RF Power Amplifiers

Iulian Rosu, YO3DAC / VA3IUL, http://www.qsl.net/va3iul

RF Power Amplifiers are used in a wide variety of applications including Wireless Communication, TV transmissions, Radar, and RF heating.

The basic techniques for RF power amplification can use classes as A, B, C, D, E, and F, for frequencies ranging from VLF (Very Low Frequency) through Microwave Frequencies.

RF Output Power can range from a few mW to MW, depend by application.

The introduction of solid-state RF power devices brought the use of lower voltages, higher currents, and relatively low load resistances.

- Most important parameters that defines an RF Power Amplifier are:
  1. Output Power
  2. Gain
  3. Linearity
  4. Stability
  5. DC supply voltage
  6. Efficiency
  7. Ruggedness

Choosing the bias points of an RF Power Amplifier can determine the level of performance ultimately possible with that PA. By comparing PA bias approaches, can evaluate the trade-offs for: Output Power, Efficiency, Linearity, or other parameters for different applications.

- The Power Class of the amplification determines the type of bias applied to an RF power transistor.
- The Power Amplifier's Efficiency is a measure of its ability to convert the DC power (Pdc) of the supply, into the signal power delivered to the load (Po).

  The definition of the Efficiency (η) can be represented in an equation form as:

$$\eta = \frac{Signal\ power\ delivered\ to\ the\ Load}{DC\ power\ Supplied\ to\ the\ output\ circuit}$$ or Power Added Efficiency: $PAE = \frac{Po - Pin}{Pdc}$

- Power that is not converted to useful signal is dissipated as heat.

  Power Amplifiers that has low efficiency have high levels of heat dissipation, which could be a limiting factor in particular design.

- In addition to the class of operation, the overall efficiency of a Power Amplifier is affected by factors such as dielectric and conductor losses. First quantify any loss in the circuit, then attempt to minimize it, and finally ensure that the mechanical and thermal design is adequate under all conditions.

- The optimum load for highest Gain is determined by the small signal parameters of the device.
- The optimum load for minimum distortion and maximum output power at the 1dB compression point (P1dB) is determined by the bias point and I-V characteristics of the device.
- The load required to obtain the maximum value of Power Added Efficiency (PAE) will be some compromise between these two conditions.
- When considering which device will provide the greatest power added efficiency at the 1 dB compression point, the answer will not necessarily be the device whose characteristics are more linear. It will be the device whose load requirements for gain and maximum value of the 1dB compression point are the most commensurate.

Power Classes

Class - A

Is defined, as an amplifier that is biased so that the output current flows at all the time, and the input signal drive level is kept small enough to avoid driving the transistor in cut-off. Another way of stating this is to say that the conduction angle of the transistor is 360°, meaning that the transistor conducts for the full cycle of the input signal. That makes Class-A the most linear of all amplifier types, where linearity means simply how closely the output signal of the amplifier resembles the input signal.

- Always have to remember this:

No transistor is perfectly linear; however, the output signal of an amplifier is never an exact replica of the input signal.

- Linear amplification is required when the signal contains AM – Amplitude Modulation or a combination of both, Amplitude and Phase Modulation (SSB, TV video carriers, QPSK, QAM, OFDM).

Signals such as CW, FM or PM have constant envelopes (amplitudes) and therefore do not require linear amplification.

- The DC-power input is constant and the efficiency of an ideal Class-A PA is 50 % at PEP.
- The DC power consumption of a Class-A amplifier is independent of the output signal amplitude.

$$P_{DC} = V_{CC}^2 / R = V_{CC} \times I_{CQ} \qquad \text{and} \qquad I_{CQ} \sim I_{MAX} / 2$$

- The amplification process in Class-A is inherently linear, hence increasing the quiescent current or decreasing the input signal level monotonically decreases IMD and harmonic levels.
- Since both positive and negative excursions of the drive affect the drain current, it has the highest gain of any PA.

- The absence of harmonics in the amplification process, allows Class-A to be used at frequencies close to the maximum capability (f_{max}) of the transistor. However, the efficiency is low. Class-A Power Amplifiers are therefore typically used in applications requiring low power, high linearity, high gain, broadband operation, or high-frequency operation.
- The efficiency of real Class-A Power Amplifier is degraded by the on-state resistance or saturation voltage of the transistor. It is also degraded by the presence of load reactance, which in essence requires the PA to generate more output voltage or current to deliver the same power to the load.

$$\eta \text{ (Efficiency\_Class-A)} = \text{Max\_Load\_Voltage} / (2 * V_{cc}^2)$$

- One important thing to mentioned is that: small signal S-parameters can be used in simulations if the large-signal amplifier is operating in Class-A.
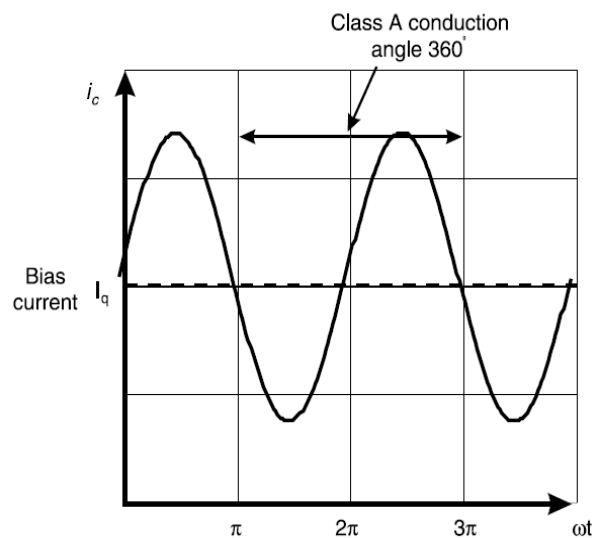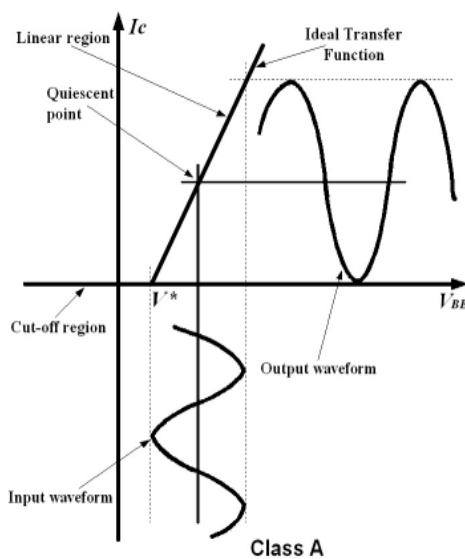
Class - B

This is an amplifier in which the conduction angle for the transistor is approximately 180°.
- Thus, the transistor conducts only half of the time, either on positive or negative half cycle of the input signal.
- The same as in Class-A, the DC bias applied to the transistor determines the Class-B operation.
- Class-B amplifiers are more efficient than Class-A amplifiers. The instantaneous efficiency of a Class-B Power Amplifier varies with the output voltage and for an ideal PA reaches $\pi/4$ (78.5 %) at PEP. However, they are much less linear. Therefore, a typical Class-B amplifier will produce quite a bit harmonic distortion that must be filtered from the amplified signal.

$P_{DC} = (2 * V_{CC} * V) / (\Pi * R)$
$P_{LOAD} = V^2 / (2 * R)$
$\eta \text{ (Efficiency\_Class-B)} = (\Pi * V) / (4 * V_{CC})$

Common configuration of Class-B amplifier is push-pull amplifier. In this configuration one transistor conducts during positive half cycles of the input signal and the second transistor conducts during the negative half cycle. In this way the entire input signal is reproduced at the output.
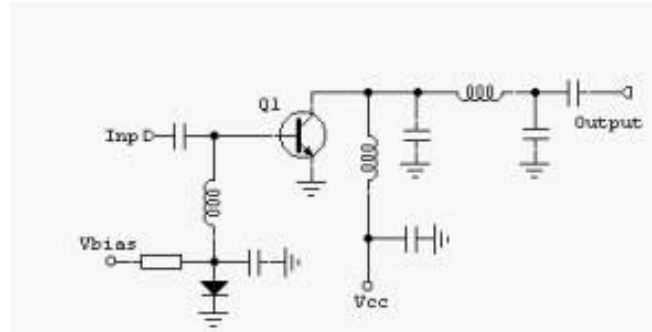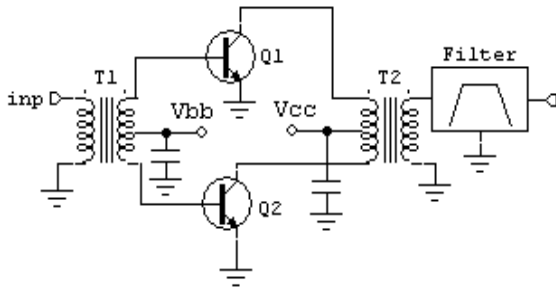
A single transistor may be used in a Class-B configuration. The only requirement in this case is that a resonant circuit must be placed in the output network of the transistor in order to "reproduce" the other half of the input signal.

- In practice, the quiescent current is on the order of 10% of the peak collector current and adjusted to minimize crossover distortion caused by transistor nonlinearities at low outputs.
- In theory 6dB or more drive power is needed to achieve Class-B compared with Class-A. In practice this 6dB reduction in power gain is lower; for BJT amplifiers is lower than FETs, approximately 2dB.
- The efficiency of the push-pull power amplifier is the same as that of the single ended power amplifier with the same conduction angle, and the output power capability of the push-pull power amplifier is twice that of the single-ended power amplifier (3dB higher).
- In the push-pull arrangement, the DC components and even harmonics cancel, but odd harmonics add, thus the output contains the fundamental only. Note that the cancellation of odd harmonics is only valid if the amplifier is not driven hard in saturation.
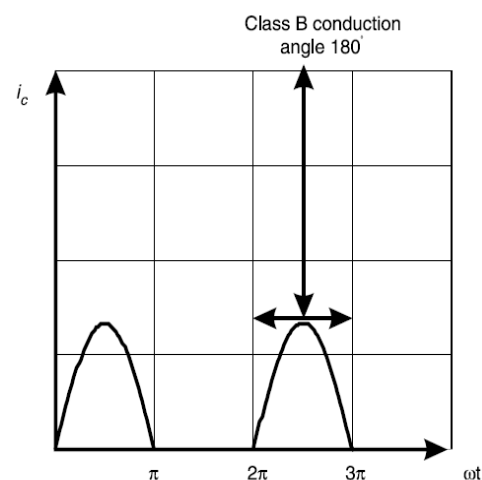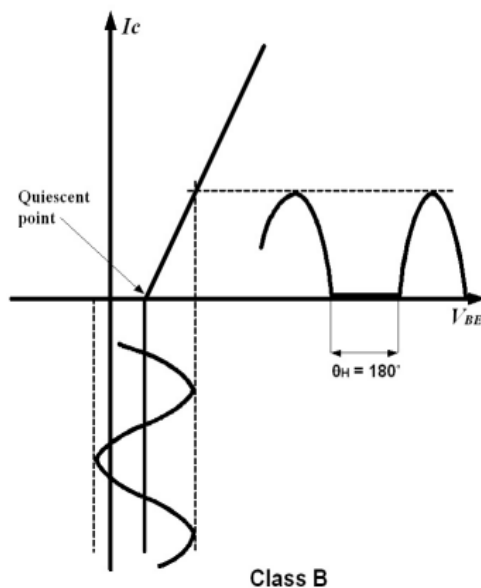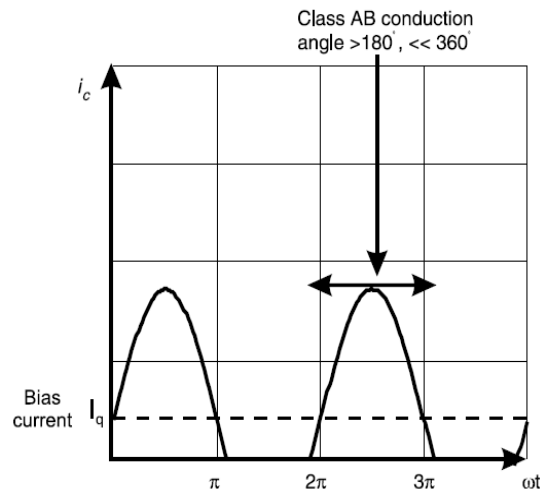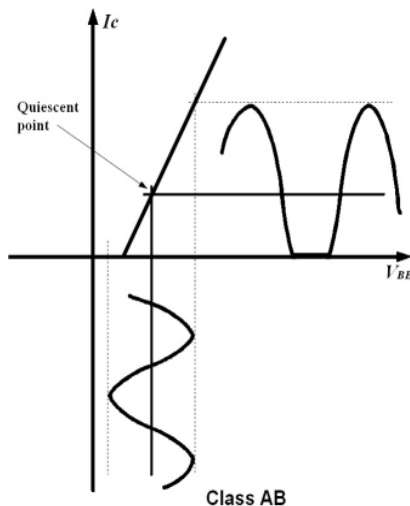
Class - AB

   This amplifier is a compromise between Class-A and Class-B in terms of efficiency and linearity.

The transistor is biased typically to a quiescent point, which is somewhere in the region between the cutoff point and the Class A bias point, at 10 to 15 percent of I_{Cmax}.

In this case, the transistor will be ON for more than half a cycle, but less than a full cycle of the input signal.

- Conduction angle in Class-AB is between 180° and 360° and Efficiency is between 50 % and 78.5 %
- Class-AB has higher efficiency than Class-A at price of linearity.
- Class-AB is not a linear amplifier; a signal with an amplitude-modulated envelope will be distorted significantly at this peak power level. The reason is in fact that in Class-AB operation the conduction angle is a function of drive level.

Experimentally was found that Class-AB often offers a wider dynamic range than either Class-A or Class-B operation. This is because gain compression in Class-AB comes from a different, and additional, source than Class-A. Saturation effects are primarily caused by the clipping of the RF voltage on the supply rails.
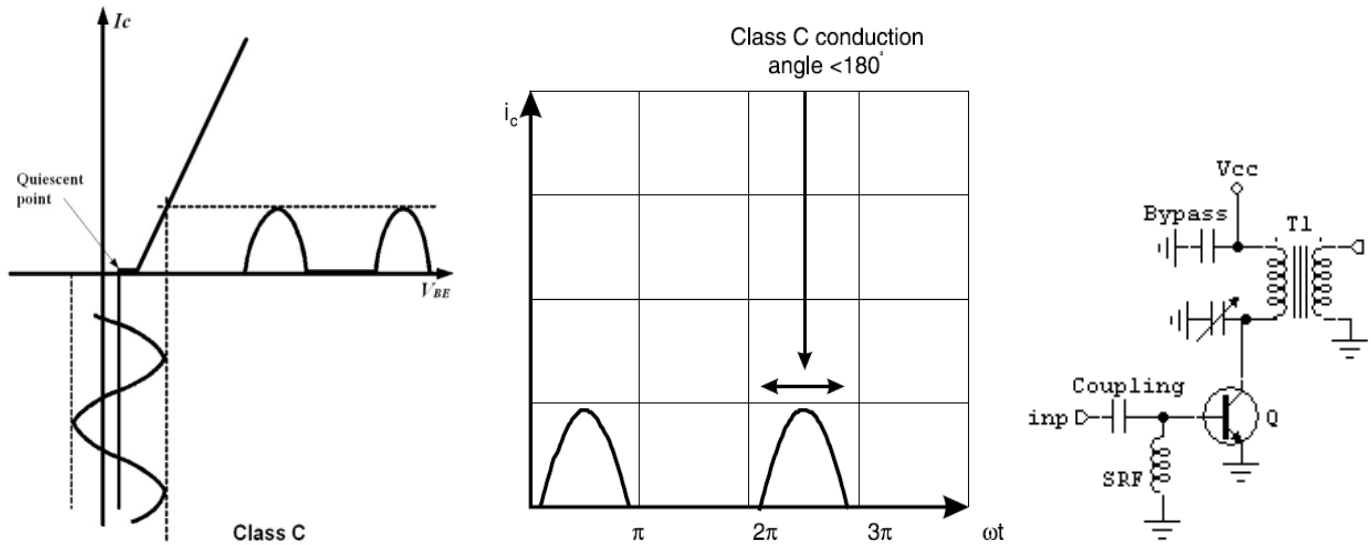

Class AB

- Linearizing the response of a BJT PA in Class-AB includes the use of specific, and very low, impedance for the base bias supply voltage. This is a very different bias design issue in comparison to the simple current bias used in small signal BJT amplifiers, or the simple high impedance voltage bias used in FET PA's.
- Running the Power Amplifier in a mid-AB class condition the power gain may be 3dB higher than Class-B.
- Conventional Class-AB operation incurs odd degree nonlinearities in the process of improving efficiency. Theoretically to increases efficiency all the way up to 78.5%, the device shall generate only even order nonlinearities. Such a device will not generate undesirable close-to-carrier intermodulation distortion.

Class - C

Is an amplifier where the conduction angle for the transistor is significantly less than 180°.
- The transistor is biased such that under steady-state conditions no collector current flows.
- The transistor idles at cut-off.   Class-C Amplifier


- Linearity of the Class-C amplifier is the poorest of the classes of amplifiers.
- The Efficiency of Class-C can approach 85%, which is much better than either the Class-B or the Class-A amplifier.
- In order to bias a transistor for Class-C operation, it is necessary to reverse bias of base-emitter junction. External biasing is usually not needed, because is possible to force the transistor to provide its own bias, using an RF choke from base to ground.

One of the major problems with utilizing Class-C in solid-state applications is the large negative swing of the input voltage, which coincides with the collector/drain output voltage peaks. This is the worst condition for reverse breakdown in any kind of transistor, and even small amounts of leakage current flowing at this point of the cycle have an important effect on the efficiency. For this reason, true Class-C operation is not often use in solid-state at higher RF and Microwave frequencies.

In order to survive Class-C operation, the transistor should have a collector voltage breakdown that is at least three times the active device's own DC voltage supply.

The reason: Class-C amplifiers have low average output power (since the transistor conducts only for short, pulse-like periods), but demand very high input drive levels. Thus, the transistor's main Class-C failure mode is the low value of the active device's own reverse breakdown voltage, which is unfortunately exacerbated by the RF input signal voltage going negative just as the transistor's collector voltage reaches its positive peak. This is especially problematic and dangerous if the load changes from design expectations, such as occurs if the system sustains a damaged or missing antenna or feed line during operation.

Class - D

The voltage mode Class D amplifier is defined as a switching circuit that results in the generation of a half-sinusoidal current waveform and a square voltage waveform. Class-D PAs use two or more transistors as switches to generate a square drain-voltage waveform. A series-tuned output filter passes only the fundamental-frequency component to the load,
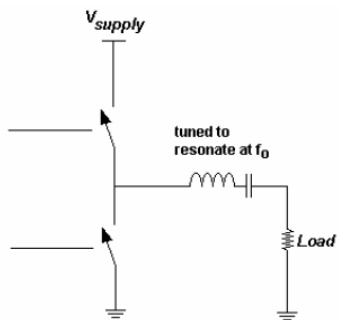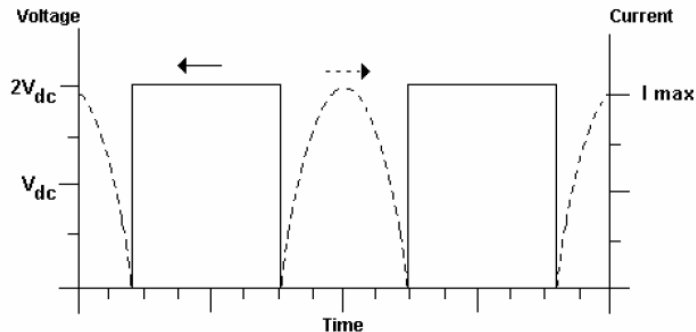
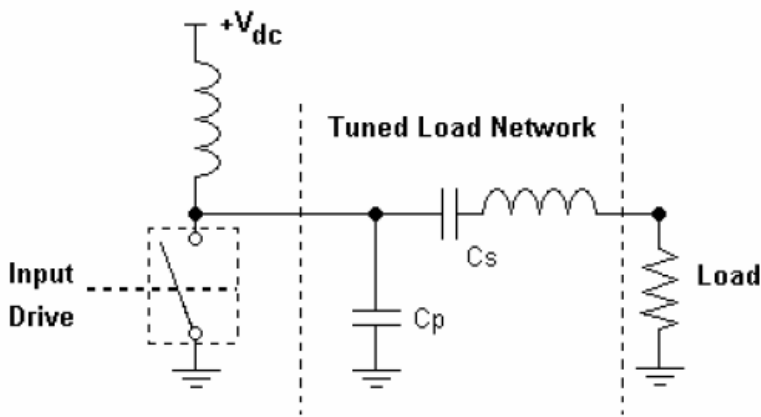Class-D amplifier   Class-D Voltage and Current waveforms

Class-D amplifiers suffer from a number of problems that make them difficult to realize, especially at high frequencies. First, the availability of suitable devices for the upper switch is limited. Secondly, device parasitics such as drain-source capacitance and lead inductance result in losses in each cycle. If realized, (they are common at low RF and audio frequencies) Class-D amplifiers theoretically can reach 100% efficiency, as there is no period during a cycle where the voltage and current waveforms overlap (current is drawn only through the transistor that is on).

- No real amplifier can be a true Class-D, as non-zero switch resistances and capacitive as well as inductive parasitics restrict the shape of the drain voltage waveform.
- A unique aspect of Class-D (with infinitely fast switching) is that efficiency is not degraded by the presence of reactance in the load.

Class - E

Class-E employs a single transistor operated as a switch. The collector/drain voltage waveform is the result of the sum of the DC and RF currents charging the drain-shunt capacitance Cp which is parallel with transistor internal capacitance c_o. In optimum class E, the drain voltage drops to zero and has zero slope just as the transistor turns on.

The result is an ideal efficiency of 100 %, elimination of the losses associated with charging the drain capacitance in class D, reduction of switching losses, and good tolerance of component variation.
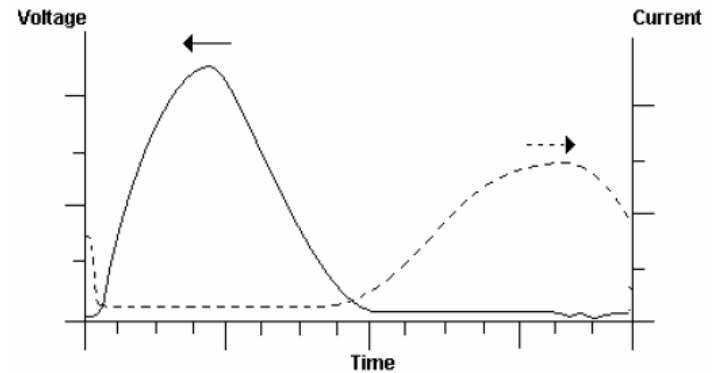
Class-E amplifier   Class-E Voltage and Current waveforms

- A Class-E amplifier will exhibit an upper limit on its frequency of operation based on the output capacitance required for the output matching circuit that produces the waveforms described and shown above.
- Specifically, a Class-E amplifier for optimum efficiency requires an upper limit on capacitance Cs.
- The radio frequency choke (RFC) is large, with the result that only DC current I_{dc} flows through it.
- The Q of the output circuit consisting of Ls and Cs is high enough so that the output current i_o and output voltage v_o consist of only the fundamental component. That is, all harmonics are removed by this filter.
- The transistor behaves as a perfect switch. When it is on, the collector/drain voltage is zero, and when it is off the collector current is zero.
- The transistor output capacitance c_o, and hence Cp, is independent of voltage.
- If a given transistor has an intrinsic capacitance c_o greater than Cp_max, it is not useable at the desired frequency. This Cs requirement implies that for high power at high frequencies, higher current densities are required, as the cross-sectional area of the switch corresponds directly to the device's intrinsic capacitance.

Class - F

Class-F boosts both efficiency and output by using harmonic resonators in the output network to shape the drain waveforms. The voltage waveform includes one or more odd harmonics and approximates a square wave, while the current includes even harmonics and approximates a half sine wave. Alternately ("inverse class F"), the voltage can approximate a half sine wave and the current a square wave.
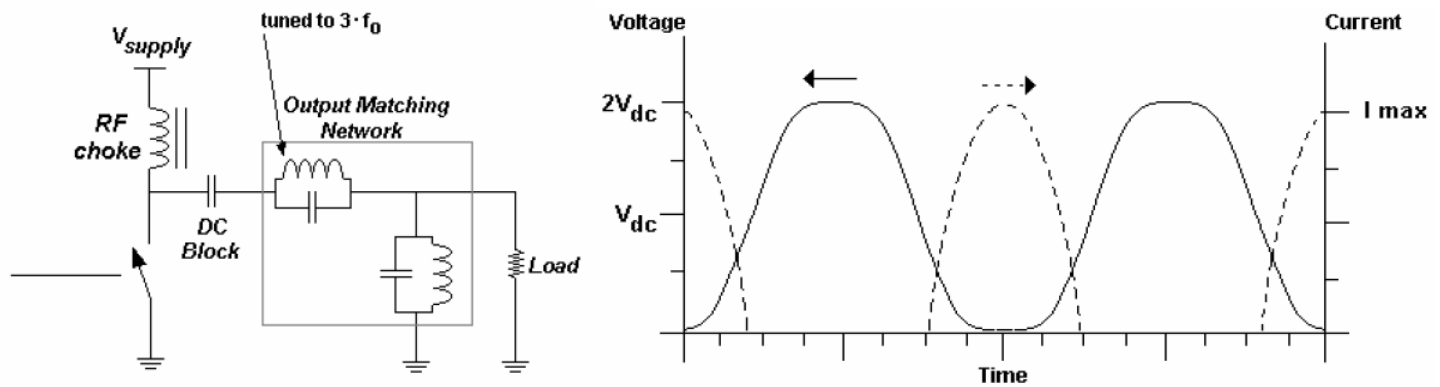
Class-F amplifier   Class-F Voltage and Current waveform**s**

- The required harmonics can in principle be produced by current source operation of the transistor. However, in practice the transistor is driven into saturation during part of the RF cycle and the harmonics are produced by a self-regulating mechanism similar to that of saturating Class-C. Use of a harmonic voltage requires creating a high impedance (3 to 10 times the load impedance) at the collector/drain, while use of a harmonic current requires a low impedance (1/3 to 1/10 of the load impedance). While Class-F requires a more complex output filter than other PAs, the impedances must be correct at only a few specific frequencies. Lumped-element traps are used at lower frequencies and transmission lines are used at microwave frequencies. Typically, a shorting stub is placed a quarter or half-wavelength away from the collector/drain.
- Class-F amplifier designs intentionally squaring the voltage waveform through controlling the harmonic content of the output waveform. This is accomplished by implementing an output matching network which provides high impedance 'open circuit' to the odd harmonics and low impedance 'shorts' to even harmonics. This results in a squared off (though for Class-F, truly squared) voltage waveform. The third harmonic only is peaked.
- Class-F amplifiers are capable of high efficiency, theoretically 88.4% for traditionally defined Class-F (or 100% if infinite harmonic tuning is used), but with about 72% efficiency in reality. Open circuit termination becomes a serious challenge at higher frequencies and higher power devices, mainly due to the escalation of the output capacitance of the active device.
- Class-F amplifier design is difficult to be implemented mainly due to the complex design of the output matching network.

A Class-F amplifier can also be built with a λ/4 transmission line as shown below:
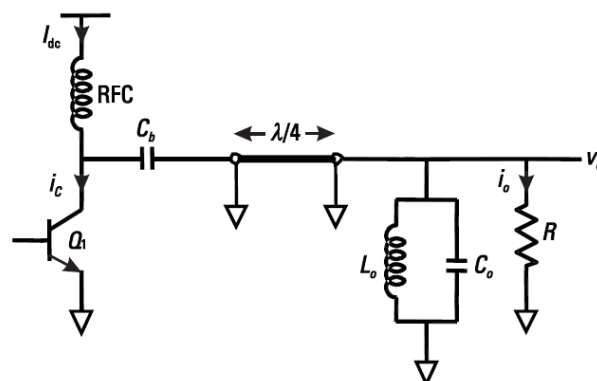
- A λ/4 transmission line transforms an open circuit into a short circuit and a short circuit into an open circuit.
- At the center frequency, the tuned circuit (L_o and C_o) is an open circuit, but at all other frequencies, the impedance is close to zero. Thus, at the fundamental frequency the impedance into the transmission line is R.
- At even harmonics, the λ/4 transmission line leaves the short circuit as a short circuit.
- At odd harmonics, the short circuit is transformed into an open circuit. This is equivalent to having a resonator at all odd harmonics, with the result that the collector voltage waveform is a square wave (odd harmonics should be at the right levels).

Power Classes definition
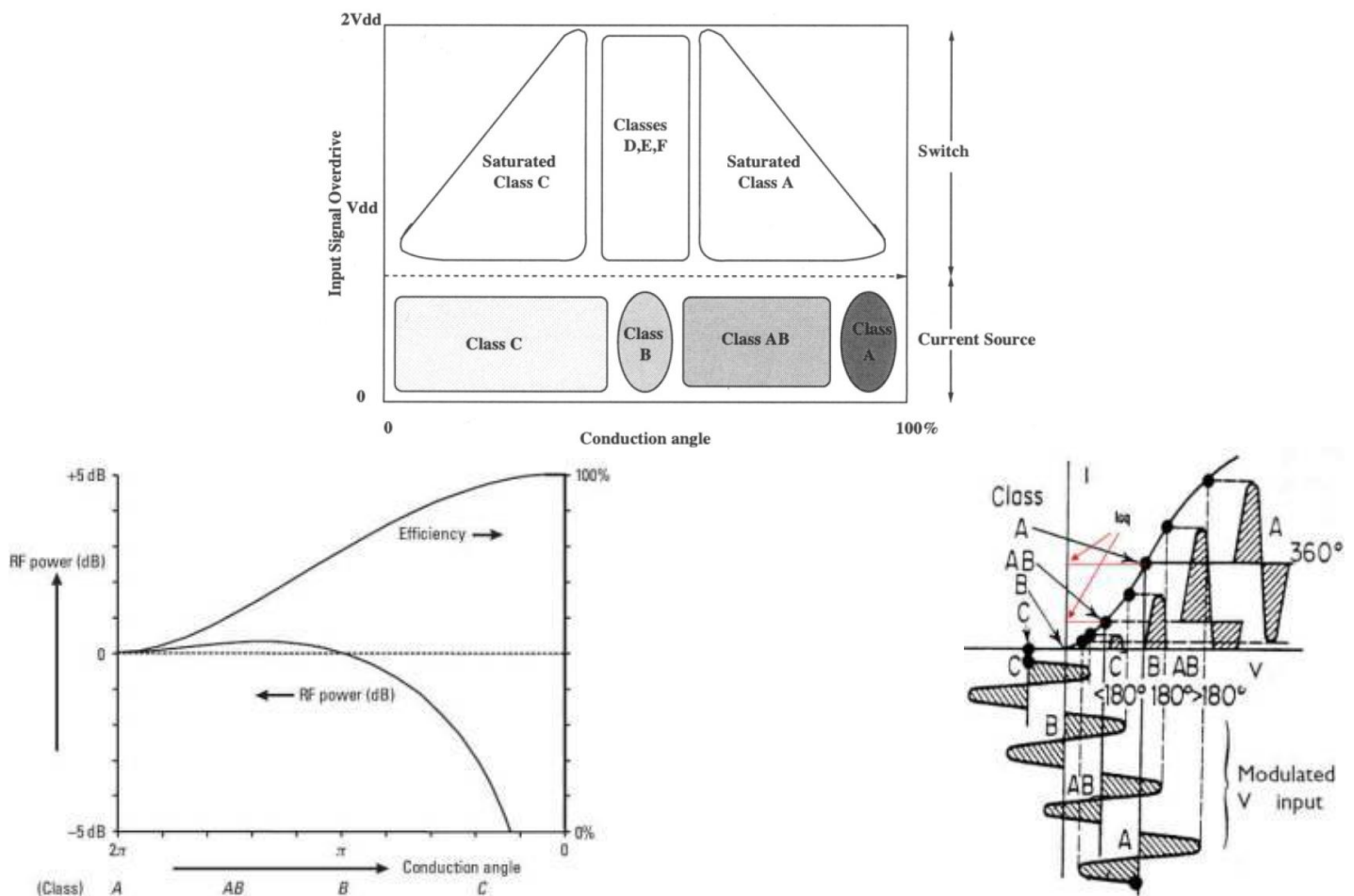
Classical definition of Power Amplifier classes

Power Amplifier Linearity

- When two or more signals are input to an amplifier simultaneously, the 2nd, 3rd, and higher-order intermodulation components (IM) are caused by the sum and difference products of each of the fundamental input signals and their associated harmonics.
- The rated PEP of a Power Amplifier is the maximum envelope power of a two-tone signal for which the amplifier intermodulation level is -30dBc.
- When two signals at frequencies f_1 and f_2 are input to any nonlinear amplifier, the following output components will result:
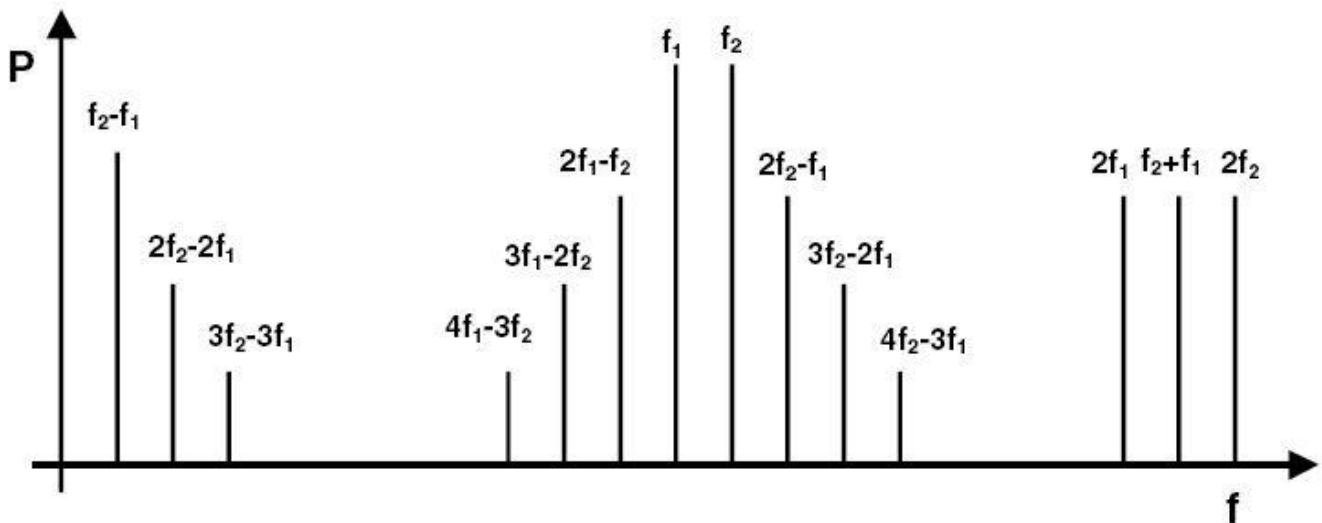
Fundamental: f_1, f_2
2nd order: $2f_1$, $2f_2$, $f_1 + f_2$, $f_1 - f_2$
3rd order: $3f_1$, $3f_2$, $2f_1 \pm f_2$, $2f_2 \pm f_1$
4th order: $4f_1$, $4f_2$, $2f_2 \pm 2f_1$
5th order: $5f_1$, $5f_2$, $3f_1 \pm 2f_2$, $3f_2 \pm 2f_1$, **+ Higher order terms**

- The odd order intermodulation products (**$2f_1-f_2$, $2f_2-f_1$, $3f_1-2f_2$, $3f_2-2f_1$, etc.**) are close to the two fundamental tone frequencies f_1 and f_2.


- The nonlinearity of a Power Amplifier can be measured on the basis of generated spectra than on variations of the fundamental signal. The estimation of the amplitude change (in dB), of the intermodulation components (IM) versus fundamental level change, is equal to the order of nonlinearity.
- For 1dB increase of the levels of the two tones (**f_1** and **f_2**), the level of IM2 will go up with 2dB, the level of IM3 will go up with 3dB, the level of IM5 goes up 5dB, and so on. The above statement is valid **only for an amplifier that is not in compression**.

- As a relation between the degree of the nonlinearity (3rd, 5th, etc.) and the frequency of the side tone (such as IM3, IM5, etc), can be mentioned that the IM5 tones are not affected by third-degree nonlinearities, but IM3 tones are functions of both, 3rd and 5th degree (and higher) nonlinearities. That means at low signal amplitudes, where the 5th order distortion products can be neglected, the amplitudes of the IM3 tones are proportional to the 3rd power of the input amplitude.
- With fairly large signal amplitude, 5th order products (which are dependent on a power of five) will start to affect the IM3 responses. As a result, the 3:1 amplitude estimate will no longer hold.

- If the phases of the 3rd and 5th degree coefficients are equal, the 5th degree nonlinearity will expand the IM3 responses. However, if the phases are the opposite, the IM3 distortion will be locally reduced. This explains why notches (sweet-spots) in the IM3 (and high-order) sidebands have been reported at certain amplitudes of output power.
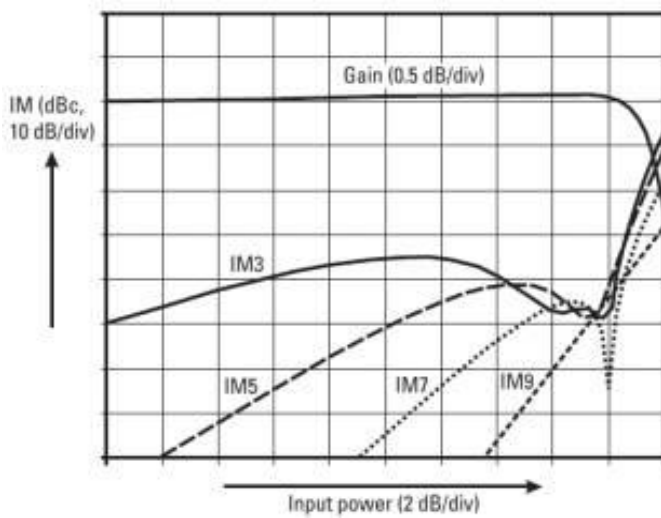
IM(n) products vs Input Power
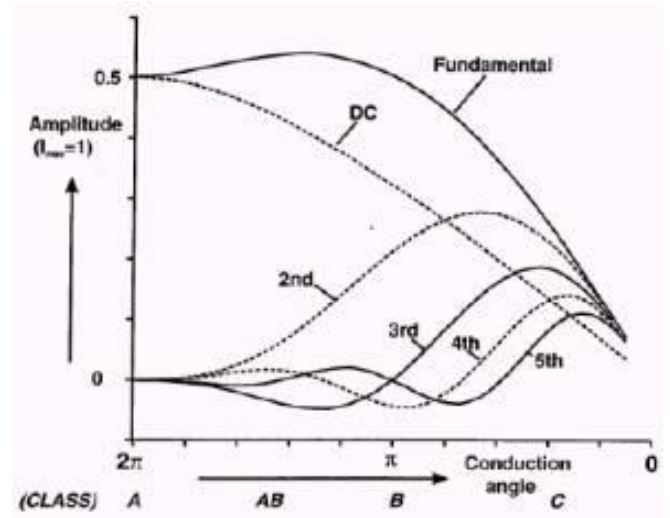
Re-growth of harmonic content vs Conduction Angle
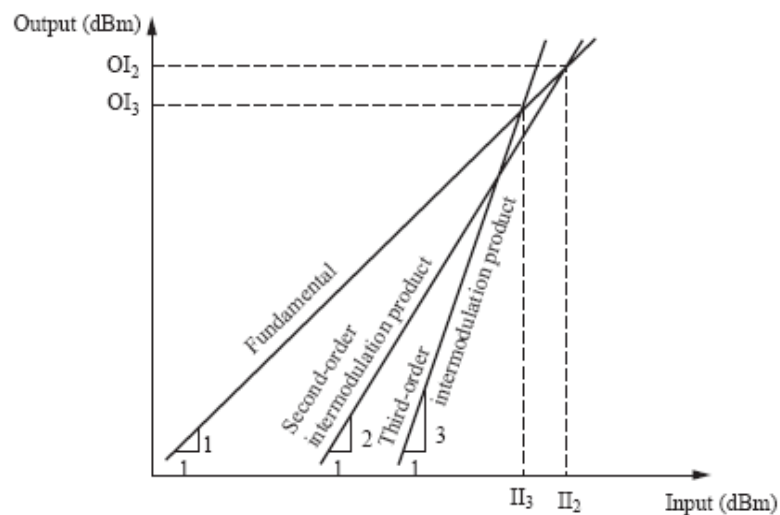
2nd and 3rd order input and output intercept points

- Since the amount of device nonlinearities cannot be changed much, distortion is most effectively minimized by optimizing the impedances seen by the distortion current sources.
- In all the Power Amplifiers, the output level is a "compressive" or "saturating" function of the input level. The gain of the Power Amplifier approaches zero for sufficiently high input levels. In RF circuits this effect is quantified by the "1dB compression point", defined as the input signal level that causes the small-signal gain to drop by 1dB. This can be plotted in a log-log scale vs input level.
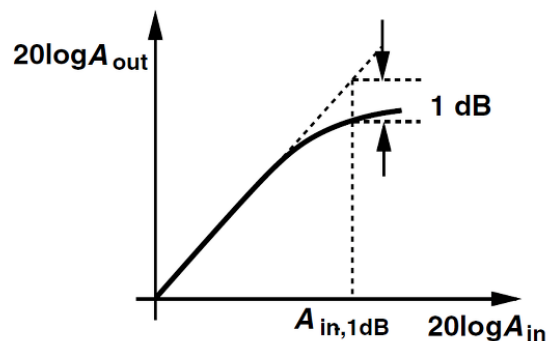
- Sometimes this Output Power vs Input Power characteristic is referred as AM-AM distortion.
- The asymmetry of side-band intermodiulation products (IM) in a two-tone test is often dependent on the carrier spacing, but not in a monotonic fashion. The effect can be explained as an interaction between AM-AM and AM-PM distortion processes. On the other hand, the mere presence of both processes does not guarantee that asymmetry will occur.
- If there is a time lag, or phase shift as measured in the envelope time domain between the AM-AM and AM-PM responses, or their individual frequency components, IM asymmetry will occur.

  Reduction of AM-PM in the PA design would alleviate the IM asymmetry issue.

- For RF power transistors, a primary cause of AM-PM effects appears to be the dynamic mistuning of the input match. Some deliberate mistuning on the high Q factor input match of RF power transistors might pay off in terms of improved AM-PM performance for the loss of a decibel or two of gain.
- Nonlinear Power Amplifiers can cause signals to be spread into adjacent channels, which can cause Cross Modulation. This is based on the same phenomena as 3rd order intermodulation for nonlinear amplifiers with two-tone inputs.
- Was mention that the level of the harmonic and intermodulation products decreases stronger than the decrease of the fundamental, with decreasing the input power. This deduces a crude method for linearization named Back-off Power Optimization for Linearity.
- Increasing of the backup ratio of an over dimensioned Power Amplifier enhances linearity at the expense of the efficiency.

Power Amplifier Linearity Test using Noise Power Ratio (NPR)

Another measurement for quantifying power amplifier linearity is **Noise Power Ratio (NPR).**
The concept of **Noise Power Ratio (NPR)** has been around since the early days of frequency division multiplexed (FDM) telephone systems.
- The NPR is simply a measure of the "quietness" of an unused channel in a multi-channel system, when there is random activity on the others.

In an NPR test system, white noise is used to simulate a multi-tone carrier signal.
An additive white Gaussian noise generator has a high Crest Factor (CF) and represents a wideband communication signal much better than a two-tone IMD stimulus.
The noise is band-limited by a filter, to either the useful bandwidth of the amplifier or the bandwidth of the expected signal. The resulting signal is passed through a notch filter with a notch typically greater than 50dB below the passband amplitude and a width of approximately 1% or less of the filtered noise bandwidth.
Several notch frequencies across the noise bandwidth (low, midband, and high) are tested to characterize the system adequately.
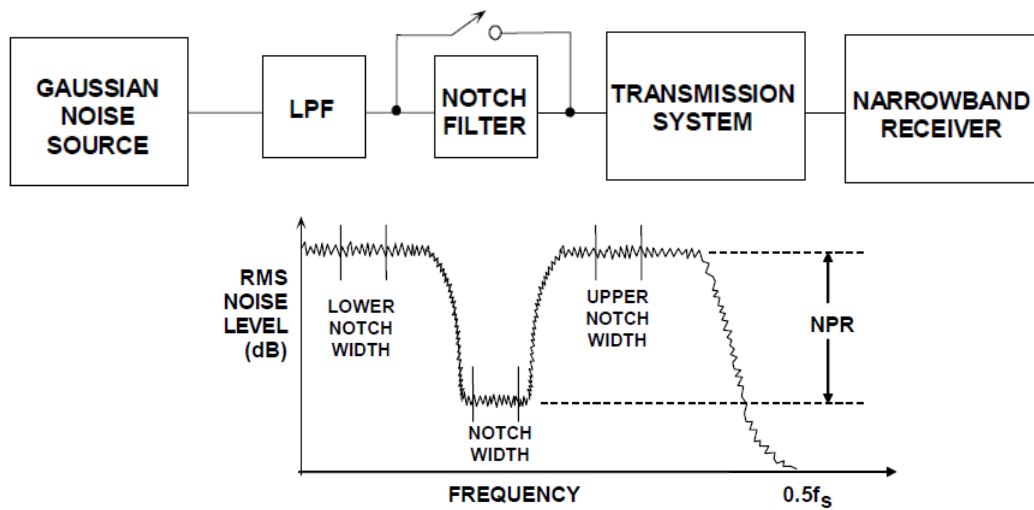
Noise Power Ratio (NPR) test setup and measurement

With the notch filter out, the rms noise power of the signal inside the notch is measured by the narrowband receiver. The notch filter is then switched in, and the residual noise inside the notch is measured. The ratio of these two readings expressed in dB is the NPR.

Crest Factor (CF) Linearity Measurement

The third approach for characterizing an amplifier's linearity is with a Crest Factor (CF) measurement. CF is the ratio of peak-to-average power.
Like the NPR measurement, the amplifier input is band-limited noise to drive it with a signal approximating what an amplifier would see in actual use.
First, using a directional coupler or signal splitter, the incident signal is measured with a wideband peak power sensor. The video bandwidth of the sensor, as well as the coupler or splitter, should be at least as wide as the bandwidth of the noise signal; otherwise, the measurement will be distorted.
A second measurement is made at the output of the amplifier, including any attenuation needed to keep the signal in the measurement range of the sensor, and the CFs of the input and output signals are compared. If the output Crest Factor (CF) is less, the amplifier is compressing the signal's highest peaks. For the most accurate measurement, a third sensor can be used to determine if a portion of the input signal is reflected rather than amplified.
By comparing the input and output average power, the amplifier's gain can be determined.
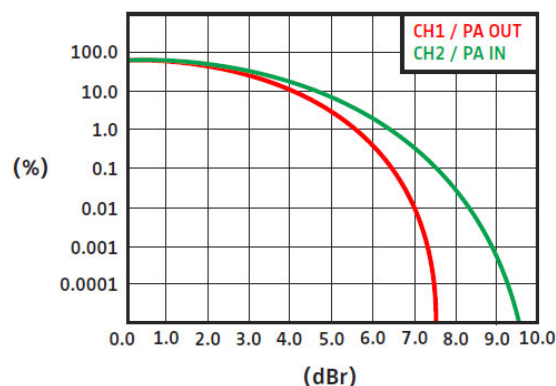
Complementary Cumulative Distribution Function (CCDF)

- The **Complementary Cumulative Distribution Function (CCDF)** can provide additional insight into amplifier performance. When the input power is increased to certain levels, the Crest Factor (CF) at the output degrades significantly, indicating a substantial degree of compression.
- For each power level, the CCDF curve shows the amount of time the signal spends above the average power level. Equivalently, the CCDF curve displays the probability that the signal power will be above the average power.

Conclusions about Amplifier Linearity

- The first order term reproduces the input signal amplified by the small signal gain.
- The second order term creates output signals which are at the second harmonic of the input signals and at all the combinations which are the sum and difference frequencies of the input signals.
- If only one input frequency is present, the second order term will produce output components at the second harmonic and a DC component. This term cannot produce outputs at the input frequency.
- The third order term creates outputs at frequencies which are very close to the input frequencies and outputs which are exactly at the input frequencies. The output components which are exactly at the input frequencies add vectorially to the output produced by the first order term. If the phase of the component created by the third order term is 180° out of phase with that produced by the first order term, the output signal at that frequency will be reduced.
- Since the output amplitude of the signal created by the 3rd order term is a function of signal level this reduction in output amplitude will increase as the signal level increases. This is known as Gain Compression.
- If the output component created by the third order term is not 180° out of phase with the component produced by the first order term, then the output amplitude and phase will be functions of the signal level.
- The change in output phase angle when the input level is changed is known as AM to PM conversion.

The intermodulation distortion components appear to be caused by the mixing of the 2nd harmonic of one signal with the fundamental of the other. This observation may result in the erroneous conclusion that if the 2nd harmonic is eliminated, intermodulation distortion will be eliminated. 2nd harmonic reduction may be accomplished by decreasing the magnitude of the coefficient of the 2nd order term. In a practical situation, this may be achieved by filtering.

- The decrease in the magnitude of the 2nd order term will not directly affect the magnitude of the 3rd order term and the outputs at the intermodulation distortion frequencies will not be affected.
- The output signals created by the third order term which are very close to the frequencies of the input signals are of great importance because they usually fall within the passband of the amplifier and cannot be filtered out.
- Commonly used methods of specifying third order distortion components are: intermodulation distortion, cross modulation distortion, modulation distortion, and triple beat distortion.

- It is obvious that they are closely associated with gain saturation since they are produced by the same third order term of the output expansion.

Memory Effect

- In a two (or multi tone) IMD test if the amplitude and/or phase of the IM signals is affected by the tone difference, the amplifier exhibits memory effects.
- Memory is caused by the storage of energy that has to be charged or discharged.
- Memory Effect could be explained as a time lag between AM-AM and AM-PM response of the amplifier. The Electrical Memory Effect is introduced by poor gate/base and drain/collector decoupling at low frequencies causing a distortion of the envelope currents which results in IMD asymmetry.
  Low frequencies mean baseband/video frequencies or the spacing between two tones. The most significant Memory Effect appears in Class AB amplifiers, with reduced conduction angle where drain/collector varies with output power. In Class A amplifiers the Memory Effect reduces.
- Smooth memory effects are not usually harmful to the linearity of the PA itself.
  - A phase rotation of 10^0 to 20^0 or an amplitude change of less than 0.5dB, as a function of modulation frequency, has no dramatic effect on the linearity of the device.
- There are two types of memory effects: Electrical and Thermal.
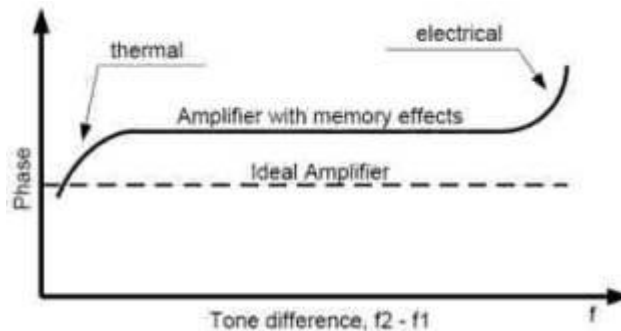
- Electrical memory effects are produced by non-constant node impedances within frequency bands as DC, Fundamental, and Harmonics. Most of these effects are generated by frequency dependent envelope impedance, and those within the DC band are the most harmful, because bias impedances are strongly frequency dependent.
- Thermal memory effects are generated by the junction temperature, which is modulated by the applied signal.
- Thermal effects will be much more prominent in a slow sweep, or a stepped CW test.
- The analysis and simulation show that the drain envelope impedance is the most important factor for reducing the memory effect and nonlinearity. A new matching topology is proposed for minimizing the drain and gate envelope impedances.
  - The matching topology consists of a series LC circuit for shorting the device at a low frequency while maintaining a matchable impedance at the operating frequency.

- The circuits are connected to the gate and drain terminals, rather than to the bias lines, since the circuit can produce a very low impedance, not limited by the quarter-wavelength bias line. The amplifier, with the reduced envelope impedances, provides drastically reduced memory effects and very linear amplification performance for wideband signals.
- Simultaneous amplitude and delayed phase modulation do generate asymmetric sidebands.

Input/Output Matching and Load Line

- The input matching configuration, including the bias circuit, has an important impact on the operation of the RF Power Amplifiers.
- The input match will show different optima for maximum gain, best linearity, and highest efficiency. Optimization of the efficiency may involve substantial reduction in power gain.
- Correct handling of harmonics is a necessary feature on the input, as well as the output, match. Device used well below its cutoff frequency may require specific harmonic terminating circuit elements on the input.
- The performance of the output matching circuit is critical for a Power Amplifier. In a Power Amplifier, impedances control how much power is delivered to the output and how much gain and noise are produced in the process, therefore matching network is critical for maximum performance.
- One aspect that's sometimes overlooked is the power dissipation in the output matching circuit. This power is lost in the capacitors, inductors, and other lossy elements that are part of the matching network. This "dissipation loss" degrades the PA's efficiency and output power capability.
- Different implementations of the output match result in different losses and there are still significant design tradeoffs to be made between bandwidth and dissipation loss. For a Power Amplifier, the loss of the output match is always a concern because of the large power levels involved.
- A capacitor's quality factor is reversal proportional to its capacitance. To minimize the dissipation loss of the output match, it's therefore necessary to design the output match with the lowest possible value of C. The tradeoff is between bandwidth and dissipation loss.
- Different capacitor technologies give different losses when used in output matching circuits.
- One way of understanding the loss mechanisms of an output match and to don't mix up mismatch loss and dissipation loss, is to simulate the match with loss-less components, then introduce loss into one component at a time.

$$\text{Mismatch Loss}_{[dB]} = 10*LOG\ (1-\Gamma^2)$$

where reflection coefficient $\Gamma = (VSWR-1)\ /\ (VSWR+1)$

- Because the dissipation loss doesn't depend on the source impedance it's possible to use S21 to find the correct dissipation loss in a circuit simulation. The procedure

involves using the complex conjugate of the simulated load line as the source impedance.

- Running at a low efficiency not only reduces talk time in a portable device, but it also creates significant problems with heating and reliability.
- The load line is set based on the needed Power Amplifier output power and available supply voltage. For example, low voltage PA's (~3.5V for mobile devices) have a load line ranging from 1 to 5 Ω.

$$R_L = V_{max} / I_{max}$$

- Matching for maximum Gain occurs when the amplifier is unconditional stable and load impedance is equal to the complex conjugate of the same source impedance (conjugate matching). Complex conjugate simply refers to complex impedance having the same real part with an opposite reactance.

e.g. - if the source impedance is Zs=R+jX, then its complex conjugate would be Zs*=R-jX
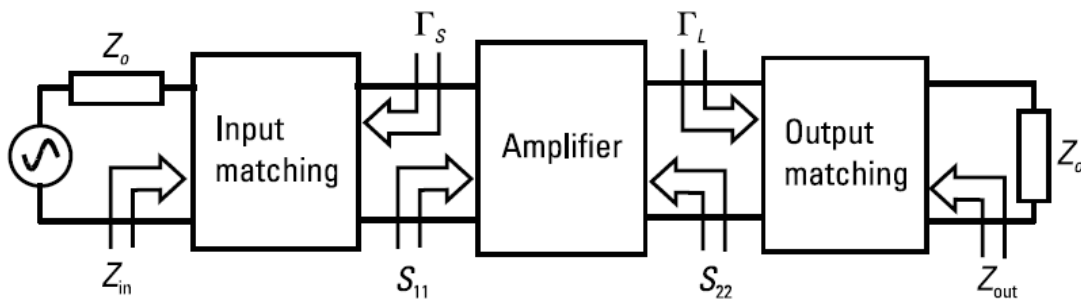
If matched: $Z_{in} = Z_o$, $\Gamma_s = S_{11}{}^*$, and $Z_{out} = Z_o$, $\Gamma_L = S_{22}{}^*$

- Matching for maximum Output Power occurs when Optimum Load impedance (R_L) is equal to Source impedance (R_{gen}). In order to obtain maximum output power, typically the power amplifier is not conjugate matched. Instead, the load is designed such that the amplifier has the correct voltage and current to deliver the required power.
- If operation is at the optimal Power Added Efficiency point, optimal-power tuning produces about 1dB to 3dB of higher power. Gain is reduced (for small Pin) typically by a slightly smaller amount.
- The transistor's input and output impedances will also decrease with an increase in frequency, which further complicates the design of a PA's matching networks, especially since these impedances can be as low as 0.5Ω. Thus, when matching a discrete driver stage to its Power Amplifier with maximum efficiency, we would normally want to implement a direct match from the true output impedance of the driver to the true input impedance of the Power Amplifier, instead of first forming a 50Ω match at the output of the driver, and then another 50Ω match for the input of the power amplifier, as this would needlessly transform the impedances from low to high, and then back from high to low.
- By selecting a transistor with a high collector voltage requirement, we can increase its output impedance over a transistor that operates at lower values of collector voltage.
- Output matching network for most high-powered amplifiers should normally consist of the T-type, rather than the Pi-type. Pi matching networks for high powered amplifiers

sometimes result in unrealistic component values when matching for the higher operating frequencies encountered into a 50Ω load. Indeed, T-networks are capable of much higher-frequency operation before this becomes a major problem. Both T and Pi networks can be used, however, if the output impedance of the transistor is higher than its load.

- It is possible to increase the bandwidth by using a higher order of output matching network. For example, instead of an L network, a double-L network can be used to convert first to an intermediate impedance, and then to the final value.
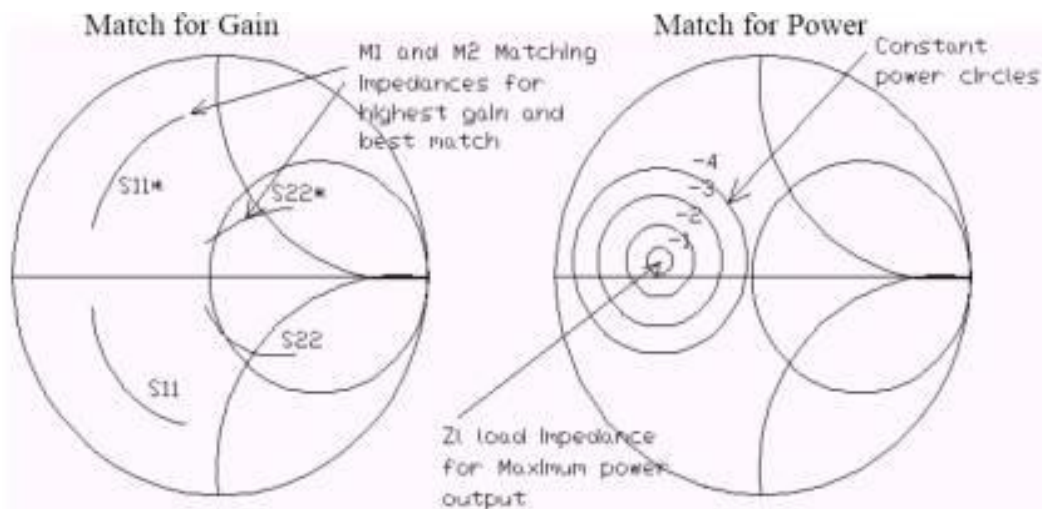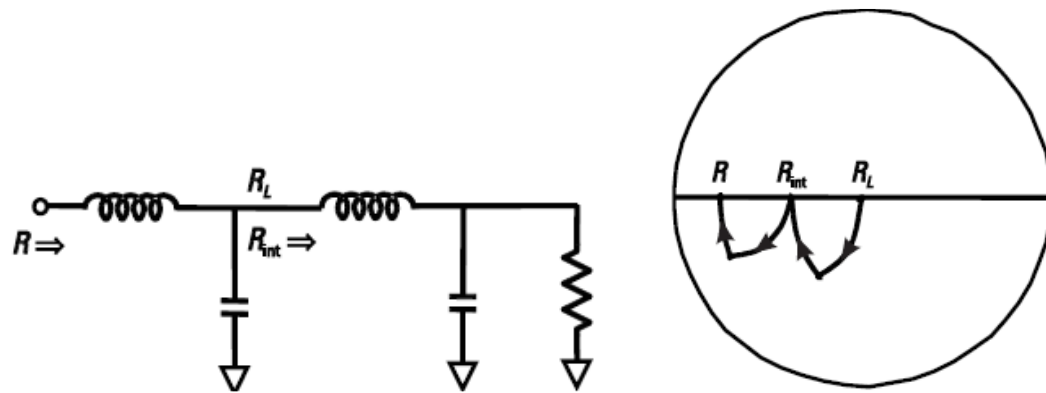
Smith Chart representation for maximum gain and power matching
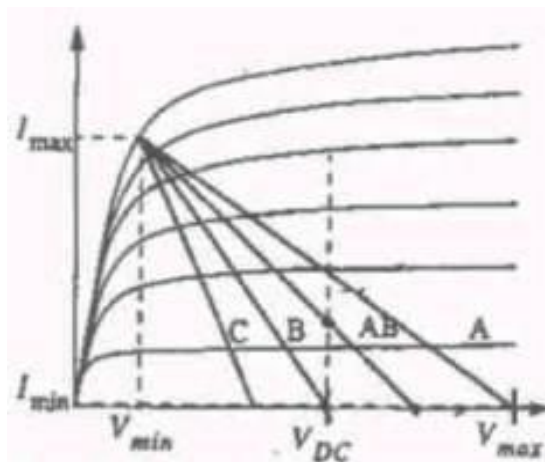
Load Line for different Power Classes

Optimum Load Resistance

In the absence of collector output resistance information on the datasheet, it becomes necessary to make a simple calculation to determine the optimum load resistance for the transistor.

The value of load resistance is dependent upon power level required and is given by:

$$R_L = \frac{(V_{CC} - V_{SAT})^2}{2P} \qquad \text{or} \qquad R_L = \frac{(V_{CC})^2}{2P} \qquad \text{(with less accuracy)}$$

where,
V_{CC} = the supply voltage
V_{SAT} = the saturation voltage of the transistor
P = the output power level required in Watts

Note that this equation provides only the load resistance, when usually in the datasheets the manufacturer provides values of shunt output capacitance vs frequency for the RF power transistor.

Power Amplifier Bias Design

- There are many different ways to bias an amplifier, depending on the required temperature stability, efficiency, cost, device, power output, linearity, and so on.

- **Power BJT transistor biasing:**

  - Must force the DC (average) value of V_{CE} and I_C to desired values and keep them constant using feedback techniques.
  - Whether employing diode or transistor bias, it is essential to thermally connect these components to the RF transistor itself. This allows the semiconductor bias components to track the power amplifier's temperature variations.
  - It is also possible to decrease input voltage as temperature increases, for example, by using a diode in the input circuit, using a current mirror, or using a more complex arrangement of thermal sensors and bandgap biasing circuits.
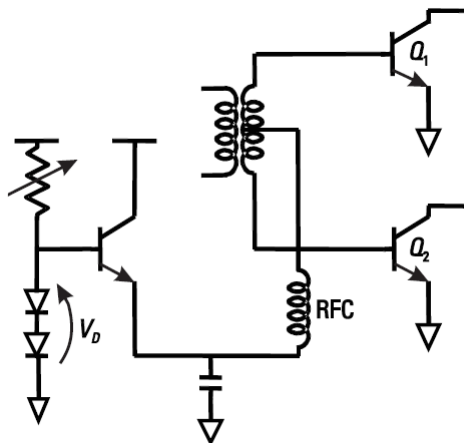
In the above example, all diodes and transistors are assumed to be at the same temperature. As temperature rises, V_D falls, reducing V_{BE} and keeping I constant.

- **Power FET transistor biasing:**
  The gate biasing circuit has several functions:
    - To maintain a constant gate-to-source voltage, Vgs.
    - To be able to supply a negative and positive gate current, Igs.
    - To protect the gate by limiting Igs when the device goes into breakdown (drain-togate or gate-to-source) or when the gate-tosource junction is biased with a positive voltage. These abnormal operating conditions for the devices can be due to an operator error, an overdrive, a system problem or ESDs.
    - To stabilize the device in case a negative resistance appears in the gate at any frequency where the device has a positive gain.
    - To filter the signal, the products and the harmonics generated by the device input from low to high frequencies without affecting the device input matching circuit.
    - To isolate the gate from any signal coming from the drain through the bias circuits.

- **Power LDMOS transistor biasing**
 The main consideration of the power LDMOS biasing is to achieve the linearity.
    - This is done DC biasing of the LDMOS transistor for optimal drain current for a given power output.
    - This bias needs to be held constant over temperature and time. Typically, the target accuracy for bias current over temperature is ±5% but ±3% is much more desirable for a high-performance design.
    - The DC Bias on LDMOS amplifiers is set by applying a DC voltage to the gate (V_{gs}) and monitoring the Drain current (I_{dd}).
    - Ideally, this I_{dd} will be constant over temperature, but since the Vgs of LDMOS amplifier devices varies with temperature, some type of temperature compensation is required. For optimal temperature compensation, in-circuit adjustments need to be made for both the temperature compensation as well as the V_{gs} bias itself.

- **Power GaN HEMT transistor biasing**

  GaN HEMTs are depletion mode devices, which means the device is normally ON when the gate source voltage is zero.
    - You need a positive and negative voltage when working with GaN - a positive high **drain** voltage, and a lower negative **gate** voltage.
    - The bias sequencing for GaN must be conducted in a certain sequence - even before the RF signal is applied to the circuit - or else you risk damaging the device.

- In contrast, LDMOS is an enhancement mode device and needs a positive, high **drain** voltage and a lower, positive **gate** voltage.
  For LDMOS, the gate voltage may be applied at the same time as drain voltage.

- For GaN Power Amplifier, the correct **bias sequences** for powering up and down are as follows:
  - To **turn-ON** the device, the sequence is:
    1. Apply Gate voltage (pinch-off voltage),
    2. Apply Drain voltage (operating voltage),
    3. Adjust Gate voltage (adjust to set operating/quiescent current),
    4. Turn RF ON.
  - To **turn-OFF** the device, the sequence is reversed:
    1. Turn RF OFF,
    2. Adjust Gate voltage (adjust to pinch-off voltage),
    3. Adjust Drain voltage (reduces to zero volts then turn-off),
    4. Turn-off Gate voltage (wait 2 seconds and then turn-off).

- In GaN PAs, the gain compression tends to start at large power back-off and compresses slowly when compared to LDMOS PAs (see plot below).
  This phenomenon is named Soft Compression.
  - The slow compression in GaN is primarily attributed to the device trans-conductance, though thermal effects also play a role.
    The rapid compression at high power is the result of knee region intrusions.
  - Lastly, the interaction between the source matching and input capacitance can affect the AM/AM, especially at higher frequencies (see plot below).
  - The nonlinear input capacitance in GaN, which includes the nonlinear gate-source capacitance and the Miller effect of the nonlinear gate-drain capacitance, affects both AM/AM and AM/PM. The impact on AM/AM is minor because the design frequency is less than the cut-off frequency of the input RC impedance. At higher frequency, the combination of the nonlinear input capacitance and the source matching can induce an amplitude expansion on the order of 1-2dB at higher input power level.
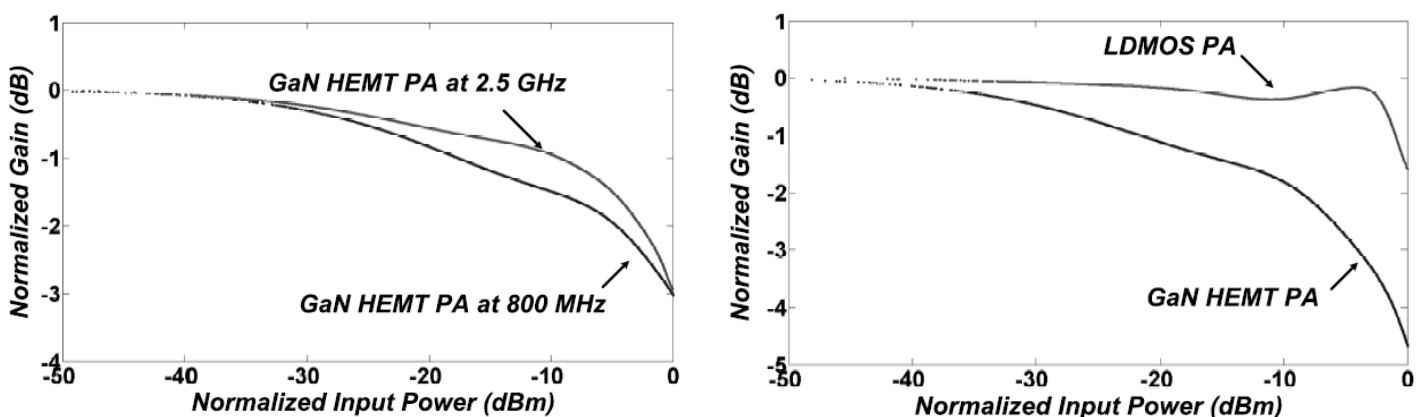
GaN HEMT PA - Soft Gain Compression vs Frequency, and comparison with LDMOS PA

- AM/PM is a key metric that provides a measure of **memory effect** in PAs.
  - For GaN Power Amplifiers, the highly nonlinear input capacitance can result in significant AM/PM distortion. Fortunately, with careful matching network design, it is possible to minimize the distortion.

- Another source of AM/PM results from the knee region interaction when the harmonic terminations are not short-circuited. This aspect of AM/PM is completely design dependent and can occur in both LDMOS and GaN PAs.
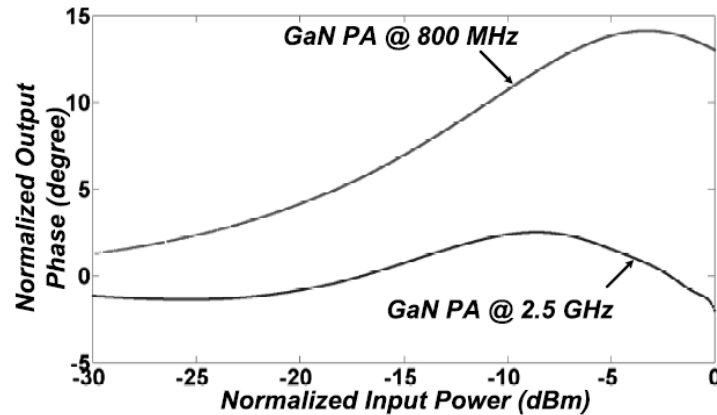
GaN Power Amplifier - AM/PM vs Frequency

Power Amplifier Design Issues

- Reflected power caused by a high VSWR condition between a PA and its load does not, in and of itself, cause a transistor's destruction or damage. Rather, a PA can be damaged or destroyed in a high VSWR environment simply because it is now looking at completely different load impedance than it was designed for.
- High device power dissipation can then produce elevated heating of the transistor and/or excessively high voltages.
- Some of the DC bias voltage would be wasted if the resistance in the chokes were too high. Also, the collector chokes must supply a very high impedance to the RF. If this impedance is not high enough, then some of the valuable RF output power generated by the PA will be wasted.
- Use low-ESR electrolytic capacitors at the PA's power supply, as this type of capacitor can immediately supply the needed current to the amplifier stage, and without pulling down the entire voltage supply during this critical transient turn-on time.
- When in saturation, a nonlinear PA's gain, PAE, and linearity are most affected by the reflections of its own harmonics back into its output port. These reflections are caused by the next stage, which will normally be a band-pass or low-pass filter, as well as an antenna.
- Instability in RF amplifiers can take the form of oscillations at almost any frequency, and may even damage or destroy the transistor. These spurious oscillations will arise at specific or very wide ranging, frequency or frequencies, and over a particular bias, drive level, temperature, or output load impedance.
- RF Power Amplifier oscillation problems can be broadly categorized into two kinds: Bias oscillations and RF oscillations.
- Bias oscillations occur at low frequencies, in the MHz to VHF range, and are caused by inappropriate and unintentional terminations at those frequencies by the bias insertion circuitry, such as the addition of a large-value decoupling capacitors.

- The oscillations have little to do with the details of the RF matching circuitry, where the RF blocking and decoupling capacitors become open circuit terminations at lower frequencies.
- RF oscillations, on the other hand, typically occur either in band or commonly out of band but still quite close to the desired bandwidth from the low frequency side.
- Decreasing the low-frequency gain of a Power Amplifier stage, which is naturally at an increased level, will assist in amplifier stability.
- The unavailability of a sufficient ground-plane, or a ground-plane that is excessively segmented, can create uncontrollable instability in a Power Amplifier.

RF Power Amplifiers for Wideband Modulations

RF Power Amplifiers for wideband modulations as CDMA or WCDMA, which operate in the linear region, are not very efficient. Only a portion of the D.C. current is used to generate the RF power; a much larger portion turns into heat.

LDMOS and GaN (Gallium Nitride) devices are best suited for the output and driver stages because of higher gain, improved linearity, and very low on-resistance. High gain reduces the number of stages needed in the amplifier to attain the same output power, compared to the old generation systems built with bipolar transistors.

In a multi-stage linear power amplifier there are various factors that need to be considered for choosing the right transistor for each of the stages of the amplifier.

- The pre-driver is biased Class-A for attaining consistent performance for minimal effect on the linearity of the device due to minor changes in bias supply. Drain efficiency is not as much of a concern for the pre-driver as it is for the latter stages in the amplifier. The driver and the output stage for such a system are typically biased Class AB, for achieving best tradeoff between linearity and efficiency of the amplifier.
- The most common method used to determine the linearity of a transistor is to characterize the Intermodulation Distortion (IMD) measured with two tones spaced. Typically, tone spacing up to 20 MHz should be used while tuning amplifiers for wide-band modulation applications.
- When transistors are used significantly backed-off from their peak power levels, it is all the more necessary that the IMD characteristics of the transistor at lower output power levels be taken into account. The profile of an IMD vs. Pout (drive-up) curve for a good transistor should have a large positive slope, even while attaining similar peak power capability, to get maximum Adjacent Channel Power Ratio (ACPR).
- The transistor used in the driver stage has similar linearity requirements as the output stage. In terms of ACPR, it needs be operated at an output power that gives a margin of at least 4 dB from the maximum allowed value for the output stage. In addition, it needs to have an input bandwidth about 2 to 2.5 times greater than the bandwidth of the modulating signal in order to maintain constant group delay and flat gain.
- One of the major factors determining the performance of the high-power transistor in the wide modulation environment is the gain flatness. The transistor needs to have a flat gain across the band for its use in multiple channel amplifiers. Very flat gain response greatly simplifies the design of linearization schemes systems.

- Fast roll-off of gain at the edges of the band causes deterioration in the ACPR performance.
- To attain the intrinsic device linearity, the 3dB bandwidth of bias networks needs to be at least two times greater than the modulation bandwidth.
- For attaining best ACPR response, it is necessary to have an excellent decoupling network at the Drain of the transistor, down to very low modulation frequencies. This can be achieved by using a high-quality shunt capacitor. This technique helps in achieving maximal gain flatness, which is very critical for wideband applications.
- It is advisable to avoid ferrite components in the biasing network and to use a series resistor on the Gate bias network to prevent instability. To achieve flat gain response across the band, the traditional inductive feed should be avoided on both the Gate and Drain. Instead, a quarter wave line at the frequency of interest, properly decoupled with a chip-capacitor, has been shown to provide very flat gain across the entire bandwidth.

References:

1. RF Circuit Design – C. Bowick
2. RF Power Amplifiers for Wireless Communications – S. Cripps
3. Advanced Techniques in RF Power Amplifier Design – S. Cripps
4. Distortion in RF Power Amplifiers – J. Vuolevi
5. Circuit Design for RF Transceivers – D. Leenaerts, J. Tang, C. Vaucher
6. Radio Frequency Transistors - N. Dye, H. Granberg
7. High Frequency Current Mode Class-D Amplifiers - A. L. Long
8. Complete Wireless Design - C. Sayre
9. Feedforward linear power amplifiers – N. Pothecary
10. Radio Frequency Integrated Circuit Design - Rogers, Plett
11. RF CMOS Power Amplifiers - Theory design and implementation - Hella, Ismail
12. Linearity of GaN HEMT RF Power Amplifiers - Sarbishaei, Wu, Boumaiza
13. Microwave Journal Magazine; 1996 – 2020
14. Portable Design Magazine; 2002 - 2005
15. High Frequency Electronics Magazine; 2002 – 2020
16. Applied Microwave & Wireless Magazine; 1989 – 2002

https://www.qsl.net/va3iul