Effect of nonlinear amplification on turbo coding gain

Eugene Grayver and P. E. Santacruz
The Aerospace Corporation
El Segundo, CA
310-336-1274
eugene.grayver@aero.org

Abstract—This paper[1,2] presents results obtained from an end-to-end, proof-of-concept system that integrates a multi-constellation modem, turbo coding, and a nonlinear traveling wave tube amplifier (TWTA). Multilevel modulation schemes make it possible to provide high-speed data communications in a limited amount of spectrum. This improvement, however, comes at the price of increased linearity requirements for the end-to-end link. This constraint is especially important for the power amplifier, which is typically a nonlinear device. Forward error correction based on turbo codes provides a solution for the problems described above, and improves the BER by increasing the noise margin by over 5 dB. This paper presents measured BER curves for different modulations and turbo codes, taken at different power levels relative to saturation. The absolute coding gain is shown to be dependent on the nonlinearity. Limitations in the standard definition of the coding gain are identified and a new metric is proposed.

TABLE OF CONTENTS

I. INTRODUCTION

Future satellite communication systems are required to support higher data rates while operating under the same spectral constraints as the lower data rate versions. These systems have proposed the use of higher-order modulations combined with pulse shaping (such as square-root raised cosine filtering) to achieve higher throughput without increasing the signal bandwidth. For example, data throughput can be increased fourfold by using a four-level modulation such as 16-QAM instead of a two-level modulation such as BPSK. Unfortunately, such modulations result in a large amplitude modulation (AM), which leads to AM/AM or AM/Phase Modulation (PM) distortions when passing through a nonlinear amplifier operated in the more power-efficient saturation region. These distortions cause significant degradation to the communication bit error rate (BER) unless the amplifier is backed off from the nonlinear region, or the effect of the nonlinearity is somehow compensated [1]. Unfortunately, back-off often leads to an unacceptable decrease in power efficiency of the power amplifier.

Error correction coding has been shown to be a very effective, and in fact essential technique for a wide range of communications systems. Since their invention in the 1990s, turbo decoders have been responsible for a revolution in wireless communications [2]. In particular, turbo coding techniques have recently been used for commercial satellites. Turbo codes rely on an iterative decoding of soft decisions combined with additional 'parity' bits to detect and correct errors in the received symbol stream. The iterative decoding is very computationally intensive and has become practical only due to advances in semiconductor technology. The additional noise margin provided by the coding gain allows error-free operation for nonlinearly amplified high-order modulations.

This paper reports measurement results obtained from a high-fidelity model of a satellite downlink channel, including a nonlinear amplifier and turbo coding. The main goal is to quantify the coding gain provided by the turbo codec as it is related to the amplifier operating point. We show that the conventional definition of coding gain is not applicable to the nonlinear channel created by the amplifier, and introduce an alternate definition.

The combination of different modulations, variable coding rates, and amplifier back-off points results in a very large system trade space. Computationally intensive operations such as turbo decoding and pulse shaping make system-level simulations in software impractically slow. The target BER for the system described here is below 10^{-10}. Thus, at least 10^{12} bits must be processed to achieve a statistically significant BER measurement. A software simulation would take weeks or months to compute a single point in the trade space. The problem is exacerbated by the need to model the nonlinear amplifier. The nonlinear behavior can be captured only by oversampling the signal, further adding to the computational requirements. All these considerations motivate the need for a real-time, hardware-based emulation system.

A system consisting of a digital modem, radio frequency up/downconverters, a nonlinear amplifier, and an AWGN channel is presented in this paper. This proof-of-concept

1
[1] 0-7803-9546-8/06/$20.00© 2006 IEEE
[2] IEEEAC paper #1001, Version 5, Updated Sept. 9, 2005

platform was developed as a risk-reduction stage in the development of the transmitter and receiver subsystems for the next-generation Geostationary Operational Environmental Satellite R-series (GOES-R) weather satellites [6].

The nonlinear amplifier is described in section II, followed by a detailed description of the test platform. Section II describes the digital modem, while section III addresses the turbo codec used by the modem. Measurement results are presented in section IV, and the paper summary is provided in section V.

II. TWTA

The nonlinear high-power amplifier for the testbed is a space-qualified traveling wave tube amplifier (TWTA) from Boeing. This device is rated for an output power of 150 W (52 dBm) at L band. Any amplifier exhibits nonlinear behavior when operating close to its maximum output power level (saturation). The nonlinearity can be characterized into two types:

1. AM/AM: An incremental change in input signal power does not correspond to a linear incremental change in the output power.
2. AM/PM: An incremental change in input signal power results in a *phase* change of the output signal.

Operating as close to saturation as possible maximizes the output power and thereby maximizes overall amplifier efficiency. However, these nonlinear effects become more severe closer to saturation and eventually degrade the performance such that the link can no longer be maintained. The AM/AM and AM/PM responses of the TWTA were measured and are presented in Figure 1. The abscissa units are defined as "output power decibels away from saturation" – dBsat.

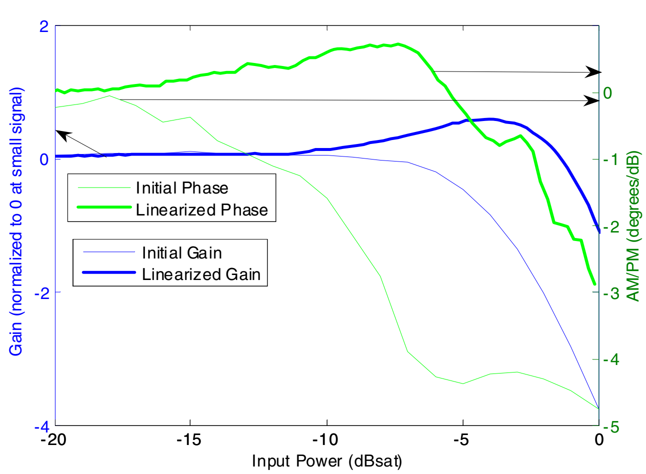

Figure 1. AM/AM and AM/PM characteristics of the TWTA (in dB)

The nonlinear behavior of the device can be partially mitigated by preprocessing the input signal through a linearizer. The linearizer is tuned to compensate for the particular device characteristics. The curves in Figure 1 show the raw performance of the TWTA, and the improved performance when the TWTA is preceded by a linearizer. Linearization can be performed at baseband in the digital domain, at baseband in the analog domain, or at RF. The system described in this document uses a commercial linearizer operating at RF (Figure 2).

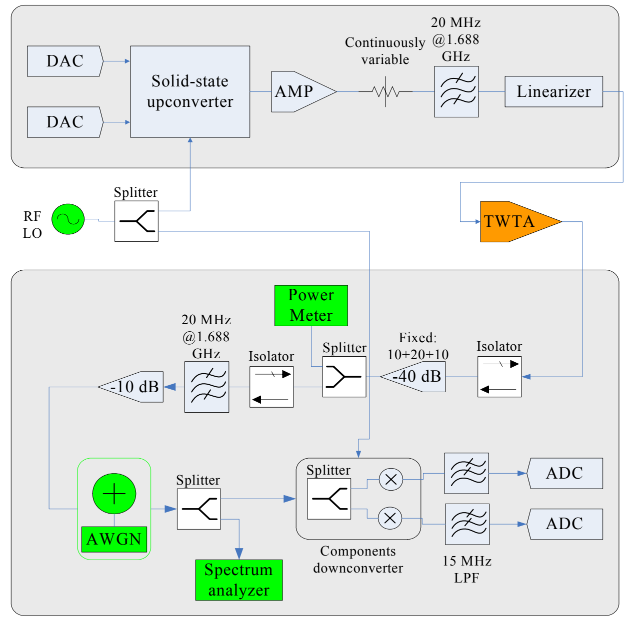

Figure 2. RF and mixed signal TWTA testbed

The effect of these nonlinearities can be observed in both time and frequency domains. Signals with variable envelope experience more corruption than do constant-envelope signals. For instance, a 16-QAM modulated signal is corrupted more than an 8-PSK signal. A multilevel constellation is impacted in two ways, as shown in Figure 3: 1) outer points are amplified less than the inner points and therefore appear compressed into the origin and 2) inner points are rotated relative to the outer points.

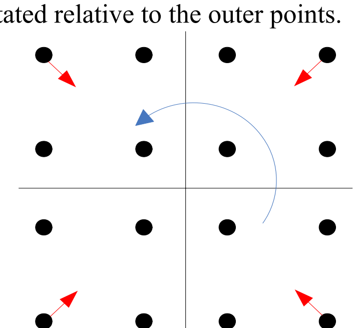

Figure 3. Schematic representation of the AM/AM and AM/PM effect on a 16-QAM constellation

Two representative 16-QAM constellation plots measured using the TWTA setup described above are presented in Figure 4a,b. Figure 4c shows that a constant-envelope constellation such as 8-PSK is less distorted than a 16-QAM constellation, even when operating closer to saturation. In addition, nonlinear distortion changes the pulse shape, making the receiver matched filter no longer 'matched' and thereby introducing intersymbol interference.

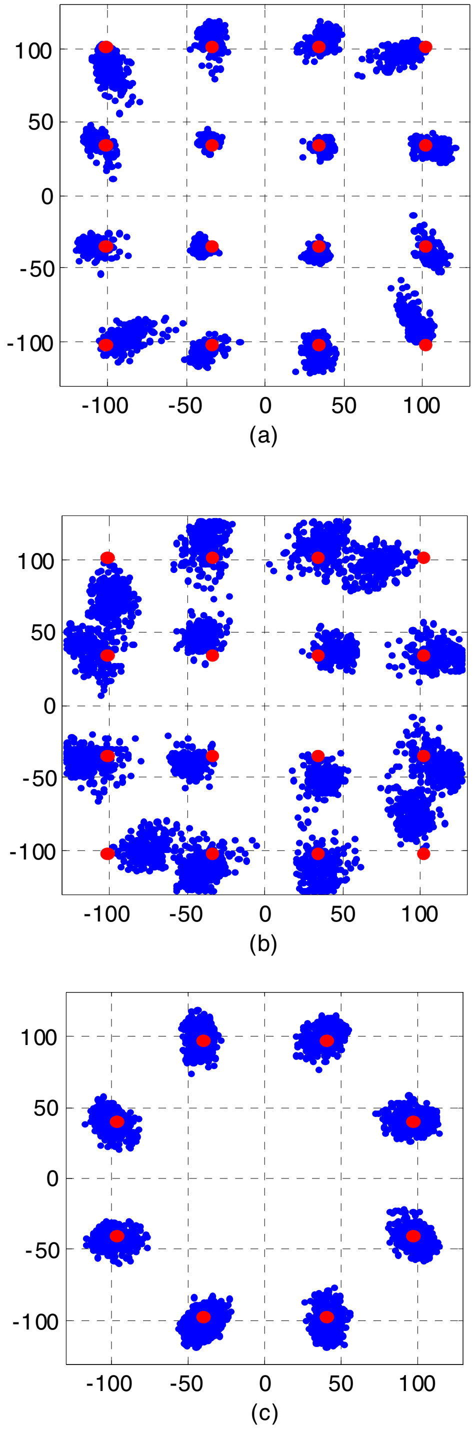

Figure 4. Measured constellations: (a) 16-QAM at –2 dBsat, (b) 16-QAM at –1.5 dBsat, (c) 8-PSK at –0.8 dBsat

In the frequency domain, the TWTA nonlinearities result in spectral regrowth. The tightly constrained spectrum of the input signal develops 'skirts.' This effect is highly undesirable since it increases the spectral occupancy of the transmitted signal and can add interference to adjacent channels. Spectrum distortion does not have a direct impact on the bit error rate, and will not be considered in this paper.

III. PROGRAMMABLE MODEM

The digital and mixed signal subsection of the TWTA testbed is provided by a flexible multi-modulation transceiver. This highly programmable design is described in detail in [7]. A summary of its capabilities is given in Table 1.

Table 1. Programmable transceiver capabilities

Category	Capabilities
Modulations	BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, 64-QAM Arbitrary constellation, up to 64-ary ½T offset option FSK, 4-FSK, 8-FSK with programmable *h,* MSK, GMSK
Symbol rate	≤10 MSPS for linear modulations ≤5 MSPS for FSK-type modulations
Data rate	≤20 Mbps using QPSK, ≤60 Mbps using 64-QAM, ≤5 Mbps using FSK
Pulse shape	Square-root raised cosine (α=0.35) Programmable 212-tap filter Gaussian filter for GMSK
Digital freq synthesis	SFDR > 60 dBc 0.1 Hz resolution
Coding	Turbo Product Codes (up to 16 kb) Code rates from 0.25 to 0.99
Implementation loss (digital)	≤0.4 dB for linear modulations ≤ 2 dB for FSK-type modulations
Mixed signal	DAC: 16-bit, 400 MSPS; ADC: 12-bit, 125 MSPS

The implementation loss of the transceiver was carefully measured and characterized to isolate individual noise contributors. The system performance was first measured entirely in the digital domain using a digital AWGN source built in to the transmitter. Another set of tests was run to characterize the loss due to the timing and carrier tracking loops. The implementation loss does not exceed 0.3 dB even at low BER. The mixed signal components introduce no measurable loss. The transceiver was then integrated with a complete RF link. The RF implementation, shown in Figure 2, uses a microwave components-based upconverter and downconverter. A COTS programmable noise source is used instead of the built-in digital source to eliminate any interaction between the noise and the upconverter. Performance of the complete system in RF loopback mode including timing and frequency tracking loops is shown in Figure 5. The RF subsystem has not been fully optimized and introduces up to 0.6 dB of implementation loss at low BER (10^{-9}).

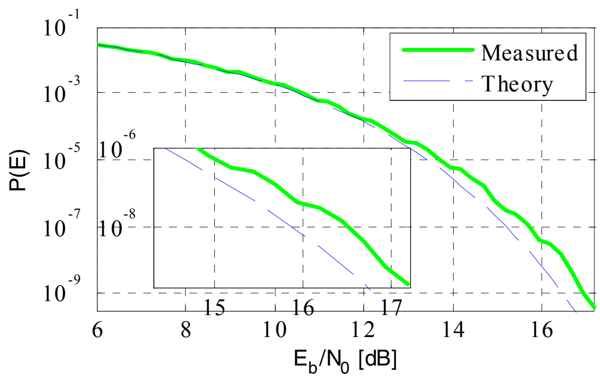

Figure 5. Transceiver calibration with RF and external noise source using 16-QAM

Turbo coding allows the system to achieve low BER even at low E_b/N_0 operation. However, the carrier, symbol, and frame tracking loops operate on the uncoded data and do not receive any benefit from the turbo decoder. The receiver was shown to provide stable and reliable tracking down to low SNR, well below the minimum to achieve the desired coded BER.

IV. TURBO CODING AND CODING GAIN

The transceiver described in this report is coupled with a commercially available forward error correction (FEC) turbo codec ASIC from AHA [8]. Error correction encoding is the addition of redundancy, e.g., parity-check symbols, to a message that is to be transmitted over a medium (the communication channel). This redundancy allows the error correction decoder to detect and/or correct erroneous data and restore the received data stream to the original data stream.

The turbo product codes (TPCs) implemented in the AHA ASIC use Extended Hamming codes (or simple parity codes) in a two-dimensional operation as follows. Encoding is performed by placing the data in a k×k array. Each row and column is then encoded with the appropriate Extended Hamming code and the parity bits are appended to the end of each row. After all rows are encoded, the columns are encoded in the same manner resulting in an n×n coded array.

This coding technique allows a very wide range of coding rates and a large number of possible code implementations. In general, the lower the coding rate (i.e., the larger the number of parity bits relative to the data bits), the greater the coding gain. Two high-rate codes are considered in this paper:

- r=0.878: (128,120) × (128,120),
- r=0.954: (64,63) × (64,62).

Coding gain is a measure frequently used to quantify the performance of a given code, and is defined by the difference in minimum E_b/N_0 required to achieve the same BER with and without the code. Coding gain is always a function of the target BER. This definition is illustrated in Figure 6 by an arrow labeled ‘a’, and will be referred to as the *horizontal* coding gain. However, this measure is undefined if one of the curves exhibits an error floor, as shown by the uncoded, nonlinear curve in Figure 6. An alternative definition of the coding gain is given by the mapping from the uncoded BER to the coded BER, as shown in Figure 6 by an arrow labeled ‘b’. This alternate definition, referred to as the *vertical* coding gain, will be shown to be more applicable across different amplifier operating points.

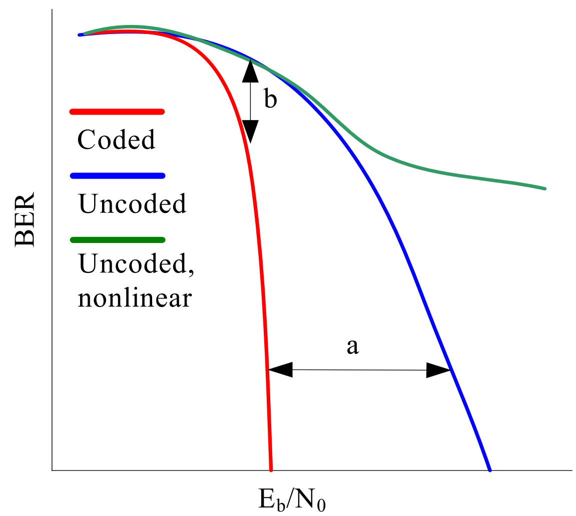

Figure 6. Two definitions of coding gain

The large slope in the coded BER curve for E_b/N_0 above the ‘cliff’ point is a feature common to all turbo codes.[3] An intuitive explanation for this behavior is based on the error correction capability of the code. The iterative processing can eliminate almost all errors as long as the initial starting point is within the ‘convergence range.’ Thus, once the uncoded BER falls below a maximum value, the turbo code ‘kicks in’ and dramatically reduces the coded BER. This argument would seem to indicate that the uncoded vs. coded BER curves for a given code should be independent of any other variables. One of the interesting results of this paper shows that it is not the case.

Figure 7 shows measured BER curves for different modulations in a pure AWGN channel. (In this, and all subsequent figures, the dashed lines correspond to uncoded BER, while the solid lines correspond to coded BER.) Linear channel is implemented via digital loopback from the transmitter to the receiver. Curves measured in the linear/loopback mode are listed as ‘loopback’ in figure 4.

[3] The E_b in a coded system can be defined as either the energy per uncoded (information) bit or per coded (channel) bit. The second definition is always used in this paper.

legends.) As expected, the uncoded curves do not exhibit an error floor and both coding gain metrics can be computed. The horizontal coding gain is shown in Figure 8. The minor variation can be attributed to measurement uncertainty and errors introduced by the different phase and timing tracking algorithms. The vertical coding gain is shown in Figure 9. This figure shows a distinct grouping of the two-level (B/QPSK) modulations and multilevel modulations. The turbo decoder used in this study can apparently tolerate 3 times the error rate for two-level modulations as for multilevel modulations. This behavior is still being investigated, and is currently attributed to the computation of LLR (Log-Likelihood Ratio) values implemented in the AHA chip [8].

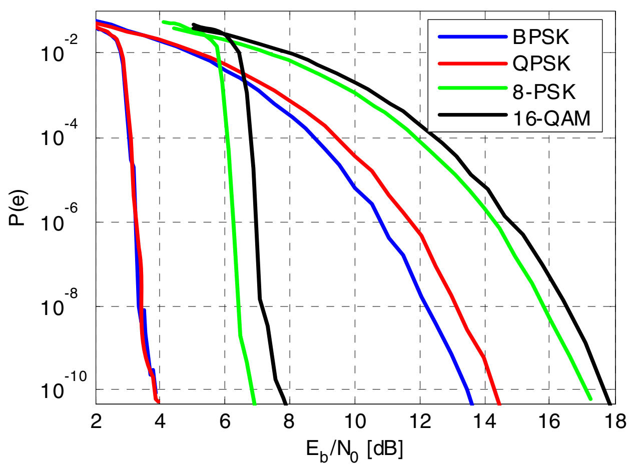

Figure 7. Coded and uncoded BER w/ linear channel

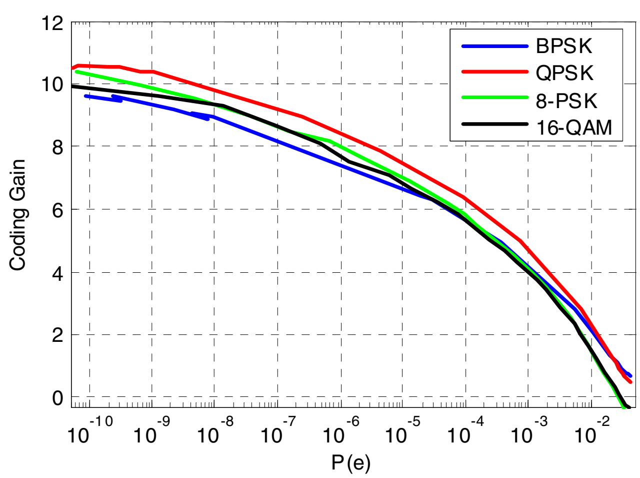

Figure 8. Horizontal coding gain w/ linear channel

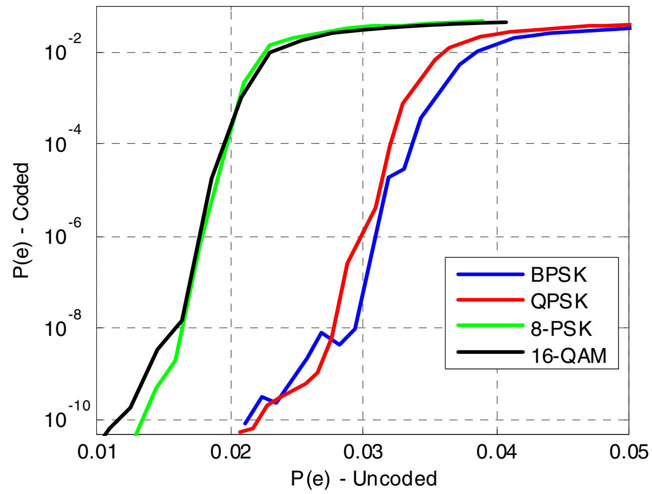

Figure 9. Vertical coding gain w/ linear channel

V. HIGH-ORDER MODULATION AND TWTA

The nonlinear behavior of the TWTA significantly degrades the BER performance of high-order modulations due to effects discussed above. The nonlinear distortion can be thought of as an additional (non-Gaussian) noise source, on top of the AWGN. Thus, as distortion increases, the maximum AWGN level to maintain a given BER decreases. This effect is demonstrated in Figure 10. As can be seen from that figure, a BER of 10^{-6} cannot be maintained even in an otherwise noise-free link when the output power back-off (dBsat) is less than 3 dB. The standard square 16-QAM is not an optimal constellation for operation close to saturation. Mappings such as 12-4-APSK and other, more circular configurations allow operation closer to saturation, but are beyond the scope of this paper [4]. The constant-envelope 8-PSK constellation is much more robust to nonlinear effects, as is evidenced by the results in Figure 11. As stated previously, the goal of this paper is analysis of the interaction between the nonlinear channel and turbo coding.

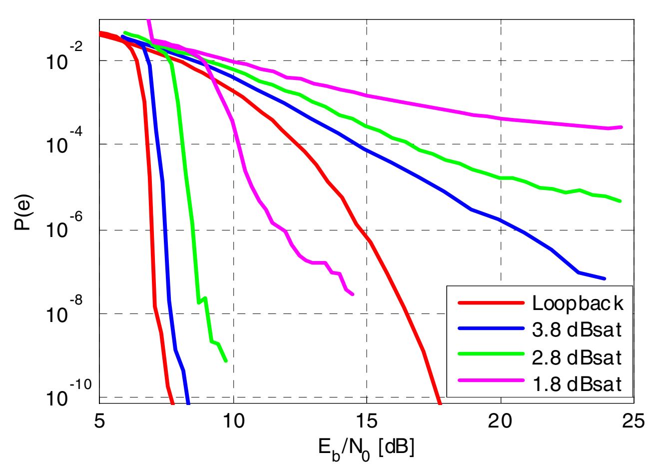

Figure 10. 16-QAM coded and uncoded BER w/ TWTA

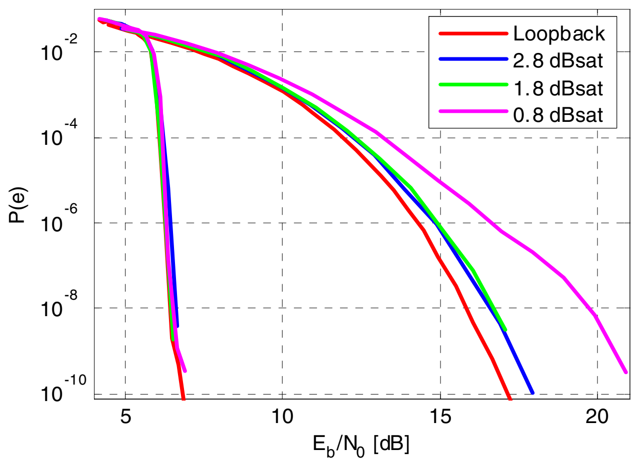

Figure 11. 8-PSK coded and uncoded BER w/ TWTA

The two coding gain measures can be computed based on the BER data for different amplifier operating points. As expected, the horizontal coding gain becomes undefined for BER values below the uncoded error floor. The horizontal coding gain for 16-QAM is shown in Figure 12.

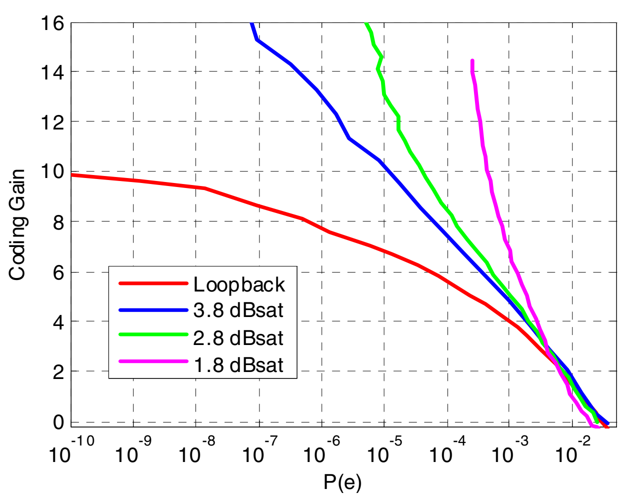

Figure 12. 16-QAM horizontal coding gain w/ TWTA

Consider the curve for 2.8 dBsat operating point. Its slope approaches infinity for BER < 10^{-5}, and therefore conveys no information if the target coded BER is 10^{-10}. It is important to notice that using the gain curve for the linear channel does not allow us to estimate the gain for the nonlinear channel (i.e., to estimate the coded BER based on the uncoded BER). For example, a naïve approach (but one still encountered in the literature) would take the coding gain of 8.2 dB at a BER of 10^{-7} and apply it to the uncoded curve at 3.8 dBsat in Figure 10 (dashed blue line). The coded BER of 10^{-7} would then be expected at E_b/N_0=23–8.2=14.8 dB. However, the measurements shown in Figure 10 demonstrate that this BER is achieved at a much lower E_b/N_0=7.2 dB. The failure of the naïve approach seems to indicate that a complete turbo-coded simulation (or hardware emulation) is required for every amplifier operating point.

The same behavior, albeit less pronounced, can be observed for the 8-PSK constellation (Figure 13). Again, estimating the coded BER based on the uncoded measurement and the gain based on the linear channel would overestimate the required E_b/N_0.

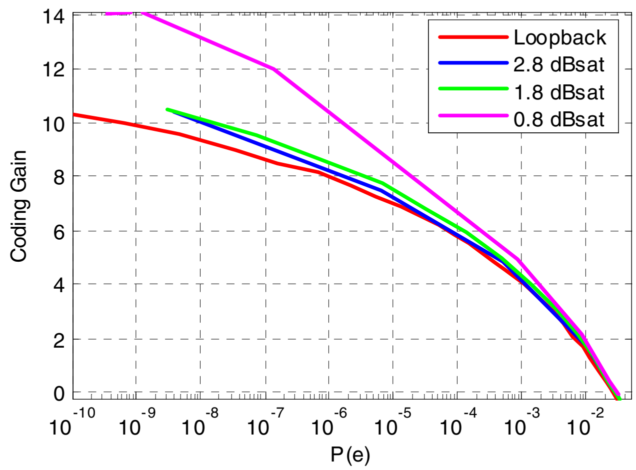

Figure 13. 8-PSK horizontal coding gain w/ TWTA

The vertical coding gain metric for 16-QAM is shown in Figure 14. Unfortunately, and somewhat unexpectedly, this coding gain is also dependent on the amplifier operating point. Indeed, three times as many errors can be tolerated in a linear channel than in the highly nonlinear channel for the same coded BER. This indicates an interaction between the code and the constellation distortion. The constellation distortion can be taken into account for computing the LLR values in the turbo decoder [3], but this technique is beyond the scope of this paper.

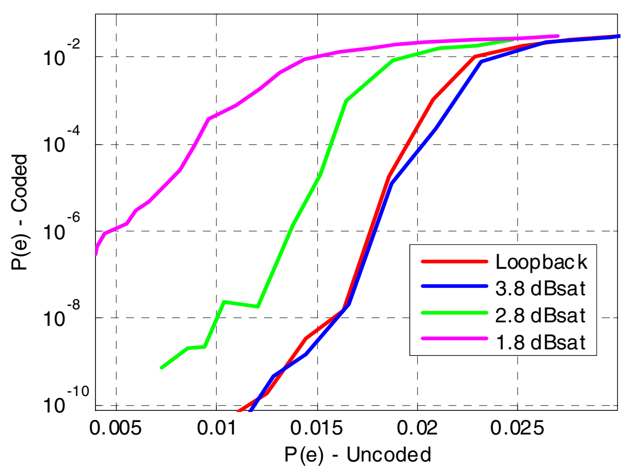

Figure 14. 16-QAM vertical coding gain w/ TWTA

The vertical coding gain exhibits less variation over the amplifier operating points for the 8-PSK constellation, as shown in Figure 15. It is interesting to note that the relative order of the curves is opposite of that for 16-QAM. That is, more errors are actually tolerated for nonlinear channel for 8-PSK. However, the difference between the curves is very small, and is within the measurement precision.

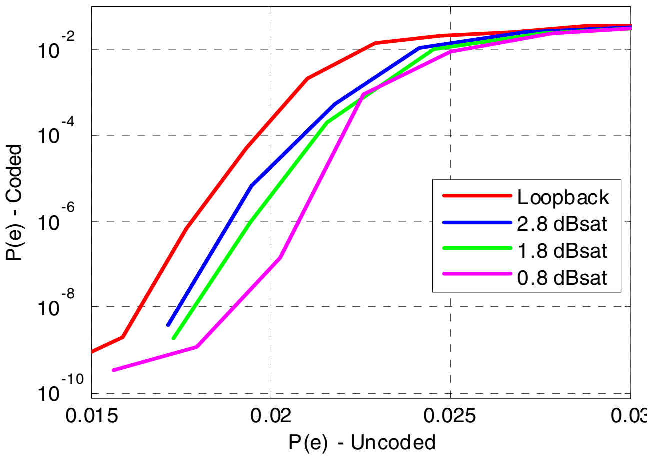

Figure 15. 8-PSK vertical coding gain w/ TWTA

It is interesting to study whether the behavior observed in the figures above is specific to that turbo code (r=0.878: (128,120)×(128,120)). For completeness, a set of results for a higher rate code (r=0.954: (64,32)×(64,62)) is presented in Figure 16 and Figure 17. As expected, the 'cliff' portion of the BER curve occurs at a lower uncoded BER for this higher rate code. The horizontal coding gain behavior is similar to that of the lower rate code for BER > 10^{-9}. The error floor of the higher rate code becomes evident for lower BER. It is also interesting to note that the higher rate code is not powerful enough to get to the 'cliff' portion at the lowest back-off level, and even the coded curve does not go below 10^{-3}. A somewhat unexpected result is that the relative positions of the vertical coding gain curves are different from those in Figure 14.

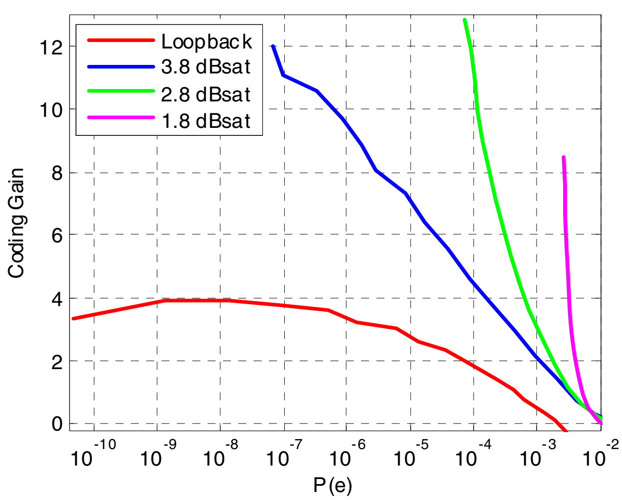

Figure 16. 16-QAM horizontal coding gain (r=0.954) w/ TWTA

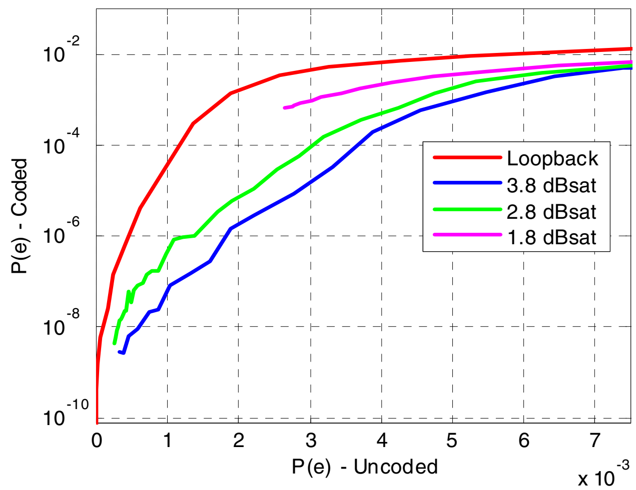

Figure 17. 16-QAM vertical coding gain (r=0.954) w/ TWTA

VI. CONCLUSION

In this paper we utilized a unique asset consisting of a multi-modulation modem that supports turbo coding and a space-qualified TWTA. This platform provides high-fidelity emulation of a satellite downlink channel. An in-depth study of the relationship between amplifier operating point (back-off) and the BER was completed for both coded and uncoded cases. The key result of this work shows that it is unreasonable to predict the performance of a communications system with coding in a nonlinear channel based on measurements taken in a linear channel. This unfortunate conclusion makes performance analysis of such systems more costly and timing consuming.

Advanced amplifier equalization techniques such as nonlinear equalization based on Volterra series expansion [9], or predistortion techniques with memory can be used to reduce the nonlinear distortion. These techniques allow operation closer to saturation, but do not change the fundamental nature of the distortion. Thus, the main result of this paper is valid even if these methods are employed.

The combination of computationally intensive turbo codec and the high sampling rates required to properly model the nonlinear amplifier result in very long simulation times. The analysis of error floor behavior of high-speed coded satellites links requires on the order of 10^{11} bits per point. These simulations would take an unreasonably long time if implemented in software even on today's top-of-the-line computers. A real-time testbed, such as the one described in this paper, is an obvious, and possibly the only solution to analyze the performance of such systems.

References

[1] L. Giugno, M. Luise, and V. Lottici, "Adaptive pre- and post-compensation of nonlinear distortions for high-level data Modulations," IEEE Transactions on Wireless Communications, V. 3, Sept. 2004 pp. 1490–1495

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes," *ICC '93*, Geneva, Switzerland, May 1993.

[3] H. Nikopour, A. K. Khandani, and S. H. Jamali, "Turbo Coded OFDM Transmission over Nonlinear Channel"

[4] R. De Gaudenzi et al., "A Novel High Power and Spectral Efficient Coded Modulation for Nonlinear Satellite Channels," 2002

[5] L. Giugno, V. Lottici, and M. Luke, "Optimizing High-speed Satellite Communication with High-Order Modulation via Adaptive Signal Processing at Both Ends of the Link"

[6] E. Grayver, P. Dafesh, M. Muha, and A. Moulthrop, "Performance Measurements of Turbo-Coded, Bandwidth Efficient Modulations in the Presence of a Nonlinear TWTA," Optics and Photonics Conference, 2005

[7] E. Grayver and P. Dafesh, "Multi-Modulation Programmable Transceiver System with Turbo Coding," IEEE Aerospace, Big Sky, MT, Feb. 2005

[8] Advanced Hardware Architectures Web Site, http://www.aha.com/. See the 4540 ASIC.

[9] S. Benedetto, E. Biglieri, and V. Castellani, Digital Transmission Theory, Prentice Hall, 1987

Biography

Eugene Grayver received a B.S. degree in electrical engineering from Caltech, and a Ph.D. degree from UCLA. He was one of the founders of a fabless semiconductor company working on low-power ASICs for multi-antenna 3G mobile receivers. In 2003 he joined The Aerospace Corporation, where he is working on flexible communications platforms. His research interests include reconfigurable implementations of digital signal processing algorithms, low-power VLSI circuits for communications, and system design of wireless data communication systems. Eugene is currently working on integrating error correction coding with high-order modulations, and on software-defined and cognitive radio research. He has published 7 journal articles, and over a dozen conference papers.

Pedro E. Santacruz received a B.S. degree in electrical engineering from the University of Texas at El Paso in 2004 and is currently working on an M.S. degree at the same institution. He has been an intern for The Aerospace Corporation in the Digital Communication Implementation Department, where he has performed work on link budget analyses for concept design tools and implementation of digital communication systems.