Performance Comparison of Selected Bandwidth-Efficient Coded Modulations

K. Andrews,[1] D. Lee,[1] F. Pollara,[1] and M. Srinivasan[1]

Bandwidth-efficient modulations may be used to reduce the possibility of future congestion in the deep-space frequency bands due to a growing demand for higher data rates and to an increasing number of simultaneous missions. High-rate error-correcting codes such as punctured turbo codes and low-density parity-check (LDPC) codes can improve bandwidth usage while still providing large coding gains essential for deep-space communications. This article examines the throughput and bit-error rate performance of various coding and modulation combinations. As the number of such combinations is unlimited, the study concentrates upon bandwidth-efficient quadrature modulations compatible with the DSN Block V Receiver and selected turbo and LDPC codes. Capacity limits are derived for modulations under bandwidth constraints, and a comparison of selected uncoded and coded modulation schemes on the basis of bandwidth-normalized throughput and bit-error rate performance, obtained via theory and/or simulation, is presented. The effect of nonlinear amplifier operation is also analyzed. Finally, candidates for high-data-rate Mars missions are identified.

I. Introduction

Communication links for deep-space missions have traditionally been severely power limited and essentially unaffected by bandwidth limitations (see Fig. 1, which illustrates how little can be gained with higher than binary modulations if power efficiency is paramount). This situation is rapidly changing with the advent of new technology developments such as very large antenna arrays on the ground, inflatable antennas on the spacecraft, and much higher power availability through radioisotope thermoelectric generators (RTGs). Even current mission designs such as Mars Reconnaissance Orbiter (MRO) stretch the bandwidth availability at 8.4 GHz (X-band) and are forced to use more bandwidth-efficient modulations or move to higher frequencies, 32 GHz (Ka-band). Future missions adopting the technologies just mentioned will be forced to use bandwidth-efficient modulations even at Ka-band due to much higher data rates and many simultaneous users.

---

[1] Communications Systems and Research Section.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

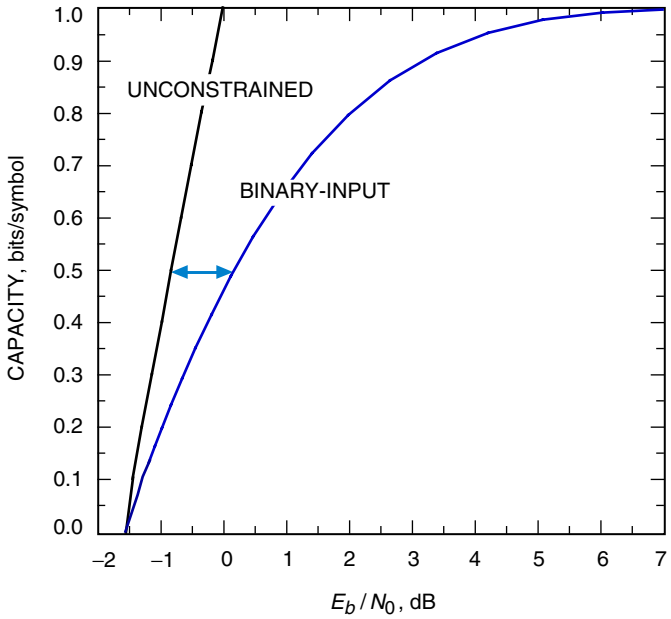

Fig. 1. Capacity of the additive white Gaussian noise channel with unconstrained and binary-input signaling.

The objective of this article is to start a systematic comparison of currently proposed coding and modulation schemes in terms of their power and spectral efficiency and of other suitable characteristics for deployment in future missions and to isolate a few recommended schemes. Besides relying on compliance with prescribed spectral masks [14], it is necessary to define appropriate metrics for performance in order to make meaningful comparisons and perform trade-off analyses. While power efficiency can be easily summarized by the single parameter E_b/N_0, representing the signal-to-noise ratio (SNR) per transmitted bit of information, the measure of spectral efficiency is open to several possibilities, as follows:

(1) *Half-power bandwidth*—The bandwidth such that the power spectrum of the signal is within 3 dB of its peak value (inside the main lobe).

(2) *Null-to-null bandwidth*—The width of the main spectral lobe.

(3) *Equivalent noise bandwidth*—Defined as

$$B_{eq} = \frac{\int_{-\infty}^{\infty} |S(f)|^2 df}{\max_f |S(f)|^2}$$

where $S(f)$ is the power spectral density (PSD). This describes a rectangle of base B_{eq} and height $\max_f |S(f)|^2$ having an area equal to the total signal power.

(4) *Fractional power containment bandwidth*—Defined as the bandwidth that contains a given percentage of the signal power (also called *essential bandwidth* in [14]).

These measures of spectral efficiency are illustrated in Fig. 2. In this article, we will base our comparisons on the last of these definitions (fractional power containment bandwidth) in accordance with the specifications in [14]. Our comparisons will be guided by the investigation of the ultimate capacity limits in the presence of bandwidth limitations, which is presented along with a discussion of Nyquist signaling in the next section. We then give an overview of selected coding and modulation schemes that

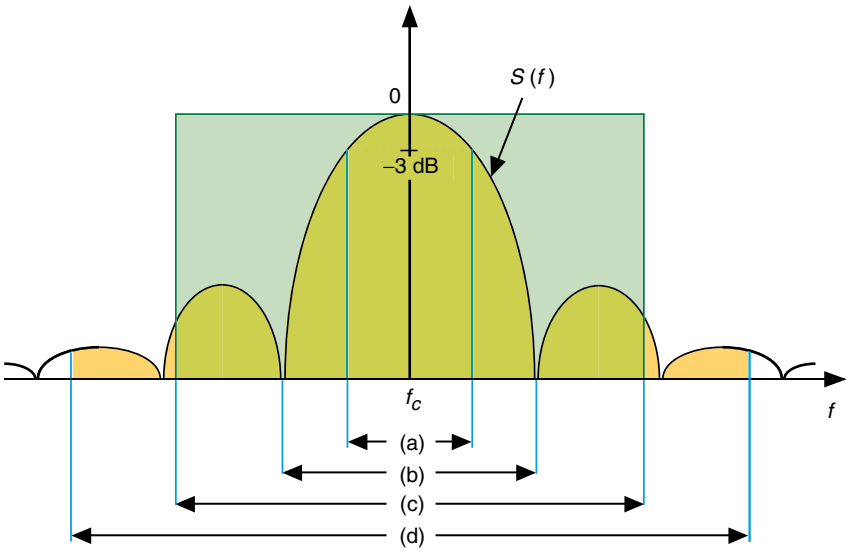

Fig. 2. Some of the possible measures for spectral efficiency: (a) half-power bandwidth, (b) null-to-null bandwidth, (c) equivalent noise bandwidth, and (d) fractional power containment bandwidth.

have been the subjects of investigation over the past several years and compare their performances in terms of bandwidth and power efficiency. As some of the candidate modulation schemes are not constant envelope, they undergo distortion when subjected to the high-efficiency but nonlinear region of power amplification. This issue is investigated by showing performance results from simulations of nonlinear amplification, as well as via a parametric study in which the peak-to-average power of each modulation is calculated along with estimates of losses due to constraining operation within the lower-efficiency, lower-power linear amplification region. Finally, conclusions regarding promising candidates among the coded modulation schemes for future high-data-rate missions are presented.

II. Signaling Scheme and Capacity Consideration

A binary phase-shift keying (BPSK) transmitter sends a sequence of binary symbols, $\{\cdots, b_{-1}, b_0, b_1, \cdots\}$, where $b_i \in \{+1, -1\}$ are independent identically distributed (i.i.d.) and zero mean, by modulating a carrier with the signal $s(t)$ obtained by convolving the impulse stream

$$b(t) = \sum_{i=-\infty}^{\infty} b_i \delta(t - iT)$$

with the impulse response $r(t)$ of the transmission filter, resulting in $x(t) = b(t) \star r(t)$. The random process $x(t)$ has PSD $S(f) \equiv (1/T)|R(f)|^2$, where $R(f)$ is the Fourier transform of $r(t)$ and T is the signaling interval. If $r(t)$ is the unit rectangle (i.e., $r(t) = 1$ for $0 \leq t < T$ and is zero elsewhere), the resulting BPSK modulation is called nonreturn to zero (NRZ).

The matched-filter receiver convolves this signal, plus added Gaussian noise, with a filter of impulse response $r^*(-t)$ and samples the result at times $t = iT$ (where the asterisk represents the complex conjugate). Note that if the receiver filter is the unit rectangle with impulse response equal to 1 for $0 \leq t < T$ and zero elsewhere, then this is the "integrate-and-dump" receiver. The impulse response of the combined transmitter and matched-filter receiver system (with no noise) is $s(t) \equiv r(t) \star r^*(-t)$, with corresponding transfer function $S(f) = (1/T)|R(f)|^2$.

If $s(t)$ is a Nyquist pulse, i.e., $s(0) = 1$ and $s(iT) = 0$ for integer $i \neq 0$, then there is no inter-symbol interference (ISI). Moreover, if $s(t)$ is Nyquist, then the noise samples are independent [12, Theorem 3.1.5, p. 87]. The ith received noise sample is the result of filtering the additive white Gaussian noise (AWGN) process with $r^*(t-iT)$, and these filters are orthogonal because $\int r(t-iT)\ r^*(t-i'T)\,dt = s\left((i-i')T\right) = 0$ when $i \neq i'$.

These arguments show that when an AWGN channel is surrounded by a BPSK transmitter with filter $r(t)$ and a matched-filter receiver, for which $r(t) \star r^*(-t)$ is Nyquist, it is converted into a discrete-time Gaussian channel without ISI. This channel has noise with variance $N_0T/2$ per transmission, and transmissions are made every T seconds.

A. Properties of Signaling Pulses

Some relevant properties of Nyquist functions $s(t)$ are discussed. Define the comb function:

$$c(t) = \sum_{i=-\infty}^{\infty} \delta(t-i)$$

Its Fourier transform is also a comb function: $C(f) = c(f)$. When $s(t)$ is Nyquist, $s(t/T)c(t/T) = \delta(t/T)$, so taking Fourier transforms, $S(fT) \star C(fT) = T$. This can only be true if $S(f)$ has support at least of length $1/T$. The Nyquist function $S(f)$ with minimum support is the unit rectangle of width $1/T$, which $R(f)$ is as well, and $r(t)$ is a sinc function.

Symmetric Nyquist functions with support up to $2/T$ also have simple transforms. It is easy to show they have point symmetry about $(1/T, 1/2)$, i.e., $S(f) = 1 - S(1/T - f)$ for $0 \leq f < 2/T$. This family of functions includes the unit rectangle just considered, the raised-cosine functions, and trapezoids, among others. All of these have a -3 dB bandwidth of exactly $1/T$.

The time domain "tails" of $r(t)$ decrease as $|t|^{-(n+1)}$, where n is the number of the first discontinuous derivative of $R(f)$. The unit rectangle is discontinuous (in the 0th derivative), so the tails decay as $1/t$. The trapezoids are discontinuous in the first derivative, and the tails decay as $1/t^2$. The raised-cosine solutions (for $\alpha \neq 0$) are discontinuous in the third derivative, so the tails decay as $1/t^4$, and for this reason they have attracted particular attention.

B. Model for Coded Modulation Systems

When the communication system uses both a suitable modulation to obtain the desired bandwidth efficiency and a coding scheme to obtain power efficiency, we can model it as in Fig. 3. In the figure, $T_b = 1/R_b$ is the information *bit* time and R_b is the bit rate, $T_s = RT_b$ is the coded *symbol* time and R is the code rate, and $T = T_s \log_2 M$ is the *baud* time corresponding to the signaling interval discussed above. Here M is the cardinality of the signal constellation: $M = 2$ for BPSK, $M = 4$ for quadrature phase-shift keying (QPSK), etc. The signal $b(t)$ corresponds to the symbols entering the modulator, and the signal $s(t)$ corresponds to bauds.

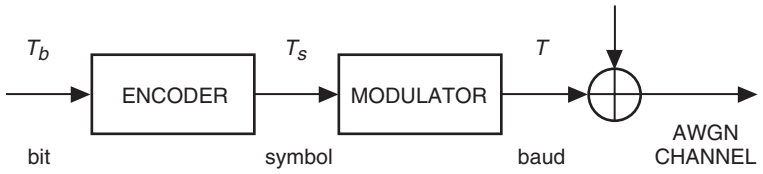

Fig. 3. Model for coded modulation systems.

C. Discrete-Time and Continuous-Time Capacity

The capacity of a discrete-time Gaussian channel is

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right) \text{ bits/symbol} \tag{1}$$

where P and N are the powers of the signal and noise, respectively. To extend this result to the continuous-time band-limited Gaussian channel, let us suppose that the signal and noise are band-limited to the interval $[-B, B]$. By the sampling theorem, we can represent each signal using at least $2B$ samples per second, each with average power P. Considering an AWGN with two-sided power spectral density $N_o/2$ sampled every $1/2B$ seconds, the noise has power N_oB and we get the capacity C of the band-limited Gaussian channel as

$$C = B\log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits/second} \tag{2}$$

This capacity is shown in Fig. 4 for one-dimensional and two-dimensional (unconstrained-input) signaling. Results on capacity loss due to filtering, shown in Fig. 4, are discussed in Subsection II.D.

Using the dimensionless quantities $\rho = P/(N_0B)$ as the signal-to-noise ratio and $\eta = C/B$ as the bandwidth efficiency in bits/second/hertz, we can rewrite Eq. (2) as

$$\eta = \log_2(1 + \rho) \text{ bits/second/hertz}$$

Conversely, the continuous-time AWGN channel can be made discrete (with transmissions every $T = 1/B$ seconds) by using a unit rectangle Nyquist filter of width B. It is straightforward to show that this system achieves the capacity of the band-limited AWGN channel.

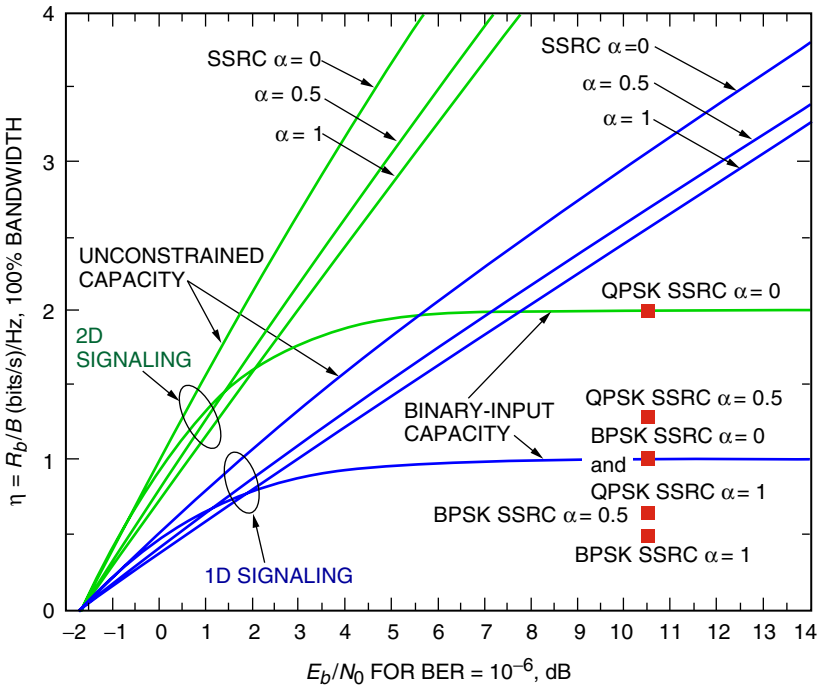

Fig. 4. Effect of filtering on unconstrained and binary-input capacity.

When constrained to binary inputs, the continuous-time AWGN channel with transmission filter $r(t)$ suffers no loss in capacity, compared to the same system under a *maximum* power constraint. This is because $r(t)$ must be low-pass at some point, and the binary input signal can be switched fast enough to approximate any analog level to the desired precision [13].

The discrete-time AWGN channel, constrained to binary inputs, does have a lower capacity than when the input has only a mean- or maximum-power constraint. With binary inputs, it transmits $H(Y)-H(N)$ bits of information, where $H(\cdot)$ is the entropy function, and N, X, and Y are random variables with distributions $N \sim \mathcal{N}(0, N_0/2)$, X equally distributed between $+\sqrt{P}$ and $-\sqrt{P}$, and $Y = X + N$. This is readily computed, is a function of $\rho = P/(N_0 B)$, and has limit

$$\lim_{\rho \to \infty} H(Y) - H(N) = 1 \text{ bit/transmission}$$

The binary input capacity is also plotted in Fig. 4 for one- and two-dimensional signaling.

D. Capacity with Bandwidth Constraints

Given an AWGN channel with input bandwidth limit B, the capacity in Eq. (2) is achieved by using Gaussian inputs with a uniform PSD:

$$S(f; f_0) = \begin{cases} \dfrac{P}{2B}, & ||f| - f_0| < \dfrac{B}{2} \\ 0, & \text{otherwise} \end{cases}$$

Because this PSD is symmetric, i.e., $S(f; f_0) = S^*(-f; f_0)$, we can simplify the notation by using the complex baseband representation:

$$S(f) \equiv \begin{cases} S(f + f_0; f_0) & \text{if } f > -f_0 \\ 0 & \text{otherwise} \end{cases}$$

Note that $\int_{-\infty}^{\infty} S(f) = 1/2 \int_{-\infty}^{\infty} S(f; f_0)$.

When the channel input is constrained to power spectral density $S(f)$, its capacity is

$$C = \frac{1}{2} \int_{-\infty}^{\infty} \log_2 \left[1 + \frac{2S(f)}{N_0} \right] df \tag{3}$$

With a constant-envelope constraint but no bandwidth constraint, capacity is not reduced. Any set of orthogonal signals, such as the sinusoids or the binary Hadamard vectors, achieves capacity [20]. With a constant-envelope constraint and a PSD constraint, capacity is reduced, although the general solution to this problem remains open [21]. With a constant-envelope constraint, a PSD constraint, and a limit on the transition rate (or transition density), the capacity is reduced further, but this also is an open problem in general [22]. In Fig. 4, we show the capacity loss as calculated from Eq. (3) due to a specific class of filters called square-root raised-cosine filters, which are described in detail in Section IV. The effect of imposing a spectral mask upon a telemetry signal has been investigated [18].[2] In [18], it was shown that certain spectral masks do not allow the use of Nyquist signaling

[2] B. Moision, "Capacity Under a Spectral Mask Constraint," JPL Interoffice Memorandum (internal document), Jet Propulsion Laboratory, Pasadena, California, October 23, 2002.

(for example, the original high-rate Space Frequency Coordination Group (SFCG) 17-2R1 recommendation [14]). Furthermore, it has been argued that all of the proposed SFCG mask definitions do not allow for a meaningful evaluation of resultant capacity, and an alternate mask definition has been proposed along with corresponding capacity calculations.[3]

III. Coding Systems

We have considered several coding systems suitable for combination with bandwidth-efficient modulations for deep-space missions. These codes are shown in Figs. 5 and 6 together with codes currently in use in the DSN and with codes recently published in the literature. An essential requirement for bandwidth efficiency is to keep the code rate high enough, say ≥ 0.5, but not too high, say ≤ 0.8, to preserve sufficient coding gain and optimality with respect to the unconstrained capacity limit.

One standard code used in the DSN whose performance is well characterized and included in our comparison is the rate 1/2, constraint length 7 convolutional inner code concatenated with the (255,223) Reed–Solomon outer code (RS+(7,1/2) c.c). In addition, we have isolated as preferred candidates two new classes of codes:

(1) Punctured turbo codes of rate 3/4 and 7/8, which are a simple extension of the present Consultative Committee on Space Standards (CCSDS) standard turbo codes. Details on these codes and the specific punctured pattern have been submitted to CCSDS for inclusion in a revised standard for coded telemetry.

(2) Low-density parity-check (LDPC) codes of rate 0.5 and 0.8 [15], as shown in Fig. 5. Figure 5 shows a variety of codes with near-capacity performance. Rates below 0.5 are suitable for low-data-rate deep-space missions; rates between 0.5 and approximately 0.8 are suitable for high-data-rate deep-space missions when used with offset QPSK (OQPSK) modulation. Code rates higher than 0.8 are not recommended for deep-space missions due to severe power efficiency reduction.

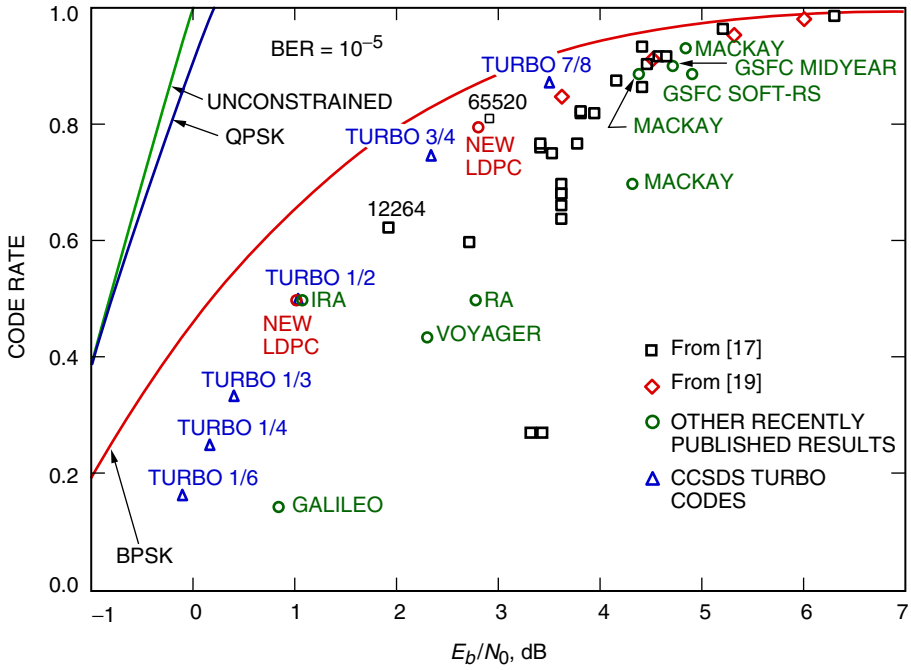

Fig. 5. Codes performance for BER = 10^{-5}.

[3] Ibid.

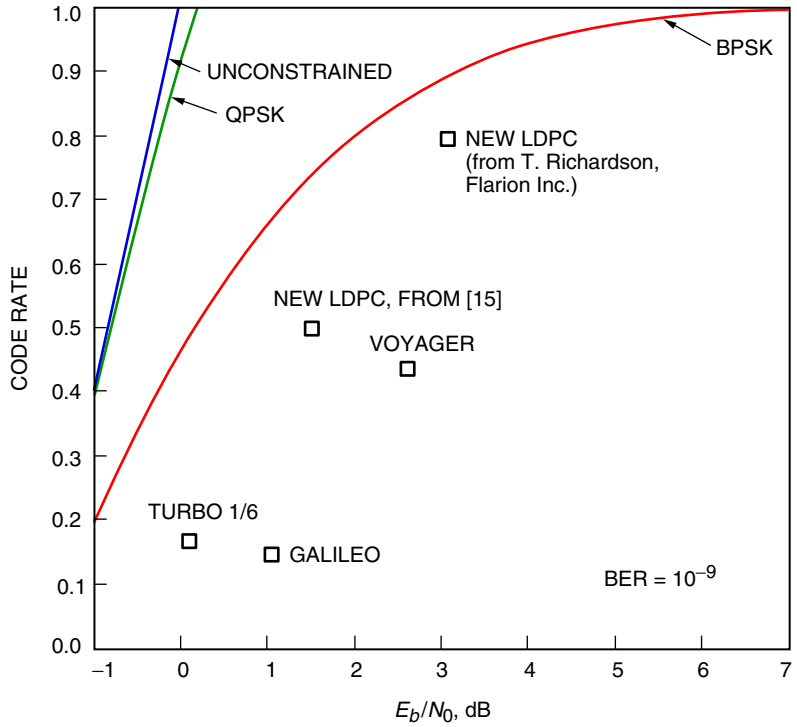

Fig. 6. Codes performance for BER = 10^{-9}.

The punctured turbo codes have the advantage of being decodable with a simple modification of the current DSN turbo decoder. The LDPC codes have two main advantages: they have lower decoding complexity and therefore are suitable for very high data rates (>10 Mb/s), and they perform better at a very low bit-error rate (BER) (10^{-9}) since their error floor can be controlled and pushed to lower BERs for a small penalty in their waterfall-region performance. For application requiring BERs lower than the traditional 10^{-6} and high data rates, LDPC codes are a suitable choice (see Fig. 6). Higher spectral efficiency is obtained by using these codes with OQPSK modulation.

IV. Modulations

The bandwidth-efficient modulations considered in this article are for the most part quadrature modulations that are compatible with the DSN Block V Receiver structure and have been recommended by the CCSDS. One exception to this is the four-dimensional (4D) 8-phase shift keying (8-PSK) trellis-coded modulation (TCM), which is being used by the European Space Agency and has been proposed for use in the Earth Exploration Satellite Service. While "modulation" in the strict sense refers to the shape of the signal constellation (M-ary phase-shift keying (M-PSK), quadrature amplitude modulation (QAM), etc.), some of the modulations discussed here actually denote specific pulse shapes:

(1) BPSK: This modulation has been used traditionally in deep-space missions with $\pm\sqrt{P}$-level rectangular-shaped baseband pulses. In our work, the use of square-root raised-cosine (SRRC) pulses, which are Nyquist pulses, has been investigated. As shown in Fig. 7, the power spectral density of the SRRC pulse is given by

$$S(f) = \begin{cases} \dfrac{PT}{2} & | \, |f| - f_0 \, | \leq \dfrac{1-\alpha}{2T} \\ \dfrac{PT}{4} \left\{ 1 + \cos \left[\dfrac{\pi T}{\alpha} \left(| \, |f| - f_0 \, | - \dfrac{1-\alpha}{2T} \right) \right] \right\} & \dfrac{1-\alpha}{2T} \leq | \, |f| - f_0 \, | \leq \dfrac{1+\alpha}{2T} \\ 0 & | \, |f| - f_0 \, | > \dfrac{1+\alpha}{2T} \end{cases} \tag{4}$$

where $T = T_s \cdot \log_2 M$ and M is the constellation size (e.g., $M = 2$ for BPSK, $M = 4$ for QPSK), with roll-off factor $0 \leq \alpha \leq 1$. This modulation has bandwidth $B = (1+\alpha)/T$ and power $\int_{-\infty}^{\infty} S(f)\, df = P$.

The time-domain pulse shapes are given by [16]:

$$s(t) = \frac{4\alpha}{\pi\sqrt{T}} \frac{\cos \dfrac{(1+\alpha)\pi t}{T} + \dfrac{T}{4\alpha t} \sin \dfrac{(1-\alpha)\pi t}{T}}{1 - \left(\dfrac{4\alpha t}{T} \right)^2} \tag{5}$$

They usually are generated using a look-up table over several symbols and satisfy the Nyquist condition for zero intersymbol interference (ISI), so that the bit-error probability is identical to that of BPSK with NRZ rectangular pulses if the receiver samples at zero-ISI points.

(2) OQPSK: In offset QPSK, the in-phase and quadrature symbols are delayed by half of the pulse duration so that phase changes are limited to 90 deg at any time. Both rectangular and SRRC pulse shaping are considered with OQPSK. Note that when pulse shaping is applied, the peak-to-average power ratio is different for BPSK and OQPSK, due to the phase constraint placed upon the OQPSK symbols.

(3) Pre-Coded GMSK: GMSK is minimum-shift keying with data filtered by a Gaussian-shaped frequency response filter, here with BT_s values of 0.25 or 0.5, producing partial-response signals. In the version of GMSK presented here, the data are pre-coded with a differential decoder in order to compensate for the inherent differential encoding in MSK-type modulations and are implemented in simulations as a frequency modulation.

(4) T-OQPSK: Trellis-coded OQPSK is an alternative implementation of staggered quadrature overlapped raised-cosine modulation in which memory is introduced in transmitted data via a rate 1/2 two-state encoder [9]. The optimal receiver for uncoded T-OQPSK consists of four matched filters followed by a Viterbi algorithm. A constant-envelope version of T-OQPSK has been obtained by passing the signal through a bandpass hard-limiter,[4] introducing cross-correlation between in-phase and quadrature channels.

(5) FQPSK: In Feher-patented QPSK, cross-correlation is introduced between in-phase and quadrature data streams and specific waveforms are used so that the envelope is nearly constant. A filtered version referred to as FQPSK-B uses patented filters in in-phase and quadrature channels following waveform generation, resulting in improved bandwidth efficiency at the expense of envelope constancy. See [8].

[4] M. K. Simon, D. Divsalar, and D. Lee, "Coded Constant Envelope T-OQPSK with Iterative Decoding," (internal document), Jet Propulsion Laboratory, Pasadena, California, October 23, 2001.

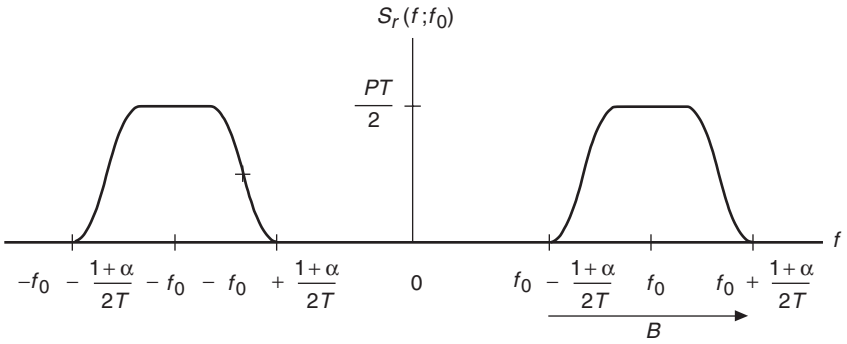

Fig. 7. Raised cosine spectrum.

V. Comparison of Coded Modulation Systems

We considered several combinations of coding and modulation schemes described in Sections III and IV. Figure 8 compares power efficiency, E_b/N_0, and spectral efficiency, $\eta = R_b/B$, measured as the fractional (99 percent) power containment bandwidth in bits/second/hertz, defined in Section I, for some of these combinations of coding and modulations on the AWGN channel. Capacity curves also are shown in the figure. When a finite error rate is acceptable, the channel capacity is higher; for typical numbers such as BER=10^{-6}, the difference is imperceptible. Because this capacity is achieved by using a uniform PSD, the curve can be raised by the fraction $1/\beta$ when measured according to a $100 \times \beta$ percent bandwidth constraint. Some of the included data points come from evaluation of analytical expressions, while other points were obtained through computer simulations. For systems employing independent coding and modulation, the spectral efficiency may be obtained by scaling the uncoded spectral efficiency by the code rate, and applying coding gains for various codes to required uncoded E_b/N_0 values for 10^{-6} bit-error rate performance. Combined coding and modulation (in which intersymbol interference from the modulation is used as an inner code of a serial concatenated pair) requires direct simulation. Figure 8 also shows the the region of interest for Mars missions, where coding gain is still of great importance, and moderate spectral efficiency is necessary to accommodate current high-data-rate X-band missions and future Ka-band missions requiring even higher data rates.

Table 1 further illustrates the same comparison of uncoded and coded modulation formats that have been studied. These modulations are compared in terms of several parameters, including throughput, bit-error rate performance, SFCG mask conformity, and effect of power amplification. Much of this information is obtained from [1–3], and further details regarding these results may be pursued therein. Other relevant evaluation criteria that were not included here are robustness to synchronization errors and receiver complexity. The table parameters that were evaluated here are explained as follows:

(1) Spectral efficiency: This is the ideal normalized throughput, i.e., the number of transmitted bits per second normalized by the 90 percent or 99 percent power containment bandwidth, in the absence of distorting factors such as nonlinear amplification. The bandwidths are obtained analytically or through spectra generated by software simulations such as Matlab and Signal Processing Workstation (SPW).

(2) E_b/N_0 at a 10^{-6} bit error rate: The required bit SNR to achieve 10^{-6} bit-error probability is obtained from theoretical formulas when available. In other cases, Monte-Carlo simulation results are given. Annotations to the table indicate when suboptimal receiver results are used.

(3) SFCG mask conformity: The SFCG has recommended a mask restricting the PSD of a telemetry signal relative to the peak of its spectrum. In the table, we refer to the mask specified by provisional Recommendation 21-4, shown in Fig. 9.

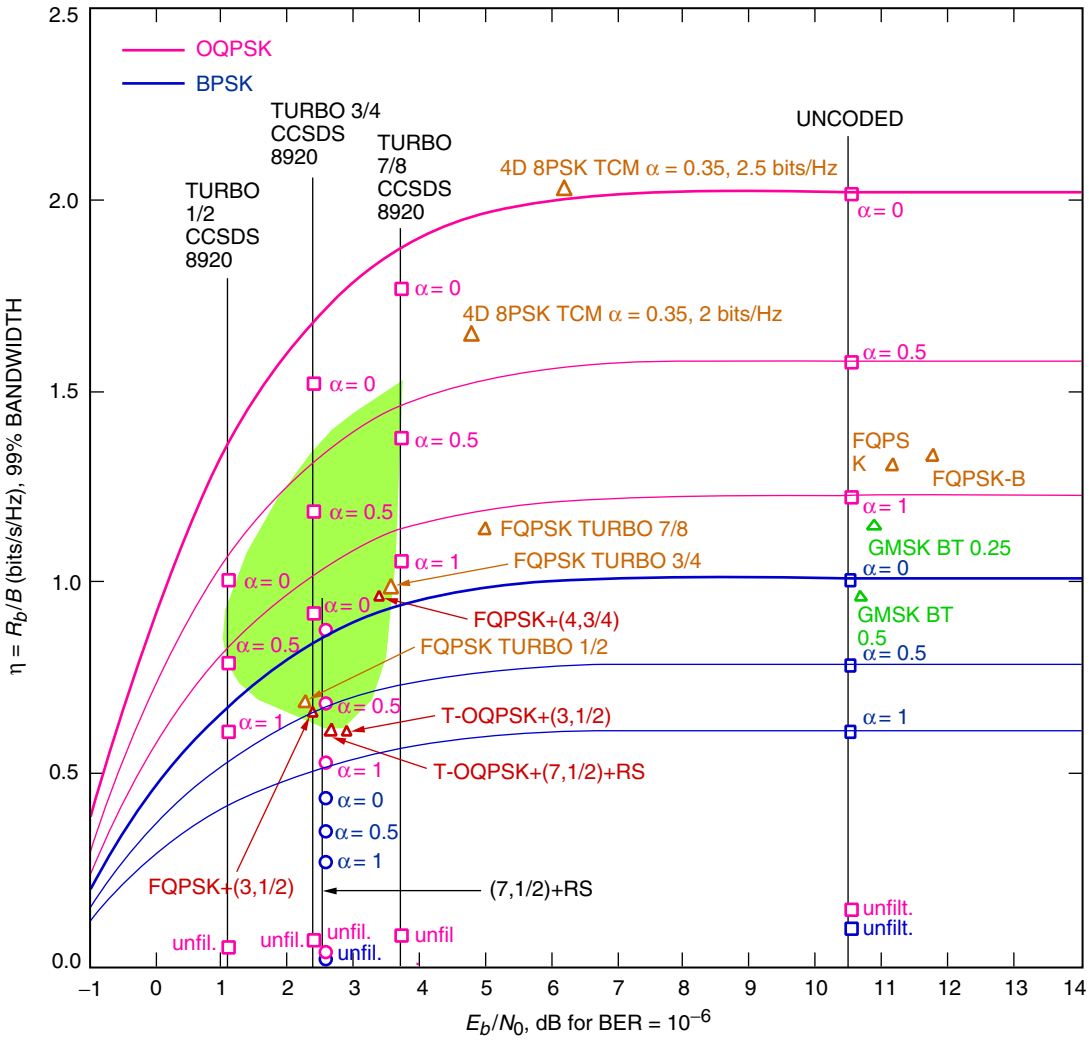

Fig. 8. Comparison of several coded modulation schemes showing the region of interest for Mars missions.

(4) Peak-to-average power ratio: The peak-to-average power ratio is given in dB. For full response signals, the peak and average powers may be obtained analytically. For partial response signals such as SRRC and GMSK, the peak and average powers are calculated from a simulated symbol stream that is thousands of symbols long.

(5) Saturated power amplifier nonlinearity: A 10-W European Space Agency (ESA) solid-state power amplifier (SSPA) was simulated using SPW via amplitude nonlinearity (AM-AM) and phase nonlinearity (AM-PM) data provided by ESA. In the simulations, the SSPA was driven at maximum output power in all cases (no back off), and resulting power spectra and bit-error probabilities were then calculated. For many of the coded schemes, the throughput was obtained by scaling the throughput for the corresponding uncoded modulations, while the required bit SNR for the nonlinear amplifier case was obtained by applying the same coding gain as in the linear case. Note that this method is an approximation and may underestimate the coded performance. For example, the coding gain for SRRC-OQPSK $\alpha = 0.5$ with the RS+(7,1/2) concatenated code is 7.9 dB. Applying this coding gain to the nonlinear channel gives us a required SNR of 3.7 dB, while an actual SPW simulation yielded a required SNR of 2.8 dB. Unfortunately, a complete set of simulations using the SSPA for all the coded modulation schemes was not available.

Table 1. Comparison of bandwidth-efficient modulations.

	Modulation and code	Spectral efficiency, (bits/s)/Hz		Ideal E_b/N_0 at 10^{-6} BER, dB	SFCG mask conformity	Peak-to-average power ratio, dB	Saturated power amplifier nonlinearity	
		90% BW	99% BW				Spectral efficiency at 99% BW	E_b/N_0 at 10^{-6}, dB
Uncoded	BPSK							
	NRZ rectangular	0.59	0.047	10.5	No	0.0	0.047	10.5
	SRRC-BPSK							
	$\alpha = 1$	0.84	0.61	10.5	Yes	3.7	0.38	10.8[a]
	$\alpha = 0.5$	1.02	0.79	10.5	Yes	3.4	0.35	11.1[a]
	$\alpha = 0$	1.11	1.01	10.5	Yes	8.2	0.31	15.8[a]
	OQPSK[b]							
	NRZ rectangular	1.18	0.094	10.5	No	0.0	0.094	10.5
	SRRC-OQPSK							
	$\alpha = 1$	1.68	1.22	10.5	Yes	2.1	1.15	10.8[b]
	$\alpha = 0.5$	2.04	1.58	10.5	Yes	3.1	1.19	11.6[b]
	$\alpha = 0$	2.22	2.02	10.5	Yes	6.0	1.11	>24[b]
	Pre-coded GMSK							
	$BT = 0.5$	1.45	0.97	10.6[c]	Yes	0.0	0.97	10.6[c]
	$BT = 0.25$	1.79	1.16	10.9[d]/11.2[c]	Yes	0.0	1.16	10.9[d]/11.2[c]
	FQPSK							
	Unfiltered	1.96	1.28	11.3[d]/11.7[c]	Yes	0.02	1.28	11.8[c]
	B (filtered)[e]	1.96	1.32	11.9[c]	Yes	0.5	1.28	11.4[d]/12.1[c]
Coded	RS+(7,1/2) convolutional code (c.c.) BPSK[f]							
	NRZ rectangular	0.26	0.021	2.6	No	0.0	0.021	2.6
	SRRC-BPSK							
	$\alpha = 1$	0.37	0.27	2.6	Yes	3.7	0.17[g]	2.9[g]
	$\alpha = 0.5$	0.45	0.35	2.6	Yes	3.4	0.15[g]	3.2[h]
	$\alpha = 0$	0.49	0.44	2.6	Yes	8.2	0.14[g]	7.9[h]
	RS+$(7, 1/2)$ c.c. QPSK[f]							
	NRZ rectangular	0.52	0.04	2.6	No	0.0	0.04	2.6
	SRRC-OQPSK							
	$\alpha = 1$	0.73	0.53	2.6	Yes	2.1	0.5[g]	2.9[h]
	$\alpha = 0.5$	0.89	0.69	2.6	Yes	3.1	0.51[g]	2.8[i]
	$\alpha = 0$	0.97	0.88	2.6	Yes	6.0	0.49[g]	>16.1[h]
	T-OQPSK with RS+$(7, 1/2)$ c.c.[f,i]	0.95	0.62	2.7	Yes	1.3	0.54[j]	2.8[c]
	Constant envelope T-OQPSK with rate 1/2, 4-state c.c.[k]	0.95	0.62	3.0	Yes	0.0	0.62	3.0

Table 1. Cont'd.

	Modulation and code	Spectral efficiency, (bits/s)/Hz		Ideal E_b/N_0 at 10^{-6} BER, dB	SFCG mask conformity	Peak-to-average power ratio, dB	Saturated power amplifier nonlinearity	
		90% BW	99% BW				Spectral efficiency at 99% BW	E_b/N_0 at 10^{-6}, dB
Coded	Unfiltered FQPSK with outer c.c.[l]							
	Rate 1/2, 4-state	0.98	0.64	2.4	Yes	0.02	0.64	2.5
	Rate 3/4, 8-state	1.47	0.96	3.4	Yes	0.02	0.96	3.5
	16+16 turbo-coded OQPSK (8920)							
	Rate 1/2 NRZ rectangular	0.59	0.047	1.1	No	0.0	0.047	1.1
	Rate 1/2 SRRC-OQPSK							
	$\alpha = 1$	0.84	0.61	1.1	Yes	2.1	0.58[g]	1.4[h]
	$\alpha = 0.5$	1.02	0.79	1.1	Yes	3.1	0.6[g]	2.2[h]
	$\alpha = 0$	1.11	1.01	1.1	Yes	6.0	0.56[g]	>14.6[h]
	16+16 turbo-coded OQPSK (8920)							
	Rate 3/4 NRZ rectangular	0.89	0.07	2.4	No	0.0	0.07	2.4
	Rate 3/4 SRRC-OQPSK							
	$\alpha = 1$	1.26	0.92	2.4	Yes	2.1	0.86[g]	2.7[h]
	$\alpha = 0.5$	1.53	1.19	2.4	Yes	3.1	0.89[g]	3.5[h]
	$\alpha = 0$	1.67	1.52	2.4	Yes	6.0	0.83[g]	>15.9[h]
	16+16 turbo-coded OQPSK (8920)							
	Rate 7/8 NRZ rectangular	1.03	0.08	3.7	No	0.0	0.08	3.7
	Rate 7/8 SRRC-OQPSK							
	$\alpha = 1$	1.47	1.06	3.7	Yes	2.1	1.01[g]	4.0
	$\alpha = 0.5$	1.79	1.38	3.7	Yes	3.1	1.04[g]	4.8[h]
	$\alpha = 0$	1.94	1.77	3.7	Yes	6.0	0.97[g]	>17.2[h]
	16+16 turbo-coded Unfiltered FQPSK (8920)							
	Rate 1/2	0.98	0.66	2.3[m]	Yes	0.5	0.64[g]	2.4[h]
	Rate 3/4	1.47	0.99	3.6[m]	Yes	0.5	0.96[g]	3.7[h]
	Rate 7/8	1.72	1.16	4.9[m]	Yes	0.5	1.12[g]	5.0[h]
	4D 8-PSK TCM with SRRC $\alpha = 0.35$ and (255, 239) Reed–Solomon (RS)[n]							
	2 bits/Hz	2.04	1.63	4.8[o]	Yes	3.4	0.79	5.3
	2.5 bits/Hz	2.55	2.04	6.4[o]	Yes	3.4	0.98	7.4

Further details regarding the data presented in Table 1 include the following points referenced in the table:

[a] The detection filter was matched to the SRRC pulse but with no equalizer to compensate for the degraded performance resulting from nonlinear amplification.

[b] In the ideal case, an OQPSK receiver that suffers no performance loss with respect to BPSK is assumed. The same pulse shapes used with BPSK may be used with OQPSK, doubling the spectral efficiency.

[c] Single-filter detection with a quasi-optimal filter (either a Wiener filter or through a computer search) [7].

[d] Viterbi maximum-likelihood receiver with unquantized branch metrics.

[e] Proprietary filter.

[f] In the Viterbi algorithm, 3-bit metric quantization and 70-bit truncation path length were used.

[g] These numbers are obtained by scaling nonlinear amplifier throughput from uncoded results.

[h] These numbers are estimated by applying coding gain in the linear channel to the nonlinear channel.

[i] Data from [2].

[j] Data from D. Lee.[5]

[k] Data from Simon et al.[6]

[l] Data from [5,6].

[m] These numbers are obtained by applying coding gain for the particular rate turbo code to the performance of uncoded FQPSK.

[n] Data from [3].

[o] Data from [4].

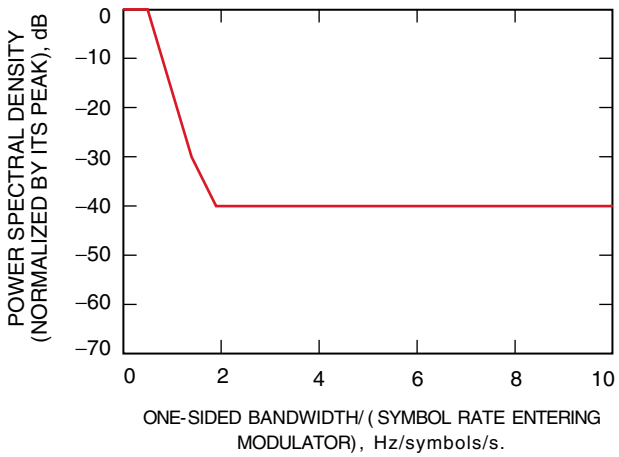

Fig. 9. Mask of modified SFCG Recommendation 21-4, provisional.

From the table we observe that, for uncoded schemes, SRRC-OQPSK with $\alpha = 0$ is most spectrally efficient under ideal conditions; under nonlinear amplification, however, SRRC-OQPSK is severely distorted and FQPSK is most spectrally efficient. In terms of power efficiency, BPSK or OQPSK with rectangular pulses ranks highest, but they do not conform to the SFCG mask. GMSK with $BT = 0.5$ performs next best (under nonlinear amplification). For coded schemes, we see that 4D 8-PSK TCM 2.5 bits/hertz is theoretically most spectrally efficient, while rate 7/8 16+16 turbo-coded unfiltered FQPSK is most spectrally efficient under nonlinear amplification. Rate 1/2 16+16 turbo-coded OQPSK with any pulse shape is theoretically most power efficient, while under nonlinear amplification, the same coded scheme with rectangular pulse shape maintains that position. Taking SFCG conformity into account, the most power-efficient scheme under nonlinear amplification is rate 1/2 16+16 turbo-coded SRRC-OQPSK with $\alpha = 1$.

The rate 1/2 LDPC code, shown in Figs. 5 and 6, with SRRC-OQPSK modulation with $\alpha = 1$ has the additional benefit of achieving lower BERs (10^{-9}) with lower decoding complexity. At BER=10^{-6}, this code has essentially the same power and spectral efficiency of the more complex rate 1/2 16+16 turbo-coded OQPSK.

VI. Effects of Amplifier Nonlinearity

In the comparison table of the prior section, the effect of power amplification upon the modulation performance was evaluated through simulations of the amplitude and phase nonlinearities for a particular amplifier at a particular point in the nonlinear region of operation (maximum output power). Alternatively, one could back off in input power and operate in the linear region, thereby not subjecting nonconstant-envelope modulations to the spectrum and pulse distortions induced by amplifier nonlinearities. By doing this, however, one pays a penalty in transmitted output power, which must be compensated for elsewhere in the link budget. Furthermore, as the operating point in saturation often occurs in a region of high amplifier efficiency, backing off from this point may also result in an additional power penalty due to lower amplifier efficiency.

Following the methodology in [10], the average total power $\overline{P}_t$ is given by

$$\overline{P}_t = \overline{P}_{dc} + \overline{P}_{in} \tag{6}$$

where $\overline{P}_{dc}$ is the average dc power supplied to the amplifier and $\overline{P}_{in}$ is the average input RF signal power of the amplifier. The power-added efficiency of the amplifier is defined as [11]

$$P_{ae}(t) = \frac{P_{out}(t) - P_{in}(t)}{P_{dc}(t)} \tag{7}$$

where P_{out} is the output RF signal power. Therefore, the average total power may be written as [10]

$$\overline{P}_t = \overline{P}_{dc} + \overline{P}_{in} = \overline{P}_{out}\left(1 + \frac{\overline{(1 - P_{ae})P_{dc}}}{\overline{P}_{out}}\right)$$

$$= \overline{P}_{out}(1 + \gamma) \tag{8}$$

[5] D. Lee, "90% and 99% Power Containment Bandwidths for Various Modulations," (internal document), Jet Propulsion Laboratory, Pasadena, California, October 2001.

[6] Simon et al., op cit.

where

$$\gamma = \frac{\overline{(1 - P_{ae})P_{dc}}}{\overline{P}_{out}} \approx \left(1 - \frac{\overline{P}_{in}}{\overline{P}_{out}}\right)\left(\frac{1 - \overline{P}_{ae}}{\overline{P}_{ae}}\right) \tag{9}$$

Figure 10 shows plots of the amplifier output power and power-added efficiency as a function of amplifier input power for an X-band 8-W SSPA provided by Centre National d'Etudes Spatiales (CNES). Phase distortion is not considered here. We may now compute the additional power necessary for transmission of a particular modulation by calculating $1 + \gamma$ using Eq. (9). For constant-envelope modulations, the operating point is the peak efficiency point, which is at approximately 19.5-dBm input power. For example, for NRZ rectangular pulse BPSK, $\overline{P}_{in} = 19.5$ dBm, $\overline{P}_{out} = 38.8$ dBm, and $\overline{P}_{ae} = 0.42$. We therefore obtain from Eq. (9) that $\gamma = 1.36$ and that the loss associated with amplifier inefficiency is 3.7 dB. Since we can operate in saturation, there is no output back-off power loss.

For nonconstant-envelope modulations, we would like the entire range of power variation to lie within the linear region of the input-to-output power curve. From Fig. 10, it appears that the linear region of operation ends at about 15.5-dBm input power. If we place the peak power at this point, then we subtract from 15.5 dBm the figure from the peak-to-average power ratio column in the modulation table in order to determine the average input power. So for example, for SRRC BPSK $\alpha = 1$, the peak-to-average power ratio is 3.7 dB, so that $\overline{P}_{in} = 11.8$ dBm, $\overline{P}_{out} = 33.3$ dBm, and $\overline{P}_{ae} = 0.18$. From these quantities, we see that the output back-off power loss is $38.8 - 33.3 = 5.5$ dB, and that $\gamma = 4.68$, so that the amplifier efficiency power loss is 7.5 dB.

In Table 2, we show the amplifier back-off and efficiency power losses for the modulations from Table 1. Note that the losses for the coded modulations are the same as for their uncoded counterparts and, therefore, are not repeated in the table. When comparing the output back-off power losses in Table 2 with the nonlinear amplifier simulation losses in Table 1, one must conclude that significantly more penalty is paid by backing off transmission power to the linear region than by subjecting the nonconstant-envelope signals to nonlinear distortions. For example, for SRRC-BPSK with $\alpha = 0.5$, nonlinear amplification results in 0.6-dB more required power from the simulation results of Table 1, whereas Table 2 shows that 5.4-dB power is lost by backing off to the linear region of operation, in addition to an efficiency loss.

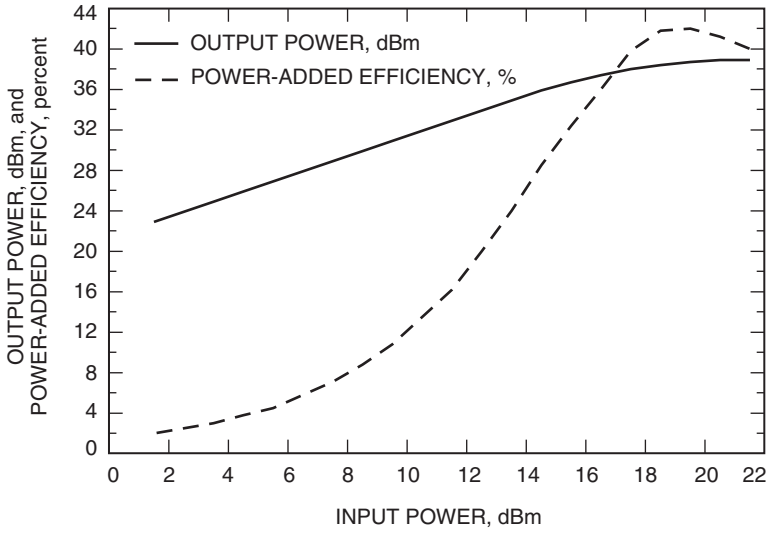

Fig. 10. SSPA input-to-output power and power-added efficiency characteristics.

Table 2. Power losses due to amplifier nonlinearities.

	Modulation and code	Peak-to-average power ratio, dB	Output back-off power loss, dB	Amplifier efficiency power loss, dB
Uncoded	BPSK: NRZ rectangular	0.0	0.0	3.7
	SRRC-BPSK $\alpha = 1$	3.7	5.5	7.5
	$\alpha = 0.5$	3.4	5.4	7.3
	$\alpha = 0$	8.2	9.8	11.6
	OQPSK: NRZ rectangular	0.0	0.0	3.7
	SRRC-OQPSK $\alpha = 1$	2.1	3.8	6.2
	$\alpha = 0.5$	3.1	4.9	7.0
	$\alpha = 0$	6.0	7.6	9.6
	Pre-coded GMSK: $BT = 0.5$	0.0	0.0	3.7
	$BT = 0.25$	0.0	0.0	3.7
	FQPSK: Unfiltered	0.02	1.8	4.8
	B (filtered)	0.5	2.4	5.1
Coded	T-OQPSK with RS+$(7, 1/2)$ c.c.	1.3	3.0	5.5
	Constant envelope T-OQPSK with rate 1/2, 4-state c.c.	0.0	0.0	3.7
	4D 8-PSK TCM with SRRC $\alpha = 0.35$ and (255,239) RS			
	2.0 bits/Hz	3.4	5.3	7.3
	2.5 bits/Hz	3.4	5.3	7.3

On the other hand, Table 1 also shows that the spectral efficiency is reduced by more than one-half by nonlinear amplifier operation.

VII. Conclusions

In this article, we defined a unified framework and specific metrics for comparison of different coded modulation systems in terms of power and spectral efficiency. Spectral efficiency is measured in terms of *essential bandwidth*, providing a much more informative method of comparison than just relying on compliance or noncompliance with a prescribed spectral mask.

Several proposed coded modulation systems have been considered and evaluated, with the goal of reducing the number of possible choices to a small subset, for a typical high-data-rate (10 to 100 Mb/s) deep-space mission, such as a Mars orbiter. We also computed the ultimate capacity limits in the presence of a spectral constraint to provide a valuable indication of the resulting capacity losses and of how close to optimal are the proposed systems.

When also taking into account future desires for lower BERs ($\leq 10^{-9}$) and low decoder implementation complexity, in order to allow high decoding speeds, rate 0.5 or 0.8 LDPC codes with OQPSK modulation and simple SRRC filtering provide a very appealing solution in the longer term for data rates exceeding 100 Mb/s. Punctured turbo codes and filtered OQPSK provide a useful solution for the near term, since the current DSN decoder can be used with minor modifications, although the decoding speed will be limited to a few Mb/s in this case. Iteratively decoded coded FQPSK is another interesting candidate due to its high power and spectral efficiency. This is mitigated by the fact that a new decoder needs to be designed and that very low BERs may be difficult to achieve.

Acknowledgments

We are thankful to Sam Dolinar for many useful discussions on constrained capacity and in particular for the normalization used in capacity calculations with a constraint on effective bandwidth. Jon Hamkins is acknowledged for providing the spectral characteristics of some modulations.

References

[1] D. Lee and T.-Y. Yan, "Performance of Selected Bandwidth Efficient Modulations," *CCSDS Panel Report*, Consultative Committee on Space Standards, October 1999.

[2] D. Lee, "Coded Bit Error Performance of Selected Bandwidth Efficient Modulations," *CCSDS Panel Report*, Consultative Committee on Space Standards, May 2000.

[3] D. Lee, T.-Y. Yan, and W. Martin, "Performance Comparison of 4D 8-PSK TCM and FQPSK-B in a Nonlinear Satellite Channel," *CCSDS Panel Report*, Consultative Committee on Space Standards, May 2000.

[4] A. Ribes, "8/16 PSK Multi-D Trellis Coded Modulation," presented at CCSDS Panel 1E meeting, October 1999.

[5] M. K. Simon and D. Divsalar, "A Reduced-Complexity, Highly Power-/Bandwidth-Efficient Coded Feher-Patented Quadrature-Phase-Shift-Keying System with Iterative Decoding," *The Telecommunications and Mission Operations Progress Report 42-145, January–March 2001,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, May 15, 2001.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-145/145A.pdf

[6] M. K. Simon and D. Divsalar, "Further Results on a Reduced-Complexity, Highly Power-/Bandwidth-Efficient Coded Feher-Patented Quadrature-Phase-Shift-Keying System with Iterative Decoding," *The InterPlanetary Network Progress Report 42-146, April–June 2001,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–7, August 15, 2001.
http://ipnpr.jpl.nasa.gov/tmo/progress_report/42-146/146I.pdf

[7] G. K. Kaleh, "Simple Coherent Receivers for Partial Response Continuous Phase Modulations," *IEEE Transactions on Selected Areas in Communications,* vol. 7, no. 9, pp. 1427–1436, December 1989.

[8] M. K. Simon and T.-Y. Yan, "Unfiltered Feher-Patented Quadrature Phase-Shift-Keying (FQPSK): Another Interpretation and Further Enhancements: Part 1," *Applied Microwave and Wireless Magazine*, vol. 12, pp. 76–96, February/March 2000.

[9] M. K. Simon, P. Arabshahi, and M. Srinivasan, "Trellis-Coded Quadrature-Phase-Shift Keying (QPSK) with Variable Overlapped Raised-Cosine Pulse Shaping," *The Telecommunications and Mission Operations Progress Report 42-136, October–December 1998,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–16, February 15, 1999.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-136/136F.pdf

[10] C.-P. Liang, J. Jong, W. E. Start, and J. R. East, "Nonlinear Amplifier Effects in Communications Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1461–1466, August 1999.

[11] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, Boston, Massachusetts: Artech House, 1999.

[12] R. E. Blahut, *Digital Transmission of Information,* Reading, Massachusetts: Addison Wesley, 1990.

[13] L. H. Ozarow, A. D. Wyner, and J. Ziv, "Achievable Rates for a Constrained Gaussian Channel," *IEEE Transactions on Information Theory,* vol. 34, issue 3, pp. 365–370, May 1988.

[14] Space Frequency Coordination Group, "Recommendation 21-4: Efficient Spectrum Utilization for Space Research Service, Deep Space (Cat. B), in the Space-To-Earth Link," October 4, 2001.
www.sfcgonline.org/handbook/rec/index.shtml

[15] Flarion Inc., Web page, www.flarion.com/products/ldpc_data_sheet.pdf

[16] E. A. Lee and D. B. Messerschmitt, *Digital Communication,* 2nd. edition, Norwell, Massachusetts: Kluwer Academic Publishers, 1994.

[17] Y. Kou, S. Lin, and M. P. C. Fossorier, "Low-Density Parity-Check Codes Based on Finite Geometries: A Rediscovery and New Results," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2711–2736, November 2001.

[18] R. McEliece and R. Palanki , "Intersymbol Interference in Pulse-Amplitude Modulation Signaling Systems Satisfying a Spectral Mask Constraint," *Interplanetary Network Progress Report 42-150, April–June 2002,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–14, August 15, 2002.
http://ipnpr/progress_report/42-150/150M.pdf

[19] "White Paper for Low Density Parity Check (LDPC) Codes for CCSDS Channel Coding Blue Book," presented by NASA Goddard Space Flight Center to CCSDS panel P1B, October 2, 2002.

[20] A. J. Viterbi, *Principles of Coherent Communications,* New York: McGraw-Hill, 1966.

[21] I. Bar-David and S. Shamai, "On Information Transfer by Envelope-Constrained Signals over the AWGN Channel," *IEEE Transactions on Information Theory,* vol. 34, no. 3, pp. 371–379, May 1988.

[22] S. Shamai and I. Bar-David, "On the Capacity Penalty due to Input-Bandwidth Restrictions with an Application to Rate-Limited Binary Signaling," *Corresp.,* vol. 36, no. 3, pp. 623–627, May 1990.