands of high-rate applications. Orthogonal frequency division multiplex (OFDM) and orthogonal frequency division multiple access (OFDMA) are two popular spectral efficient systems. However, OFDM and OFDMA modulated signals exhibit high peak-to-average power ratio (PAPR), thus suffering from severe nonlinearity effects [6], [7] as illustrated in Fig. 1(b), in which PA output power $P_{\rm out}$ is shown over the PA input power $P_{\rm in}$. Two commonly used models, namely the Rapp model [8] and the soft limiter model [9], are shown in Fig. 1(b). They are used to describe the nonlinear amplitude (i.e., signal power) distortion, especially at a high power region, while the phase is assumed to be undistorted (the details will be given in Section III, and for more nonlinearity models, refer to the references in [2]). In practice, to circumvent the resulting performance degradation, input backoff (IBO) is implemented by reducing the power of the input signal at the PA, so that the amplification stays within

Parts of this work have been presented at the *IEEE Global Telecommunications Conference (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012 [1], and the *Asia Pacific Signal and Information Processing Association (APSIPA)*, Annual Summit and Conference, Hollywood, CA, USA, Dec. 2012 [2].

The authors are with the Institute for Infocomm Research ($\rm I^2R$), A*STAR, Singapore 138632 (e-mail: {jgjoung, hock, sunsm}@i2r.a-star.edu.sg)

the linearity region as much as possible. While IBO allows high SE to be achieved, it can reduce the EE, because the PA efficiency is typically designed to peak near the saturation point and it usually drops rapidly as the input power decreases [10]. Hence, a tradeoff between SE and EE is inevitable while optimizing with respect to the PA. It is thus important to jointly characterize the role that a PA plays in both SE and EE of wireless communication systems. Recently, circuit power consumption has been taken into consideration for energy efficient system designs [11]–[13], but without consideration of the nonlinearity of the PA.

In this paper, the tradeoff of SE and EE for OFDM systems is analyzed by taking into account the impact of practical PAs that is both *inefficient* and *nonlinear*. To provide tractable results, we assume that the nonlinearity of the PA is modeled by a soft limiter. To capture the PA inefficiency, we propose a nonlinear transmit power model depending on the PA types. We further provide theoretical results to achieve maximum SE and maximum EE from our analysis, and verify the theoretical results through simulations using real-life device parameters. Consequently, it is shown that the practical SE-EE tradeoff increases before a turning point and decreases rapidly after the turning point. In other words, the PA can support a narrow SE-EE tradeoff with only a limited range of SE. In cellular communications, however, a wide range of SE-EE tradeoff is desired because the BSs need high data rates intermittently, yet need to save energy whenever possible to save operation costs. To achieve a wide Pareto-optimal SE-EE tradeoff region, we propose a PA switching (PAS) technique, in which one or more PAs are switched on at any time to maximize the EE while satisfying the required SE, resulting in a high EE over a wide SE range. For example, with 15% SE reduction for low SE demand, the PAS between a low power PA ($25\,\mathrm{W}$ maximum power) and a high power PA ($100\,\mathrm{W}$ maximum power) can improve EE by 323%, while a single high power PA transmitter improves EE by only 68%. Specifically, our key contributions are summarized as follows:

- *Practical SE*: We obtain a closed-form expression of SE with consideration of PA nonlinearity, and show that its approximation is a concave function with a unique maximum with respect to the input power of the PA.
- *Practical EE*: We establish a PA-dependent nonlinear power consumption model from various recent studies on empirical power measurement and parameters for cellular and wireless local area networks. We show that the EE is a piecewise quasi-concave function with a unique maximum point if the PA is perfectly linear.
- *PAS*: We observe that the practical SE-EE tradeoff decreases rapidly after a turning point, i.e., the limited SE-EE tradeoff for dynamic traffic conditions. To circumvent this, we propose a PAS technique. Numerical results show that the SE-EE tradeoff improvement is significant even though practical losses are considered, such as switch insertion loss and switching time overhead.

II. Prologue

This paper attempts to quantify analytically and numerically the degradation of both SE and EE caused by the practical nonlinearities and energy consumption of the PA. Specifically, we define SE, in $\mathrm{b/s/Hz}$, as the amount of bits that are reliably decoded per channel use (i.e., per unit time and per unit bandwidth). We define EE, in $\mathrm{b/J}$, as the total amount of reliably decoded bits normalized by the energy. Thus, SE and EE are given respectively by [14], [15]

$$\mathsf{SE} = \frac{I(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{Y}})}{N} \tag{1a}$$

$$\mathsf{EE} = \frac{T\Omega\,\mathsf{SE}}{TP_{\mathrm{c}}} = \frac{\Omega\,\mathsf{SE}}{P_{\mathrm{c}}}. \tag{1b}$$

Here, $I(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{Y}})$ is the mutual information in $\mathrm{b/s/Hz}$ given the length-N transmitted and received vectors $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{Y}}$, representing an achievable sum rate over N channel uses [16]; Ω is the total bandwidth used; T is the total time used; and P_{c} is the total power consumption including the PA power consumption P_{PA}.

For illustration, consider an *ideal* system without system overhead power consumption, i.e., $P_{\mathrm{c}} = P_{\mathrm{PA}}$. Furthermore, consider the *ideal PA* which is always *perfectly efficient* (dotted line in Fig. 1(a)), i.e,. $P_{\mathrm{PA}} = P_{\mathrm{out}}$, and always *perfectly linear* (dotted line in Fig. 1(b)). Using Gaussian signalling, which is optimal for the ideal PA, we get $\mathsf{SE} = \log_2(1 + P_{\mathrm{out}}/\sigma^2)$, where σ^2 is the noise power [16]. For the ideal system with ideal PA, therefore, asymptotically as P_{c} increases, SE increases proportionally with $\log_2(P_{\mathrm{c}})$ and hence EE decreases proportionally with $\log_2(P_{\mathrm{c}})/P_{\mathrm{c}}$. In other words, the SE-EE tradeoff region is a decreasing convex as observed in [14], [15] when the system and PA are ideal. In contrast, for a practical system, asymptotically as P_{c} increases, the output saturates and so SE saturates to some upper limit, hence EE decreases proportionally with $1/P_{\mathrm{c}}$; moreover, significant overhead power exists, because $P_{\mathrm{c}} > P_{\mathrm{PA}} > P_{\mathrm{out}}$. To account for the degradation of both SE and EE in practice, it is essential to consider practical overhead in system power consumption P_{c} and have a sufficiently accurate, yet tractable, model for the PA (i) on its energy consumption to specify the relationship of P_{PA} and P_{out}, and (ii) on its nonlinearity behavior. In the sequel, we shall address both issues when we determine the SE and EE in (1).

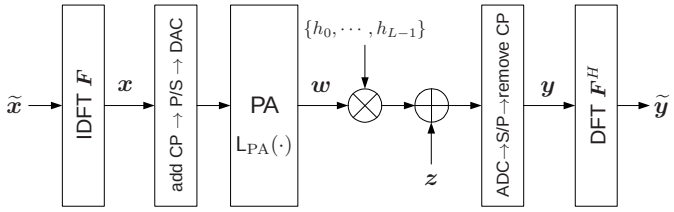

Fig. 2. An OFDM system with a nonlinear memoryless PA represented by function $\mathsf{L}_{\mathrm{PA}}(\cdot)$, assuming perfect synchronization.

III. System Model

Consider the OFDM system with a nonlinear memoryless PA shown in Fig. 2. Without loss of generality (w.l.o.g.), we consider the transmission of one OFDM symbol which consists of N complex-valued data symbols, denoted by the

data vector $\widetilde{\boldsymbol{x}} = [\widetilde{x}_0, \cdots, \widetilde{x}_{N-1}]^T$. The data symbol $\widetilde{x}_n$ is sent on the nth orthogonal subcarrier. These N subcarriers occupy a total frequency band of Ω Hz. The data symbols are assumed to be identical and independently distributed (i.i.d.) subject to the power constraint $\mathrm{E}[|\widetilde{x}|^2] \leq P_{\mathrm{in}}$; here and subsequently, we drop the subcarrier or time index if there is no dependence on it or when there is no ambiguity. We transform $\widetilde{\boldsymbol{x}}$ to the time domain signal vector $\boldsymbol{x} = [x_0, \cdots, x_{N-1}]^T$ according to $\boldsymbol{x} = \boldsymbol{F}\widetilde{\boldsymbol{x}}$, where $\boldsymbol{F}$ is an N-by-N unitary inverse discrete Fourier transform (IDFT) matrix. Thus, $\mathrm{E}[|x|^2] \leq P_{\mathrm{in}}$. Then a cyclic prefix (CP) of length N_{CP} is added to $\boldsymbol{x}$ and passed to a parallel-to-serial (P/S) converter, followed by a digital-to-analogue converter (DAC). We assume the DAC, the analogue-to-digital converter (ADC) and subsequent processing (such as timing and frequency synchronization) are ideal such that, w.l.o.g., we use x_t to represent the output of the DAC at discrete time index t. We rewrite $x_t = a_t e^{j\theta_t}$ where $a_t \triangleq |x_t|$ is the amplitude and θ_t is the phase of x_t where $0 \leq \theta_t < 2\pi$.

Next, the DAC output x_t is amplified through a memoryless PA described by a nonlinear function $\mathsf{L}_{\mathrm{PA}}(\cdot)$ to give the output $w_t = \mathsf{L}_{\mathrm{PA}}(x_t)$, denoted collectively by the vector $\boldsymbol{w} = [w_0, \cdots, w_{N+N_{\mathsf{CP}}-1}]^T$. Under the Rapp and soft limiter models illustrated in Fig. 1(b), we can write $w_t = b_t e^{j\theta_t}$ where $b_t \triangleq |w_t|$ while the phase remains the same as that of x_t. Specifically, the Rapp model describes the amplitude distortion according to

$$\mathsf{L}_{\mathrm{PA}}(a_t) = \sqrt{g}a_t \left(1 + \left(\frac{\sqrt{g}a_t}{b_{\mathrm{sat}}}\right)^{2p}\right)^{-\frac{1}{2p}},$$

where $\sqrt{g} \geq 1$ is a parameter interpreted as the desired linear gain; b_{sat} is the saturation amplitude when $a_t \to \infty$; and p controls the smoothness of the transition from the linear region to the saturation region. Thus, the gain is nonlinear for all input signals. For the soft limiter model, the amplitude distortion follows

$$\mathsf{L}_{\mathrm{PA}}(a_t) = \begin{cases} \sqrt{g}a_t, & \text{if } a_t < a_{\max} \\ b_{\max}, & \text{if } a_t \geq a_{\max}, \end{cases}$$

where $a_{\max} \triangleq \sqrt{P_{\mathrm{in}}^{\max}}$ and $b_{\max} \triangleq \sqrt{P_{\mathrm{out}}^{\max}}$. Thus, the output of soft limiter is clipped to a constant $b_{\max}$ if the input signal exceeds a threshold value $a_{\max}$, and experiences a linear scaling of its input with gain $\sqrt{g}$ otherwise.

Finally, the PA output is transmitted through an L-tap multipath channel $\{h_0, h_1, \cdots, h_{L-1}\}$. Assuming $L \leq N_{\mathsf{CP}}$ and perfect timing synchronization, the CP is removed and the received signal is given by

$$y_t = h_t \otimes w_t + z_t \triangleq r_t e^{j\phi}, \tag{2}$$

for $t = 0, \cdots, N-1$ (for convenience, we shift the time indices to start from 0). Here, $\otimes$ is the circular convolution operator, $z_t \sim \mathcal{CN}(0, \sigma_z^2)$ is an additive white Gaussian noise (AWGN), and r_t and ϕ_t represent the amplitude and phase of y_t. The received signal vector $\boldsymbol{y} = [y_0, \cdots, y_{N-1}]^T$ is transformed via a DFT (i.e., a Hermitian transpose of $\boldsymbol{F}$) to give the frequency domain signal vector $\widetilde{\boldsymbol{y}} = [\widetilde{y}_0, \cdots, \widetilde{y}_{N-1}]^T = \boldsymbol{F}^H \boldsymbol{y}$.

In practice, the time-domain signal after IDFT typically produces a Gaussian-like signal with a high PAPR. It is well known that the nonlinearity of a PA can thus result in significant degradation of the achievable rate of the signal [6]. To analytically model the high PAPR and the nonlinearities, we make the following assumptions:

$A1$: We assume that the data symbols are i.i.d. with complex normal distribution with zero mean and P_{in} variance, denoted as $\widetilde{x} \sim \mathcal{CN}(0, P_{\mathrm{in}})$. Hence the time-domain signals are also i.i.d. with distribution $x \sim \mathcal{CN}(0, P_{\mathrm{in}})$. The time domain signals have very high PAPR and thus they are representatives of the scenario when a high-order modulation is used or when N is large.

$A2$: For tractability of subsequent analysis, we employ the soft limiter model for the PA. A good approximation of the maximum power output $P_{\mathrm{out}}^{\max}$ is given by the one-dB input compression output, where the output power drops 1 dB below the desired power output if the gain is linear as illustrated in Fig. 1(b). Thus, the maximum power input is $P_{\mathrm{in}}^{\max} = P_{\mathrm{out}}^{\max}/g$. We shall use data sheets of commercially available products (e.g., Table II in Appendix A) to extract suitable parameters for g and $P_{\mathrm{in}}^{\max}$ to obtain numerical results. The soft limiter model is analytically tractable, and it can capture the clipping effect in the high power region as the Rapp model (the Rapp model approaches the soft limiter model as p increases). Nonlinearity in low power region of the soft limiter can be assumed to be mitigated by applying linearization techniques, such as feedforward, feedback, and predistortion (refer to the references in [2]), which is the same as the Rapp model.

The assumption $A1$ is independent to the assumption of the soft limiter model in $A2$, because the probability density functions (pdfs) of x and $\widetilde{x}$ do not change regardless of the PA model. In this paper, we focus on point-to-point communications. The spectral regrowth arisen from the nonlinearity of the PA, which increases the adjacent channel interferences to neighboring bands, is not considered explicitly.

Typically, an IBO is performed to mitigate the degradation resulting from PA nonlinearities, by reducing the input signal power P_{in} such that it is much less than $P_{\mathrm{in}}^{\max}$. To reflect this, we write $P_{\mathrm{in}} = \xi P_{\mathrm{in}}^{\max}$, where $\xi \geq 0$ is a power loading factor and is related to the IBO as $\mathrm{IBO} \triangleq 10\log_{10}(\xi^{-1})\,\mathrm{dB}$. By varying ξ, we can then perform IBO to tradeoff between EE and SE.

Based on assumptions $A1$ and $A2$, we shall obtain tractable results which offer insights on how the PA affects the SE and EE in Sections IV and V, respectively. Then we study how this leads to the analysis of a new PA architecture in Section VI, which improves SE and EE tradeoff.

IV. Spectral Efficiency

In this section, we determine the SE in (1a) under assumptions $A1$ and $A2$. To this end, we obtain the mutual information $I(\widetilde{\boldsymbol{X}}; \widetilde{\boldsymbol{Y}})$ for flat fading channels in Section IV-A, and for multipath channels in Section IV-B. For simplicity, we ignore the throughput loss due to the addition of the

CP. We fix the following PA-related parameters: the power loading factor ξ, the gain g in the linearity region and the maximum power output $P_{\text{out}}^{\max}$. Thus the maximum input power $P_{\text{in}}^{\max} = g^{-1}P_{\text{out}}^{\max}$ is also fixed; for convenience, let $\gamma \triangleq P_{\text{out}}^{\max}/\sigma_z^2 > 0$ be the maximum power output normalized by the noise variance σ_z^2.

We use upper case letters to represent random variables, such as X, W, and Y, and lower case letters to represent their realizations, such as x, w, and y. The pdf of random variable X is denoted by $f_X(\cdot)$. Recall that the signals are written in terms of their amplitudes and phases as $x = ae^{j\theta}$, $w = be^{j\theta}$, and $y = re^{j\phi}$.

A. Mutual Information in Flat Fading Channel

Consider the flat fading channel where the number of multipath is $L = 1$. Let $h_0 = 1$, w.l.o.g., as the actual channel attenuation and any fixed energy losses incurred can be reflected by adjusting the noise variance such that the signal-to-noise ratio (SNR) is maintained. Given input $X = Ae^{j\theta}$, the channel model at time index t is

$$Y_t \;=\; W_t + Z_t, \;\; \text{where } W_t = \mathsf{L}_{\text{PA}}(A_t)e^{\theta_t}. \tag{3}$$

The SE, which is given by the achievable rate averaged over N transmissions, is

$$\begin{aligned} I(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{Y}})/N &\overset{(a)}{=} I(\boldsymbol{X};\boldsymbol{Y})/N \overset{(b)}{=} \sum_{t=0}^{N-1} I(X_t;Y_t)/N \\ &\overset{(c)}{=} I(X;Y) \overset{(d)}{=} H(Y) - \log_2 \pi e\sigma_z^2 \text{ [b/s]}. \end{aligned} \tag{4}$$

Here, (a) follows from the facts that the frequency-domain signals (transmitted and received vectors $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{Y}}$) and time-domain signals (transmitted and received vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$) are related by a unitary transform, which does not change the mutual information; (b) follows from the independence of the signals in the time domain (because of the memoryless PA and the i.i.d. transmitted signals and noise); (c) follows from the fact that the mutual information is identical over time, and so the time index can be dropped; and (d) follows from the facts that $I(X;Y) = H(Y) - H(Y|X)$, the conditional entropy $H(Y|X) = H(N)$, and $H(N)$ is the differential entropy of a complex Gaussian random variable with variance σ_z^2 derived by $\log_2 \pi e\sigma_z^2$. The entropy of Y in (4) is given by [16]

$$H(Y) = -\int_y f_Y(y)\log_2 f_Y(y)dy. \tag{5}$$

Nonlinear distortion at the transmitter makes it difficult to derive $f_Y(y)$ in (5) directly. To tackle this problem, we define a binary random variable S that denotes whether clipping at the PA occurs, i.e., $S = 0$ if $A \le a_{\max}$ and $S = 1$ otherwise, and rewrite the pdf of y as $f_Y(y) = \sum_{i=0,1} f_Y(y, S = i)$. Since $X = Ae^{j\theta} \sim \mathcal{CN}(0, P_{\text{in}})$, the random variable A follows the Rayleigh distribution. Thus, we get the probability of S as

$$\begin{aligned} \Pr(S=0) &= \Pr\left(A \le a_{\max}\right) = 1 - \exp\left(-a_{\max}^2 P_{\text{in}}^{-1}\right) \\ &= 1 - \exp\left(-\xi^{-1}\right) \\ \Pr(S=1) &= 1 - \Pr(S=0) = \exp\left(-\xi^{-1}\right). \end{aligned} \tag{6}$$

The numerical computation of the entropy (5) is straightforward with a closed-form expression of $f_Y(y, S = 0)$ and $f_Y(y, S = 1)$, which are derived respectively as follows (see Appendix B):

$$f_Y(y, S=0) \;=\; \mathsf{N}_0(y)\left[1 - \mathsf{Q}_1\left(\sqrt{\mu(y)}, \sqrt{\rho_{\max}}\right)\right] \tag{7a}$$

$$\begin{aligned} f_Y(y, S=1) \;=\; \mathsf{N}_1(y)\Bigg[&\Pr(S=1)\exp\left(\frac{-2b_{\max}y_{\text{Re}}}{\sigma_z^2}\right) \\ &\times \mathsf{I}_0\left(\frac{2b_{\max}|y|}{\sigma_z^2}\right)\Bigg] \end{aligned} \tag{7b}$$

where $\mathsf{N}_0(y)$ denotes the pdf of $\mathcal{CN}\left(0, gP_{\text{in}} + \sigma_z^2\right)$; $\mathsf{Q}_1(\cdot,\cdot)$ is the Marcum-Q-function [17] with parameters $\rho_{\max} \triangleq \frac{2(gP_{\text{in}}+\sigma_z^2)}{gP_{\text{in}}}\sqrt{b_{\max}}$ and $\mu(y) \triangleq \frac{8gP_{\text{in}}(gP_{\text{in}}+\sigma_z^2)}{\sigma_z^4}|y|^2$; $\mathsf{N}_1(y)$ is the pdf of $\mathcal{CN}\left(b_{\max}, \sigma_z^2\right)$; y_{Re} is the real part of y; and $\mathsf{I}_0(\cdot)$ is the modified Bessel function of first kind [17].

B. Mutual Information in Multipath Channel

We now consider the general case of an L-tap multipath channel, where $1 \le L \le N_{\text{CP}}$. The received signal in the time domain is given by (2). If the amplification of PA is perfectly linear, then the mutual information is given equivalently in the frequency domain as $I(\boldsymbol{X};\boldsymbol{Y})/N = \sum_{k=0}^{N-1} I(\widetilde{X}_k;\widetilde{Y}_k)/N = \sum_{k=0}^{N-1}\log_2(1+|\widetilde{H}_k|^2 gP_{\text{in}}/\sigma_z^2)/N$, where $\widetilde{H}_k$ is the frequency domain channel, see e.g., [16]. In our case of interest, however, the nonlinear PA makes the exact analysis of the mutual information intractable, because the PA nonlinearities result in a correlated interference in the frequency domain which is not formulated as a closed-form expression. Instead, we obtain a lower bound for the mutual information (see Appendix C):

$$\begin{aligned} I(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{Y}})/N \;\ge\; &\sum_{t=0}^{L-1} I(X_t; Y_t, \cdots, Y_{t+L-1}|X_1, \cdots, X_{t-1})/N \\ &+ \sum_{t=L}^{N-1} I_t^{\text{LB}}/N \end{aligned} \tag{8}$$

where I_t^{LB} is the mutual information of flat fading channel (3) with the SNR given by the equivalent channel (C.4). As $N \to \infty$, the first term approaches zero, while the second term equals approaches I_t^{LB} which is in fact independent of t (we drop the index subsequently). Thus, the lower bound in (8) is given asymptotically by I^{LB} for $N \gg L$. Note that I^{LB} can be computed from (4) directly. Numerical results (not included) show that the bound is typically tight if the power of the multipath decreases exponentially over the channel delay.

C. Analytical Results on SE

Using (7a) and (7b) into (5), we find $H(Y)$ and get $I(X;Y)$ from (4) in flat fading channels. Similarly, from (C.3), we can obtain the mutual information of the signals in multipath channels. Accordingly, we derive the SE in (1a) as a function of ξ as

$$\mathsf{SE}(\xi) = H(Y) - \log_2 \pi e\sigma_z^2, \tag{9}$$

where note that the entropy $H(Y)$ is a function of ξ as the conditional probabilities in (7a) and (7b) are functions of $P_{\rm in} = \xi P_{\rm in}^{\max}$ and $b_{\max} = \sqrt{gP_{\rm in}\xi^{-1}}$.

If the PA is perfectly linear, i.e., $b_{\max} \to \infty$ and thus $a_{\max} \to \infty$, it can be easily checked that $f_Y(y, S=0) = \mathsf{N}_0(y)$ from (B.1)–(B.3) and $f_Y(y, S=1) = 0$ from (B.5) as $P(S=0) = 1$ and $P(S=1) = 0$ in (6). Thus, $H(Y) = \log_2 \pi e \left(gP_{\rm in} + \sigma_z^2\right)$ and we recover the well-known SE for ideal PA as

$$\mathsf{SE}^{\rm ideal}(\xi) = \log_2\left(1+\gamma\xi\right). \tag{10}$$

For tractable analysis, the SE in (9) is approximated under the assumption of low power input signal to PA, i.e., small ξ. If $\xi \ll 1$, we can approximate the joint pdfs in (7a) and (7b) as follows:

$$\begin{aligned} f_Y(y, S=0) &\approx \mathsf{N}_0(y) \\ f_Y(y, S=1) &\approx \mathsf{N}_1(y)\Pr(S=1). \end{aligned} \tag{11}$$

The approximations comes from the observation that $\xi \ll 1$ implies that the received signal y is also around zero with high probability, i.e., $f_Y(y)$ is significant only for $|y| \ll 1$. Thus, $\mu(y) \ll 1$, $\mathsf{Q}_1(\sqrt{\mu(y)}, \sqrt{\rho_{\max}}) \approx 0$ in (7a), $\exp(\cdot) \approx 1$, and $\mathsf{I}_0(\cdot) \approx 1$ in (7b), which leads to (11). Thus, we approximate (5) as

$$\begin{aligned} \widetilde{H}(Y) &= -\int_y \left(\mathsf{N}_0(y) + \mathsf{N}_1(y)\Pr(S=1)\right) \\ &\qquad \times \log_2\left(\mathsf{N}_0(y) + \mathsf{N}_1(y)\Pr(S=1)\right) dy \\ &\approx -\int_{y_0} \mathsf{N}_0(y_0)\log_2 \mathsf{N}_0(y_0) dy_0 \\ &\quad -\int_{y_2} \mathsf{N}_1(y_1)\Pr(S=1) \\ &\qquad \times \log_2 \mathsf{N}_1(y_1)\Pr(S=1) dy_1 \qquad \text{(12a)} \\ &= \log_2 \pi e\left(1+\gamma\xi\right) - e^{-\frac{1}{\xi}}\log_2 e^{-\frac{1}{\xi}} \\ &\quad + e^{-\frac{1}{\xi}}\log_2 \pi e\sigma_z^2 \qquad \text{(12b)} \end{aligned}$$

where the approximation in (12a) follows from the further observation that the domains of $\mathsf{N}_0(y_0)$ and $\mathsf{N}_1(y_1)$ are approximately disjoint as the gap of their mean values is much larger than their variances, i.e., $b_{\max} \gg \{gP_{\rm in} + \sigma_z^2, \sigma_z^2\}$. For example, see Fig. 3 where $\mathsf{N}_0(y_0)$ and $\mathsf{N}_1(y_1)$ are shown for $\phi = \{0, \pi\}$. We note that typically this holds if $\xi \ll 1$, when IBO is used. We thus call the resulting SE as $\mathsf{SE}^{\rm IBO}$ which is obtained by substituting (12b) to (9) as

$$\mathsf{SE}^{\rm IBO}(\xi) = \widetilde{H}(Y) - \log_2 \pi e\sigma_z^2. \tag{13}$$

The following theorems for the approximated SE, $\mathsf{SE}^{\rm IBO}(\xi)$, allow us to obtain insights on the structured properties of the actual SE, $\mathsf{SE}(\xi)$, at least for $\xi \ll 1$. The proofs are given in Appendix D.

Theorem 1: The approximated SE, $\mathsf{SE}^{\rm IBO}(\xi)$, is a concave function over $\max\left(0, -\frac{1}{\ln \pi\sigma_z^2}\right) < \xi \le \frac{1}{2}$.

Using *Theorem 1*, we obtain the SE-aware optimal power loading factor $\xi_{\mathsf{SE}}^\star$.

Theorem 2: The SE-aware optimal power loading factor $\xi_{\mathsf{SE}}^\star$ which maximizes $\mathsf{SE}^{\rm IBO}(\xi)$ is obtained by the solution of the following equality:

$$\frac{\gamma}{1+\gamma\xi} = e^{-\xi^{-1}}\xi^{-2}\left(-\xi^{-1} + 1 - \ln \pi e\sigma_z^2\right). \tag{14}$$

Proposition 3: A closed form approximation of $\xi_{\mathsf{SE}}^\star$ is given by

$$\xi_{\mathsf{SE}}^\star \approx \widetilde{\xi}_{\mathsf{SE}}^\star \triangleq \frac{-1}{\mathsf{W}\left(\frac{1}{\ln(\pi e\sigma_z^2)}\right)}, \tag{15}$$

where $\mathsf{W}(\cdot)$ denotes the Lambert W function[1] that satisfies $q = \mathsf{W}(q)e^{\mathsf{W}(q)}$ [18].

Interestingly, the approximated $\widetilde{\xi}_{\mathsf{SE}}^\star$ depends only on σ_z^2; intuitively, this is because we assume $\xi \ll 1$. This makes $\widetilde{\xi}_{\mathsf{SE}}^\star$ independent of other PA parameters. The typical values of IBO are between $8\,\mathrm{dB}$ and $12\,\mathrm{dB}$ for large (e.g., macro) and small (e.g., femto) cell base stations, respectively, which include an additional margin for fading channels [19], [20]. The numerical results in the subsequent subsection show that the analytical results with $\xi \ll 1$ are accurate for $\xi \le 0.3$, i.e., $\mathrm{IBO} \ge 5\,\mathrm{dB}$.

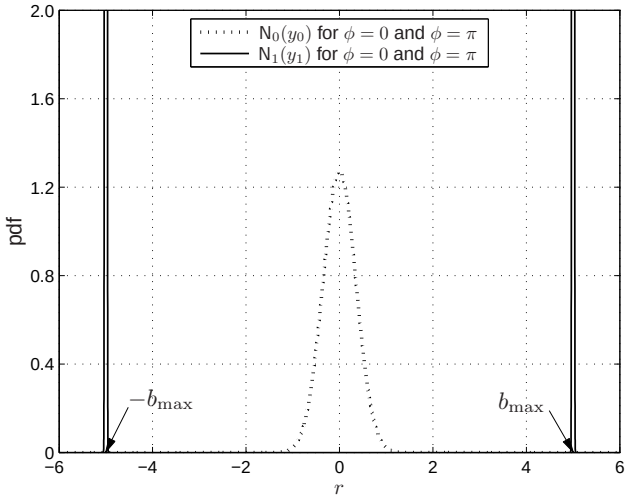

Fig. 3. The pdfs of $\mathsf{N}_0(y_0)$ and $\mathsf{N}_1(y_1)$ for $\phi = 0$ and $\phi = \pi$, where $\xi = 0.01$. The details for the simulation environment are given in Section IV-D.

D. Numerical Results on SE

To verify the analytical results on SE, we evaluate $\mathsf{SE}(\xi)$ with respect to the power loading factor ξ. The bandwidth is set to $10\,\mathrm{MHz}$. For simplicity, Rayleigh fading channel is assumed with zero mean and unit variance. A more realistic multipath channels as given in [21] may also be used for verifying the results obtained in Section IV-B. The channel attenuation is modeled as follows [22]: $G - 128 + 10\log_{10}(d^{-\alpha})\,\mathrm{dB}$ where G includes the transceiver feeder loss and antenna gains; and $d^{-\alpha}$ is the path loss where d is the distance in kilometers between a transmitter and a receiver and α is a path loss exponent. In simulations, we set $G = 5\,\mathrm{dB}$, $\alpha = 3.76$, $d = 200\,\mathrm{m}$, and $\sigma_z^2 = -174\,\mathrm{dBm/Hz}$ and use a PA SM2122-44L ($P_{\rm out}^{\max} = 44\,\mathrm{dBm} = 25\,\mathrm{W}$ and $g = 55\,\mathrm{dB}$) in Table II.

[1]For $q < 0$, $\mathrm{W}(q)$ can take multiple values. We assume $\mathsf{W}(\cdot) \le -1$ which is known as the lower branch of $\mathsf{W}(\cdot)$, so that $\widetilde{\xi}_{\mathsf{SE}}^\star \le 1$. This gives a unique value for $\mathsf{W}(\cdot)$.

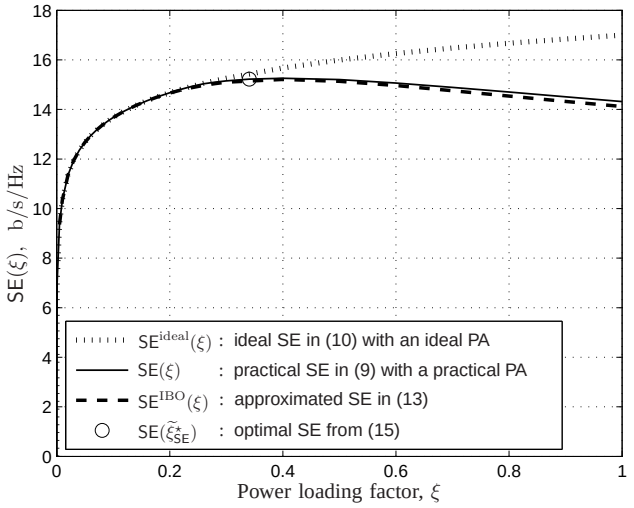

Fig. 4. Spectral efficiency evaluation with $P_{\rm out}^{\max} = 25\,\mathrm{W}$ and $g = 55\,\mathrm{dB}$.

Fig. 4 shows the numerical evaluation of SE. As expected, $\mathsf{SE}^{\rm ideal}(\xi)$ in (10) achieved by a perfectly linear PA is an increasing concave (log-shape) function, while the practical SE $\mathsf{SE}(\xi)$ in (9) is a concave function with a unique maximum when $\xi \leq \frac{1}{2}$. The approximated SE $\mathsf{SE}^{\rm IBO}$ in (12b) matches well with practical SE $\mathsf{SE}(\xi)$ if ξ is low. The optimal $\widetilde{\xi}^{\star}_{\mathsf{SE}}$ in (15) found from $\mathsf{SE}^{\rm IBO}(\xi)$ yields almost the highest SE $\mathsf{SE}(\widetilde{\xi}^{\star}_{\mathsf{SE}})$ as marked by '$\circ$' (*Theorem 2*). This illustrates the tightness of the approximation made to obtain $\mathsf{SE}^{\rm IBO}(\xi)$, at least for obtaining the optimal $\xi^{\star}_{\mathsf{SE}}$. On the other hand, the discrepancy between the practical SE $\mathsf{SE}(\xi)$ and the approximated SEs, $\mathsf{SE}^{\rm ideal}(\xi)$ and $\mathsf{SE}^{\rm IBO}(\xi)$, increases as ξ (i.e., the PA input or output power) increases.

V. Energy Efficiency

To derive the EE, we first model the power consumption at the transmitter. As shown in Fig. 5, power consumption and losses at the transmitter can occur in five modules: a direct current (DC) power supply (PS) module, a base band (BB) module, a radio frequency (RF) module, a PA module, and an active cooler and battery backup (CB) module. Power consumption at BB, RF, PA, and CB modules are denoted by $P_{\rm BB}$, $P_{\rm RF}$, $P_{\rm PA}$, and $P_{\rm CB}$, respectively, see details in [19], [23]–[25]. After introducing two known power consumption models, we will introduce a new model by taking the PA types (efficiency) and power loading factor ξ into consideration, and subsequently derive the corresponding EE.

A. Existing Power Consumption Models

One *empirical linear* model given in many recent studies, such as [11], [19], and [23], is

$$P_{\rm c}(\xi') = P_{\rm fix} + c\xi' P_{\rm out}^{\max}, \tag{16}$$

where ξ' is a frequency loading factor in OFDMA systems ($0 < \xi' \leq 1$); $P_{\rm fix}$ is a power consumption which is independent of the PA output signal power, i.e., $\xi' P_{\rm out}^{\max}$; and c is a scaling coefficient for the power loading dependency. If $\xi' = 0$, i.e., at the idle mode, $P_{\rm c}(\xi') = P_{\rm idle}$. In Table I, we summarize the parameters $P_{\rm fix}$, $P_{\rm out}^{\max}$, $P_{\rm idle}$, and c for various types of networks. The power coefficient in [23] is modeled as $c = \frac{1}{\eta} + \frac{P_{\rm BB}}{P_{\rm out}^{\max}} + \frac{P_{\rm RF}}{P_{\rm out}^{\max}}$. The parameters depend on the various practical factors, such as the transmitter configuration, the network structure, and the semiconductor technologies employed. For further information, refer to [20].

TABLE I
Power Model Parameters from [11]†, ([19], [20])‡, [23]§.

BS type	$P_{\rm out}^{\max}$ W	$P_{\rm fix}$ W	$P_{\rm idle}$ W	c
Macro	$20^{\ddagger,\S}$	$130^{\ddagger}$ $354.44^{\dagger}$, $405^{\S}$	$75^{\ddagger}$	$4.7^{\ddagger}$, $21.45^{\dagger}$, $17.8^{\S}$
RRH$^{\diamond}$	$20^{\ddagger}$	$84^{\ddagger}$	$56^{\ddagger}$	$2.8^{\ddagger}$
Micro	$2^{\S}$, $6.3^{\ddagger}$	$56^{\ddagger}$, $71.5^{\dagger}$, $106^{\S}$	$39^{\ddagger}$	$2.6^{\ddagger}$, $7.84^{\dagger}$, $108.3^{\S}$
Pico	$0.13^{\ddagger}$	$6.8^{\ddagger}$	$4.3^{\ddagger}$	$4^{\ddagger}$
Femto	$0.05^{\ddagger}$	$4.8^{\ddagger}$	$2.9^{\ddagger}$	$8^{\ddagger}$

$^{\diamond}$ remote radio head or remote radio unit (RRU)

Since the model in (16) is obtained from empirical measurements, it gives a reasonable indication of power consumption; however, no accurate indication is given for the specific PA type used. Furthermore, as shown in [19], there is a nonlinear relationship between the loading factor and the actual power consumption, especially in high power transmission, e.g., at the macro BS. To address these limitations, a PA-dependent model is given by [24], [25]

$$P_{\rm c} = (1 + C_{\rm PS})(1 + C_{\rm CB})(P_{\rm BB} + P_{\rm RF} + P_{\rm PA}) \tag{17}$$

where $C_{\rm PS}$ is a PS coefficient (typically $0.1 \leq C_{\rm PS} \leq 0.15$) and $C_{\rm CB}$ is an CB coefficient (typically less than 0.4). We can modify the model in (17) according to the PA types and the power loading factor ξ because the PA power consumption is modeled explicitly and separately from the other power consumption factors. Note that the frequency loading factor ξ' in (16) can be interpreted as the power loading factor ξ in the time domain.

B. Proposed PA-dependant Nonlinear Power Consumption Model

Though the PA power consumption $P_{\rm PA}$ depends on many factors including the specific hardware implementation, DC bias condition, load characteristics, operating frequency and PA output power, the component that consumes the majority of the power is given by the DC power fed to the PA [26]. Since the drain efficiency η depends on the PA types, we can express $P_{\rm PA}$ for different types of PA as a function of ξ [27]. For the ℓ-way Doherty PA, where ℓ is a fixed positive integer that depends on the implementation, the PA power consumption is expressed as

$$P_{\rm PA}(\xi) = \frac{4P_{\rm out}^{\max}}{\ell\pi} \times \begin{cases} \sqrt{\xi}, & 0 < \xi \leq \frac{1}{\ell^2} \\ (\ell+1)\sqrt{\xi} - 1, & \frac{1}{\ell^2} < \xi \leq 1. \end{cases} \tag{18}$$

Henceforth, we assume the use of the ℓ-way Doherty PA which has widespread use [28], [29]. The Doherty PA includes the special case of the class B PA with $\ell = 1$. The PA modeled in (18) can be considered to be an one-stage PA, which is relevant typically for low power transmission. We can obtain $P_{\rm PA}(\xi)$ similarly for other PA types, e.g., multi-stage PA combining class-A and Doherty, for high power transmission.

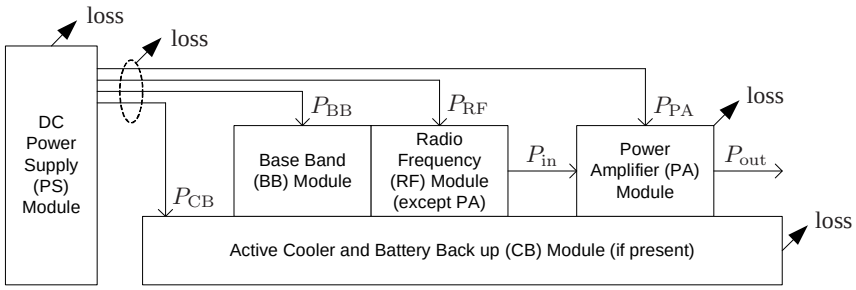

Fig. 5. Power consumption block diagram including DC power supply (PS), base band (BB), radio frequency (RF), power amplifier (PA), and active cooler and battery back up (CB) modules.

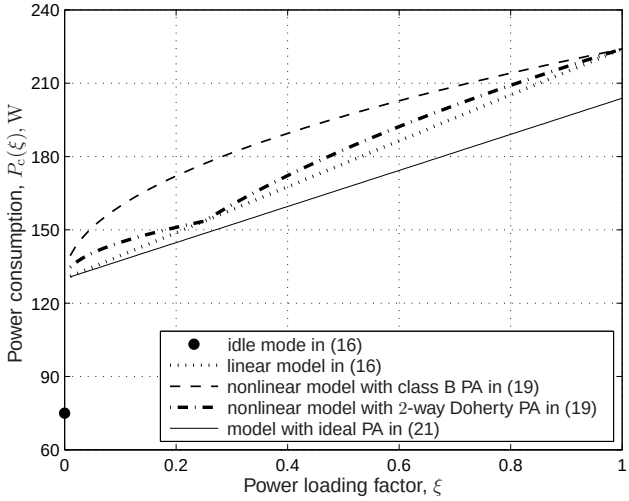

Fig. 6. Power consumption of microcell BS with $P_{\rm fix} = 130\,{\rm W}$ and $c = 4.7$.

It is straightforward to generalize to a multi-stage PA, in which the PA efficiency will change and (18) will be slightly modified accordingly with more levels. However, the EE analysis in the paper will remain without changes in the low power region.

Substituting (18) to (17), we get a *PA-dependant nonlinear* power consumption model as

$$P_{\rm c}(\xi) = P_0 + c_0\left(c_1 + c_2\sqrt{\xi}\right)P_{\rm out}^{\max} \tag{19}$$

for $0 < \xi \leq 1$, where $P_0 = (1+C_{\rm PS})(1+C_{\rm CB})(P_{\rm BB}+P_{\rm RF})$, and

$$(c_1, c_2) = \frac{4}{\ell\pi} \times \begin{cases} (0,1)\,, & 0 < \xi \leq \frac{1}{\ell^2}, \quad (20a) \\ (-1, \ell+1)\,, & \frac{1}{\ell^2} < \xi \leq 1. \quad (20b) \end{cases}$$

Comparing (19) with the model in (16), we also see that the new model in (19) reflects the PAs' characteristics. However, since $P_{\rm RF}$ is actually related to ξ, there are degrees of freedom to determine P_0 and c_0. In this work, we set $P_0 = P_{\rm fix}$ and $c_0 = \frac{\pi}{4}c$, so that (19) matches to (16) when $\xi = 1$. In other words, this alignment allows us to match the power consumption in (19) with that of (16) at the critical points of ξ, namely, $\xi = 0$ (idle), $\xi = \frac{1}{\ell^2}$, and $\xi = 1$ as shown in Fig. 6. In Fig. 6, we use a macrocell setup in Table I where $P_{\rm fix} = 130\,{\rm W}$ and $c = 4.7$. Following the same procedure of modeling in this subsection, any PA can be reflected in (19).

If a PA is ideal, namely, the PA is perfectly linear and efficient[2], then $P_{\rm out} = gP_{\rm in}$ and $P_{\rm PA} = P_{\rm out} - P_{\rm in}$, respectively. Thus, $P_{\rm PA} = (1-g^{-1})\xi P_{\rm out}^{\max}$. From (17), we can model the PA power consumption with the ideal PA as follows ($0 < \xi \leq 1$):

$$P_{\rm c}^{\rm ideal}(\xi) = P_{\rm fix} + c_0\left(1-g^{-1}\right)\xi P_{\rm out}^{\max}. \tag{21}$$

From Fig. 6, $P_{\rm c}^{\rm ideal}(\xi)$ gives a lower bound for the power consumption of the other models, as expected.

C. Analytical Results on EE

Using the practical SE in (9) and PA-dependent nonlinear power consumption $P_{\rm c}(\xi)$ in (19), we obtain the *practical* EE given by (1b) as

$$\mathsf{EE}(\xi) = \frac{\Omega \mathsf{SE}(\xi)}{P_{\rm c}(\xi)}. \tag{22}$$

An upper bound of $\mathsf{EE}(\xi)$ is obtained assuming an ideal PA with perfect linearity and efficiency as

$$\mathsf{EE}^{\rm ideal}(\xi) \triangleq \frac{\Omega \mathsf{SE}^{\rm ideal}(\xi)}{P_{\rm c}^{\rm ideal}(\xi)}. \tag{23}$$

However, the bound $\mathsf{EE}^{\rm ideal}(\xi)$ is not tight enough. Furthermore it does not reflect the PA types. Thus, we remove the perfect efficiency assumption from (23) and get a PA-dependant tighter bound as follows:

$$\mathsf{EE}^{\rm linear}(\xi) \triangleq \frac{\Omega \mathsf{SE}^{\rm ideal}(\xi)}{P_{\rm c}(\xi)} \tag{24}$$

where we retain the assumption of a perfectly linear PA. Using $\mathsf{EE}^{\rm linear}(\xi)$, we can obtain the following theorems, which allow us to obtain insights on the structured properties of the practical EE, $\mathsf{EE}(\xi)$, at least for $\xi \ll 1$. The proofs are given in Appendix D.

Theorem 4: $\mathsf{EE}^{\rm linear}(\xi)$ is a piecewise quasi-concave function over $\xi \geq \zeta \triangleq \left(v + \sqrt{1+v^2}\right)^2/\gamma^2$, where $v = P_{\rm out}^{\max}c_0c_2/(P_0 + P_{\rm out}^{\max}c_0c_1)$. Specifically, $\mathsf{EE}^{\rm linear}(\xi)$ is quasi-concave over $\zeta \leq \xi \leq 1/\ell^2$ and also over $1/\ell^2 < \xi \leq 1$.

We denote $\xi_{\mathsf{EE}}^{\star}$ as the optimal power loading factor that maximizes $\mathsf{EE}^{\rm linear}(\xi)$, which in general depends on the PA

[2]Power-added efficiency (PAE) and overall efficiency are defined as $\frac{P_{\rm out}-P_{\rm in}}{P_{\rm PA}}$ and $\frac{P_{\rm out}}{P_{\rm in}+P_{\rm PA}}$, respectively [26].

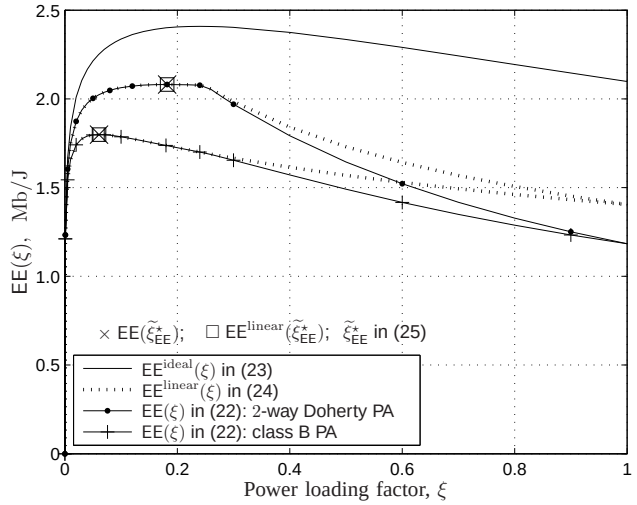

Fig. 7. Energy efficiency evaluation with $P_{\rm fix} = 130\,\mathrm{W}$ and $c = 4.7$.

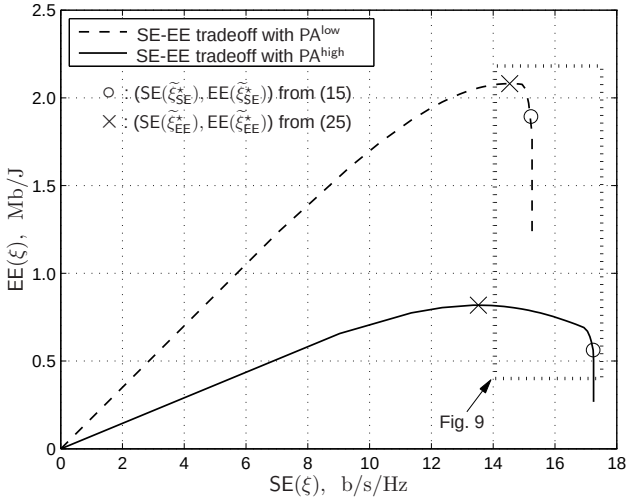

Fig. 8. SE-EE tradeoff with 2-way Doherty PAs. $\mathsf{PA}^{\mathsf{low}}$ is a low power PA with $P_{\rm out}^{\max} = 25\,\mathrm{W}$ and $g = 55\,\mathrm{dB}$, and $\mathsf{PA}^{\mathsf{high}}$ is a high power PA with $P_{\rm out}^{\max} = 100\,\mathrm{W}$ and $g = 50\,\mathrm{dB}$.

parameters. Typically, $\zeta \approx 0$ as $\gamma \triangleq P_{\rm out}^{\max}/\sigma_z^2$ is large, see e.g., the numerical results in Section IV-D. Assuming $\xi_{\mathsf{EE}}^\star \geq \zeta$, *Theorem 5* states the solution for $\xi_{\mathsf{EE}}^\star$.

Theorem 5: Assuming $\xi_{\mathsf{EE}}^\star \geq \zeta$, $\xi_{\mathsf{EE}}^\star$ equals either $[\xi_1^\star]_\zeta^{1/\ell^2}$ or $[\xi_2^\star]_{1/\ell^2,}^1$, where $\xi_1^\star$ and $\xi_2^\star$ are the solutions of $\frac{\partial \mathsf{EE}^{\rm linear}(\xi)}{\partial \xi} = 0$ in (D.1) with (c_1, c_2) defined as (20a) and (20b), respectively. Here, notation $[x]_a^b = a$ if $x < a$, $[x]_a^b = b$ if $x > b$, and $[x]_a^b = x$ otherwise.

Assuming $\xi_{\mathsf{EE}}^\star \geq \zeta$, *Theorem 5* shows that there are at most two candidates for $\xi_{\mathsf{EE}}^\star$. Hence, $\xi_{\mathsf{EE}}^\star$ can be obtained easily by checking which candidate maximizes $\mathsf{EE}^{\rm linear}(\xi)$. Moreover, *Proposition 6* shows that an approximation of $\xi_{\mathsf{EE}}^\star$ can be obtained in closed form. For the special case of class B PA (i.e., $\ell = 1$), it follows from *Theorem 5* that the optimal solution is given exactly by $\xi_{\mathsf{EE}}^\star = \xi_1^\star$, assuming $\xi_{\mathsf{EE}}^\star \geq \zeta$.

Proposition 6: A closed form approximation of $\xi_{\mathsf{EE}}^\star$, denoted by $\widetilde{\xi}_{\mathsf{EE}}^\star$, equals either $[\widetilde{\xi}_1^\star]_\zeta^{1/\ell^2}$ or $[\widetilde{\xi}_2^\star]_{1/\ell^2,}^1$, where $\widetilde{\xi}_i^\star$ ($i \in \{1,2\}$) approximates $\xi_i^\star$ and is given by

$$\xi_i^\star \approx \widetilde{\xi}_i^\star = \frac{1}{\gamma} \exp\left(2 + 2\mathsf{W}\left(\frac{\sqrt{\gamma}}{ev}\right)\right). \tag{25}$$

In (25), v is defined in Theorem 4, and $\mathsf{W}(\cdot) > 0$ as $\frac{\sqrt{\gamma}}{ev} > 0$, so that $\mathsf{W}(\cdot)$ is unique.

Numerical results in next subsection show the tightness of the EE bound and that $\widetilde{\xi}_{\mathsf{EE}}^\star$ is a near maximizer of the practical EE, $\mathsf{EE}(\xi)$.

D. Numerical Results on EE

To verify the analysis on EE, we evaluate the EE numerically. For power consumption parameters, the macrocell setup in Section V-B is employed. Other parameters are the same as environment given in Section IV-D.

Fig. 7 shows the EE for class B and 2-way Doherty PAs. Though the PA specifications, such as the maximum output power and gain, are identical, each of them has different efficiency resulting in different PA parameters in (20). From Fig. 7, we observe that the EE functions are concave (*Theorem 4*), and that the Doherty PA achieves the closest EE to the ideal PA's. The EEs, $\mathsf{EE}(\widetilde{\xi}_{\mathsf{EE}}^\star)$ and $\mathsf{EE}^{\rm linear}(\widetilde{\xi}_{\mathsf{EE}}^\star)$, are illustrated by '$\times$' and '$\square$,' respectively. As shown in *Theorem 5*, $\widetilde{\xi}_{\mathsf{EE}}^\star$ yields the maximum $\mathsf{EE}^{\rm linear}(\widetilde{\xi}_{\mathsf{EE}}^\star)$ and it is almost identical to $\mathsf{EE}(\widetilde{\xi}_{\mathsf{EE}}^\star)$ (which are overlapped in the figure). This is because the practical EE is maximized in the linear region, and the practical SE is also maximized in linear region as shown numerically in the previous section. From the results, we can surmise that the optimal $\widetilde{\xi}_{\mathsf{EE}}^\star$ in (25) is a good approximation of the maximizer of $\mathsf{EE}(\xi)$.

VI. SE-EE Tradeoff and PA Switching Strategy

To obtain the practical SE-EE tradeoff, Fig. 8 is regenerated from the results of SE in Fig. 4 and EE in Fig. 7. In addition to the SE and EE of PA SM2122-44L in subsection IV-D, we include the results obtained from a PA SM1720-50 ($P_{\rm out}^{\max} = 50\,\mathrm{dBm} = 100\,\mathrm{W}$ and $g = 50\,\mathrm{dB}$) in Table II. The former and the latter PAs are denoted by $\mathsf{PA}^{\mathsf{low}}$ and $\mathsf{PA}^{\mathsf{high}}$, respectively. We use a 2-way Doherty PA for all results.

In contrast to the SE-EE tradeoff for an ideal PA which is a decreasing convex function, the EE in the practical SE-EE tradeoff drops rapidly when the SE exceeds beyond a threshold that corresponds to the maximum EE. The closed-form analysis of the SE-EE tradeoff appears intractable. Instead, we focus on the analysis of the tradeoff based on the approximated SE and EE defined (13) and (24), respectively.

Proposition 7: The Pareto-optimality of the approximated SE-EE tradeoff is characterized as follows:

i) For $\min\{\widetilde{\xi}_{\rm EE}^\star, \widetilde{\xi}_{\rm SE}^\star\} \leq \xi \leq \max\{\widetilde{\xi}_{\rm EE}^\star, \widetilde{\xi}_{\rm SE}^\star\}$, the corresponding approximated SE-EE tradeoff is Pareto-optimal: to increase the approximated SE, the approximated EE must decrease, and vice versa.

ii) For $\xi < \min\{\widetilde{\xi}_{\rm EE}^\star, \widetilde{\xi}_{\rm SE}^\star\}$, both the approximated SE and EE increase as ξ increases.

iii) For $\xi > \max\{\widetilde{\xi}_{\rm EE}^\star, \widetilde{\xi}_{\rm SE}^\star\}$, both the approximated SE and EE decrease as ξ increases.

From *Proposition 7*, it is sufficient to consider only the region in i), because the remaining regions do not lead to the approximated Pareto-optimal SE-EE tradeoff. In Fig. 8, the approximated Pareto-optimal SE-EE tradeoff is narrow, which lies between the maximum SE and the maximum EE as indicated by '$\circ$' and '$\times$,' respectively. In cellular communications, however, a wide range of SE-EE tradeoff such as that illustrated by the dotted box in Fig. 8 is desired. This motivates us to use multiple PAs, where one or more PAs are switched on at any time. We call this technique *PA switching (PAS)*. Although PAS incurs a switch insertion loss of G_S and an overhead of switching time ϵ which decrease the SE and EE, we may obtain a better tradeoff of SE-EE from the degree of freedom of choosing different PAs.

For simplicity of description, we consider two PAs, PA-1 and PA-2; subsequent results are readily extended to multiple PAs. Let the SE and EE of PA-i, $i \in \{1, 2\}$, including the switch insertion loss G_S, be $\mathsf{SE}_i'(\xi)$ and $\mathsf{EE}_i'(\xi) = \frac{\Omega \mathsf{SE}_i'(\xi)}{P_c^i(\xi)}$, respectively, where $P_c^i(\xi)$ is the total power consumption with PA-i. In the following subsections, we apply the PAS technique to two systems, namely, frequency division duplex (FDD) and time division duplex (TDD) systems, and derive their SEs and EEs.

A. PA Switching for FDD Systems

Consider K FDD frames each with length of T. For PAS, we assume PA-1 is used for the first k frames, then PA-1 is switched to PA-2 which consumes ϵ seconds, and finally PA-2 is used for the remaining $K - k$ frames. Defining the time sharing factor as $\kappa \triangleq \frac{k}{K}$, $0 \le \kappa \le 1$, the achievable SE and EE from PAS can be derived as follows:

$$\begin{aligned} \mathsf{SE}_{\mathrm{s}}^{\mathrm{FDD}}(\xi, \kappa) &= \frac{kT\mathsf{SE}_1'(\xi) + (K-k)T\mathsf{SE}_2'(\xi)}{KT + \epsilon} \\ &= \frac{KT}{KT+\epsilon}\left(\kappa \mathsf{SE}_1'(\xi) + (1-\kappa)\mathsf{SE}_2'(\xi)\right) \end{aligned} \tag{26}$$

and

$$\begin{aligned} \mathsf{EE}_{\mathrm{s}}^{\mathrm{FDD}}(\xi, \kappa) &= \frac{KT\Omega \mathsf{SE}_{\mathrm{s}}^{\mathrm{FDD}}(\xi, \kappa)}{kTP_c^1(\xi) + (K-k)TP_c^2(\xi)} \\ &= \frac{KT}{KT+\epsilon}\left(\frac{\mathsf{EE}_1'(\xi)\mathsf{EE}_2'(\xi)\left(\kappa \mathsf{SE}_1'(\xi) + (1-\kappa)\mathsf{SE}_2'(\xi)\right)}{\kappa \mathsf{SE}_1'(\xi)\mathsf{EE}_2'(\xi) + (1-\kappa)\mathsf{SE}_2'(\xi)\mathsf{EE}_1'(\xi)}\right), \end{aligned} \tag{27}$$

where $\epsilon = 0$ if $\kappa = 0$ or if $\kappa = 1$ (i.e., no switching), and $\epsilon > 0$ otherwise. Here, we ignore the switch power consumption as it is relatively negligible compared to $P_c^i(\xi)$.

B. PA Switching for TDD Systems

In TDD systems, the downlink (DL) and uplink (UL) frames are transmitted alternately from BS to UE and from UE to BS. Here, we assume that ϵ is less than UL frame length which is typically true. For example, one LTE frame consumes a time period of $10\,\mathrm{ms}$ [22], while the switching time is much less than $1\,\mathrm{ms}$ (refer to the PA turn-on time in Table II which consumes most of the switching time). We therefore can switch the PAs between consecutive DL frames while receiving UL frame, without switching time overhead. The corresponding SE and EE can be readily obtained from (26) and (27) by setting ϵ to be zero. Note that the switching insertion loss is still incurred.

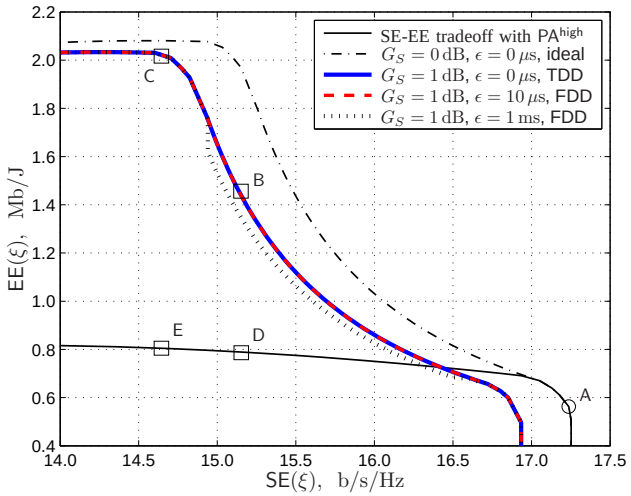

Fig. 9. SE-EE tradeoff with switching 2-way Doherty PAs, $\mathsf{PA}^{\mathsf{low}}$ and $\mathsf{PA}^{\mathsf{high}}$, when $\epsilon = 10\,\mu\mathrm{s}$, $L_S = 1\,\mathrm{dB}$, $T = 10\,\mathrm{ms}$, and $K = 20$.

C. Numerical Results and Discussion on PAS

The PAS is useful for adaptive systems where the traffic and channel conditions change dynamically. Fig. 9 shows the SE-EE tradeoff with PAS between $\mathsf{PA}^{\mathsf{low}}$ and $\mathsf{PA}^{\mathsf{high}}$. For comparison, we include the results of a single PA $\mathsf{PA}^{\mathsf{high}}$ and an ideal switching, namely, $G_S = 0\,\mathrm{dB}$ and $\epsilon = 0$. For practical switching, the switch insertion loss G_S is set to $1\,\mathrm{dB}$, and ϵ is set to $0\,\mu\mathrm{s}$ for TDD frame, while $10\,\mu\mathrm{s}$ and $1\,\mathrm{ms}$ are used for FDD frames. We consider $K = 20$ frames with $T = 10\,\mathrm{ms}$ for each frame length. From Fig. 9, we can verify that an SE-EE tradeoff is substantially improved by PAS. For example, let us consider the TDD system. The EE can be improved by around 210% (323%) if we reduce SE by 12% (15%) from A to B (C), respectively, as marked in Fig. 9. In contrast, if a single PA $\mathsf{PA}^{\mathsf{high}}$ is used instead, the EE is improved by only around 64% (68%) with the same reduction of SE from A to D (E). Next, consider the FDD system. Even with a switching time that is 10% of the frame size, i.e., $\epsilon = 1\,\mathrm{ms}$, a better SE-EE tradeoff is observed for most of the tradeoff region.

To implement the PAS in practice, the network overhead to obtain full channel state information at the transmitter can be significant, but it can be resolved by limiting the PA numbers with limited feedback information. Other issue is the increased form factor; however, this may not be significant issue in cellular networks where the BSs are already large in form factor due to other circuits. Furthermore, even with a small number of PAs, as we show in our recent work [30], significant performance gain can be achieved. In the near future, advancement of semiconductor technology will help further reduce the related concerns with form factor and hardware cost, making the proposed PAS an even more convincing technology for any type of transmitters.

TABLE II
POWER AMPLIFIER CHARACTERISTICS (ASCENDING ORDER OF $P_{\text{out}}^{\max}$).
(a)

PA#	Model	$P_{\text{out}}^{\max}$(dBm)	g(dB)	V_{PA} (Volt)	C_{PA} (mA)	$P_{\text{in}}^{\max}$(dBm)	Frequency (GHz)	turn-on time (μs)	Institution
1	MAX2242	5	28.5	3.3	50	10	2.4 – 2.5	1.5	MAXIM
2	FMPA2151	7	31	3.3	280	0	2.4 – 2.5	–	FAIRCHILD Semiconductor
3	FMPA2151	7	33	3.3	375	0	4.9 – 5.9	–	FAIRCHILD Semiconductor
4	PA1137	8	17	2	20	10	2 – 2.2	–	tyco Electronics
5	ADL5570	10	29	3.5	100	–	2.3 – 2.4	1.0	Analog Devices
6	MAX2242	13	28.5	3.3	90	10	2.4 – 2.5	1.5	MAXIM
7	MAX2840	15	22.8	3.3	155	–	5.15 – 5.35	1.5	MAXIM
8	BGA6289	15	13	4.1	88	–	1.95 – 2.5	–	NXP Semiconductors
9	AWT6134	16	23.5	3.4	130	10	1.75 – 1.78	–	ANADIGICS
10	AWT6138	16	15	3.4	57	10	1.85 – 1.91	–	ANADIGICS
11	AWT6252	16	24.5	3.4	54	10	1.92 – 1.98	–	ANADIGICS
12	AWT6252	16	20.5	3.4	54	10	1.92 – 1.98	–	ANADIGICS
13	RF2192	16	22	3.4	150	10	0.824 – 0.849	40	EF MICRO-DEVICES
14	RF2196	16	20	3.4	160	10	1.85 – 1.91	–	EF MICRO-DEVICES
15	RF3163	16	24	3.4	125	10	0.824 – 0.849	46	EF MICRO-DEVICES
16	RF3164	16	28	3.4	130	10	1.85 – 1.91	46	EF MICRO-DEVICES
17	RF3165	16	27	3.4	130	10	1.75 – 1.78	46	EF MICRO-DEVICES
18	RF6100-1	16	26	3.4	135	10	0.824 – 0.849	46	EF MICRO-DEVICES
19	AP172-317	17	33	3.3	140	20	1.8 – 2.5	–	SKYWORKS
20	SKY65006	17	30	3.3	110	10	2.4 – 2.5	–	SKYWORKS
21	RMPA5255	18	33	3.3	230	–	4.9 – 5.9	1.0	FAIRCHILD Semiconductor
22	MAX2841	18	22.5	3.3	260	–	5.15 – 5.35	1.5	MAXIM
23	FMPA2151	19	31	3.3	600	0	2.4 – 2.5	–	FAIRCHILD Semiconductor
24	FMPA2151	19	33	3.3	600	0	4.9 – 5.9	–	FAIRCHILD Semiconductor
25	RMPA2458	19	31.5	3.3	103	5	2.4 – 2.5	1.0	FAIRCHILD Semiconductor
26	RF5117	21	26	3	200	10	1.8 – 2.8	–	EF MICRO-DEVICES
27	RF5189	21	25	3	220	10	2.4 – 2.5	–	EF MICRO-DEVICES
28	SST13LP01	21	34	3.3	340	–	4.9 – 5.8	0.2	Silicon Storage Technology, Inc.
29	RF5117	22	26	3	500	10	1.8 – 2.8	–	EF MICRO-DEVICES
30	RMPA2455	22	30	5	195	10	2.4 – 2.5	1.0	FAIRCHILD Semiconductor
31	MAX2242	22	28.5	3.3	300	10	2.4 – 2.5	1.5	MAXIM
32	AP172-317	22.5	33	3.3	220	20	1.8 – 2.5	–	SKYWORKS
33	MAX2247	23	29.5	3	305	5	2.4 – 2.5	1.5	MAXIM
34	SST12LP00	23	27	3.3	115	–	2.4 – 2.5	–	Silicon Storage Technology, Inc.
35	SST12LP14	23	31	3.3	290	–	2.4 – 2.5	0.1	Silicon Storage Technology, Inc.
36	SST13LP01	23	28	3.3	260	–	2.4 – 2.485	0.1	Silicon Storage Technology, Inc.
37	AP178-321	23.5	19	3.3	186	20	1.8 – 2.5	–	SKYWORKS
38	MAX2247	24	29.5	3.3	307	5	2.4 – 2.5	1.5	MAXIM
39	AP172-317	24	33	3.3	240	20	1.8 – 2.5	–	SKYWORKS
40	CX65003	24.5	11.5	5	138	15	1.4 – 2.5	–	SKYWORKS
41	ADL5570	25	29	3.5	440	–	2.3 – 2.4	1.0	Analog Devices
42	ADL5571	25	29	3.3	450	–	2.5 – 2.7	1.0	Analog Devices
43	RF2163	25	19	3.3	378	15	1.8 – 2.5	–	EF MICRO-DEVICES
44	MAX2247	25	30.5	4.2	345	5	2.4 – 2.5	1.5	MAXIM
45	SST12LP14	25	31	3.3	340	–	2.4 – 2.5	0.1	Silicon Storage Technology, Inc.
46	NE552R479A	26	11	3	217	19	2.45 – 2.45	–	CEL California Eastern Laboratories
47	PA1153	26.4	28.5	15	250	15	1.8 – 2	–	tyco Electronics
48	PA1133	26.5	29	15	200	15	1.85 – 1.91	–	tyco Electronics
49	ADL5571	27	27.5	5	620	–	2.5 – 2.7	1.0	Analog Devices
50	RF2114	27	36	6.5	300	12	0.001 – 0.6	0.1	EF MICRO-DEVICES
51	RF2161	27	30	3	477	6	1.85 – 2	–	EF MICRO-DEVICES
52	RF2186	27	31	3	668	6	1.85 – 2	–	EF MICRO-DEVICES
53	RF5117	27	26	5	500	10	1.8 – 2.8	–	EF MICRO-DEVICES
54	RF5176	27	26	3	476	6	1.85 – 2	–	EF MICRO-DEVICES
55	AWT6252	27.5	26.5	3.4	423	10	1.92 – 1.98	–	ANADIGICS
56	AWT6134	28	26	3.4	475	10	1.75 – 1.78	–	ANADIGICS
57	AWT6134	28	25	3.4	462	10	1.75 – 1.78	–	ANADIGICS
58	AWT6138	28	26	3.4	487	10	1.85 – 1.91	–	ANADIGICS
59	RF3163	28	28.5	3.4	455	10	0.824 – 0.849	46	EF MICRO-DEVICES
60	RF3164	28	28	3.4	460	10	1.85 – 1.91	46	EF MICRO-DEVICES
61	RF3165	28	28	3.4	460	10	1.75 – 1.78	46	EF MICRO-DEVICES
62	RF6100-1	28	29	3.4	465	10	0.824 – 0.849	46	EF MICRO-DEVICES
63	RF2132	28.5	29	4.8	327	12	0.824 – 0. 849	0.1	EF MICRO-DEVICES
64	RF2146	28.5	18.5	4.8	393	12	1.5 – 2	0.55	EF MICRO-DEVICES
65	RF6100-4	28.5	28	3.4	535	10	1.85 – 1.91	46	EF MICRO-DEVICES
66	CX65105	28.5	25	5	470	7	1.7 – 2.2	–	SKYWORKS
67	SKY65162-70LF	28.8	20	5	306	–	0.869 – 0.96	–	SKYWORKS
68	RF2192	29	30	3	715	10	0.824 – 0.849	40	EF MICRO-DEVICES
69	RF2196	29	27	3	755	10	1.85 – 1.91	–	EF MICRO-DEVICES
70	MGA-43228	29.2	38.5	5	500	–	2.3 – 2.5	–	Avago Technologies

VII. CONCLUSION

In this paper, we provided a theoretical analysis of the spectral efficiency and energy efficiency (SE-EE) tradeoff of OFDM systems by taking into account the practical non-ideal effects of the power amplifiers (PAs). Optimal power loading factors of PA are derived to achieve the maximum SE and EE. We identified the problem of a narrow SE-EE tradeoff region due to the nonlinearity and inefficiency of the practical PAs, and proposed a PA switching that is a useful technique to achieve a wide SE-EE tradeoff. Future studies include the SE-EE analysis of multiuser communication systems, MIMO systems, and a more accurate PA model with a memory effect and nonlinearity at low power regime.

APPENDIX A

See Table II.

APPENDIX B
DERIVATION FOR (7)

Joint pdf $f_Y(y, S=0)$: Given $S=0$, i.e., $A \le a_{\max}$, we have $W = \mathsf{L}_{\text{PA}}(A)e^{j\theta} = \sqrt{g}Ae^{j\theta}$. From assumption $A1$, the conditional pdf of W is a truncated complex Gaussian as

$$f_W(w|S=0) = \begin{cases} \frac{1}{\Pr(S=0)} \frac{1}{\pi g P_{\text{in}}} \exp\left(-\frac{|w|^2}{gP_{\text{in}}}\right), & |w| < b_{\max} \\ 0, & |w| \ge b_{\max}. \end{cases} \quad \text{(B.1)}$$

The pdf of the AWGN Z is

$$f_Z(z|S=0) = f_Z(z|S=1) = f_Z(z) = \frac{1}{\pi\sigma_z^2} \exp\left(-\frac{|z|^2}{\sigma_z^2}\right). \quad \text{(B.2)}$$

From (3), we can express the joint pdf as follows:

$$\begin{aligned} f_Y(y, S=0) &= \Pr(S=0)\left(f_W(w|S=0) * f_Z(z|S=0)\right) \\ &= \Pr(S=0)\int_{\tau\in\mathbb{C}} f_W(\tau|S=0) f_Z(y-\tau)\, d\tau \\ &= \Pr(S=0)\int_0^{\infty}\int_0^{2\pi} f_W\left(re^{j\phi}|S=0\right) \\ &\qquad \times f_Z\left(y - re^{j\phi}\right)|J(\phi, r)|\, d\phi dr \end{aligned} \quad \text{(B.3)}$$

where '$*$' is the linear convolution operator and $|J(\phi, r)| = r$ is the Jacobian [17]. Noting $(f_W|S=0) = 0$ if $|w| \ge b_{\max}$,

(b) *Continued*

PA#	Model	$P_{\text{out}}^{\max}$(dBm)	g(dB)	V_{PA} (Volt)	C_{PA} (mA)	$P_{\text{in}}^{\max}$(dBm)	Frequency (GHz)	turn-on time (μs)	Institution
71	MGA-43328	29.3	37.3	5	470	–	2.5 – 2.7	–	Avago Technologies
72	AWT6104M5	30	30	3.5	714	10	1.85 – 1.91	–	ANADIGICS
73	HMC457QS16G/E	30.5	25	5	500	15	2.01 – 2.17	–	Hittite Microwave corporation
74	HMC453QS16G/E	33	8	6.5	725	–	2.01 – 2.17	–	Hittite Microwave corporation
75	SM0825-33/33H	33	20	12	1100	4	0.8 – 2.5	–	Stealth Microwave
76	SM1025-36DMQ2	33	10	12	800	–	1 – 2.5	–	Stealth Microwave
77	SM1025-37MQ2	33	10	12	1600	–	1 – 2.5	–	Stealth Microwave
78	PA1110	33	10	10	725	28	1.8 – 2	–	tyco Electronics
79	PA1132	33	22	12	725	15	1.8 – 2	–	tyco Electronics
80	SM1727-34HS	34	33	12	1200	1	1.7 – 2.7	–	Stealth Microwave
81	SM1727-34HSQ	34	36.5	12	1200	1	1.7 – 2.7	–	Stealth Microwave
82	SM2023-34HS	34	33	12	1200	1	2 – 2.3	–	Stealth Microwave
83	PA1157	36	24.5	10	1350	15	2 – 2.2	–	tyco Electronics
84	PA1159	36.2	23.5	10	1700	28	2.3 – 2.4	–	tyco Electronics
85	PA1162	36.2	30	10	1450	11	0.8 – 0.96	–	tyco Electronics
86	SM04060-37HS	37	36	12	1800	1	0.4 – 0. 6	–	Stealth Microwave
87	SM04093-36HS	37	34	12	1600	1	0.4 – 0. 925	–	Stealth Microwave
88	SM5659-37S	37	20	12	2300	20	5.6 – 5.9	1.0	Stealth Microwave
89	SM5759-37HS	37	39	12	2300	2	5.7 – 5.9	1.0	Stealth Microwave
90	PA1182	37.5	23	28	1000	15	2.3 – 2.4	–	tyco Electronics
91	PA1223	37.5	25	28	1000	15	2.11 – 2.17	–	tyco Electronics
92	PA1224	37.5	25	28	1000	15	2 – 2.2	–	tyco Electronics
93	PA1186	38	29	28	1000	15	0.8 – 0.96	–	tyco Electronics
94	XD010-42S-D4F/Y	39	30	28	930	20	0.869 – 0.894	–	Sirenza Micro Devices
95	SM0822-39	39	45	12	3500	-4	0.8 – 2.2	1.0	Stealth Microwave
96	SM0825-40Q	40	39	12	5500	1	0.8 – 2.5	–	Stealth Microwave
97	SM2023-41	41	55	12	4500	-13	2 – 2.3	–	Stealth Microwave
98	SM2027-41LS	41	51	12	6000	-7	2 – 2.7	–	Stealth Microwave
99	SM4450-41L	41	55	12	5000	-13	4.4 – 5	1.0	Stealth Microwave
100	SM1822-42LS	42	52	12	5500	-8	1.8 – 2.2	–	Stealth Microwave
101	SM3338-43	43	50	12	8500	-6	3.3 – 3.8	1.0	Stealth Microwave
102	SM5053-43L	43	55	12	9200	-7	5 – 5.3	1.0	Stealth Microwave
103	SM7785-43A	43	48	12	9500	–	7.725 – 8.5	–	Stealth Microwave
104	SM1923-44L	44	55	12	8200	-8	1.9 – 2.3	–	Stealth Microwave
105	SM2025-44L	44	55	12	8500	-10	2 – 2.5	–	Stealth Microwave
106	**SM2122-44L**	**44**	**55**	**12**	**8200**	**-9**	**2.1 – 2.2**		**Stealth Microwave**
107	SM2325-44	44	55	12	8000	-10	2.3 – 2.5	–	Stealth Microwave
108	SM2025-46L	46.3	52	12	15000	-7	2 – 2.5	–	Stealth Microwave
109	SM04548-47L	47	55	12	14000	-8	0.45 – 0.48	–	Stealth Microwave
110	SM2023-47L	47	55	12	15000	-7	2 – 2.3	–	Stealth Microwave
111	SM3134-47L	47	55	12	15000	-6	3.1 – 3.4	–	Stealth Microwave
112	SM3436-47L	47	56	12	15000	-6	3.4 – 3.6	–	Stealth Microwave
113	**SM1720-50**	**50**	**50**	**12**	**27000**	**2**	**1.7 – 2**	**–**	**Stealth Microwave**
113	SM2325-50L	50	59	12	31000	-9	2.3 – 2.5	–	Stealth Microwave
115	SM1819-52LD	52	45	30	11000	–	1.8 – 1.9	–	Stealth Microwave

and using (B.1) and (B.2) to (B.3), we further derive

$$\begin{aligned}
&f_Y(y, S=0)\\
&= \frac{1}{\pi^2 g P_{\text{in}} \sigma_z^2} \int_0^{b_{\max}} \int_0^{2\pi} r \exp\left(-\frac{r^2}{gP_{\text{in}}} - \frac{|y - re^{j\phi}|^2}{\sigma_z^2}\right) d\phi dr\\
&= \frac{1}{\pi^2 g P_{\text{in}} \sigma_z^2} \int_0^{b_{\max}} r \exp\left(-\frac{r^2}{gP_{\text{in}}} - \frac{r^2 + |y|^2}{\sigma_z^2}\right)\\
&\qquad \times \int_0^{2\pi} \exp\left(\frac{2r|y|\cos(\theta - \phi)}{\sigma_z^2}\right) d\phi dr\\
&= \frac{2}{\pi g P_{\text{in}} \sigma_z^2} \int_0^{b_{\max}} r \exp\left(-\frac{r^2}{gP_{\text{in}}} - \frac{r^2 + |y|^2}{\sigma_z^2}\right) \mathsf{I}_0\left(\frac{2r|y|}{\sigma_z^2}\right) dr\\
&= \frac{1}{\pi\left(gP_{\text{in}} + \sigma_z^2\right)} \exp\left(-\frac{|y|^2}{gP_{\text{in}} + \sigma_z^2}\right)\\
&\qquad \times \int_0^{\rho_{\max}} \left[\frac{1}{2} \exp\left(-\frac{\mu(y) + \rho}{2}\right) \mathsf{I}_0\left(\sqrt{\mu(y)\rho}\right)\right] d\rho
\end{aligned}$$

where θ denotes the angle of y and $\mathsf{I}_0(\cdot)$ is a modified Bessel function of first kind. The last equality is obtained after some mathematical manipulations, for which we observe the function outside the integral is a pdf of a normal distribution with a zero mean and variance $(gP_{\text{in}} + \sigma_z^2)$ and the function inside the integral is a pdf of a noncentral chi-squared random variable with one degree of freedom. Since the cdf of a noncentral chi-squared random variable is obtained as $1 - \mathsf{Q}_1(\sqrt{\mu(y)}, \sqrt{\rho_{\max}})$, we can readily arrive at (7a).

Joint pdf $f_Y(y, S=1)$: Given $S=1$, i.e., $A > a_{\max}$, we have $W = \mathsf{L}_{\text{PA}}(A)e^{j\theta} = \sqrt{g}a_{\max}e^{j\theta} = b_{\max}e^{j\theta}$. Thus, the amplitude is a constant while the phase θ is uniformly distributed. Therefore, we can express the conditional pdf of W given $S=1$ as

$$f_W(w|S=1) = c\delta(|w| - b_{\max}) \tag{B.4}$$

where $\delta(\cdot)$ is the Dirac delta function and $c = (2\pi b_{\max})^{-1}$ is a constant obtained from the normalization $\int_{w\in\mathbb{C}} f_W(w|S=1)\, dw = 1$. Then the joint pdf is similarly to give

$$\begin{aligned}
&f_Y(y, S=1)\\
&= \Pr(S=1) \int_0^{\infty} \int_0^{2\pi} r f_W\left(re^{j\phi}\middle|S=1\right)\\
&\qquad \times f_Z\left(y - re^{j\phi}\right) d\phi dr.
\end{aligned} \tag{B.5}$$

Substituting (B.4) and $f_Z(z|S=1) = f_Z(z)$ in (B.2) into (B.5), and integration over r, we get

$$\begin{aligned}
&f_Y(y, S=1)\\
&= \frac{\Pr(S=1)}{2\pi^2\sigma_z^2} \int_0^{2\pi} \exp\left(-\frac{\left|y - b_{\max}e^{j\phi}\right|^2}{\sigma_z^2}\right) d\phi\\
&= \frac{\Pr(S=1)}{2\pi^2\sigma_z^2} \exp\left(-\frac{|y|^2 + b_{\max}^2}{\sigma_z^2}\right) 2\pi \mathsf{I}_0\left(\frac{2b_{\max}|y|^2}{\sigma_z^2}\right)\\
&= \frac{1}{\pi\sigma_z^2} \exp\left(-\frac{|y - b_{\max}|^2}{\sigma_z^2}\right)\\
&\qquad \times \left[\Pr(S=1) \exp\left(\frac{-2b_{\max}y_{\text{Re}}}{\sigma_z^2}\right) \mathsf{I}_0\left(\frac{2b_{\max}|y|}{\sigma_z^2}\right)\right]
\end{aligned}$$

where the function outside $[\cdot]$ is a pdf of a normal distribution $\mathcal{CN}(b_{\max}, \sigma_z^2)$. Thus, we get (7b).

APPENDIX C
PROOF OF (8)

Henceforth, we take the indices in the subscript to be modulo N, e.g., $X_{N+i} = X_{(N+i) \bmod N} = X_i$. We express

the mutual information as

$$\begin{aligned} I(\widetilde{\boldsymbol{X}};\widetilde{\boldsymbol{Y}}) =& I(\boldsymbol{X};\boldsymbol{Y}) \\ &\overset{(a)}{=} \sum_{t=0}^{N-1} I(X_t; Y_0,\cdots,Y_{N-1}|X_0,\cdots,X_{t-1}) \\ &\overset{(b)}{\geq} \sum_{t=0}^{N-1} I(X_t; Y_0,\cdots,Y_{t+L-1}|X_0,\cdots,X_{t-1}) \end{aligned} \tag{C.1}$$

where (a) follows from the chain rule of mutual information and (b) follows from the data processing inequality (by discarding the received signals $Y_{t+L},\cdots,Y_N$). The inequality in (b) is typically tight from numerical experiments, because X_t is only present in $Y_t,\cdots,Y_{t+L-1}$; therefore, intuitively the discarded signals do not directly contribute to the information on X_t (although they contribute to the information on the interfering terms in $Y_t,\cdots,Y_{t+L-1}$). The summand in (C.1) for $t \geq L$ can be lower bounded as follows:

$$\begin{aligned} &I(X_t; Y_0,\cdots,Y_{t+L-1}|X_0,\cdots,X_{t-1}) \\ &\quad\overset{(a)}{=} I(X_t; Y_t,\cdots,Y_{t+L-1}|X_0,\cdots,X_{t-1}) \\ &\quad\overset{(b)}{=} I(X_t; Y'_t,\cdots,Y'_{t+L-1}|X_0,\cdots,X_{t-1}) \\ &\quad\overset{(c)}{=} I(X_t; Y'_t,\cdots,Y'_{t+L-1}) \\ &\quad\overset{(d)}{\geq} I(X_t; Y''_t) \triangleq I_t^{\mathrm{LB}} \end{aligned} \tag{C.2}$$

where (a) follows from the independence of X_t and $\{Y_1,\cdots,Y_{t-1}\}$ given $\{X_0,\cdots,X_{t-1}\}$; (b) follows from the definition $Y'_{t+i} = Y_t - \sum_{j=i+1}^{L-1} h_j X_{t+i-j}$ for $i = 0,\cdots,L-1$; (c) follows from the fact that $Y'_t,\cdots,Y'_{t+L-1}$ consist only of the signal terms $X_t,\cdots,X_{N-1}$ and noise; and (d) follows from the data processing inequality where Y''_t is the maximum ration combining (MRC) of $Y'_t,\cdots,Y'_{t+L-1}$. For additive independent interference with variance σ^2, the mutual information is lower bounded by the channel where the noise is treated as AWGN with the same variance σ^2 [16]. Hence, a lower bound of $I(X_t;Y''_t)$, denoted by I_t^{LB}, is given by the mutual information of the following channel:

$$Y'''_t \quad = \quad \mathsf{L}_{\mathrm{PA}}(A_t)e^{\theta_t} + Z'_t \tag{C.3}$$

where $Z'_t \sim \mathcal{CN}(0, \sigma_z^2/|h'_t|^2)$ and h'_t is derived by

$$h'_t = \sqrt{\frac{gP_{\mathrm{in}}|h_0|^2}{\sigma_z^2} + \sum_{i=1}^{L-1} \frac{gP_{\mathrm{in}}|h_i|^2}{\sigma_z^2 + gP_{\mathrm{in}}\sum_{j=0}^{i-1}|h_j|^2}}. \tag{C.4}$$

The channel in (C.4) is the equivalent channel after MRC of the current and $(L-1)$ future received signals, where we take yet-to-be decoded, transmitted signals as interferences. The channel in (C.3) is the flat fading channel (3) with equivalent noise variance given by the original noise variance divided by the equivalent channel gain. Hence, we can obtain I_t^{LB} from (4) directly. In summary, from (C.1) and (C.2), the mutual information is lower bounded by (8).

Appendix D
Proof of Theorems and Propositions

Proof of Theorem 1: After substituting (12b) into (13), we can derive the second derivative of $\mathsf{SE}^{\mathrm{IBO}}(\xi)$ with respect to ξ as follows:

$$\begin{aligned} \frac{\partial^2 \mathsf{SE}^{\mathrm{IBO}}(\xi)}{\partial^2 \xi} = \frac{-1}{\ln 2}\Bigg(&\frac{\gamma^2}{(1+\gamma\xi)^2} + e^{-\xi^{-1}}\xi^{-4} \\ &+ e^{-\xi^{-1}}\xi^{-3}\left(\xi^{-1}-2\right)\left(\ln\frac{1}{\pi\sigma_z^2} - \frac{1}{\xi}\right)\Bigg) \end{aligned}$$

$\mathsf{SE}^{\mathrm{IBO}}(\xi)$ is concave over ξ if $\max\left(0, \frac{-1}{\ln(\pi\sigma_z^2)}\right) \leq \xi \leq \frac{1}{2}$ as the second derivative is negative. ■

Proof of Theorem 2: Since $\mathsf{SE}^{\mathrm{IBO}}(\xi)$ is concave assuming $\max\left(0, \frac{-1}{\ln(\pi\sigma_z^2)}\right) \leq \xi \leq \frac{1}{2}$ from *Theorem 1*, we can find the maximizer of $\mathsf{SE}^{\mathrm{IBO}}(\xi)$ by solving $\frac{\partial \mathsf{SE}^{\mathrm{IBO}}(\xi)}{\partial \xi} = 0$, which gives (14). ■

Proof of Proposition 3: Since $\gamma\xi \gg 1$ with practical value of $\gamma \triangleq P_{\mathrm{out}}^{\max}/\sigma_z^2$ and typical value of ξ, we can approximate $1+\gamma\xi \approx \gamma\xi$ in (14), and get $\xi e^{\frac{1}{\xi}} = -\frac{1}{\xi} + 1 - \ln\left(\pi e \sigma_z^2\right) \approx -\ln\left(\pi e \sigma_z^2\right)$ where we discard relatively small terms to obtain the closed form solution $\widetilde{\xi}_{\mathsf{SE}}^{\star}$ in (15). ■

Proof of Theorem 4: First, assume (c_1, c_2) is fixed over all ξ (the dependence according to (20) will be considered shortly). The first derivative of EE with respect to ξ in (24) is given by

$$\begin{aligned} \frac{\partial \mathsf{EE}^{\mathrm{linear}}(\xi)}{\partial \xi} &= \frac{\Omega}{2\sqrt{\xi}\left(v_1 + v_2\sqrt{\xi}\right)^2} \\ &\times \left(\frac{2}{\ln 2}\frac{\gamma}{1+\gamma\xi}\left(v_1\sqrt{\xi} + v_2\xi\right) - v_2\log_2(1+\gamma\xi)\right) \end{aligned} \tag{D.1}$$

where $v_1 = P_0 + P_{\mathrm{out}}^{\max} c_0 c_1$, and $v_2 = P_{\mathrm{out}}^{\max} c_0 c_2$. Clearly $v_2\log_2(1+\gamma\xi)$ is increasing in ξ. It can be shown that $\frac{\gamma}{1+\gamma\xi}\left(v_1\sqrt{\xi} + v_2\xi\right)$ is decreasing in ξ if $\xi \geq \zeta$, where ζ is defined in Theorem 4. Assuming $\xi \geq \zeta$, $\mathsf{EE}^{\mathrm{linear}}(\xi)$ then has at most one turning point and thus $\mathsf{EE}^{\mathrm{linear}}(\xi)$ is a quasi-concave function; it cannot be a quasi-convex function because $\mathsf{EE}^{\mathrm{linear}}(\xi)$ is decreasing in ξ for sufficiently large ξ. Now note that (c_1, c_2) is in fact constant for $0 \leq \xi \leq \frac{1}{\ell^2}$, and also constant for $\frac{1}{\ell^2} < \xi \leq 1$. Thus, $\mathsf{EE}^{\mathrm{linear}}(\xi)$ is a quasi-concave function for $\zeta \leq \xi \leq \frac{1}{\ell^2}$, and also for $\frac{1}{\ell^2} < \xi \leq 1$. ■

Proof of Theorem 5: First, consider the optimal solution that maximizes $\mathsf{EE}^{\mathrm{linear}}(\xi)$ over $\zeta \leq \xi \leq \frac{1}{\ell^2}$. From *Theorem 4*, $\mathsf{EE}^{\mathrm{linear}}(\xi)$ is a quasi-concave function. Thus the optimal solution is given by the turning point $\xi_1^{\star}$. In other words, the solution of $\frac{\partial \mathsf{EE}^{\mathrm{linear}}(\xi)}{\partial \xi} = 0$ with (c_1, c_2) defined as (20a), if $\xi_1^{\star}$ lies between ζ and $\frac{1}{\ell^2}$. Since $\xi_1^{\star}$ is a turning point, the optimal solution must be ζ if $\xi_1^{\star} < \zeta$, while the optimal solution must be $\frac{1}{\ell^2}$ if $\xi_1^{\star} > \frac{1}{\ell^2}$. More concisely, the optimal solution is $[\xi_1^{\star}]_{\zeta}^{1/\ell^2}$. Similarly, the optimal solution that maximizes $\mathsf{EE}^{\mathrm{linear}}(\xi)$ over $\frac{1}{\ell^2} \leq \xi \leq 1$ can be shown to be $[\xi_2^{\star}]_{1/\ell^2}^{1}$. ■

Proof of Proposition 6: The same approximation used in the proof of *Theorem 2*, i.e., $1+\gamma\xi \approx \gamma\xi$, gives an equality $\frac{2}{\ln 2}\frac{1}{\xi}\left(v_1\sqrt{\xi} + v_2\xi\right) = v_2\log_2(\gamma\xi)$ to make (D.1) be a zero. After some mathematical manipulations of the equality, we can find the unique solution given by (25). ■

Proof of Proposition 7: Suppose $\widetilde{\xi}_{\mathrm{EE}}^{\star} \leq \widetilde{\xi}_{\mathrm{SE}}^{\star}$. From *Theorems* 1 and 4 and *Propositions* 3 and 6, the approximated SE and EE are concave functions over ξ, and $\widetilde{\xi}_{\mathrm{SE}}^{\star}$ and $\widetilde{\xi}_{\mathrm{EE}}^{\star}$

are their maximizers, respectively. Thus, we can show the following. i) For $\xi \geq \widetilde{\xi}^{\star}_{\mathrm{EE}}$, $\mathsf{EE}(\xi)$ decreases as ξ increases. For $\xi \leq \widetilde{\xi}^{\star}_{\mathrm{SE}}$, $\mathsf{SE}(\xi)$ increases as ξ increases. Thus, for $\widetilde{\xi}^{\star}_{\mathrm{EE}} \leq \xi \leq \widetilde{\xi}^{\star}_{\mathrm{SE}}$, $\mathsf{EE}(\xi)$ decreases, while $\mathsf{SE}(\xi)$ increases as ξ increases. ii) For $\xi < \widetilde{\xi}^{\star}_{\mathrm{EE}}$, both $\mathsf{EE}(\xi)$ and $\mathsf{SE}(\xi)$ increase as ξ increase. iii) For $\xi > \widetilde{\xi}^{\star}_{\mathrm{SE}}$, both $\mathsf{EE}(\xi)$ and $\mathsf{SE}(\xi)$ decrease as ξ increases. The proof is completed by showing the similar analysis for the case when $\widetilde{\xi}^{\star}_{\mathrm{EE}} < \widetilde{\xi}^{\star}_{\mathrm{SE}}$. ■

REFERENCES

[1] J. Joung, C. K. Ho, and S. Sun, "Tradeoff of spectral and energy efficiencies: Impact of power amplifier on OFDM systems," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Anaheim, CA, USA, Dec. 2012.

[2] ——, "Green wireless communications: A power amplifier perspective," in *Proc. Int. Conf. Smart Grids, Green Commun. and IT Energy-aware Technol.*, Anaheim, CA, USA, Dec. 2012.

[3] W. Vereecken, W. V. Heddeghem, M. Deruyck, B. Puype, B. Lannoo, W. Joseph, D. Colle, L. Martens, and P. Demeester, "Power consumption in telecommunication networks: overview and reduction strategies," *IEEE Commun. Mag.*, vol. 49, pp. 62–69, Jun. 2011.

[4] J. Baliga, R. Ayre, K. Hinton, and R. S. Tucker, "Energy consumption in wired and wireless access networks," *IEEE Commun. Mag.*, vol. 49, pp. 70–77, Jun. 2011.

[5] H. Bogucka and A. Conti, "Degree of freedom for energy savings in practical adaptive wireless systems," *IEEE Commun. Mag.*, vol. 49, pp. 38–45, Jun. 2011.

[6] X. Li and J. L. J. Cimini, "Effects of clipping and filtering on the performance of OFDM," *IEEE Commun. Lett.*, vol. 2, pp. 131–133, May 1998.

[7] A. R. S. Bahai, M. Singh, A. J. Goldsmith, and B. R. Saltzberg, "A new approach for evaluating clipping distortion in multicarrier systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1038–1046, Jun. 2002.

[8] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPS/OFDM signal for a digital sound broadcasting system," in *Proc. Second European Conf. on Satellite Commun.*, Lihge, Belgium, Oct. 1991, pp. 179–184.

[9] J. Tellado, L. M. C. Hoo, and J. M. Cioffi, "Maximum-likelihood detection of nonlinearly distorted multicarrier symbols by iterative decoding," *IEEE Trans. Commun.*, vol. 51, pp. 218–228, Feb. 2003.

[10] A. Shirvani, D. K. Su, and B. A. Wooley, "A CMOS RF power amplifier with parallel amplification for efficient power control," *IEEE J. Solid-State Circuits*, vol. 27, pp. 684–693, Jun. 2002.

[11] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Anchorage, AK, USA, Sep. 2009.

[12] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Miami, FL, USA, Dec. 2010.

[13] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink OFDMA," *IEEE Trans. Commun.*, vol. 60, pp. 112–120, Jan. 2012.

[14] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[15] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, pp. 30–37, Jun. 2011.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New Jersey: John Wiley & Sons, 2006.

[17] A. Papoulis and S. Pillai, *Probability, random variables and stochastic processes*, 4th ed. McGraw-Hill, 2002.

[18] F. Chapeau-Blondeau and A. Monir, "Numerical evaluation of the Lambert W function and application to generation of generalized gaussian noise with exponent 1/2," vol. 50, pp. 2160–2165, Sep. 2002.

[19] G. Auer, V. Giannini, C. Desset, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, pp. 40–49, Oct. 2011.

[20] Energy aware radio and network technology (EARTH) project. [Online]. Available: https://www.ict-earth.eu

[21] C.-X. Wang, M. Pätzold, and Q. Yao, "Stochastic modeling and simulation of frequency-correlated wideband fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1050–1063, May 2007.

[22] "LTE; E-UTRA; RF requirements for LTE pico node B," ETSI, Tech. Rep. 136 931 V9.0.0, 2011. [Online]. Available: http://www.etsi.org/deliver/

[23] M. Deruyck, W. Joseph, and L. Martens, "Power consumption model for macrocell and microcell base stations," *Trans. Emerging Telecommunications Technologies*, Aug. 2012, doi:10.1002/ett.2565.

[24] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *IEEE Future Netw. and Mobile Summit*, Florence, Italy, Jun. 2010.

[25] R. V. R. Kumar and J. Gurugubelli, "How green the LTE technology can be?" in *Proc. IEEE Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE)*, Chennai, India, Feb./Mar. 2011.

[26] F. H. Raab, P. Asbeck, S. C. Cripps, P. B. Kenington, Z. B. Popović, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microw. Theory Tech.*, vol. 50, pp. 814–826, Mar. 2002.

[27] F. H. Raab, "Efficiency of Doherty RF power-amplifier systems," *IEEE Trans. Broadcast.*, vol. BC-33, pp. 77–83, Sep. 1987.

[28] B. Kim, I. Kim, and J. Moon, "Advanced Doherty architecture," *IEEE Microw. Mag.*, vol. 11, pp. 72–86, Aug. 2010.

[29] H. S. Bennett, J. J. Pekarik, and M. Huang, "Radio frequency and analog/mixed-signal technologies for wiireless communications," *National Institute of Standards and Technology (NIST)*, vol. 36, Jan. 2011.

[30] J. Joung, C. K. Ho, and S. Sun, "Power amplifier switching (PAS) for energy efficient systems," *IEEE Wireless Commun. Lett.*, 2013, to be published.