[6] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[7] D. Catrein, L. Imhof, and R. Mathar, "Power Control, Capacity, and Duality of Up- and Downlink in Cellular CDMA Systems," RWTH Aachen, Germany, Tech. Rep., 2003.
[8] C. Sung, "Log-convexity property of the feasible SIR region in power-controlled cellular systems," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 248–249, June 2002.
[9] S. Stanczak, "On sequence sets for symbol-asynchronous CDMA channels with fixed time offsets," in *Proc. 37th Annu. Conf. Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2003.
[10] H. Boche and S. Stańczak, "Iterative algorithm for finding resource allocation in symbol-asynchronous CDMA channels with different SIR requirements," in *Proc. 36th Asilomar Conf. Signals, Systems, and Computers*, Monterey, CA, Nov. 03–06, 2002, pp. 1909–1913.
[11] H. Boche and M. Schubert, "Solution of the SINR downlink beamforming problem," in *Proc. Conf. Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2002.
[12] ——, "The maximal throughput of cellular systems with multiple basestation antennas," in *Proc. 57th IEEE Semiannual Vehicular Technology Conf.*, Jeju, Korea, Apr. 22–25, 2003.
[13] M. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication-Part I: System analysis," *IEEE Trans. Commun.*, vol. COM-25, no. 8, pp. 795–799, Aug. 1977.
[14] H. Boche and S. Stańczak, "Optimal allocation of resources in an asynchronous CDMA channel with identical SINR requirements for all users," *IEICE Trans. Commun.*, vol. E86-B, no. 1, pp. 397–405, Jan. 2003.
[15] R. Cruz and A. Santhanam, "Optimal routing, link scheduling and power control in multi-hop wireless networks," in *Proc. 22nd IEEE Conf. Computer Communications (INFOCOM)*, San Francisco, CA, Mar. 30– Apr. 3, 2003, pp. 702–711.
[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[17] E. Seneta, *Non-Negative Matrices and Markov Chains*. Berlin, Germany: Springer Verlag, 1981.

Coefficient Rate and Lossy Source Coding

Wenye Yang, *Member, IEEE*, Jerry D. Gibson, *Fellow, IEEE*, and Tao He, *Student Member, IEEE*

***Abstract*—Campbell derived and defined a quantity called the coefficient rate of a random process that involves the process spectral entropy. In this correspondence, his interpretation is substantiated with two new derivations. One derivation tightens the connection to source bandwidth, while the second derivation implies a specific approach to adaptive coefficient selection in realization-adaptive approaches to compression. After a discussion on the role the coefficient rate plays in adaptive source coding, a quantity called Campbell bandwidth is defined based on its connection to source bandwidth and is contrasted with Fourier bandwidth and Shannon bandwidth. The connection between coefficient rate and reverse water-filling from rate distortion theory is also demonstrated.**

***Index Terms*—Adaptive coefficient selection, Campbell bandwidth, coefficient rate, Shannon bandwidth.**

Manuscript received February 8, 1999; revised December 18, 2002. The material in this correspondence was presented in part at the IEEE Data Compression Conference, Snowbird, UT, March 2003.

W. Yang is with Conexant Systems, Inc. Red Bank, NJ 07701 USA.

J. D. Gibson is with Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106-6065 USA.

T. He is with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275 USA.

Communicated by P. A. Chou, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2004.839531

I. Introduction

The coefficient rate of a random process was first derived and defined by Campbell in 1960 [1]. Campbell considered the product of N sample functions of a random process, and showed, using an asymptotic equipartition property (AEP)-like argument, that a Karhunen–Loéve expansion (K-L expansion) of this product could be separated into two sets—one set with average power very close to that of the product and the other set having very low average power. Asymptotically, in the number of sample functions forming the product and in the support interval of the process, he showed that the average number of terms in the high-power set approached a quantity that he interpreted as a coefficient rate given by

$$Q_2 = \exp\left[-\int_{-\infty}^{\infty} S(f)\log S(f)df\right] \tag{1}$$

where we denote the quantity in the exponent as the spectral entropy (we use Q_2 for the coefficient rate and reserve Q_1 for Shannon's entropy rate power [2]). The implications of coefficient rate and spectral entropy for source compression were not explored by Campbell and no coding theorems were presented. Shortly thereafter, Abramson [3] examined Campbell's coefficient rate and noted that an approach to source compression drawing on Campbell's result did not exist and was not immediately evident.

Thirty years later, the spectral entropy began to attract attention for speech and image compression. Motivated by Campbell's paper, Gibson, *et al* [4] compared Campbell's coefficient rate and Shannon's entropy rate power for autoregressive (AR) processes and for speech classification, and McClellan and Gibson [5]–[7] used spectral entropy to develop adaptive speech coders. Around the same time, other researchers [8]–[10] were investigating spectral entropy and a quantity similar to coefficient rate for compression applications, but apparently without knowledge of Campbell's paper.

In particular, for a two-dimensional discrete cosine transform (DCT) based coding scheme, Mester and Franke [8] use two factors to classify data blocks so that different types of data blocks can use different coding strategies. For two-dimensional transform coefficients $C(i,j)$, they define an activity measure

$$A = \sum_{i,j} |C(i,j)|$$

and the spectral entropy

$$E = -\sum_{i,j} a(i,j)\log a(i,j)$$

where $a(i,j) = \frac{|C(i,j)|}{A}$, the normalized coefficient magnitude. These two quantities are used to classify the data block. The activity measure A reflects the total energy of the data block, which may indicate that the number of coefficients to be coded is large when A is high. On the other hand, the spectral entropy E acts as a measure of the distribution of the energies, so if energy is concentrated in a single coefficient, the spectral entropy will be small, which indicates that the number of coefficients to be coded should be small.

A similar approach is taken by Coifman and Wickerhauser in [9], although the goal of their method is different. For transform coefficients x_n, Coifman and Wickerhauser define the *theoretical dimension* of signals as

$$d = \exp\left(-\sum_n p_n \log p_n\right)$$

where $p_n = \frac{|x_n|^2}{\|x\|^2}$, and the exponent is the spectral entropy. They use d as a measure of the number of coefficients to be coded in a wavelet transform, and by minimizing d, the best wavelet packet basis can be chosen as the best transform basis since it produces the minimum number of coefficients. The same method is extended to other transformations for speech processing in [10].

Perhaps the most significant development in the field of image compression in recent years has been the discovery of wavelet-based, tree-structured methods for still-image compression [11], [12]. These methods have three notable features: 1) They are input, or sample function adaptive; 2) they are nonlinear approximation methods in that the best n basis functions are chosen for the current input, rather than the first n [13]; and 3) they do not require the explicit coding of side information.

Ortega and Ramchandran [14] call such approaches "input-by-input" coding and they state, "However, even a narrowly defined 'class' of inputs will likely show significant variations among inputs (...), and thus techniques that allow an 'input-by-input' parameter selection are likely to be superior. . . ." Further, Effros [15] states, ". . .source-dependent components of a compression system are designed to do well on *average* across the training set but may not achieve, on any particular member of that set, performance as good as the performance achievable with a code designed specifically for that data."

Realization-adaptive compression and nonlinear approximation do not fit neatly into classical rate distortion theory. An additional challenge to applying classical rate distortion theory is that many practical coders operate in the low bit rate or high distortion region, where much less is known about rate distortion optimal compression [16].

In this correspondence, we return to Campbell's original work and substantiate his interpretation of Q_2 as a coefficient rate with two new derivations of this quantity. One derivation tightens the connection to source bandwidth, while the second derivation implies a specific approach to adaptive coefficient selection in realization-adaptive approaches. We present an example demonstrating the role coefficient rate plays in realization-adaptive methods. Based upon the coefficient rate connection to source bandwidth, we define a quantity called Campbell bandwidth and contrast Campbell bandwidth with Fourier bandwidth and Shannon bandwidth, as defined by Massey [17]. Finally, a connection between coefficient rate and reverse water-filling is established.

II. Campbell's Minimum Coefficient Rate [1]

The idea behind Campbell's approach is intuitively quite straightforward. When a sequence of samples of a random variable is to be coded, it is generally true that concatenating samples and coding them together can achieve a lower code rate. When the number of samples concatenated gets very large, the code rate can approach its lower bound, the *entropy* of the random variable, as close as possible. The same idea is used by Campbell when a number of sample functions of a random process are considered.

Consider a zero-mean stationary continuous-time random process $X(t)$. Using the K-L expansion, in time duration $[0, T]$, the process can be decomposed as

$$X(t) = \sum_{i=1}^{\infty} C_i \phi_i(t) \tag{2}$$

where $\phi_i(t)$'s are normalized eigenfunctions of the expansion and C_i's are uncorrelated random variables with zero mean and $\boldsymbol{E}[C_i^2] = \lambda_i$. In other words, the random process can be represented by a random vector $\{C_1, C_2, \ldots\}$. Also, the total average energy of the process is $T = \sum_i \lambda_i$.

Let

$$x_1(t_1), x_2(t_2), \ldots, x_N(t_N), \qquad 0 \le t_j \le T,\ j = 0, 1, \ldots, N$$

be N independent sample functions of the process, and assuming there is only a finite number M of components in the K-L expansion, these sample functions can be expressed as

$$x_j(t_j) = \sum_{i=1}^{M} c_{ij} \phi_i(t_j), \qquad j = 1, 2, \ldots, N$$

where $c_{ij}, j = 1, 2, \ldots, N$ are N samples of random variable C_i.

Let

$$y(t_1, t_2, \ldots, t_N) = x_1(t_1) x_2(t_2) \cdots x_N(t_N)$$

be the product of these N sample functions, and since these sample functions are independent, $y(t_1, t_2, \ldots, t_N)$ can be expressed in terms of products of the eigenfunctions of the K-L expansion

$$\begin{aligned} y(t_1, t_2, \ldots, t_N) &= x_1(t_1) x_2(t_2) \cdots x_N(t_N) \\ &= \prod_{j=1}^{N} \left[\sum_{i=1}^{M} c_{ij} \phi_i(t_j) \right] \\ &= \sum_{i_1, i_2, \ldots, i_N = 1}^{M} c_{i_1} \cdots c_{i_N} \phi_{i_1}(t_1) \cdots \phi_{i_N}(t_N) \\ &= \sum_{k=1}^{M^N} c^{(k)} \phi^{(k)}(t_1, t_2, \ldots, t_N) \end{aligned} \tag{3}$$

where $c^{(k)}$ is the product of c_{i_j} and these coefficients are arranged in decreasing order of their variances, and $\phi^{(k)}(t_1, t_2, \ldots, t_N)$ is also reordered accordingly.

Campbell approximated $y(t_1, t_2, \ldots, t_N)$ by choosing $\mu (< M^N) c^{(k)}$'s with largest variances such that the average energy loss of the approximation is small. Since the $c^{(k)}$'s are ordered according to their variances, the approximation is just

$$y_\mu(t_1, t_2, \ldots, t_N) = \sum_{k=1}^{\mu} c^{(k)} \phi^{(k)}(t_1, t_2, \ldots, t_N) \tag{4}$$

and the average energy loss of the approximation satisfies

$$\begin{aligned} &\frac{1}{T^N} \boldsymbol{E}[y(t_1, t_2, \ldots, t_N) - y_\mu(t_1, t_2, \ldots, t_N)]^2 \\ &\qquad = \frac{1}{T^N} \sum_{k=\mu+1}^{M^N} \mathrm{E}[c^{(k)}]^2 < \epsilon. \end{aligned} \tag{5}$$

Using an AEP-like argument, Campbell showed that the number μ satisfies the following asymptotic relation:

$$\frac{1}{N} \log \mu \to H\left(\frac{\lambda}{T}\right) = -\sum_{i=1}^{M} \frac{\lambda_i}{T} \log \frac{\lambda_i}{T} \tag{6}$$

where $H\left(\frac{\lambda}{T}\right)$ is the spectral entropy of the source in discrete form. Using a limiting theorem from Toeplitz integral equations [18], as $N \to \infty$ and $T \to \infty$, (6) becomes

$$\lim_{T \to \infty} \lim_{N \to \infty} \log \frac{\mu^{1/N}}{T} = -\int_{-\infty}^{\infty} S(f) \log S(f) df \tag{7}$$

where $S(f)$ is the normalized power spectral density of the random process and the right-hand side of (7) is the differential entropy of the

spectrum, or the spectral entropy of the random process, which we denote as $h(S)$. Campbell then argued that $[\mu^{1/N}]/T$ is the number of coefficients per dimension per unit time, so he defined the minimum coefficient rate of a random process as

$$Q_2 = \exp\left[-\int_{-\infty}^{\infty} S(f)\log S(f)df\right]. \tag{8}$$

A simple example of Campbell's result is when the power spectral density of a source is rectangular for $|f| \leq W$, then Q_2 is just the Nyquist sampling rate, i.e., $Q_2 = 2W$. Furthermore, if $S(f)$ is triangular, the coefficient rate is $Q_2 = \sqrt{e}W = 1.648W$. This shows that the coefficient rate is somehow related to the signal bandwidth. Using a more direct AEP approach in Section III, we get a similar result that directly relates Q_2 with signal bandwidth.

It is also important to point out the relationship of coefficient rate and classical rate distortion theory results. For continuous time, band-limited sources, classical rate distortion theory assumes Nyquist sampling and then proceeds to specify the minimum number of bits per sample (or bits per coefficient here) required to represent the source with the desired fidelity. On the other hand, the coefficient rate is not a source coding result at all. It is a statement about the minimum number of coefficients per second or samples per second required to represent a random process in the sense of the number of terms needed to approximate the energy in the product of sample functions, as described in this section. There is a tradeoff between sampling rate and coding accuracy when encoding continuous-time sources, and Abramson [3] comments on this tradeoff in his paper. Interestingly, coefficient rate as specified by Campbell gives us an analytical indicator of the required minimum rate that was previously unavailable.

III. The Equivalent Bandwidth Explanation

Consider the same stationary random process $X(t)$ with a normalized spectrum $S(f)$. Since $S(f)$ is normalized, it can be treated as a probability density function and probability can be defined as

$$P(F) = \int_F S(f)df \tag{9}$$

where F is a set defined along the frequency axis, and $P(F)$ is the power of the process in the frequency band defined by F. According to the AEP for continuous random variables [19], the volume of a N-dimensional set $F^{(N)} \in \boldsymbol{R}^N$ is defined as

$$\mathrm{Vol}(F^{(N)}) = \int_{F^{(N)}} df_1 df_2 \ldots df_N. \tag{10}$$

Given the physical meaning of $S(f)$, the volume of set $F^{(N)}$ has a meaning related to signal bandwidth. For example, in one dimension, if F is a continuous set, $\mathrm{Vol}(F)$ is just the bandwidth of the signal; if F consists of several separate subsets along the frequency axis, then $\mathrm{Vol}(F)$ is the sum of the bandwidths of these subsets. Now let the support of $S(f)$ be the edges of $F^{(N)}$ in all dimensions, then $\mathrm{Vol}(F^{(N)}) = 1$.

The typical set $F_\epsilon^{(N)}$ is then defined such that

$$P\left(F_\epsilon^{(N)}\right) > 1 - \epsilon \tag{11}$$

that is, $F_\epsilon^{(N)}$ contains most of the power in set $F^{(N)}$. Using the AEP, the volume of the typical set $F_\epsilon^{(N)}$ that contains most of the power satisfies

$$(1-\epsilon)\mathrm{e}^{N(h(S)-\epsilon)} \leq \mathrm{Vol}\left(F_\epsilon^{(N)}\right) \leq \mathrm{e}^{N(h(S)+\epsilon)} \tag{12}$$

where $h(S)$ is the spectral entropy as defined in (7). As $N \to \infty$, $(1-\epsilon)^{\frac{1}{N}} \to 1$, so on a per-dimension basis, the equivalent bandwidth of the random process, $W_e = (1/2)\left[\mathrm{Vol}\left(F_\epsilon^{(N)}\right)\right]^{1/N}$, satisfies

$$\mathrm{e}^{h(S)-\epsilon} \leq 2W_e \leq \mathrm{e}^{h(S)+\epsilon}. \tag{13}$$

If we call $2W_e$ the equivalent rate of the random process, i.e., $R_e = 2W_e$, then the equivalent rate of a random process is within a small range around $\mathrm{e}^{h(S)}$, which is the coefficient rate derived by Campbell.

IV. Dominant Terms in Campbell's Approximation

We now explore Campbell's result further to estimate asymptotically the variance of the surviving terms after truncation in (4) as $N \to \infty$ and to calculate the number of these terms. Since all of the sample functions are independent, for any random variable

$$C^{(k)} = C_{i_1} C_{i_2} \cdots C_{i_N}$$

in which each C_{i_j} is one of the random variables in the K-L expansion,

$$\mathrm{E}[(C^{(k)})^2] = \mathrm{E}\left[C_{i_1}^2\right]\mathrm{E}\left[C_{i_2}^2\right]\cdots\mathrm{E}\left[C_{i_N}^2\right] = \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_N}.$$

So the total average energy of the product

$$\begin{aligned}\mathrm{E}\left[y(t_1,t_2,\ldots,t_N)^2\right] &= \sum_{k=1}^{M^N}\mathrm{E}\left[(C^{(k)})^2\right] \\ &= \sum_{i_1,i_2,\ldots,i_N=1}^{M}\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_N} \\ &= \prod_{j=1}^{N}\sum_{i=1}^{M}\lambda_i = T^N.\end{aligned} \tag{14}$$

There are a total of M^N terms in $\mathrm{E}\left[y(t_1,t_2,\ldots,t_N)^2\right]$, but some of them are very small compared to others. As $N \to \infty$, the large energy terms will dominate the sum and small energy terms can be thrown away without significantly affecting the total energy. On the other hand, each of these M^N terms is a product of N energy terms chosen from the M K-L expansion components. Since $M < N$, the product can be rewritten as $\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}$ and we can see that there is some repetition among those M^N terms; therefore, (14) can be rewritten as

$$\begin{aligned}\mathrm{E}\left[y(t_1,t_2,\ldots,t_N)^2\right] &= \prod_{j=1}^{N}\sum_{i=1}^{M}\lambda_i \\ &= \left(\sum_{i=1}^{M}\lambda_i\right)^N \\ &= \sum_{\left\{n_i:\sum_{i=1}^{M}n_i=N\right\}}\frac{N!}{n_1!\cdots n_M!}\lambda_1^{n_1}\cdots\lambda_M^{n_M}.\end{aligned} \tag{15}$$

The coefficient before the energy term is the count of repetitions of that energy term. We want to find the largest term of $\frac{N!}{n_1!n_2!\cdots n_M!}\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}$ subject to the constraint that

$$\sum_{i=1}^{M} n_i = N.$$

As $N \to \infty$, this term grows much faster than any other term, and it dominates the total energy of $y(t_1,t_2,\ldots,t_N)$. So the problem becomes finding the maximum of the functional

$$I = \log\left\{\frac{N!}{n_1!n_2!\cdots n_M!}\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}\right\} + \alpha\sum_{i=1}^{M} n_i. \tag{16}$$

Using the approximation $\log N! = N \log N - N$ for large N [20], we have

$$I = \log N! - \sum_i \log n_i! + \sum_i n_i \log \lambda_i + \alpha \sum_i n_i$$
$$= N \log N - \sum_i n_i \log n_i + \sum_i n_i \log \lambda_i + \alpha \sum_i n_i. \quad (17)$$

Taking partial derivatives of I with respect to n_i and setting them to zero, we have

$$\frac{\partial I}{\partial n_i} = -\log n_i - 1 + \log \lambda_i + \alpha = 0$$

so, $n_i = \mathrm{e}^{\alpha - 1}\lambda_i$. By using the constraint $\sum_i n_i = N$, we obtain $\alpha - 1 = \log(N/T)$. So[1]

$$n_i = \frac{\lambda_i}{T} N, \qquad i = 1, 2, \ldots, M. \quad (18)$$

That is, the number of λ_i in $\lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M}$ is proportional to λ_i. Let $p_i = \lambda_i/T$, since $\sum_i \lambda_i = T$, $\sum_i p_i = 1$, and we have $n_i = p_i N$.

We can find the value of these energy terms and the number of them. We have

$$\sigma^2 \approx \lambda_1^{n_1}\lambda_2^{n_2}\cdots\lambda_M^{n_M} = \exp\left[\sum_i n_i \log \lambda_i\right]$$
$$= T^N \mathrm{e}^{-NH(S)} \quad (19)$$

and the number of such terms is

$$\mu \approx \frac{N!}{n_1! n_2! \cdots n_M!}$$
$$= \exp\left[-N \sum_i \frac{n_i}{N} \log \frac{n_i}{N}\right]$$
$$= \exp\left[-N \sum_i \frac{\lambda_i}{T} \log \frac{\lambda_i}{T}\right]$$
$$= \mathrm{e}^{NH(S)} \quad (20)$$

where $H(S) = -\sum_i (\lambda_i/T)\log(\lambda_i/T)$, the spectral entropy in discrete form. This agrees with Campbell's result. We also have the asymptotic relation $T^N = \mu\sigma^2$, so we can see that when $N \to \infty$, we can use these μ equal energy terms to approximate $y(t_1, t_2, \ldots, t_N)$, and the energy loss introduced by truncation is negligible.

V. Adaptive Source Compression

We have presented two alternative derivations of Campbell's coefficient rate results, and each provides insight into the physical situation of interest. To elaborate on these ideas, it is useful to recall the approach used in two-dimensional discrete cosine transform coding of images. The basic approach in classical transform coding is to take samples of an image, apply a two-dimensional discrete transform to a particular image block, assign the number of bits to be used to encode each coefficient, and then quantize and code each coefficient using the allocated number of bits [13], [21]. Coefficients allocated zero bits are not coded at all. According to rate distortion theory, when the distortion is measured using mean-square error (MSE), an optimal bit-allocation rule should be based on the so-called reverse "water-filling" result, wherein bits are assigned to a coefficient in direct proportion to the coefficient variance—a higher variance receives a greater share of the bits to be allocated. If the bit allocations are determined once and then held fixed for all encodings, the decoder can be sent this information once, and the rate required for this information is asymptotically negligible.

The bits allocated to each coefficient for a block is called the *side information*, and a binary indicator showing which coefficients are encoded and which are not transmitted at all is called the *significance map*. Most transform (or wavelet-based) coders used today adaptively allocate bits on a block-to-block basis so the side information and significance map are continuously changing. Thus, this information needs to be provided to the receiver relatively often (or surmised from the data structure).

The equivalent bandwidth result in (13) gives a connection between coefficient rate and the signal bandwidth, but this does not necessarily imply that the coefficient rate is the number of samples one should use for the random process. Instead, the sampling rate may still be the Nyquist rate, but the importance of the different samples may be different; thus, the significance map should be adjusted according to the importance of the samples.

The results on dominant terms have a more direct connection to coder design, and, in fact, imply a novel coder structure. First, observe that the results in Section IV have exactly the kind of interpretation that is expected from an AEP approach. Namely, the number of terms in the high-power set is related to the entropy of the (spectral) density as given by (20), and each coefficient in the high-power set is about the same and can be found from (19). The new implication for coding comes from (18). Equation (18) says that in a sequence of N samples of a particular coefficient, the number of coefficient samples that should be coded is proportional to the variance of the coefficient!

VI. Realization-Adaptive Coefficient Selection

The basic approach to source compression implied by the spectral entropy results is illustrated in Fig. 1. In Fig. 1(a), we show M transform coefficients or coefficients of basis functions for N frames of source data, denoted c_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, N$. The coefficients in the first frame or block are c_{i1}, $i = 1, 2, \ldots, M$, while the coefficients for the second block are c_{i2}, $i = 1, \ldots, M$, and so on. Therefore, the block index is indicated by the second subscript (j) and the coefficient index is indicated by the first subscript (i). In classical transform-based coding, coefficient bit allocation is accomplished on a frame-by-frame or block-by-block basis as illustrated in Fig. 1(b). That is, given a particular frame or block (fixed j), a fixed number of bits is allocated across the M coefficients according to their relative energies. The spectral entropy approach implies that each transform coefficient should be considered as a separate sequence, c_{ij}, $j = 1, \ldots, N$, as shown in Fig. 1(c), and the significant values of that coefficient in the sequence should be determined by comparing to a threshold derived from the coefficient energy.

To select which coefficients are to be coded, the coefficient rate theory implies a method that selects coefficients according to the energy level of coefficients, that is, the higher the energy of a coefficient, the more likely it is to be coded. However, such a scheme is not optimal in the rate-distortion sense. This is not hard to understand: in the optimal coding scheme, the distortion level for each coded coefficient is the same for all the coefficients, which is what optimal bit allocation tries to achieve. If we select coefficients based on their energy by encoding more high-energy coefficients, assuming we still keep the same distortion level for all the coded coefficients, since we are coding fewer of those low-energy coefficients, we are actually increasing the distortion level of the low-energy coefficients if we take into account those discarded samples. Therefore, such a scheme is not rate-distortion optimal.

[1]We have only demonstrated a stationary point here. The divergence inequality can be used to show (18) yields a maximum.

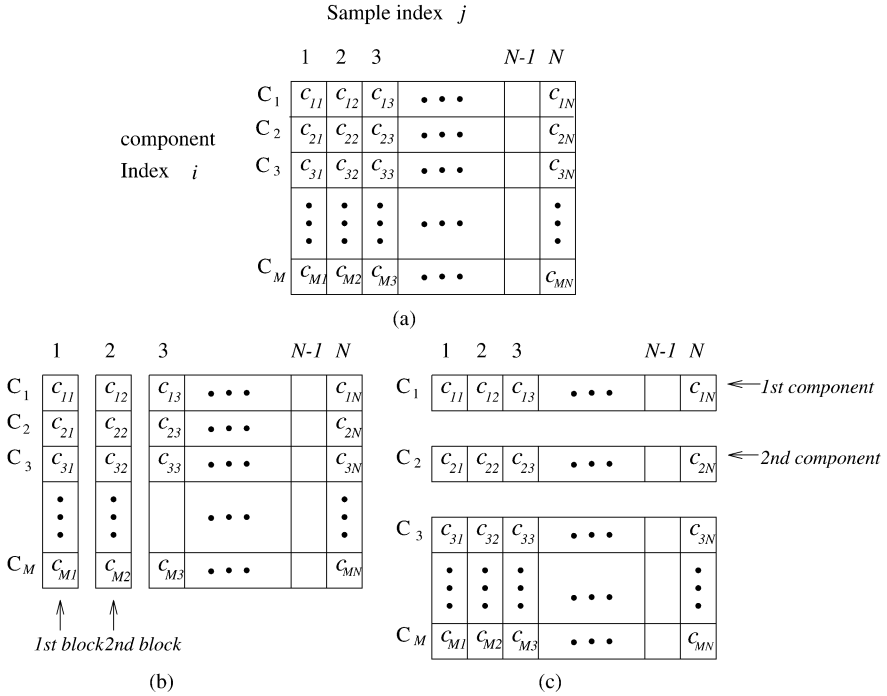

Fig. 1. Encoding transform coefficients. (a) Coefficients of N sample functions. (b) Encoding sample function by sample function. (c) Encoding component by component.

VII. Fourier Bandwidth, Shannon Bandwidth and Campbell Bandwidth

The equivalent bandwidth derivation implies a connection between the coefficient rate and bandwidth. We define the *Campbell bandwidth* as $W_c = Q_2/2$. In this section, we compare the Campbell bandwidth with two other bandwidth-related quantities: the Fourier bandwidth W, which is the most commonly used bandwidth in communication and reflects the actual frequency range of a signal; and the Shannon bandwidth, which is defined as follows: for a random process $X(t)$ with K-L expansion

$$X(t) = \sum_{i=1}^{\infty} C_i \phi_i(t)$$

if a sample function $x(t)$ over the time interval $0 \le t \le T$ can be represented or very well approximated by

$$x(t) = \sum_{i=1}^{M} c_i \phi_i(t)$$

then the *Shannon bandwidth* of the sample function $x(t)$ is $W_s = M/2T$.

For a process with rectangular spectrum, the three quantities are the same, i.e., $W_c = W_s = W$. The following theorem gives the general relationship of the three bandwidths.

Theorem 1: For each sample function of a random process, its Campbell bandwidth, Shannon bandwidth, and Fourier bandwidth, which are denoted by W_c, W_s, and W, respectively, satisfy following relation:

$$W_c \le W_s \le W.$$

The equality holds if and only if the sample function has a rectangular spectrum over the band $-W \le f \le W$.

Proof:

Part 1: Shannon gives a conceptually simple proof of $W_s \le W$ [22], and there are many other proofs.

Part 2: Following Campbell's development in Section II, we have $\mu^{\frac{1}{N}} \le M$, the coefficient rate Q_2 would be

$$Q_2 = \lim_{T\to\infty} \lim_{N\to\infty} \frac{\mu^{1/N}}{T} \le \frac{M}{T}. \tag{21}$$

So we have

$$W_c = \frac{Q_2}{2} \le \frac{1}{2}\frac{M}{T} = W_S. \tag{22}$$

Equality holds if and only if all the coefficients c_i are equal so we have to keep all coefficients in the K-L expansion of $X(t)$.

Combining Part 1 and Part 2, we prove that

$$W_c \le W_s \le W. \tag{23}$$

VIII. Water-Filling

In the rate-distortion sense, the number of coefficients in the expansion that are coded can vary depending on the distortion level. According to the reverse water-filling theory, for a Gaussian process $X(t)$ with transform coefficients $\{C_i, i = 1, 2, \ldots\}$, the number of coefficients retained is given by the number of coefficients with energy greater than a constant distortion level σ, i.e., $(1/2T)\sum_i \mathcal{X}\left(E\left(C_i^2\right) > \sigma\right)$, where $\mathcal{X}(x) = 1$ for $x > 0$, and otherwise is zero. The Campbell bandwidth, however, does not have an obvious physical meaning at a certain distortion level; but if the distortion level σ varies, the interpretation of Campbell bandwidth is clearer.

For the same Gaussian process $X(t)$, let $\lambda_i = E\left(C_i^2\right)$; if we encode the transform coefficients $\{C_i, i = 1, 2, \ldots\}$ at different distortion levels σ_j, $j = 1, 2, \ldots, n$, then at each distortion level, the number of coded coefficients is given by $(1/2T)\sum_i \mathcal{X}(\lambda_i > \sigma_j)$, and σ_j has a range of $[0, \lambda_{\max}]$, where $\lambda_{\max} = \max(\lambda_i)$. For n very large so that σ_j can be treated as uniformly distributed in $[0, \lambda_{\max}]$, the probability that C_i is selected is proportional to the magnitude of λ_i, so we have a probability function $p_i = \lambda_i / \sum_i \lambda_i$.

Following the same argument in Section IV, $T = \sum_i \lambda_i$, we have the spectral entropy

$$H(S) = -\sum_i \frac{\lambda_i}{T} \log \frac{\lambda_i}{T}.$$

Then the Campbell bandwidth is $W_c = (1/2)e^{H(S)}$, and we can say the Campbell bandwidth is the minimum average bandwidth for encoding the process across all possible distortion levels.

IX. Conclusion

We have presented two new derivations of the coefficient rate introduced by Campbell. One derivation solidifies its interpretation as a coefficient rate, and shows that the spectral entropy of a random process is proportional to the logarithm of the equivalent bandwidth of the smallest frequency band that contains most of the energy. The second derivation implies that the number of samples of a particular component should be proportional to the variance of that component. We discussed the implications of the latter result for realization-adaptive source coding and provided a connection with the familiar reverse water-filling result from rate distortion theory. From the coefficient rate, we defined a quantity called the Campbell bandwidth of a random process, and we contrasted Fourier bandwidth, Shannon bandwidth, and Campbell bandwidth.

Acknowledgment

The authors are indebted to the referees for their constructive comments and insights.

References

[1] L. Campbell, "Minimum coefficient rate for stationary random processes," *Inf. Contr.*, vol. 3, no. 4, pp. 360–371, Dec. 1960.
[2] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1968.
[3] N. Abramson, "Information theory and information storage," in *Proc. Symp. System Theory*, New York, Apr. 1965, pp. 207–213.
[4] J. D. Gibson, S. P. Stanners, and S. A. McClellan, "Spectral entropy and coefficient rate for speech coding," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Nov. 1993, pp. 925–929.
[5] S. McClellan and J. D. Gibson, "Variable-rate CELP based on subband flatness," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 120–130, Mar. 1997.
[6] J. D. Gibson, M. L. Moodie, and S. A. McClellan, "Variable rate techniques for CELP speech coding," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 29 –Nov. 1, 1995.
[7] S. McClellan and J. D. Gibson, "Variable rate tree coding of speech," in *Proc. 1994 IEEE Wichita Conf. Communications, Networking, and Signal Processing*, Wichita, KS, Apr. 1994.
[8] R. Mester and U. Franke, "Spectral entropy-activity classification in adaptive transform coding," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 5, pp. 913–917, Jun. 1992.
[9] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992.
[10] E. Wesfreid and M. V. Wickerhauser, "Adapted local trigonometric transforms and speech processing," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3596–3600, Dec. 1993.
[11] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
[12] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
[13] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.
[14] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.
[15] M. Effros, "Optimal modeling for complex system design," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 50–73, Nov. 1998.
[16] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.
[17] J. L. Massey, "Toward an information theory of spread-spectrum systems," *Code Division Multiple Access Commun.*, pp. 29–46, 1995.
[18] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, CA: Univ. Calif. Press, 1958.
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[20] L. Rade, *Beta Mathematics Handbook*. Boca Raton, FL: CRC, 1992.
[21] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic, 1991.
[22] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

Regular and Irregular Progressive Edge-Growth Tanner Graphs

Xiao-Yu Hu, *Member, IEEE*, Evangelos Eleftheriou, *Fellow, IEEE*, and Dieter M. Arnold, *Member, IEEE*

***Abstract*—We propose a general method for constructing Tanner graphs having a large girth by establishing edges or connections between symbol and check nodes in an edge-by-edge manner, called progressive edge-growth (PEG) algorithm. Lower bounds on the girth of PEG Tanner graphs and on the minimum distance of the resulting low-density parity-check (LDPC) codes are derived in terms of parameters of the graphs. Simple variations of the PEG algorithm can also be applied to generate linear-time encodeable LDPC codes. Regular and irregular LDPC codes using PEG Tanner graphs and allowing symbol nodes to take values over GF(q) ($q > 2$) are investigated. Simulation results show that the PEG algorithm is a powerful algorithm to generate good short-block-length LDPC codes.**

***Index Terms*—Girth, low-density parity-check (LDPC) codes, LDPC codes over GF(q), progressive edge growth (PEG), PEG Tanner graphs.**

I. Introduction

Codes on graphs [1]–[13] have attracted considerable attention owing to their capacity-approaching performance and low-complexity iterative decoding. The prime examples of such codes are the low-density parity-check (LDPC) codes. It is known that the belief-propagation (BP) or sum–product algorithm (SPA) over cycle-free Tanner graphs [1] provides optimum decoding. Hence, it is natural to try to minimize the influence of the cycles in the iterative decoding process. This approach has been adopted for both LDPC [14] and turbo codes [15]

Manuscript received September 2, 2002; revised July 18, 2004. The material in this correspondence was presented in part at 2001 IEEE Global Telecommunications Conference, San Antonio, TX, November 2001.

X.-Y. Hu and E. Eleftheriou are with IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland (e-mail: xhu@zurich.ibm.com; ele@zurich.ibm.com).

D. M. Arnold was with IBM Research, Zurich Research Laboratory, Rüschlikon, Switzerland (e-mail: d.m.arnold@ieee.org).

Communicated by K. A. S. Abdel-Ghaffar, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2004.839541