The Theoretical Limits of Source and Channel Coding

S. J. Dolinar and F. Pollara
Communications Systems Research Section

This article presents the theoretical relationship among signal power, distortion, and bandwidth for several source and channel models. It is intended as a reference for the evaluation of the performance of specific data compression algorithms.

I. Introduction

The theoretical limits on the performance of source and channel coding are well known for several source and channel models [1,3]. In this article these limits are calculated for the Gaussian channel, used as a model of the deep-space channel, and for some simple sources, potentially useful as models of planetary images. The formulas underlying these calculations are well known; the aim of this article is to collect and graphically display the results. The performance of specific data compression algorithms can be compared to the ultimate limits shown in these graphs. Similar results were presented in [2] for a binary symmetric source with probability of error distortion criterion.

The results show the basic tradeoffs that must be made among signal power, distortion, and bandwidth, assuming that the system designer is free to design encoders and decoders but has no control over the source, the channel, or the user. These results also suggest that improvements in information transmission in future missions should be sought primarily through better source encoding rather than by pushing channel coding gain closer to its limit.

II. The Communication System

Consider a communication system as shown in Fig. 1. The source produces *source symbols* x with average energy $\mathcal{E}_x$ at the rate of R_x source symbols per second. The source encoder outputs bits b with average energy $\mathcal{E}_b$ at the rate of R_b bits/sec. The channel encoder produces *channel symbols* y with average energy $\mathcal{E}_y$ at the rate of R_y channel symbols per second. The channel produces noisy outputs $\hat{y}$; the channel decoder and source decoder produce outputs $\hat{b}$ and $\hat{x}$, attempting to reproduce the values of b and x, respectively.

A. Channel Models

Channel capacity is defined as

$$C(\cdot) = \max I(y;\hat{y}) \tag{1}$$

where $I(y;\hat{y})$ is the average mutual information provided by the output $\hat{y}$ about the input y. The maximization is performed over all probability densities $q(y)$ satisfying the desired cost constraint $E[y^2] \leq \mathcal{E}_y$. The capacity is written as a function of an unspecified argument; it depends on the energy constraint $\mathcal{E}_y$ and the transition statistics $p(\hat{y}|y)$ that characterize the channel.

Case C1. For a discrete-time additive white Gaussian noise channel with continuous amplitude input and output, and with noise distribution $N(0,\mathcal{E}_n)$, Eq. (1) yields the capacity function in bits per channel symbol [3]:

$$C(\rho) = \frac{1}{2}\log_2(1+\rho)$$

where $\rho = \mathcal{E}_y/\mathcal{E}_n = 2\mathcal{E}_y/N_0$, and $N_0/2$ is the two-sided spectral density of the noise.

As a comparison consider two additional channel models obtained by constraining the input and/or the output of the additive white Gaussian noise channel. One channel, Case C3, is obtained by restricting the channel input to binary symbols, while leaving the channel output unconstrained. The other case, Case C2, is obtained by constraining the channel output as well as its input to be binary, that is, by hard-quantizing the Gaussian channel output to two levels.

Case C2. A hard-quantized Gaussian channel with binary signaling and signal levels $\pm\sqrt{\mathcal{E}_y}$ is a binary symmetric channel (BSC) with crossover probability $\epsilon = Q(\sqrt{\rho})$, where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-u^2/2}du$. The capacity for such a channel in bits per channel symbol is [3]

$$C(\rho) = 1 - \mathcal{H}[Q(\sqrt{\rho})]$$

where $\mathcal{H}(\epsilon) = -\epsilon\log_2\epsilon - (1-\epsilon)\log_2(1-\epsilon)$ is the binary entropy function.

Case C3. For the binary input Gaussian channel, the capacity can be written [2]

$$C(\rho) = 1 - E_u[\log_2(1+e^{-2u})]$$

where u is a random variable with distribution $N(\rho,\rho)$, and E_u represents expectation over u.

B. Source Models

A discrete-time, continuous-amplitude stationary source is considered. Its rate distortion function is defined as

$$R(\cdot) = \min I(x;\hat{x})$$

The minimization is over all conditional probability densities $p(\hat{x}|x)$ satisfying a distortion constraint $E[d(x,\hat{x})] \le D$, where $d(\cdot,\cdot)$ is a distortion measure. As with channel capacity, $R(\cdot)$ is written as a function of an unspecified argument; it depends on the distortion constraint D and on the source statistics $p(x)$.

Case S1. For a Gaussian memoryless source with mean square error (MSE) distortion constraint $E[z^2] \le \mathcal{E}_z$, where $z = \hat{x} - x$, the rate distortion function can be expressed in bits per source symbol as [1]

$$R(\delta) = \frac{1}{2}\log_2(1/\delta),\ 0 < \delta < 1 \quad (2)$$

where $\delta = \mathcal{E}_z/\mathcal{E}_x$ is the normalized MSE distortion.

Case S2. As a comparison consider a memoryless binary symmetric source (BSS) under the probability of error distortion measure, $\text{Prob}(\hat{x} \ne x) \le P_e$. This source has the rate distortion function [1]

$$R(P_e) = 1 - \mathcal{H}(P_e),\ 0 \le P_e \le 1/2$$

Case S3. A simple example of a source with memory is a stationary Gaussian source with MSE distortion constraint as in Case S1. In this case only a parametric form for $R(\delta)$ is known [1]:

$$\delta(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\min[\theta,\Phi_0(\omega)]d\omega$$

$$R(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\max[0,\log_2\frac{\Phi_0(\omega)}{\theta}]d\omega$$

where $\Phi_0(\omega) = \sum_{k=-\infty}^{\infty}\frac{\phi(k)}{\phi(0)}e^{-jk\omega}$, and $\frac{\phi(k)}{\phi(0)}$ is the normalized autocorrelation function of the Gaussian process. Of particular interest is the case of a first-order Markov source with autocorrelation $\phi(k) = \mathcal{E}_x\gamma^{|k|}$, $0 < \gamma < 1$, which provides a good approximation to the autocorrelation function of actual images, as shown in the example of Fig. 2. The rate distortion function for a Gauss–Markov source is shown in Fig. 3 for several values of γ. Note that, at any given distortion level, the rate R varies insignificantly with γ for moderate values of the correlation coefficient ($\gamma < 0.5$) but is reduced dramatically as γ approaches 1.

III. The Source–Channel Coding Theorem

Shannon's channel coding theorem and source coding theorem can be merged into the source–channel coding theorem [3], which is the central result of information theory. The source–channel coding theorem answers the following fundamental question: Given a source with rate distortion function $R(\cdot)$ and a channel with capacity function $C(\cdot)$, under what conditions is it possible, with sufficient coding, to achieve source distortion and channel cost that do not exceed the constraints used in the definitions of $R(\cdot)$ and $C(\cdot)$, respectively? According to the theorem, this is always possible if the capacity $C(\cdot)$ is greater than the rate distortion $R(\cdot)$ (where the two functions must be measured in consistent units), and never possible if the rate distortion is greater than the capacity.

In general, the source symbol rate R_x need not correspond to the channel symbol rate R_y. Therefore the requirement on the information transmission system is normalized as

$$C(\cdot)R_y > R(\cdot)R_x$$

or

$$C(\cdot) > rR(\cdot) \tag{3}$$

where $r = R_x/R_y$ is the number of source symbols per channel symbol. A large r corresponds to a small channel bandwidth requirement relative to the bandwidth of the source. To get an idea of the practical range of interest of r, consider the following examples: A system with an 8-bit/symbol source and no source compression on a channel encoded with code rate 1/6 yields $r \simeq 0.02$; at another extreme, the same source with a source compression ratio of $8:1$ and no channel coding yields $r = 1$; the Voyager concatenated system with a source compression of $2:1$ yields $r \simeq 0.11$.

This requirement can now be applied to all the combinations of source and channel models introduced in Section II. For all three channels considered in Section II, the capacity $C(\cdot)$ is a function of a single variable ρ, which is a signal-to-noise ratio involving the energy per encoded channel symbol. Usually it is more meaningful to measure the channel cost in terms of a signal-to-noise ratio, $\rho_b = \mathcal{E}_b/\mathcal{E}_n = 2\mathcal{E}_b/N_0$, involving the energy per channel encoder input bit, or a signal-to-noise ratio, $\rho_x = \mathcal{E}_x/\mathcal{E}_n = 2\mathcal{E}_x/N_0$, involving the energy per source symbol. The variables ρ, ρ_b, and ρ_x are related by

$$r\rho_x = \frac{R_x}{R_y}\rho_x = \rho = \frac{R_b}{R_y}\rho_b = r_c\rho_b = rr_s\rho_b$$

where $r_c = R_b/R_y$ is the rate of the channel code, and $r_s = R_b/R_x$ is the rate of the source code. Therefore the capacity constraint, Eq. (3), can be equivalently written in terms of ρ_x as

$$C(r\rho_x) > rR(\cdot)$$

or, in terms of ρ_b as

$$C(rr_s\rho_b) > rR(\cdot)$$

With r_s taken to equal its minimum allowable value, $R(\cdot)$, the latter requirement is

$$C(rR(\cdot)\rho_b) > rR(\cdot)$$

These formulas give the required relationships among r, ρ_b or ρ_x, and the unspecified distortion variable in $R(\cdot)$.

Case S1.C1. A Gaussian channel and a Gaussian memoryless source under the MSE distortion measure gives the requirement

$$\rho > \delta^{-r} - 1 \tag{4}$$

For a given source and channel the designer seeks a realization with small δ and ρ, and large r. Of course these are conflicting requirements, limited by Eq. (4), and illustrated in Figs. 4(a), 4(b), and 4(c), where the feasible region is above each curve with a given parameter r. When the source and channel have the same clock rate ($r = 1$) and $\rho + 1 = 1/\delta$, the source and channel are perfectly matched, and ideal performance $C = R$ is obtained without any need for source or channel coding, assuming proper scaling of the signal levels. If the channel clock rate is much higher than the source clock rate, the required $\mathcal{E}_b/N_0$ approaches the well known infinite bandwidth expansion limit of -1.59 dB.

Case S2.C1. For a BSS and an unconstrained Gaussian channel the following requirement, shown in Figs. 5(a), 5(b), and 5(c), is obtained:

$$\rho > 2^{2r[1-\mathcal{H}(P_e)]} - 1$$

In this case the distortion P_e is extremely sensitive to small changes in $\mathcal{E}_b/N_0$ or r. Conversely, extremely low error probabilities can be obtained as long as $\mathcal{E}_b/N_0$ is greater than the channel capacity limit for any fixed channel code rate, $r_c = r$.

Case S1.C2. A Gaussian memoryless source and a BSC gives the requirement

$$2^{[1-\mathcal{H}(\epsilon)]} > \delta^{-r/2}$$

which relates the channel cost $\epsilon = Q(\sqrt{\rho})$ to the normalized MSE distortion δ, as in Figs. 6(a), 6(b), and 6(c). The asymptotic behavior of these graphs for large signal-to-noise ratio is due to the hard quantization introduced by the BSC. The quantization loss can be seen by comparing Fig. 6(a) to Fig. 4(a), Fig. 6(b) to Fig. 4(b), or Fig. 6(c) to Fig. 4(c). The required $\mathcal{E}_b/N_0$ at infinite bandwidth expansion is 0.37 dB, or 1.96 dB higher than the corresponding requirement for the unquantized channel.

Case S2.C2. For a BSS and BSC under the P_e distortion measure, the requirement is

$$1 - \mathcal{H}(\epsilon) > r[1 - \mathcal{H}(P_e)]$$

as illustrated in Figs. 7(a), 7(b), and 7(c). When $r = 1$ and $\epsilon = P_e$, another case of perfect matching between source and channel is obtained, which yields ideal performance $C = R$ without any need for source or channel coding. The loss due to the hard quantization of the channel can be seen by comparing Fig. 7(a) to Fig. 5(a), or Fig. 7(b) to Fig. 5(b), or Fig. 7(c) to Fig. 5(c).

Case S2.C3. The cases involving the binary input Gaussian channel do not yield a simple analytical expression. The results are shown in Figs. 8(a), 8(b), and 8(c), for this channel combined with the BSS. This source–channel combination is the same case considered in [2], and Fig. 8(c) is directly comparable to Fig. 1 of [2]. (Note that E_b/N_0 in [2] refers to the source symbol signal-to-noise ratio denoted $\mathcal{E}_x/N_0$ in this article.)

Case S3.C1. Finally, the case of a Gauss–Markov source and Gaussian channel is shown in Figs. 9, 10, and 11, for three selected values of the rate r and several values of the correlation coefficient γ. It is apparent that a Gauss-Markov source with $\gamma = 0.99$, which is a good model for some highly correlated planetary images, offers large savings in signal power due to the high redundancy of the source.

IV. Conclusion

The results presented provide a reference against which the performance of specific data compression algorithms can be measured. Choosing an appropriate model from Figs. 4–11, and a desired ratio r of source symbol rate to channel symbol rate, one can determine the minimum channel symbol signal-to-noise ratio $\mathcal{E}_y/N_0$, bit signal-to-noise ratio $\mathcal{E}_b/N_0$, or source symbol signal-to-noise ratio $\mathcal{E}_x/N_0$ required to produce a normalized source distortion that is less than δ or P_e.

References

[1] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, New Jersey: Prentice Hall, 1971.

[2] S. A. Butman and R. J. McEliece, "The Ultimate Limits of Binary Coding for a Wideband Gaussian Channel," *DSN Progress Report 42-22*, vol. May–June 1974, Jet Propulsion Laboratory, Pasadena, California, pp. 78–80, August 15, 1974.

[3] R. J. McEliece, *The Theory of Information and Coding*, Reading, Massachusetts: Addison-Wesley, 1977.

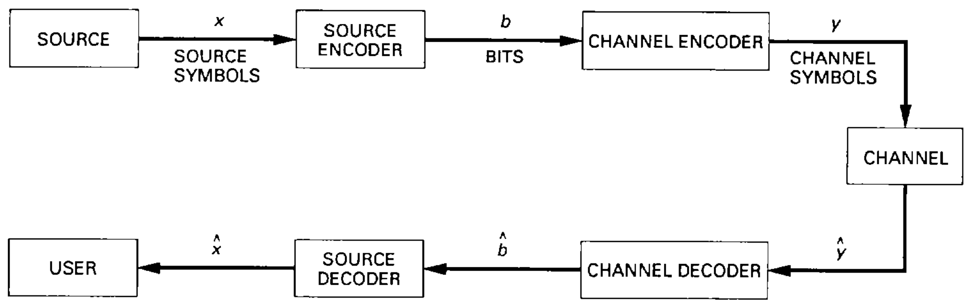

Fig. 1. A communication system model.

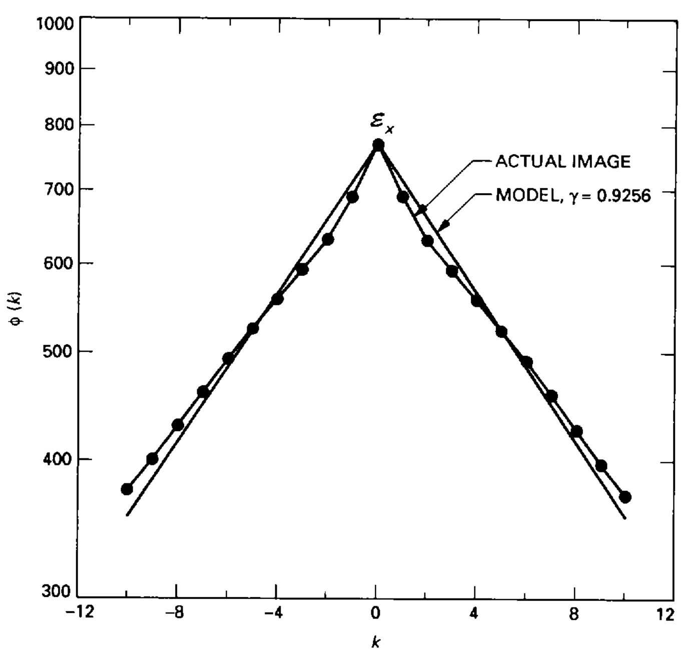

Fig. 2. Autocorrelation functions of a first-order Markov model and of an image of the moon.

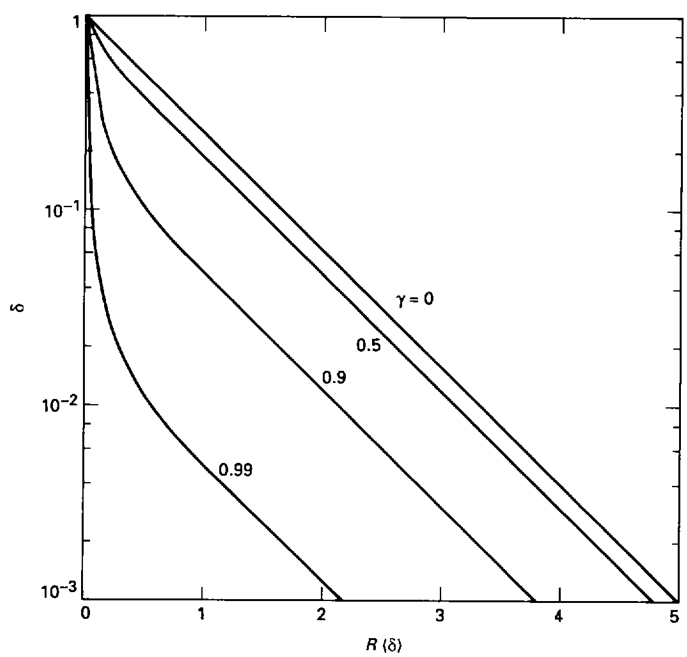

Fig. 3. Rate distortion function for a Gauss–Markov source with correlation γ.

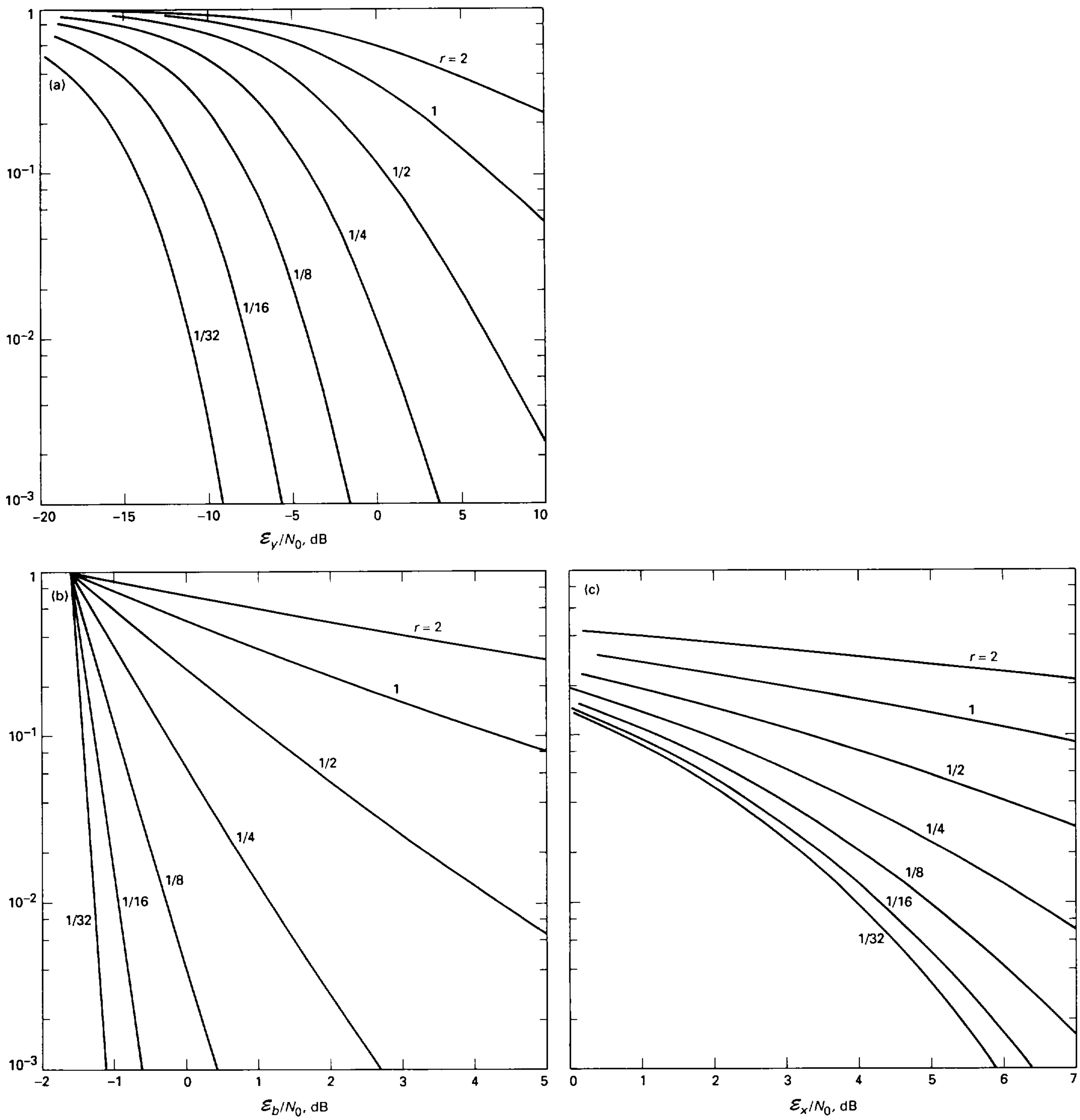

Fig. 4. Source-channel coding limits for a Gaussian source and a Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

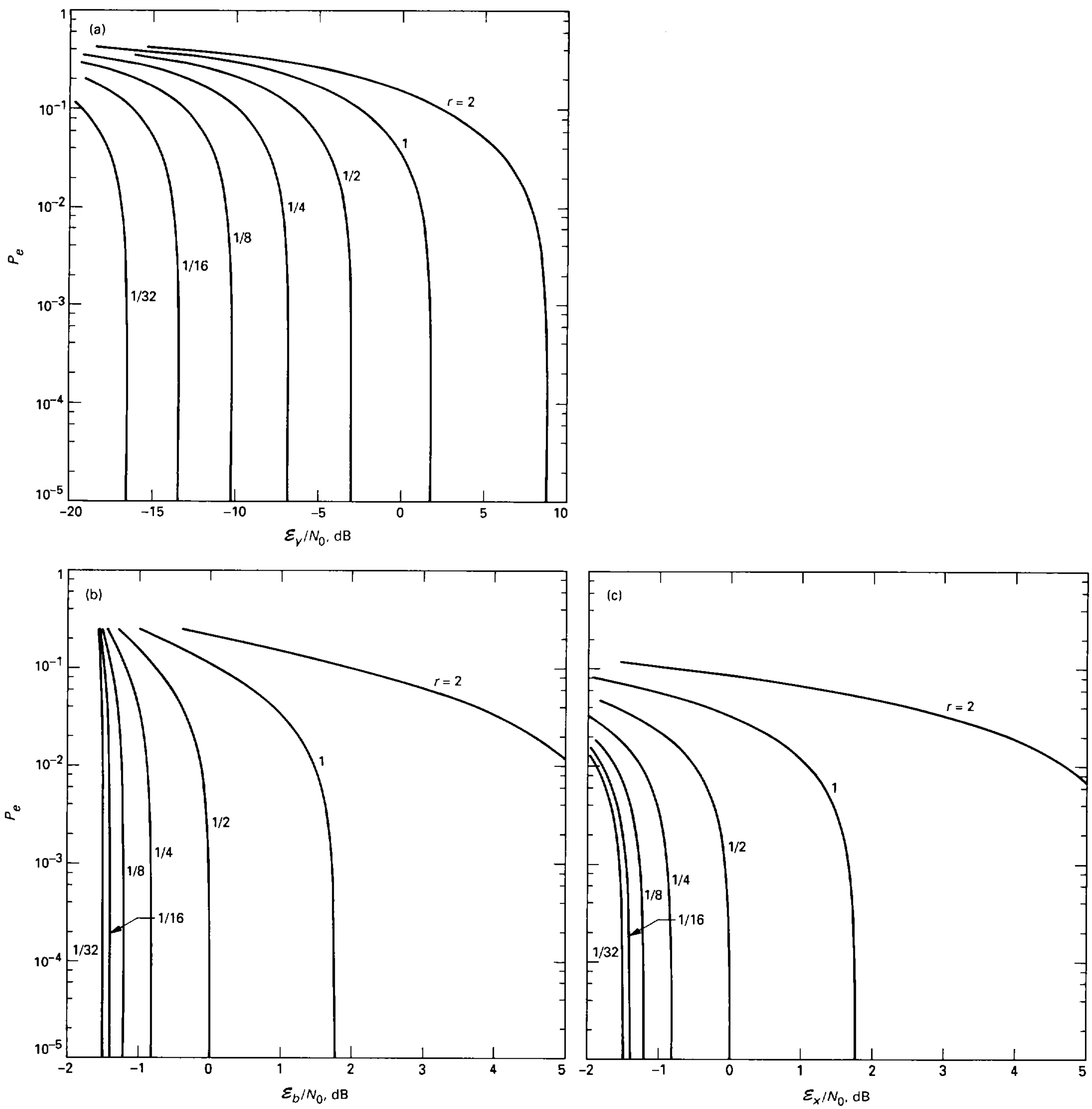

Fig. 5. Source-channel coding limits for a binary symmetric source (BSS) and a Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

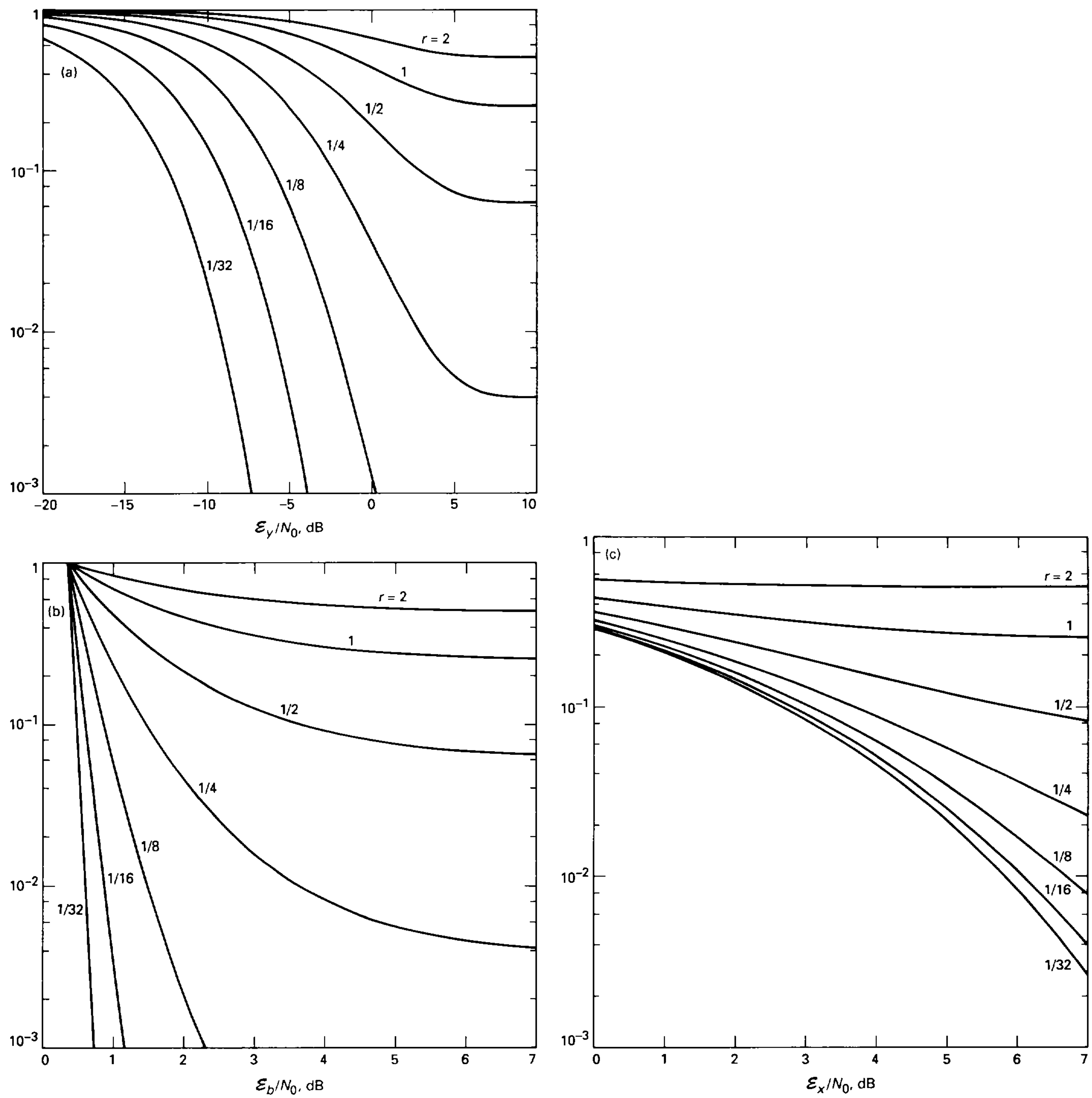

Fig. 6. Source-channel coding limits for a Gaussian source and a hard-quantized Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

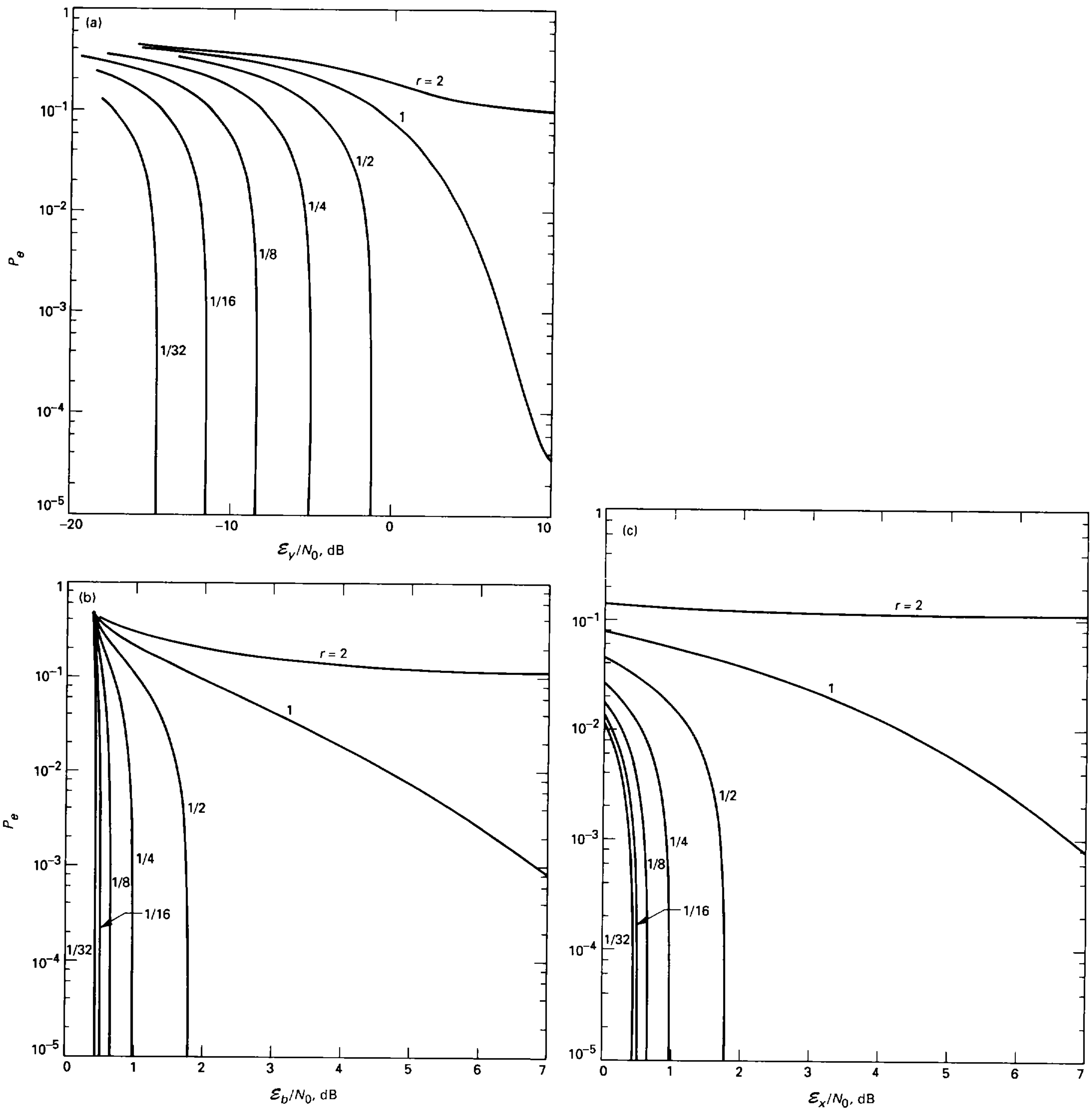

Fig. 7. Source-channel coding limits for a binary symmetric source (BSS) and a hard-quantized Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

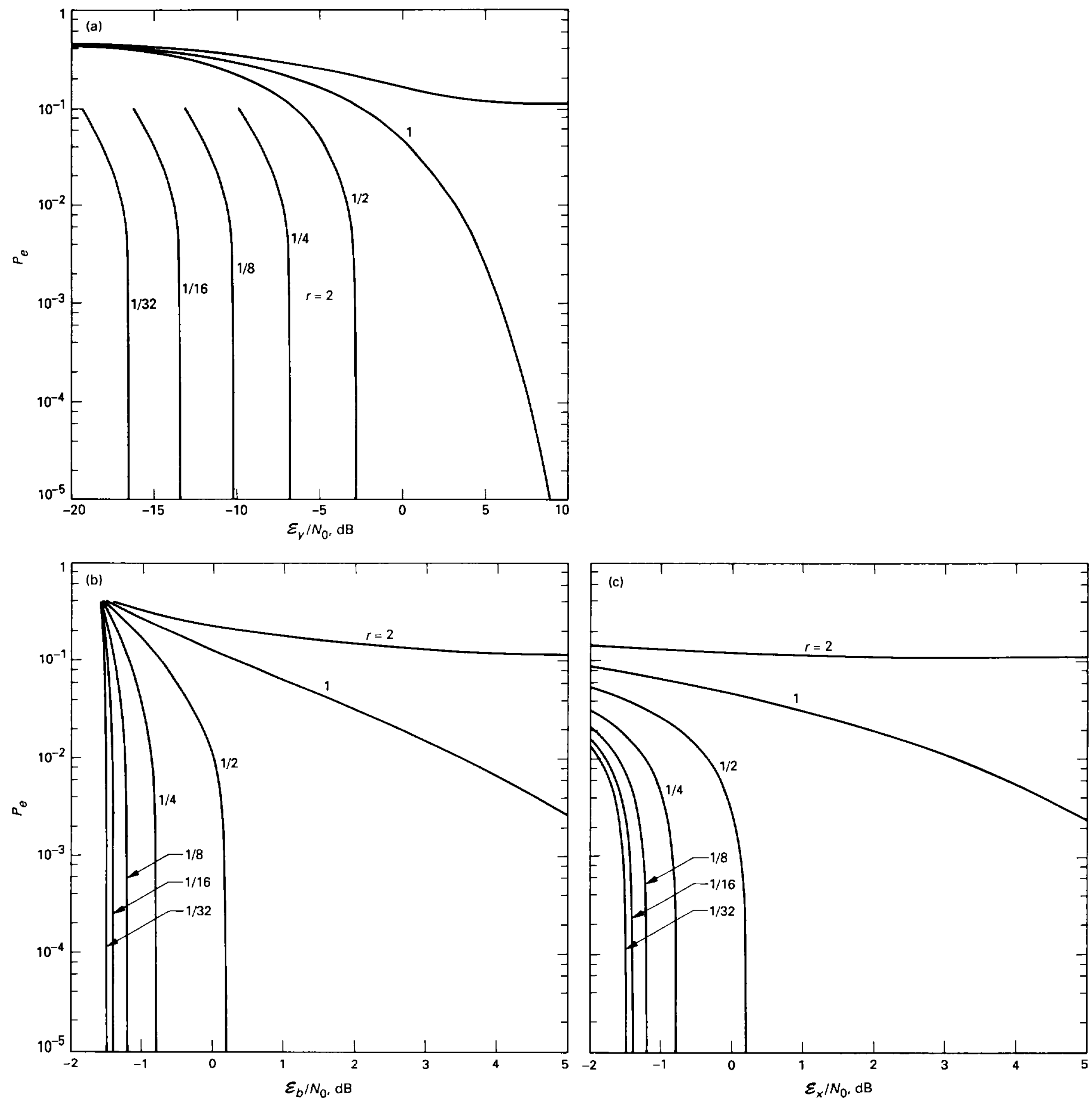

Fig. 8. Source-channel coding limits for a binary symmetric source (BSS) and a binary input Gaussian channel: distortion versus (a) channel symbol signal-to-noise ratio; (b) source-encoded bit signal-to-noise ratio; and (c) source symbol signal-to-noise ratio.

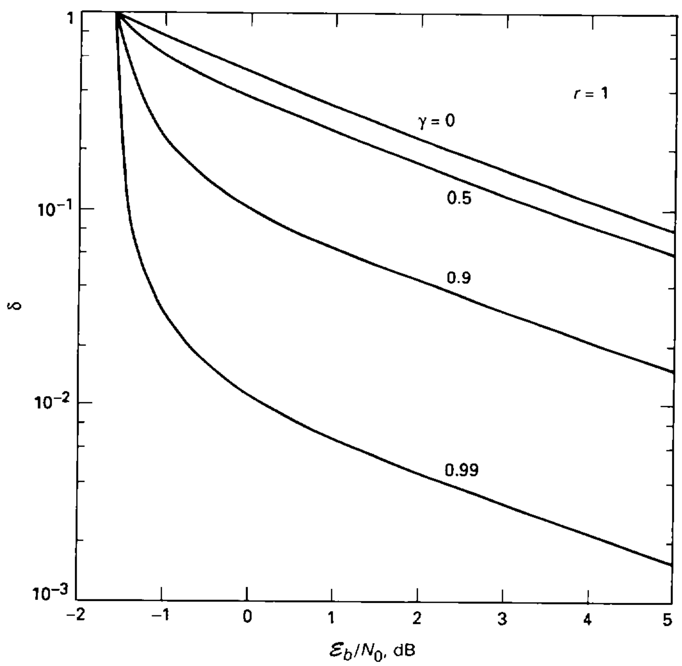

Fig. 9. Source-channel coding limits for a Gauss–Markov source and a Gaussian channel ($r = 1$): distortion versus source-encoded bit signal-to-noise ratio.

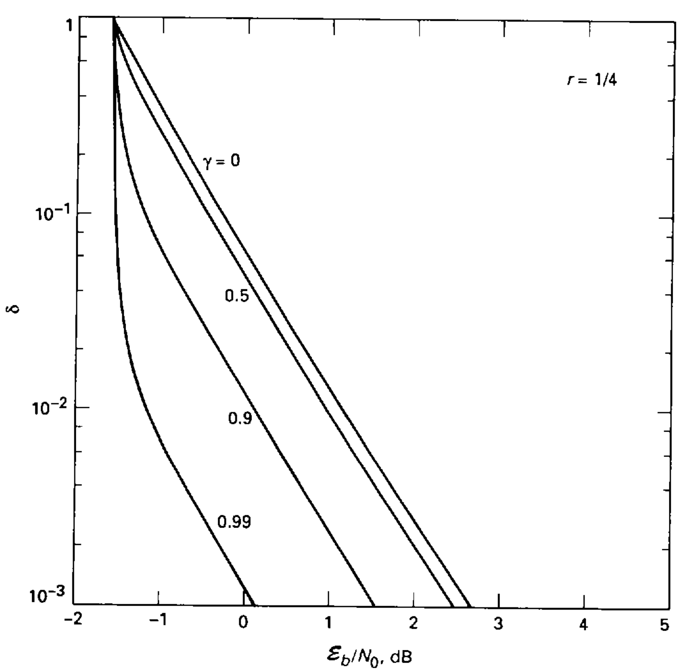

Fig. 10. Source-channel coding limits for a Gauss–Markov source and a Gaussian channel ($r = 1/4$): distortion versus source-encoded bit signal-to-noise ratio.

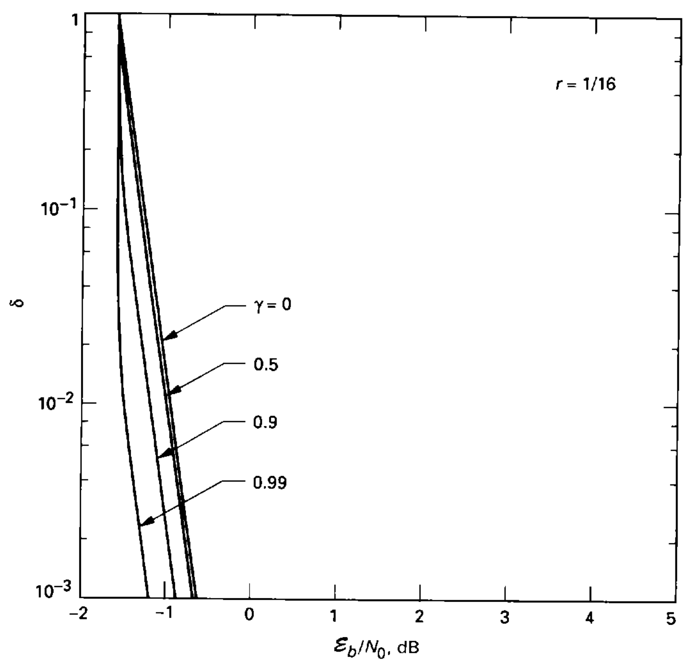

Fig. 11. Source-channel coding limits for a Gauss–Markov source and a Gaussian channel ($r = 1/16$): distortion versus source-encoded bit signal-to-noise ratio.