Energy Efficiency Maxima for Wireless Communications: 5G, IoT, and Massive MIMO

Earl McCune,  CTO, Eridan Communications

Abstract – Maximizing the operating energy efficiency of any wireless communication link requires a global optimization not only across the entire system block diagram, but also including the signal modulation and aspects of the link operating protocol. Achieving this global optimization is first examined for the transmitter, receiver, and baseband circuitry. Then the important aspects of signal modulation necessary to access these circuit optimizations, with examples from existing systems along with proposed signals for “5G” including Internet of Things (IoT) and massive MIMO, are presented. This is followed by the correspondingly important protocol aspects needed for link efficiency. Operating temperature consequences of signal modulation selection, and its corresponding PA efficiency ceiling, are explored for the envisioned massive-MIMO antenna arrays. A realization of this global energy efficiency optimization in actual circuitry with measurements completes this paper.

I. INTRODUCTION

Any power amplifier (PA) is a peak-power limited circuit. It must be designed to provide whatever the signal maximum peak envelope power (PEP) may ever be. In contrast, communication coverage and range are based on signal root-mean-square (rms) power, usually called the average signal power. To the extent that the PEP exceeds this average power there exists a peak-power to average-power ratio (PAPR) that is greater than 1. The average output power reduction from signal PAPR therefore reduces available communication range from a particular sized PA. Large PAPR values cause several important economic and realization problems, including: a) the communication range available from a particular PA is proportionally reduced, b) the PA needed to provide a required communication range must be proportionally larger, c) the maximum output power from the PA is not available to the channel for communication range and coverage, and so on. Signal PAPR is therefore economically expensive, and high values must be accepted only when a value to the communication system justifies its cost [1].

Development goals for the present “Fifth Generation” (5G) communications network are presented in Fig. 1. Ten specific development topics are shown. Among these ten topics, relative importance to three application areas are illustrated by the three contours intersecting the ten topics. Of particular interest here is the energy efficiency topic.

The general block diagram of a wireless communication transceiver is shown in Fig. 2. Around this block diagram are shown the top level power flows associated with its operation. At this level there is effectively no signal input power (\(P_{IN}\)), but it remains shown in Fig. 2 for completeness. Energy efficiency is high when the output power (\(P_{OUT}\)) is close to the DC power (\(P_{DC}\)). Whatever difference there is between \(P_{OUT}\) and \(P_{DC}\) is dissipated as heat (\(P_{DISS}\)). Achieving the objective of high energy efficiency is therefore equivalent to minimizing \(P_{DISS}\).

The operation of conventional linear amplifiers is summarized in Fig. 3a, where the traditional transistor characteristic curves and load line are shown. New here are the addition of constant power dissipation curves on top of the amplifier design information. The linear (class-A) bias point along the load line, mid-way between the cut-off and compression boundaries, is very near to the highest value power dissipation curve. A linear amplifier therefore maximizes power dissipation, which is opposite of what is needed for high energy efficiency. Any effort to improve amplifier efficiency is therefore equivalently an exercise in tolerating circuit nonlinearity, which includes class-AB and class-B operation shown in Fig. 3b. This challenge is more than a century old, and is fundamental.
II. CIRCUITRY OPTIMIZATION FOR EFFICIENCY

Optimizing the circuitry for energy efficiency starts from individual examinations of the three important blocks in Fig. 2: transmitter (TX), receiver (RX), and baseband (BB). Of these, the transmitter is usually the most important. We begin with the TX optimization for energy efficiency.

There are two major methods available to improve PA energy efficiency, which is identical to having the PA transistor operate at, or along, power dissipation curves of low value. One option is presented in Fig. 5, where the value of the PA supply voltage is reduced when the required output signal power is low. This shifts the load line while the PA still operates linearly, seen in Fig. 5a using full power and reduced power scales of a LTE uplink signal. The corresponding reduction in $P_{\text{Diss}}$ is substantial, as seen in Fig. 5b. This is the average power tracking (APT) technique, a first step toward the envelope tracking technique [3].

A second technique is to operate the PA only at the endpoints of its load line. This intersects the minimum possible value power dissipation curves, which maximizes the available PA energy efficiency. But it also eliminates all PA linearity. This is called a switch-mode power amplifier (SMPA) with operating characteristics shown in Fig. 6a. The output signal has a larger range because there

---

Fig. 3. Conventional linear amplifier characteristics: a) transistor characteristic curves with a load line including transistor knee profile and power dissipation curves; b) power dissipation along the load line in the transistor and the load resistance.

Additionally, any transistor operates as the intended controlled current source only when the voltage at its controlled port exceeds a minimum value called the knee voltage, also known as the device compliance voltage. The presence of this knee voltage reduces the available output signal magnitude from the amplifier, and its corresponding power available to the amplifier load resistance also falls. This further reduces the available amplifier efficiency. The impact of the knee voltage to amplifier available output power and energy efficiency is presented in Fig. 4.

---

Fig. 5. Improving energy efficiency with variable supply voltage: a) intersecting lower value power dissipation curves with a lower supply voltage for smaller output signals (LTE uplink is shown); b) corresponding power dissipation profiles at these two supply voltages.

---

Fig. 6. Switch-mode power amplifier (SMPA) with operating characteristics.
is no restriction to operating the transistor as a current source. This means that more output power is available from the same transistor operating as a SMPA than as a linear PA. To successfully achieve SMPA operation and its associated efficiency it is necessary that the transitions between transistor ON and OFF states be fast compared to the signal frequency cycle time. This is related to the transistor transition frequency \( f_T \) specification as shown in Fig. 6b [4]. It is best for the transistor \( f_T \) to exceed the operating frequency by at least 20x.

![Fig. 6. Switching power amplifier operation and requirements: a) output signal swing increase from the maximum available from linear operation; b) achieving switching operation requires the transistor transition frequency \( f_T \) to exceed the operating frequency by 20x or more](image)

Between these two extremes of amplifier transistor operation, there are many options with different trade-offs between circuit linearity (with minimum efficiency) and switching operation (having no circuit linearity). Two of these middle cases are presented in Fig. 7 along with the aforementioned extreme cases. The baseline linear amplifier is shown in Fig. 7a with a simplified version of that in Fig. 3a. The use of load impedance modulation with a fixed supply voltage is shown in Fig. 7b, where the improvement in efficiency is evident from the load lines intersecting contours of lower power dissipation as load impedance increases. Envelope tracking (ET) is shown in Fig. 7c, where the PA remains linear at all times and the action of the dynamic supply variation intersects contours of lower power dissipation as the signal envelope falls. Direct polar (DP) modulation [5], requiring SMPA operation, is presented in Fig. 7d. The difference between ET and DP operation is very significant.

![Fig 7. PA strategies for improved energy efficiency: a) standard linear; b) Load modulation; c) envelope tracking; d) direct polar modulation](image)

Comparing the transmitter efficiency enhancement techniques of ET and DP to that of conventional linear amplifiers in Fig. 8 shows the superiority of switching operation when energy efficiency must be maximized.

![Fig. 8. Available efficiency increases using envelope tracking (linear) and polar (switching) power amplifiers and dynamic power supplies.](image)

The receiver block has essentially very small input power, which is negligible with respect to its power dissipation, and no output power, as seen in Fig. 9. It
inherently therefore has zero efficiency, and the objective becomes minimizing its power dissipation.

![Fig. 9. Receiver power flows show that RX energy efficiency is effectively zero.](image)

Minimizing receiver power dissipation has two parts, which are 1) keeping the amount of circuitry to the least possible, and 2) having that circuitry not be linear if possible. Unfortunately the RX is not as free to use switching circuitry as the TX is, because any time there is circuit nonlinearity the Fourier Transform requires there to be intermodulation among multiple signals present in the circuit. The TX has only one signal (unless that signal is OFDM, LTE, or a carrier aggregation (CA)). The RX will be minimum power if most of its gain can be realized with a limiter (nonlinear high gain amplifier). This requires both that the signal channelization be complete ahead of the limiter so that only one signal is present, and also that the signal modulation be tolerant of the switching behavior of the limiter.

Power flows in the BB are shown in Fig. 10, and like the receiver the effective energy efficiency is zero. Also as seen for the RX, this means that optimizing the BB for overall system efficiency means design to minimize its power dissipation.

![Fig. 10. Baseband power flows show that BB energy efficiency is effectively zero.](image)

Basebands are universally implemented in CMOS technology, which means that the CMOS power dissipation model is directly applicable here. CMOS power is traditionally minimized by operating at the lowest possible voltage and clock frequency, and also by minimizing the number of active gates. This latter point means that overall energy efficiency favors a single ‘dimension’ modulation where the BB only needs to provide and receive a single modulation signal. Thus modulation types such as OOK, ASK, FSK and PSK in their ‘pure’ form [6] are preferable to any quadrature modulation, which inherently requires two BB modulation signals both into the TX and from the RX along with the circuitry to implement them. The presence of static power in modern nanometer CMOS causes additional energy efficiency complications [7].

III. SIGNAL MODULATION REQUIREMENTS

Signal modulation type appears as a part of the circuit optimizations for each of the circuitry blocks. This is a very important part of this global optimization and is included here. In Fig. 11 the linear amplifier of Fig. 3a is shown with projections parallel to the load line of signal envelope value probabilities for four popular signal modulation types: constant envelope (CE), the QAM used for 3G, OFDM as used for WLAN, and LTE. The height of each of these envelope probability density functions (pdf) represents how much time the corresponding signal has that value. Each pdf is normalized to its signal envelope peak value, which is then scaled to fully span the linear range of this amplifier. Thus, each of these cases represents maximum output power from this PA design.

![Fig. 11. Signal (peak normalized) envelope voltage probability density functions along a linear amplifier load line with corresponding transistor power dissipation curves](image)

Both of the multicarrier (MC) signals (OFDM and LTE) spend the majority of their time near zero magnitude, where the transistor power dissipation is maximum. The CE signals spend their time at the load line limits which are at lower transistor power dissipation. The 3G signal is between these extremes, but is closer to the CE profile than the OFDM profile. Signals that provide for maximum overall communication energy efficiency therefore have following characteristics: low PAPR (high rms power), and no time near zero envelope magnitude.

Using 64QAM as an example, shown in Fig. 12a, meeting the first objective means that the arrangement of the constellation points should be circular instead of
spreading the outer points of the square around the circle actually spreads them further apart and improves tolerance of noise by 0.9 dB. The remaining inside points can be rearranged with equal or greater spacing into two more inner rings as seen in Fig 12b, forming one of the possible 64-state polar ring (PR) constellations [8]. The rms value of the ring arrangement for these 64 constellation points is 2.2 dB greater than for the square, for the same signal peak value. This becomes “free” power; more power into the channel from the same size PA. The overall link benefit is $2.2 + 0.9 = 3.1$ dB, just from rearranging the constellation points.

![Image of 64-QAM and PR constellations](image)

Fig. 12. Example 64 state signal constellations: a) conventional 64-QAM constellation and measured Nyquist filtered vector diagram; b) rearranged into three rings for one of the 64-PR constellations and its corresponding measured QAM- and polar-Nyquist filtered vector diagram.

Meeting the objective of eliminating envelope zero crossings is achieved here by changing the signal filtering domain from quadrature to polar. The result is seen in Fig. 12b, where the polar-domain Nyquist filtered PR signal (64-PRPF) has no envelope value less than 40% of the peak. This signal can be readily implemented with a switching PA at nearly full efficiency.

The impact on PA energy efficiency of particular signal modulation designs is predictable using the technique of modulation-available energy efficiency (MAEE). MAEE analysis [9] also accounts for the knee voltage effect [10] of Fig. 4. MC modulation types must have cross-modulation products generated in the presence of any circuit nonlinearity, and so these signals are generally not tolerant of much signal distortion. This means that it is necessary to amplify these signals with a very linear, class-A amplifier. PA technologies that have small knee voltage values are needed to keep these PA efficiencies up as high as possible, as seen in Fig. 4.

Results of MAEE analysis for 11 different popular signal modulation types are presented in Fig. 13. When a SMPA can be used, the MAEE is very high. When the PA must be linear, then the efficiency limit of Fig. 4 applies, shown by the ‘+’ characters. Available efficiency drops below this limit as the signal PAPR increases. For MC signals such as OFDM and the generalized frequency-division modulation (GFDM) proposed for 5G, the MAEE is near to, or even below, 10%. Thus, the result reported in [11] of 7.5% PA efficiency for OFDM-based WLAN is actually a very good result. But, definitely not what we desire for energy efficient communications.

![Image of MAEE chart](image)

Fig. 13. MAEE chart showing the progressively lower PA efficiency ceiling due to signals needing PA linearity and having progressively higher PAPR values. When such signals are also MC, the ability to avoid class-A is further reduced.

The link effectiveness improvement discussed above relating 64-QAM and 64-PR can be generalized by comparing the rms value of any signal constellation (TX power into the channel) with its minimum Euclidean distance (minimum constellation point spacing, corresponding to noise tolerance at the RX). When these parameters are used as coordinates in a graph, the result is provided in Fig. 14. It is apparent that a constellation figure of merit (FoM) presents itself. Specifically, the product of these two signal design properties shows that when this product is maximized, the link benefit is also maximized. In this manner it is possible to trade off minimum Euclidean distance and constellation point arrangement for the most effective communication at a particular distance.

For one example, we observe that from a link perspective 16-PSK is better than 16-QAM in that the rms power available from the TX with the constant modulus constellation is more of a benefit than the larger Euclidean distance of 16_QAM by 0.9 dB. And as seen before, 64-PR improves on 64-QAM by having both a larger TX rms power and a larger Euclidean distance. This link benefit
exceeds 3 dB – a huge improvement from just rearranging the constellation points. When PAPR extension resulting from Nyquist filtering is accounted for, the actual link benefit between 64-QAM and 64-PR increases to 4 dB. This is a very significant benefit to achieving energy efficient communications. This link benefit immediately translates to needing less TX power for the same communications quality of service.

In order to take full advantage of the electronic communication capabilities provided us by physics, it is necessary to review the propagating wave equation derived from Maxwell’s Equations. The governing differential equation for the propagating electromagnetic wave (focusing on the electric field E) is

$$\nabla^2 E - \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = 0$$  \hspace{1cm} (1)

Taking the plane wave case propagating along the z direction for simplicity, (1) reduces to

$$\frac{\partial^2 B_z}{\partial z^2} - \mu \epsilon \frac{\partial^2 E_z}{\partial t^2} = 0$$  \hspace{1cm} (2)

which has as its solution

$$E_z(z,t) = \rho e^{i \omega t - \frac{\omega}{c} z}$$  \hspace{1cm} (3)

We observe that at any specific place in space, such as at \(z=0\) for convenience, this solution is a phasor which means that physics is fundamentally polar in nature. The three signal parameters that directly follow from (3) are magnitude (not amplitude) (\(\rho\)), frequency (\(\omega = \text{d}\theta/\text{d}t\)), and phase-shift. Mathematics in these parameters, specifically in phase and its derivatives, is inherently nonlinear which is not convenient for the application of linear algebra for signal processing. Our insistence on using quadrature signal representations so that we can use linear algebra actually loses signal capability that physics makes available (the quadrature–polar transformation is NOT unique, due to restrictions on the arctangent to make it an invertible function). To take full advantage of what physics makes available, we must make peace with polar-based signal processing. This is a significant challenge, which it is well worth taking on and mastering.

So-called bandwidth expansion of polar signals compared to quadrature signals is actually an artifact of the modulation design and not fundamental to the polar technique. Returning to the signal solution from Maxwell’s equations, we note that at an envelope zero crossing two things happen: the envelope waveform \(\rho(t)\) has an inflection point, and the signal phase in \(\theta(t)\) at that point is undefined. The Fourier transform holds for \(\rho(t)\) and \(\theta(t)\) as well as for \(I(t)\) and \(Q(t)\). Bandwidth necessary to resolve these envelope zero-crossings does increase due to the corresponding discontinuities in \(\rho(t)\) derivatives and the \(\theta(t)\) direct discontinuity. But if the signal modulation is designed to never have envelope zero crossing events, then these waveform discontinuities disappear from the polar signal coordinates. Much of the bandwidth expansion disappears too.

![Fig. 14. Comparing signal constellations by (TX) constellation rms power vs. constellation point spacing (RX constellation point Euclidean distance).](image1)

IV. PROTOCOL REQUIREMENTS

Once the hardware power dissipation is minimized when it operates, as allowed by an appropriate signal modulation type selection, then the conventional technique of duty cycling the communication system operation has its best effect. When the communication is active, its power dissipation is minimum to get the needed job done. When the system is OFF, power dissipation should be zero. This duty cycling is important, but it is the last step in this process. Not the first.

While power consumption does scale with duty cycle, implementation cost does not. Whatever cost that must be incurred in designing to meet a particular standard does not change if the standard compliant circuitry operates 100% of the time or at 0.01% of the time. Linear circuitry is generally more expensive to design and build, largely because of the sensitivities to environment (temperature and manufacturing variation) and the need for unconditional stability. While it is not universally true, in general the cost of nonlinear circuitry is lower than that for

![Fig. 15. Envelope and phase waveforms at a signal envelope zero crossing: envelope signal is continuous with discontinuous derivatives, and the phase waveform has a direct discontinuity.](image2)
linear circuitry, which holds for design time, manufacturing time and materials, and in operating expense (reduced by the increased efficiency).

Present practice is to use well-known signals with equally well-known very low energy efficiency, such as LTE [12] to implement systems requiring low energy consumption such as IoT. Such is the present path that 3GPP is on for its standardization of 5G. We can, and should, do much better.

V. MASSIVE MIMO REQUIREMENTS

The primary intent of the massive MIMO proposal is to use an array of antennas where the antenna count is well above 100, and to use these antennas in combinations to direct transmitted energy only to specific mobile devices. In order to have full generality to generate separate beams to each individual user, each antenna must have its own independent transmitter with independently settable RF magnitude, frequency, and phase shift. The consequence is that each antenna has its own local transceiver, including its own PA. Considering only the heat from PA power dissipation, the electronics behind the antenna array will get increasingly hot toward the center of the array as shown in Fig. 17.

With a densely packed array like is envisioned for massive MIMO, the only place to effectively locate a heatsink is around the outside of the array electronics. Thus, the outer layer of electronics has at least one surface touching the heat sink (each corner element has two). The next inner layer must dissipate its heat through the heat dissipated by the outer layer, raising the temperature of the second layer. Layers of antenna elements and associated electronics that are further toward the array center have progressively more “electric blankets” keeping them warm. One result of this is shown in Fig. 17. Here the array has 128 elements distributed in an 8x16 array. The signal is assumed to be conventional LTE, which by [12] is known to have a PA efficiency around 9%. Working through the power dissipation and corresponding temperature rise from conventional thermal resistances we conclude that the internal array elements will operate at a temperature above 300°C, shown in Fig. 17a. This is not practical. The power dissipation must be reduced.

In order for the center elements to operate at a temperature below 100°C, analysis shows that the PA efficiency must get above 50% [14]. From Fig. 4 this means that the PA cannot operate as a linear amplifier. This then precludes the use of any MC signal modulation type. Unless, hopefully, that the technology to implement MC signals using SMPA techniques is refined to generate these signals with sufficient fidelity that standards requirements are fully met.

The primary goal then is to find a way (or ways) for polar modulation to handle envelope zero crossings so that it is completely compatible with the presently standardized signal modulations which contain envelope zero-crossing events (as seen in Fig. 11, the OFDM signal emphasizes zero-crossing events because its envelope pdf is maximum at zero). There is considerable research going on in this specific area, and promising results are being seen.

VI. CIRCUIT REALIZATION

Building and operating high efficiency communication systems is where we realize their value. Fig. 7 shows that to achieve maximum energy efficiency a switching PA is required, which must be placed within a polar architecture to allow envelope variations and modulation generality [15]. This architecture is shown in Fig. 16a, where the SMPA is powered through a dynamic power supply (DPS) which sets the output signal magnitude. A measurement of its envelope modulation operation is shown in Fig. 16b. This standard AM signal is modulated to a depth of 99.9% with very good modulation quality. Envelope control dynamic range exceeds 60 dB. For this measurement, the SMPA is manufactured using the gallium-nitride (GaN) MMIC from [15]. Similar precision is achieved using GaAs E-pHEMT based SMPA technology [16]. Polar techniques are also successful in silicon technologies as summarized in [17].
Fig. 16 Maximum energy efficiency requires a switching PA operated in a polar architecture: a) block diagram; b) measured dynamic-supply modulation to 100% AM

VII. SUMMARY

For the “Internet of Things”, widespread wireless sensor networks, and any energy harvested application it is essential that energy efficiency be maximized to take full benefit of each electron available to the wireless application. This requires that both energy efficient operation when operating be maximized to the extent that physics allows, along with maximum use of duty cycling to keep operating time to a minimum. All of these problems are addressable using the efficiency maximizing techniques described here.

This analysis shows that the present communications standards are not written in favor of achieving the available energy efficiency that circuitry can provide. All present standards though are set and will not change. As the opportunity arises for new standards to address new applications, adopting analysis techniques outlined here will enable fundamental improvements in energy efficiency and economic viability.

Key to overall success is adopting signal modulations that allow the hardware to operate at its natively available energy efficiency. This analysis fortunately provides a recipe to do this. This recipe includes the steps:

- Select a signal modulation with no envelope zero crossings
- PAPR to less than 6 dB
- Signal order at 64 or less (for low costs)
- Single carrier operation

The IEEE Standards Association has recently established the Working Group (WG) IEEE 1923.1 to standardize this recipe on what is necessary to fundamentally achieve maximum energy efficiency communications, and ultimately to provide a tool set for all other standards WGs, and to the community at large, to use in the development of future communications standards.

It is possible to simultaneously achieve bandwidth efficiency of 6 bps/Hz and exceed 60% energy efficiency in a single signal. The keys to this result require paying close attention to the foundations of electromagnetic communication, as they are provided, by physics. This process yields signals and circuit structures that are not widely used today, even though their existence is well-known. This situation shows that a lot of communication system performance improvement over presently implemented systems remains available to achieve. We need to change our development processes, and master the mathematics of polar-domain signal processing.

References