Decade Bandwidth Agile GaN Power Amplifier Exceeding 50% Efficiency

E. McCune, Senior Member IEEE; D. Babić, Senior Member IEEE; R. Booth;
Douglas Kirkpatrick, Member IEEE
Eridan Communications, Santa Clara, California, 95050, USA

Abstract — Minimizing size, weight, and power consumption (SWaP) in any communication system requires that the transmitter power amplifier operate at peak possible efficiency at all times and at all frequencies. Here a GaN watt-level transmitter is presented that operates across 0.2 to 3 GHz, with no matching network reconfiguration, and broadband efficiency exceeding 50%. Besides minimizing power consumption, this design provides the required power simultaneously with having the smallest possible transistor size, power supply, and heat sink. Frequency agility is across the entire PA bandwidth, and no circulator is required to operate in phased-array applications. Signal modulation compatibility is also discussed.

Index Terms — amplifier; compression; efficiency; gallium nitride; power amplifiers

I. INTRODUCTION

When meeting design requirements for maximum energy efficiency in transmitters, it is known that circuit linearity must not only be compromised, it must actually be suppressed [1]. In these maximally nonlinear operation modes [2] the usual measures of circuit gain cannot remain in their common indiscriminate use, but rather must be evaluated separately for their unique information [3]. By intentionally performing maximal-efficiency circuit design, additional requirements are identified that additionally provide wideband operation. Such amplifiers are also inherently compatible with wideband phased-array systems because of their suppressed linearity. All of these traits directly benefit desired reductions in size, weight, and power (SWaP) for field, airborne, and spaceborne communications and radar systems.

Need for wideband transmitter operation is driven by several factors, including for example the need for low probability of intercept (LPI), and compatibility among multiple system types in a particular deployment. Agility, defined as the ability to operate at any particular frequency at any particular time, should not be limited by any reconfiguration of the transmitter circuitry across frequency bands.

This paper is organized as follows. Following this introduction is an evaluation of the physical circuit principles necessary for wideband amplifier efficiency at any power. Section III presents additional design constraints necessary to simultaneously achieve wideband frequency agility. Section IV reports on the tolerance of the resulting wideband high efficiency and frequency agile transmitter to reverse signals such as those encountered in phased-array operation. Section V examines the compatibility of signal modulations with this high efficiency wideband transmitter. Implications for improved satisfaction of SWaP needs are addressed in Section VI. Conclusions are drawn in Section VII.

II. WIDEBAND PA EFFICIENCY

Efficiency is often evaluated as a conservation-of-power relationship, which is illustrated in Fig. 1a. Physics though does not demand that power be conserved, but rather that energy be conserved. The ability to use a conservation-of-power approach described by

\[ P_{\text{IN}} \left( \frac{J}{\text{sec}} \right) + P_{\text{DC}} \left( \frac{J}{\text{sec}} \right) + P_{\text{OUT}} \left( \frac{J}{\text{sec}} \right) + P_{D,\text{PA}} \left( \frac{J}{\text{sec}} \right) = 0 \]  

is legitimate only when the unit time part of each term in (1) has equal interpretation: at average values. Then the time factors normalize out and the relationship among the power numbers is identical to the required conservation of energy.

Supply power and dissipated power normalized to unit output power across the possible values of amplifier efficiency is provided in Fig. 1b. When the amplifier efficiency exceeds 50% the power absorbed by the heatsink is less than that delivered to the antenna. But when the amplifier efficiency is less than 33%, the power absorbed by the heatsink is more than double that same power being provided to the antenna. All of the additional dissipated power must be provided to the amplifier by the power supply, which the amplifier converts directly to heat. At these lower efficiency values, both the heatsink and the power supply must become larger than the minimum required to provide the intended RF output power.

Having validated the constraints by which a conservation of power calculation can be used for evaluating amplifier efficiency, this efficiency is derived from (1) to be

\[ \frac{P_{\text{OUT}}}{P_{\text{DC}}} = 1 - \frac{P_{D} - P_{\text{IN}}}{P_{\text{DC}}} \approx 1 - \frac{P_{D}}{P_{\text{DC}}} \]  

where each power term is defined in Fig. 1a, and the final approximation is valid for high amplifier gain; \( P_{\text{OUT}} \gg P_{\text{IN}} \).

The important observation from (2) is that the consideration of amplifier energy efficiency is equivalent to the consideration of its power dissipation; specifically that efficiency is maximized when power dissipation is minimized. Design of maximally efficient amplifiers is therefore...
appropriately done using Fig. 2, which is an overlay of transistor characteristic curves (here for a FET) and hyperbolic contours of constant power dissipation for this transistor. Also included are a typical amplifier load line and an open circle identifying the lower $V_{DS}$ boundary of linear amplifier operation along this load line.

Fig. 1. Importance of amplifier energy efficiency: a) power flows of an amplifier; b) impact of efficiency on power supply size (solid curve) and heatsink size (dashed curve) scaled to TX output power = 1.

Fig. 2. Overlay of power dissipation curves on transistor characteristic curves, with a typical amplifier load line (graph shows an example load line of $4\Omega$). The open circle marks the lower $V_{DS}$ boundary for linear operation along this load line.

Minimizing power dissipation in the amplifier transistor requires operating the transistor in locations along its characteristic curves that intersect with power dissipation contours of low value. These low dissipation contours are located close to either the current or voltage axis, highlighted by the ellipses in Fig. 3. Also in Fig. 3 is a dotted curve that identifies the boundary of linear and compressed operation: the knee voltage profile $V_{k}$ [4]. As expected, the necessary
transistor operating conditions for minimum power dissipation, and therefore maximum energy efficiency, are located away from the linear amplifier operating region. This is the reason for the linearity vs. efficiency tradeoff for any linear amplifier. Ohm’s Law requires that in order to have the highest amplifier efficiency, any amplifier must operate its finite-resistance active device in a strongly nonlinear fashion. Indeed, any improvement in efficiency requires nonlinear transistor operation, even for class-AB (which uses nonlinear cut-off operation).

Fig. 3. Minimizing power dissipation requires operation close to the current axis (within the ellipse) and well below the transistor knee voltage profile, well outside the linear region.

Transition times between these highest efficiency operating regions must be short to achieve this high efficiency. The transition must follow the load line as shown in Fig. 4 because $V_{GS}$ is the controlling parameter. This transition never settles with the transistor operating as a current source, so the transistor never regulates the current flowing in the load. Instead, it operates as a selector of when current flows through the load, and when is does not: the transistor must operate in C-mode [2], behaving as a switch.

Fig. 4. Transitions between the minimum power dissipation regions must be rapid.

An appropriate circuit model for this operation is shown in Fig. 5, explicitly identifying the resistance exhibited by this switch when it is conducting. Power dissipation in this ON
resistance ($R_{ON}$) is bad as it reduces amplifier efficiency. Power dissipation in the load resistance ($R_L$) is good, since this is the output signal. Corresponding signal waveforms are presented in Fig. 6 where $V_{IN}$ is the voltage signal driving the gate of the transistor, $V_{OUT}$ is the voltage appearing across the drain-source terminals of the transistor, $P_{LOAD}$ the power dissipated by the load $R_L$, and $P_{FET}$ is the power dissipated by the FET resistance $R_{ON}$ which peaks during the transitions from ON to OFF states and settles during the ON state.

For wideband operation in this switching mode, it is necessary to insure that the transistor is capable of the necessary fast transitions even at the highest operating frequency. The usual metric for transistor speed is the transition frequency ($f_T$), defined for a FET as

$$f_T = \frac{g_m}{C_{GS} + C_{GD}}$$

where $g_m$ is the FET transconductance, $C_{GS}$ is the gate to source (input) capacitance, and $C_{GD}$ is the gate to drain (reverse transfer) capacitance. This metric is not directly applicable to the output switching design objective because it does not consider the output capacitance $C_{DS}$. To the extent that the capacitances in (3) are greater than or less than $C_{DS}$, predictions of switching behavior using $f_T$ will be correspondingly pessimistic or optimistic. Applying $f_T$ to this problem is useful though when the characteristics of $C_{DS}$ are known in comparison to the capacitances in (3).

Fig. 5. Model of the maximum efficiency transistor circuit: a switch with finite ON resistance.

Assuming that $1/f_T$ is a useful estimate of the output transition time (when the amplifier transistor is in its highest power dissipation condition) then a first-order analysis of available energy efficiency with respect to the ratio of $f_T$ to the operating frequency $f_0$ provides the curves in Fig. 7. This simplified analysis does not take into account the transistor-external characteristic of the $R_L C_{DS}$ time constant which governs the turn-OFF transition.

Several important results are seen in Fig. 7. First, the closer that the operating frequency gets to $f_T$, the lower the available efficiency becomes because the amplifier spends more time proportionally in the high power dissipation transitions. Second, actual switching operation is seen when the available energy efficiency becomes independent of the operating frequency, which signifies that the transition time is now negligible compared to the stable ON and OFF times seen in Fig. 6. This independence is best when $f_T / f_0 > 50$, or when the maximum operating frequency is less than 2% of $f_T$. By accepting some efficiency degradation it is possible to operate up to 10% of $f_T$. At any higher operating frequency the efficiency penalty increases dramatically.

Third, the available energy efficiency increases as the ratio of $R_L/R_{ON}$ increases. This is identical to saying that power dissipation in the load resistance is good, and from the circuit model in Fig. 5 this corresponds to lower values of $R_{ON}$ and/or higher values of $R_L$.

When evaluating amplifier efficiency a usual metric is the power-added efficiency (PAE)

$$PAE = \frac{(P_{OUT} - P_{IN})}{P_{DC}}.$$  \hspace{1cm} (4)

While this is a perfectly valid calculation its physical premise, that the input signal power transfers to the load and that the amplifier adds power to it, is physically false for any common-source (CS) amplifier. There is no physical path for the input signal current to flow through the load. This means that the input signal power for any CS amplifier must be internally dissipated, giving total amplifier efficiency (TAE) as a different model for amplifier efficiency that is physically consistent:
\[ TAE = \frac{P_{\text{out}}}{P_{\text{DC}} + P_{\text{in}}} \]  

Both TAE and PAE reduce to drain efficiency (2) when gain is high enough to neglect the input signal power.

Initial TAE measurements for a wideband high efficiency amplifier are presented in Fig. 8a. Over the frequency band of 200 MHz through 3000 MHz the TAE remains above 50%. The measured \( f_t \) for this GaN HEMT PA transistor is 64 GHz, which allows plotting this same data on top of the theoretical results in Fig. 7. This result is provided in Fig. 8b. While the shape of the measured data is consistent with this simplified analysis, the roll-off in measured efficiency is much faster than predicted. This shows that approximations made in the theoretical work behind Fig. 7 yield a model that is too simplistic, making the predicted results optimistic. The shape of the measured data curve implies that the basic analysis principles are likely sound.

\[ f_t = \frac{Q}{\ln(2)} \]  

where the network output impedance is set here to 50 ohms. Evaluating (6) for the network bandwidth \( BW \) at a center frequency \( f_0 = 1600 \text{ MHz} \) provides the curve shown in Fig. 9. Large impedance transformation ratios correspond to narrower operating bandwidth.

\[ BW = \frac{R_L}{R_{\text{in}}} \times \frac{f_0}{Q} \]  

\[ f_0 = 1600 \text{ MHz} \]  

Fig. 9. Wide agility bandwidth operation requires a high load resistance at the transistor. For an operation bandwidth exceeding 3 GHz centered at 1600 MHz the load resistance must exceed 40 ohms (dashed ellipse).

To exceed a 3 GHz operating bandwidth, Fig. 9 shows that the transistor load impedance must exceed 40 ohms when the intended amplifier output impedance is 50 ohms. This is a small impedance transformation ratio. This resistance at the transistor drain forces high supply voltage operation in order to develop even moderate output signal power. These supply voltage minimums are presented in Fig. 10, showing two curves. The solid curve is the minimum supply voltage needed for linear operation, and the dashed curve corresponds to the same output power in switching operation.

Supply voltage is lower for any required output power in switching operation for two reasons:

1) linear operation requires the output signal to always be above the transistor knee voltage, meaning that the peak to peak magnitude of the saturated power square-wave is larger than the maximum available linear signal excursion (as seen in Fig. 6), and

2) the fundamental frequency signal component of the power saturated square-wave from Fourier Analysis is 2 dB
greater than for a linear sinewave of equal peak to peak excursion. With both of these effects increasing the switching PA output power from the same transistor, in order to normalize the output power for both techniques the supply voltage needs to be lowered in power saturated operation.

Simultaneously achieving decade bandwidth beyond the UHF bands, switching operation, and watt level power requires the transistor process to support high voltage and have $f_t$ near 100 GHz. Modern GaN HEMT transistor technology with 0.15 micron gate width meets these requirements to about 3 watts. The GaN transistor used in this PA development is shown in Fig. 11. At higher powers the voltage capability of GaN HEMTs may not be enough, and the amplifier must be designed to have a lower load resistance to achieve higher power. Any reduction in the load resistance requires an output matching network and its consequential reduction in operation bandwidth seen in Fig. 9.

Fig. 11. Die photograph of the initial GaN power transistor used for this decade bandwidth high efficiency amplifier.

IV. REVERSE SIGNAL TOLERANCE

In any application where multiple transmitters are located close together, the radiated signal from each transmitter will be ‘received’ by each other nearby antenna due to the reciprocity of antenna operation. When the transmitters are operating simultaneously, this results in a reverse signal flowing back into the output of each transmitter which is the sum of all the received signals from nearby transmitters. This reverse signal interacts at the transmitter output with the intended signal, often changing the output signal characteristics in undesirable ways [5].

When the reverse signal is at a different frequency from the transmitter, then this interaction produces intermodulation products (reverse intermod (RIM)) that can be severe enough to require control in a system specification [6]. Switching amplifiers are already known to exhibit low reverse intermod in cellular applications [7] compared to linear amplifiers.

This issue is particularly important in phased-arrays, where the reverse signals are coherent with phase offset to the intended transmitter signal. This coherent reverse signal case is called load pulling. Traditional phased array designs solve this load pulling problem by absorbing the reverse signal in an RF isolator, therefore protecting the PA. This approach limits operating bandwidth far more severely than the effects of matching networks due to the bandlimited nature of isolators.

The present decade-bandwidth power amplifier is evaluated for its ability to tolerate reverse signals without any RF isolation, with results provided in Fig. 12. Peak-to-peak phase excursion shown in Figure 12 is the total phase change (minimum to maximum) measured on the fundamental tone of the output signal while the reverse signal phase is varied across 360°. These measurements were performed at 950 MHz. Identical testing is performed on linear class-A and Class-AB amplifiers to provide a comparison reference. The data shows that in all cases this switching amplifier provides a much more stable output signal phase than either of the reference amplifiers. For the switching amplifier (labeled SE Switch in Fig. 12) when the reverse signal is -20 dBc or lower the measured total phase variation from desired is one degree or less.

This small phase shift has a very minor impact on array pointing performance. It is expected that adopting such a wideband switching amplifier would remove the need for TX isolators in the array, also removing their operating bandwidth limitations. And as a side benefit the efficiency of each TX element improves from both the elimination of the isolator loss and the PA switching operation.

Fig. 12. Measured output phase ‘stiffness’ for linear and switching transmitters. Smaller output phase excursions are best.

V. MODULATION COMPATIBILITY

Because amplifier linearity is suppressed in order to gain transmitter efficiency, any signal that requires amplifier linearity is fundamentally incompatible with these efficient transmitters. For reasons discussed in Section II, it is not physically possible to operate any linear amplifier with high efficiency. Thus, any signal modulation that requires a linear amplifier is fundamentally incompatible with this decade-bandwidth high efficiency circuitry.

The broad class of constant-envelope (CE) modulations though is directly compatible with this transmitter approach. Such signals include continuous-phase modulation (CPM) [8] which includes minimum-shift keying (MSK) as a subset. All frequency shift keyed (FSK) signals are also compatible, as is
ON-OFF keying (OOK) as long as transition shaping is not required. All of these signals have a limited bandwidth efficiency, often less than 1 bps/Hz [9]. Improved bandwidth efficiency that approaches that provided by quadrature amplitude modulation (QAM) is available from the constant-envelope pure phase shift keyed signals (pPSK) [10] [11].

In radar applications all CE signals also directly apply. These include wideband chirp, and signals modulated in frequency or phase only. Pulsed CW is also applicable, as this is equivalent to the OOK communications modulation.

VI. SIZE, WEIGHT, AND POWER (SWAP) ISSUES

SWaP for a radio communications system is minimized when the signal is generated with the smallest size transmitter, including its power supply and heatsink. Fig. 1b shows that power supply and heatsink sizes are smallest when the transmitter energy efficiency exceeds 50%. Conversely, SWaP is significantly increased (a bad thing) when the transmitter efficiency is 40% or less.

Output power from a transistor of any size is maximum when the transistor operates at output power saturation. This corresponds to switching operation, and also to the minimum sized transistor for that particular power. Switching operation simultaneously operates at minimum amplifier power dissipation, for smallest heatsink and supply size – all good for SWaP objectives.

In a radar system all of the above considerations are also applicable. Of further interest to phased-array radar applications is the ability to further reduce SWaP by eliminating RF isolators at each PA output. Recovering the bandwidth lost through the frequency restriction of RF isolators opens the opportunity to converge communications and radar operations in the same array aperture. This reduces the number of radios needed to be carried, a further benefit to reduced SWaP.

VII. CONCLUSION

Simultaneous realization of decade frequency bandwidth and energy efficiency exceeding 50% is a new capability with application in both communications and radar.

Realizing true switching operation requires use of a process that provides a speed capability measured by $f_T$ that exceeds the operating frequency by preferably a factor of 50. Having a high voltage and high $f_T$ process enables direct connection to a 50 ohm load and eliminates the bandwidth limiting matching network, thereby enabling a wide instantaneous bandwidth.

ACKNOWLEDGEMENT

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings contained in this paper are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Approved for Public Release, Distribution Unlimited.

REFERENCES