Weigh Amplifier Dynamic-Range Requirements

Although signals for digital communications systems may seem simple, amplifier requirements for such signals can be quite complex and demanding of high linearity.

Daniel M. Dobkin, Director, Technical Marketing, e-mail: daniel.dobkin@wj.com,
Walter Strifler, Staff Scientist, Gleb Klimovitch, Senior Member of the Technical Staff
WJ Communications, Inc., 401 River Oaks Parkway, San Jose, CA 95134; (408) 577-6200, FAX: (408) 577-6620, Internet: www.wj.com.

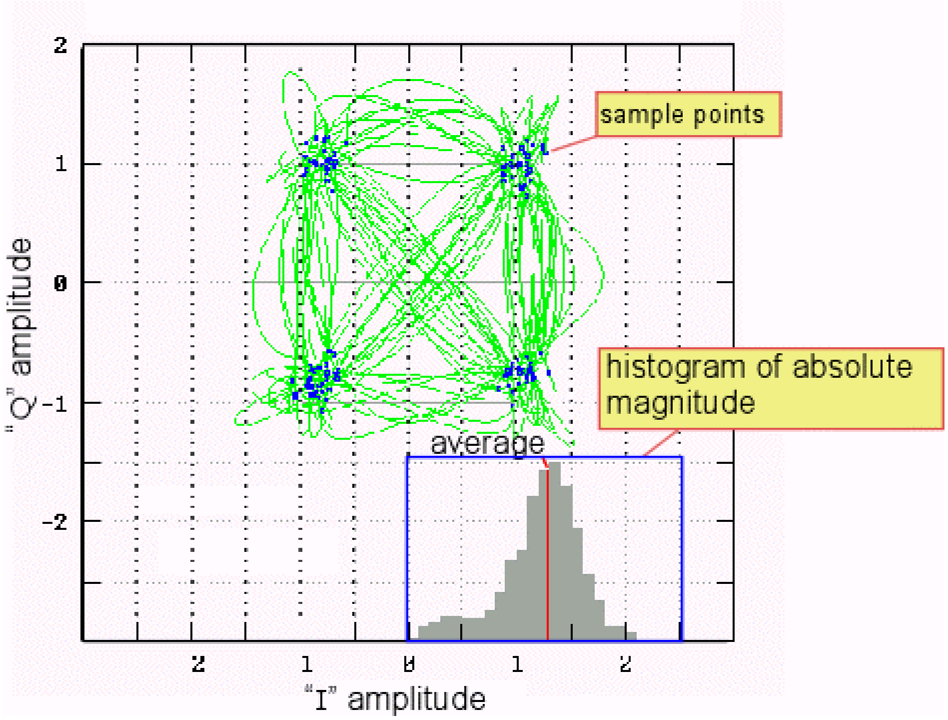
Amplifiers are often characterized in terms of gain, noise figure, and maximum output power when used in analog applications. For modern data communications systems, however, designers are often more concerned with nonlinearity and distortion levels. Such quantities as input or output third-order intercept points, spurious-free dynamic range, composite second-order (CSO) distortion, composite-triple-beat (CTB) distortion, and cross-modulation become critical specifications for digital communications systems.

The purpose of digital transmission is to move a sequence of digital ones and zeros from one location to the next, and the representation of intermediate states (the linearity of the system) would seem irrelevant. But linearity is critical to the success of a digital communications system, chiefly due to the limited frequency spectrum available to signals in wireless communications systems, and the complex characteristics of the communications channel that uses that spectrum. Any distortion is equivalent to changes in the Fourier transform of the signal. That is, unintended frequencies are radiated which may interfere with neighboring channels. Thus, radio designers must avoid operating at amplitudes at which such distortion is significant.

Due to the typically large ratio of the peak-to-average power levels of digital signals (the crest factor), the low-distortion transmission of digital signals can become quite complex. A large crest factor can lead not only to interference with adjacent channels, but to additional in-band distortion, and consequent increases in the bit error rate (BER).

SPECTRAL EFFICIENCY

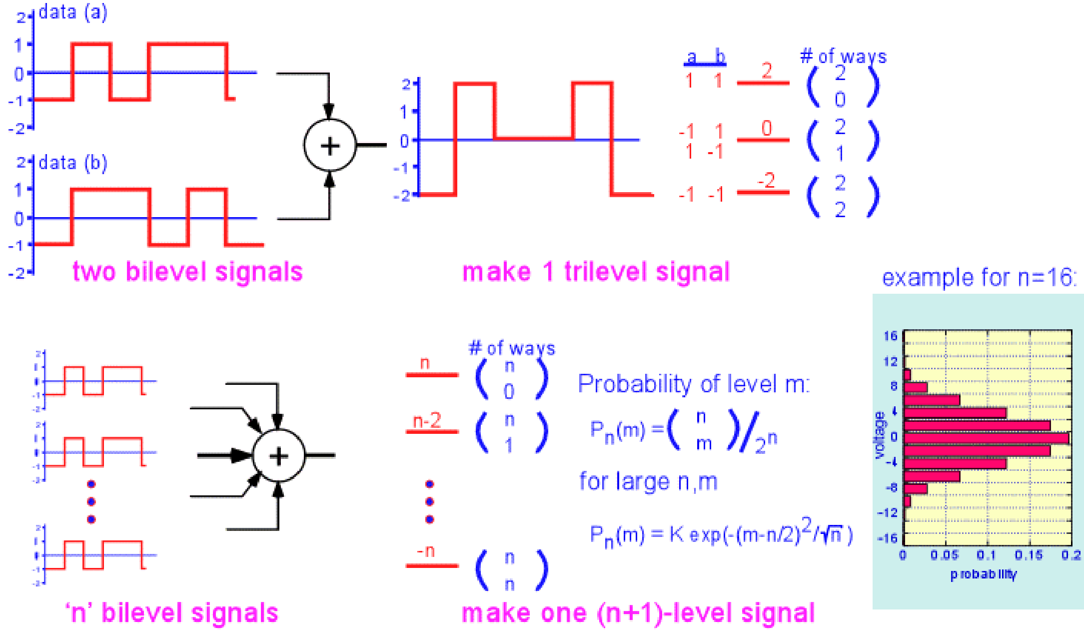
To make efficient use of available spectrum, wireless communications channels filter digital data to smooth the transitions between bits as much as possible while still maintaining the integrity of the signal at the “sample points.”[1] Modulation schemes, which transmit more than one bit per symbol, are used to maximize the data rate for a given slice of spectrum at a given signal-to-noise ratio (SNR).[2] The net result is that the actual analog signal transmitted is considerably more complex than the binary sequence it represents. **Figure 1** offers an example, a quadrature-phase-shift-keying (QPSK) signal. Shown is the modeled path of a filtered QPSK signal, composed of two pseudo-random bitstreams, represented in the phase/amplitude plane. Sample points (i.e. the location of the signal in the I/Q plane at the sample times) are also shown, as well as a histogram showing relative probability of various values of signal amplitude.



1. This plot shows a simulated phase/amplitude path and amplitude distribution for a random digital bitstream with QPSK modulation.

The histogram shows that the average value of the signal amplitude is approximately 1.3 (in normalized units), whereas rare excursions to much larger amplitudes also occur. The ratio of the peak power to the average power for the QPSK signal is approximately 4.3 dB (a factor of 2.7) where the peak is taken from a sample of 256 trajectories (sample to sample), and thus, is at a probability level of roughly 10^{-3}.

In addition to the challenges of filtered complex constellation paths, real signals often transmit more than one channel simultaneously. This inevitably leads to peak levels that are sometimes much higher than the average signal power. **Figure 2** shows this schematically: the addition of multiple uncorrelated bitstreams produces a final signal which can have many possible levels.



2. The sum of a large number of binary signals gives rise to a normal distribution.

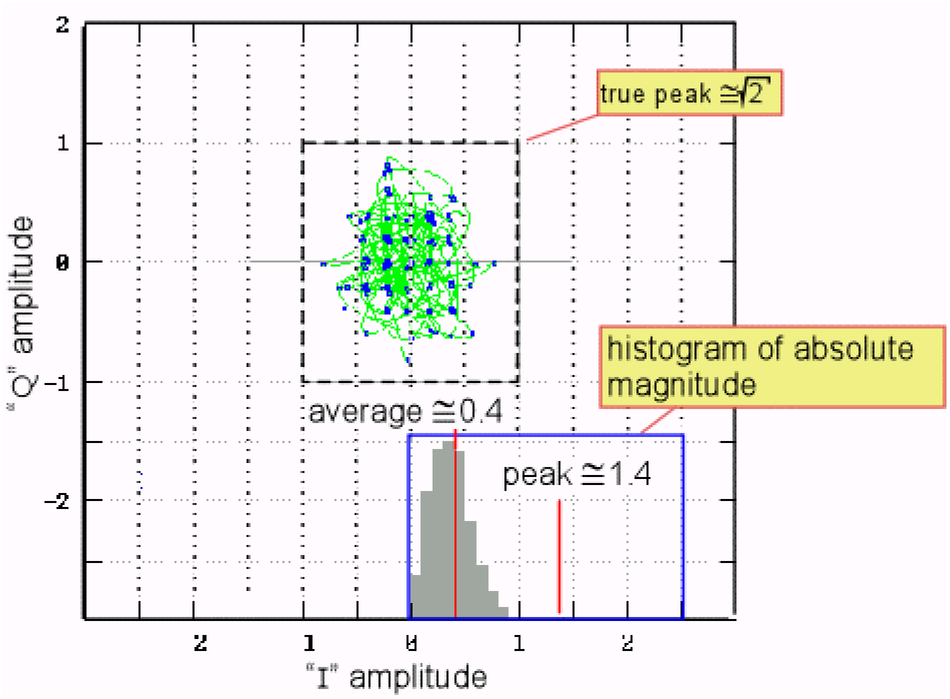
The probability of each "voltage" level is proportional to the number of ways in which that voltage can result. For example, in the case in which two bilevel signals are combined, there is only one way to make a level of +2 and only one way to make a level of –2, but there are two ways to make a level of 0. Thus, a zero is twice as likely to occur as either of the extreme cases. In the general case, if equal likelihoods of 1 or -1 in the incoming data signals are assumed (see equation), the probability of obtaining a level "m" from n bilevel signals is described by the binomial coefficient:

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!} \tag{1}$$

As the number (n) of signals grows large, the distribution of voltages closely approximates a normal or "Gaussian" distribution, with standard deviation of $\sqrt{N}/2$. The distribution of signal power is a chi-squared distribution of order 1 for the case where each signal is either on or off, or of order 2 in the case where there are two orthogonal components, the in-phase (I) and quadrature (Q) signals, combining to form the final signal. The peak-to-average power ratios of the order-1 and order-2 signals are 12.8 and 13.3 dB, respectively, at a probability of 10^{-5}.

A practical example of the superposition of multiple uncorrelated data streams to produce a complex analog signal is encountered in code-division-multiple-access (CDMA) schemes used for mobile communications. In this approach, signals can be sent to many users at the same time using the same frequency without significant interference. Each user's binary bit stream is multiplied by a code consisting of a sequence of very short "chips." The resulting signal has a higher effective bit rate and is thus spread out in frequency, but if codes are chosen to be orthogonal to each other or nearly so, multiple signals can be sent using the same frequency. Each user multiplies the total signal by their individual code, thereby extracting only his or her data stream. These multiple streams are added together and sent by the base station (downstream transmission); users send simultaneous transmissions of which the sum must be received by the base station (upstream transmission). Thus, the base station must handle a signal composed of a large number of uncorrelated separate data streams (Fig. 2).

Figure 3 shows a simulated signal resulting from summing 10 independent QPSK datastreams, as would be encountered in a CDMA basestation (downstream) transmission. Note that the signal spends almost all its time near the center of the phase plane at small amplitudes, but at rare intervals an extreme value occurs. In this simulation, the peak-to-average ratio is about 6.8 dB (a factor of 5), again for a sample of 256 trajectories, so there is a probability of approximately 10^{-3}. Note that this value is a lower bound on the peak-to-average ratio that would be obtained with a larger sample of points. (A real signal would go a bit farther, as noted before, due to filtering.) For a more realistic symbol error requirement of around 10^{-5}, the peak-to-average ratio would approximately be 11 dB, already close to the limiting values corresponding to a Gaussian distribution in each axis.



3. This plot shows a simulated phase/amplitude path and amplitude distribution for 10 superimposed QPSK-modulated signal streams. Note that in this limited sample of 250 data points, the outer points of the possible constellation are never accessed.

Orthogonal-frequency-division-multiplexing (OFDM) modulation as used in fixed-wireless communications systems and in proposed wireless local-area-network (WLAN) standards also combines a number of uncorrelated data streams, but at distinct carrier frequencies. The resulting signal displays rare excursions to amplitudes much larger than the average, and thus a large peak-to-average power ratio. Cable television (CATV) systems combine simultaneous signals from as many as 110 separate carriers, often mixing digital and conventional NTSC analog signals. CATV signals are also characterized by peak-to-average power ratios of approximately 11 to 13 dB.
(CATV systems also span multiple octaves, so that many distortion products are in-band, and represent a significant linearity challenge for the designer.)

The ratio of the power in the adjacent channel to that in the intended channel--the adjacent channel power ratio (ACPR)--is often a key specification that wireless communications systems must meet to avoid interference between users in different channels. Since distortion is highly dependent on signal amplitude, the ACPR expected with a given average signal power depends strongly on the type of signal being sent.

It is important to note that while the ratio of in-band distortion to input signal goes as the square of the input power, the ratio of adjacent-channel distortion to adjacent-channel power goes as the cube of the input power. The in-band channel power increases when the distortion increases, but the adjacent channel is uncorrelated with the channel generating the distortion, and, thus, might operate at low power when the interference becomes high, degrading the signal-to-noise ratio (SNR) and increasing the likelihood of bit errors. Thus, it is often the case that even though the in-band distortion is larger than the adjacent-channel distortion, it is the ACPR that causes problems with meeting performance specifications.

SOURCES OF DISTORTION

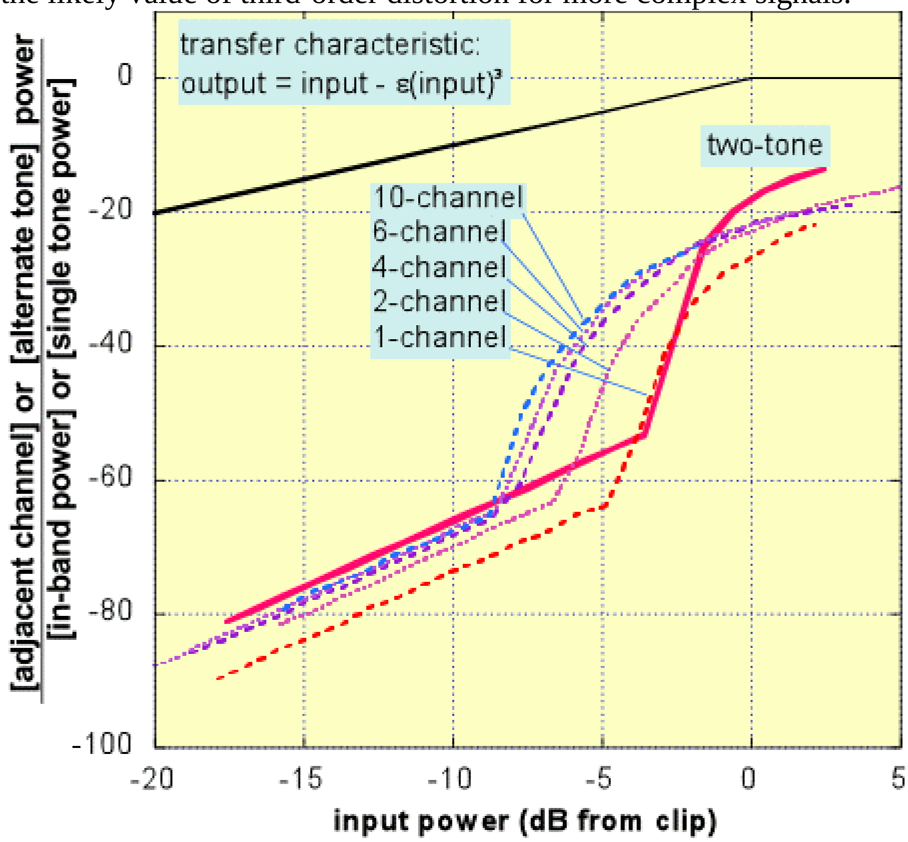
Third-order distortion is usually the most important small-signal distortion in amplifiers, since it generates distortion products within and close to the band of interest. Third-order distortion is usually characterized by the (output) power at the third-order intercept point (OIP3), defined as that power where one of the two spurious products generated from the mixing of two nearby frequencies (tones) is equal in amplitude to one of the tones. This point is extrapolated from data acquired at low signal power levels. Spurious power resulting from third-order distortion varies as the cube of the signal power. Thus, the ACPR resulting from third-order distortion varies as the square of the power, resulting in a slope of 2 on a logarithmic plot.

Any real amplifier can only supply a finite output voltage. For input signal levels beyond its capacity, the output becomes “clipped” or distorted. For multi-channel digital signals with large peak-to-average power ratios, clipping will act first on the rare excursions to high amplitude. In the limit where the distribution of signal power is nearly Gaussian, the likelihood of a clipping event will be described by a complementary error function and, thus, the total power generated by clipping distortion will vary exponentially with input power, falling rapidly when the signal power is backed off from the clipping power by more than the peak-to-average power ratio.

As the signal becomes strongly clipped, the signal will become a square wave with a sin x/x spectrum. Thus, the adjacent-channel power will converge to a fixed value. For a Gaussian signal, the ACPR is about -12 dBc at input powers well beyond the clipping threshold.

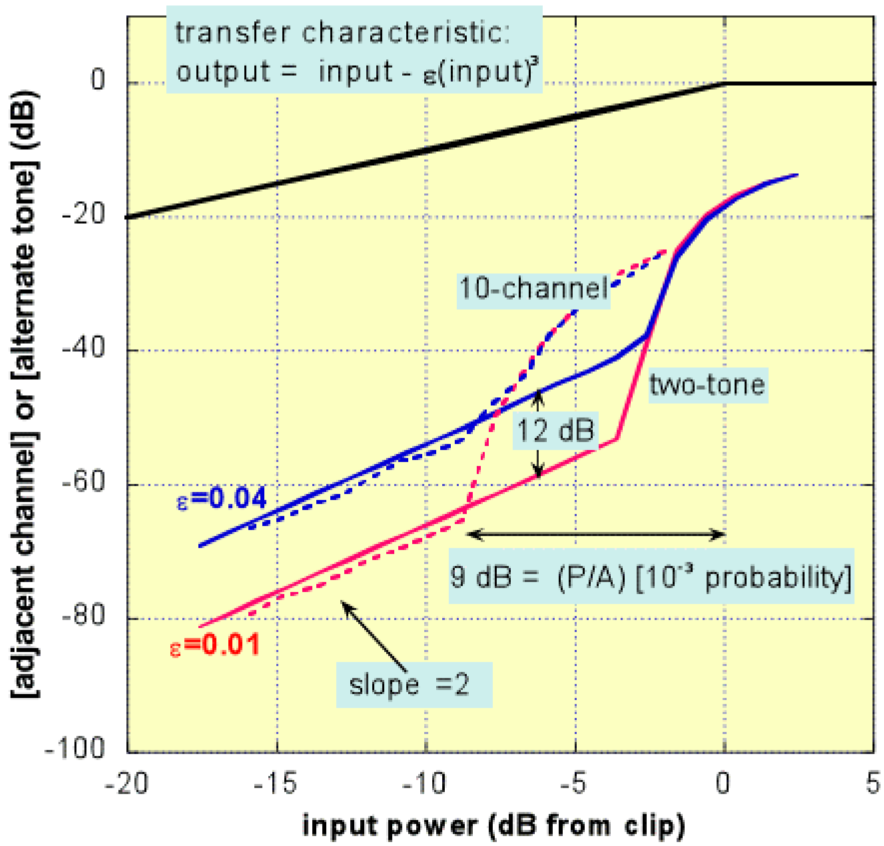
Figure 4 shows the modeled adjacent-channel power for the same 10-channel QPSK signal in Fig. 3, using a simplified transfer curve including third-order distortion and hard clipping at a relative power of 0-dB input. Figure 4 demonstrates how the level of clipping distortion and third-order spurious power, at an average signal power, are strongly influenced by the type of signal employed. It can be seen that as the number of channels superimposed in the signal increases, the distortion behavior rapidly converges:

four channels produce nearly the same peak-to-average ratio and distortion results as 10 channels. Note that the third-order distortion from a signal with high peak-average ratio is nearly equal to the two-tone third-order intermodulation, as predicted in ref. 3. This fortunate circumstance explains why a simple two-tone measurement is a useful guide to the likely value of third-order distortion for more complex signals.



4. Adjacent-channel power ratio (ACPR) is shown as a function of relative input power, for signals constructed from a varying number of QPSK input channels (the dashed lines), and a piecewise-cubic transfer characteristic. The two-tone third-order intermodulation product relative to the tone amplitude (the solid line) is included for comparison.

Figure 5 shows the impact of varying the third-order distortion for a fixed input signal type. For input powers closer to the clipping power than the peak-to-average power ratio, distortion is dominated by clipping and third-order distortion behavior has no effect on the ACPR. However, as one input signal is "backed off" sufficiently, the third-order distortion takes over and sets a limit on the achievable ACPR.



5. ACPR is shown in Fig. 4. The solid lines represent the two-tone signals, the dashed lines represent 10-channel QPSK signals with (blue) high cubic distortion or (red) low cubic distortion.

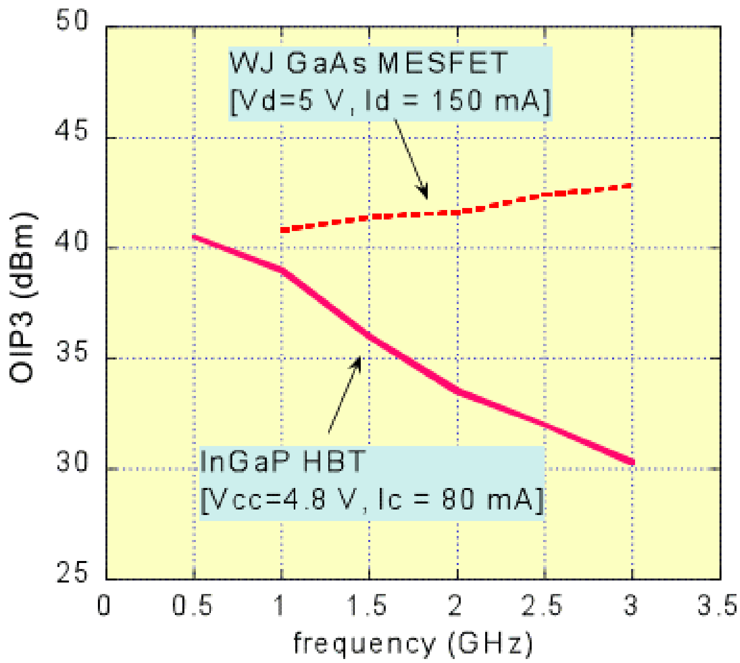
Note that these simplified models do not take higher-order curvature of the transfer characteristic, or the detailed "shape" of the saturation behavior into account. Actual device behavior will differ from the predictions of the simple models, particularly at the intersection between the third-order and clipped characteristics, where higher-order terms can play a role and changes in the relative phase of the different contributions can cause fluctuations of several decibels in ACPR with modest changes in input power.

To fabricate amplifiers with good linear efficiency, component designers can employ processes and devices designed specifically for enhanced dynamic range, or seek circuit topologies which minimize the deleterious effects of distortion, or both. For example, at WJ Communications (San Jose, CA), the company's GaAs MESFETs have been optimized for low third-order distortion. This is accomplished by careful adjustment of the channel doping and geometry of the gate recess, ensuring that signal-dependent variations in device transconductance are almost perfectly nullified by the signal dependence of the device output conductance. Most of the company's current high-dynamic range MESFETs are optimized for operation at zero gate bias ($I_{ds} = I_{dss}$), allowing operation from a single-voltage supply.

An advantage of MESFET technology is that the dominant nonlinear device elements are conductances, determined by doping concentration and electron mobility. Thus, the distortion behavior of MESFET amplifiers is relatively insensitive to variations in operating frequency and ambient temperature. But, MESFET nonlinear modeling is poorly understood compared to the nonlinear behavior of bipolar junction transistors (BJTs).

For high gain in limited semiconductor area, it is possible to use (BJTs) instead of FETs. The transconductance of a BJT is approximately proportional to the collector current: $I_c/(kT/q) = I_c/40$ at room temperature. For reasonable current densities, bipolar transistors can provide much higher transconductance per unit chip area than comparable MESFETs. The high gain supports the use of copious amounts of negative feedback while still preserving acceptable overall amplifier gain, achieving low third-order distortion and good dynamic range. A Darlington configuration provides a low-impedance source (the first transistor) to drive the voltage gain stage (the emitter follower), decreasing sensitivity to parasitics, particularly the Miller capacitance. Heterojunction bipolar transistors (HBTs), with their heavily-doped base regions, also exhibit very low output conductance, making the output matching design simpler than for a conventional bipolar transistor circuit.

The feedback capacitance of a BJT is much larger than the corresponding capacitance in a MESFET, so the Miller effect magnifies the input capacitance in the second stage of a Darlington pair to produce significant gain roll-off. The presence of a significant nonlinear capacitance and the strong frequency dependence of the intrinsic gain cause the distortion behavior of BJT circuits to be more frequency dependent than MESFETs **(Fig. 6)**. Bipolar circuits require a resistor in the collector path, adding to power consumption. Finally, MESFETs tend to be more robust than BJTs when operated at high channel temperatures. The FET channel current decreases with increasing channel temperature due to reduced electron mobility, whereas BJT collector current increases with increasing temperature due to reduced electron mobility, whereas BJT collector current increases with increasing temperature due to increased injection from the emitter, requiring careful design for good thermal stability.



6. The third-order intercept point is plotted versus operating frequency for generally similar GaAs HBT and MESFET amplifiers.

References

1. H. Nyquist, “Certain Topics in Telegraph Transmission Theory”, Transactions of the AIEE, Vol. 47, p. 617, April 1928.
2. E. Wesel, *Wireless Multimedia Communications*, Addison-Wesley, Boston, 1998, Chap. 4.
3. J. Pedro and N. de Carvalho, “On the Use of Multitone Techniques for Assessing RF Component’s Intermodulation Distortion”, IEEE Transactions on Microwave Theory & Techniques, Vol. MTT-47, 1999, p.2393 (especially Fig. 5 therein).