FACTA UNIVERSITATIS (NIŠ)
Series: Electronics and Energetics vol. 11, No.1 (1998), 71-85

TRELLIS SHAPING AND SPECTRAL NULLS

Josef Forster

Abstract. A line coding scheme based on trellis shaping is presented. The trellis shaping algorithm is used to obtain a signal with a controlled running digital sum as well as to achieve a Gaussian–like amplitude density distribution. The former causes a first order spectral null at zero frequency, while the latter allows to preserve a significant part of the gain attained by pure trellis shaping. Spectral nulls at arbitrary frequencies as well as second order nulls can easily be realized. The complexity of a slightly modified four state Viterbi algorithm is sufficient to achieve the major part of the possible gain.

1. Introduction

From the early beginning of digital transmission line coding has been an important subject. Although the most important aim is to obtain a spectral null at zero frequency, e.g. for AC–coupled baseband transmission, there are as well applications that require nulls at other frequencies in the signal spectrum (HDTV, magnetic recording). While line coding focuses on the transmitter side, there is also a possibility to compensate an AC–coupled channel at the receiver using decision feedback techniques. Due to significant losses [1], depending on the noise spectrum, and a quasi–catastrophic error propagation this method will not be discussed further.

The ideal line code produces a first or second order spectral null at the desired frequency, doesn't affect the distance properties of the original signal sequences, exhibits a very low decoding complexity, and doesn't increase signal power. Even if especially the latter requirement cannot be fulfilled in an ideal way, it is a challenging task to keep the loss as small as possible. Moreover, one usually wishes to have some influence on the width of the spectral null, i.e. the cut–off frequency.

Most of the traditional approaches introduce an expanded signal constellation together with some kind of encoder or discrete–time filter to shape

Manuscript received January 11, 1998.

Dipl.-Ing. J.K. Forster is with Federal Armed Forces University Munich, Institute for Communication Engineering, D–85577 Neubiberg, Germany.

the transmitted sequence [3][4][5]. In this case, the encoder frequently improves the distance properties of the transmitted sequence. Unfortunately, high decoding complexity is necessary to make use of this improvement. If, on the other hand, one decides to drop this advantage for sake of a simpler decoder, that means not to track the encoder states in the decoder, the gain in euclidean distance (which costs an increase in average power of the transmitted signal emanating from the signal set expansion) is lost.

Forney showed in his work about *Trellis Shaping* [6] how trellis codes can be used to shape the amplitude distribution of a signal, which results in a decrease in average signal power. As in all trellis–coded schemes the first step is to expand the signal constellation. The signal points are selected from this expanded constellation according to a Gaussian–like probability distribution, leading to a gain of up to 1.53 dB compared to the unshaped signal. The Gaussian–like distribution results from an optimization of the transmitted sequences with respect to their average energy.

In this paper, it is shown that it is possible to obtain a spectral null at an arbitrary frequency if the power of the running digital sum is used as an optimization criterion instead of the average signal power, and that both a true spectral null and a significant part of the shaping gain can be obtained if both criteria are combined.

2. A short review of Trellis shaping

This section gives a short summary of the basic ideas of trellis shaping as discussed in detail in [6]. In this context we concentrate on aspects that are of special interest for spectral shaping. In order to distinguish Forney's introduction of trellis shaping to obtain a Gaussian–like signal distribution (and thus to minimize the average power of the signal) and the approach presented here we call the first one *power oriented shaping* (POSH) and the latter one *spectral oriented shaping* (SOSH). Then we use the term trellis shaping for the technique itself, regardless what properties of the signal are influenced.

In this paper sequences are represented by their D–transforms and marked with small letters: $a = a(D) = \ldots + a_{\nu-1}D^{\nu-1} + a_\nu D^\nu + a_{\nu+1}D^{\nu+1} + \ldots$. A single symbol at time ν is referred to as a_ν. All sequences are assumed to be wide sense stationary and calculations are done in $GF(2)$ in their binary representation. Matrices of polynomials in D are represented by bold capital letters: $\mathbf{B} = \mathbf{B}(D)$. A code C_S is given by a generator matrix $\mathbf{G_S}$.

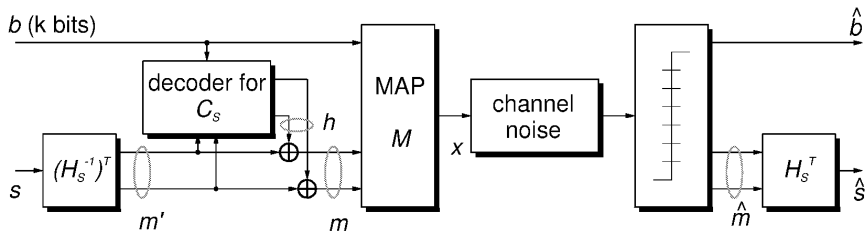

Figure 1. A trellis shaping system supporting $k+1$ (uncoded or coded) bits per two dimensions.

In the following we discuss the trellis shaping system depicted in figure 1. This scheme supports $k+1$ bits per two–dimensional symbol using a square constellation of 2^{k+2} signal points from a translated version of the lattice $\mathbf{Z}^2$. Throughout this paper a two–dimensional symbol is physically represented by two consecutive one–dimensional symbols, but generally may also be chosen as inphase and quadrature component in amplitude/phase modulation schemes. With this constellation we have a redundancy of 1 bit per two dimensions, which is used to select a sequence from a given shaping code C_S with respect to a given criterion.

The initial sequence m' (2 bit) stems from the input sequence s (1 bit) passed through an *inverse syndrome–former* $(\mathbf{H}_\mathbf{S}^{-1})^T$ for the shaping code C_S. The initial sequence m' is modified by adding a shaping sequence h selected by the shaping decoder. Since any sequence h from C_S is orthogonal to any sequence produced by an inverse syndrome former for C_S, both sequences are separable. The mapping function is, similar to Forney's sign bit shaping such that the modified bits m_v choose the sign, i. e. the quadrant of the 256 point square constellation. Thus the sequence s can easily be recovered by passing the estimated sequence $\hat{m}$ obtained by a decision on the sign of the channel output through the *syndrome–former* $\mathbf{H}_\mathbf{S}^T$.

Formally, a syndrome–former $\mathbf{H}_\mathbf{S}^T$ for a convolutional code C_S with generator matrix $\mathbf{G}_\mathbf{S}$ is defined by

$$\mathbf{G}_\mathbf{S}\,\mathbf{H}_\mathbf{S}^T = \mathbf{0}. \tag{1}$$

For any sequence $h \in C_S$

$$h\,\mathbf{H}_\mathbf{S}^T = 0. \tag{2}$$

An inverse syndrome–former is defined by

$$\left(\mathbf{H}_\mathbf{S}^{-1}\right)^T\ \mathbf{H}_\mathbf{S}^T = \mathbf{1}. \tag{3}$$

Hence we can write for the noiseless case

$$\hat{x} = \hat{m}\,\mathbf{H}_{\mathbf{S}}^{T} = \left[s\left(\mathbf{H}_{\mathbf{S}}^{-1}\right)^{T} + h\right]\mathbf{H}_{\mathbf{S}}^{T} = s. \tag{4}$$

In order to avoid unnecessary error propagation the syndrome–former in the receiver should be the minimal feedback–free version, which exists for every binary linear convolutional code, see [7][8]. The inverse syndrome–former, which needs not necessarily be feedback–free (and which is not unique), can be chosen arbitrarily.

Hence the shaping decoder can select a sequence h with respect to an arbitrary criterion without affecting the sink sequence $\hat{s}$. For the classical application of trellis shaping (POSH), i.e. minimizing the average energy of the transmitted signal, the decoder is an implementation of the Viterbi algorithm (VA) [9] with a metric function

$$\Gamma_{POSH}(\mu) = E(\mu) = \sum_{\nu=-\infty}^{\mu} \gamma_{\nu} = \sum_{\nu=-\infty}^{\mu} \left|M\left(b_{\nu} + (m'_{\nu} \oplus h_{\nu})\right)\right|^{2}, \tag{5}$$

where M is the mapping function (cf. figure 1) and h_{ν} are the modification bits belonging to the considered state transition in the trellis. Searching a path with minimal Γ_{POSH} is equivalent to search for a sequence h from C_S that (added in GF(2) to the initial sequence m') minimizes the average energy of the transmitted signal. The shaping gain achieves about 1 dB for trellis codes with four to eight states.

Note: The implementation of the VA must ensure that the decoder selects only valid sequences h from C_S. Otherwise the syndrome former output $\hat{s}$ will no longer be unaffected by h since $h\,\mathbf{H}_{\mathbf{S}}^{T}$ becomes $\neq 0$ (compare with (4)).

3. Spectral nulls

The following subsection is a short overview of mathematical properties of sequences exhibiting a spectral null at zero or at an arbitrary frequency. In the second subsection it is shown how trellis shaping can be used to encode random sequences in such a way that they exhibit these properties. The last subsection deals with the question how to combine POSH and SOSH to obtain both, a spectral null at an arbitrary frequency and a considerable part of the possible shaping gain.

3.1 Discrete–time sequences and spectral nulls

There seems to be a rather obvious connection between the DC component of a sequence and its *running digital sum* (RDS)

$$RDS(\mu) = \sum_{\nu=-\infty}^{\mu} x_\nu. \tag{6}$$

Using (6) the DC–component can be written as

$$\begin{aligned} \bar{x} &= \lim_{N\to\infty} \frac{1}{N} \sum_{\nu=-\infty}^{N} x_\nu \\ &= \lim_{N\to\infty} \frac{1}{N} RDS(N). \end{aligned} \tag{7}$$

Hence it is apparently a sufficient (and necessary, cf. [11], [12]) condition for a DC–free spectrum that the RDS is bounded.

Nevertheless a better criterion can be found. Justesen [11] showed that first order mean power spectra can be described by

$$S_x(f) = \frac{\sigma_x^2(1+\beta)(1-\cos(2\pi fT))}{1+\beta^2-2\beta\cos(2\pi fT)}, \tag{8}$$

depending solely on a single parameter β, where $1/T$ is the symbol rate and σ_x^2 is the average power of the output signal. The parameter β, which is the optimal coefficient of a linear predictor for first order low pass spectra (cf. [5]) is given by

$$\beta = 1 - \frac{\sigma_x^2}{2\,\sigma_z^2} \tag{9}$$

where σ_z^2 is the average power of the RDS signal z obtained by filtering the output signal x with

$$\begin{aligned} G_{RDS}(f) &= \frac{1}{1-e^{-j2\pi fT}} \\ &= \frac{1}{1-D}. \end{aligned} \tag{10}$$

The cut–off-frequency f_c for this spectrum is

$$\begin{aligned} f_c \cdot T &= \frac{1}{2\pi} \arccos\left(\frac{2\beta}{1+\beta^2}\right) \\ &\approx \frac{1}{2\pi} \frac{\sigma_x^2}{2\,\sigma_z^2}. \end{aligned} \tag{11}$$

The approximation on the right hand side of (11) shows that the width of the spectral nulls depends on the ratio of the output signal variance to the variance of the RDS sequence. Thus for a wide spectral null the ratio σ_x^2/σ_z^2 is to be maximized. This is done by minimizing the average power of the RDS σ_z^2. This criterion proves to perform better than the usual minimization of the absolute RDS.

To obtain a spectral null at a normalized frequency $\theta_0 = 2\pi f_0 T$ other than zero the whole RDS spectrum is shifted by θ_0. Thus the transformed RDS sequence

$$RDS_{\theta_0}(\mu) = \sum_{\nu=-\infty}^{\mu} x_\nu \, e^{-j\nu\theta_0} \tag{12}$$

is obtained by filtering the output sequence with

$$\begin{aligned} G_{RDS,\theta_0}(f) &= \frac{1}{1 - e^{-j2\pi(f-f_0)T}} \\ &= \frac{1}{1 - D \cdot e^{-j2\pi f_0 T}}. \end{aligned} \tag{13}$$

Monti and Pierobon [13] have shown that a necessary and sufficient condition for a second order spectral null (i.e. a fourth order null of the power density) is that the running sum of the RDS, the running digital sum sum RDSS,

$$RDSS(\mu) = \sum_{\nu=-\infty}^{\mu} RDS(\nu), \tag{14}$$

obtained by filtering the output sequence with

$$\begin{aligned} G_{RDSS}(f) &= \frac{1}{\left(1 - e^{-j2\pi f T}\right)^2} \\ &= \frac{1}{\left(1 - D\right)^2}, \end{aligned} \tag{15}$$

is constrained to a certain interval. A better criterion for second order spectral nulls is again to minimize the variance of the RDSS sequence.

Moreover, arbitrary spectra can be achieved, if a suitable (more complex) filter is chosen instead of the RDS or RDSS filter, respectively. The global optimization rule is again to minimize the power of the filtered sequence. However, due to the general structure of such a filter, a metric function similar to (16) (below) might involve future symbols. Hence the Viterbi Algorithm is no longer applicable.

3.2 Controlling the RDS by Trellis shaping

The idea is to let the shaping decoder control the RDS of the output sequence. In other words, the decoder has to search the sequence h from C_S that minimizes the average RDS energy σ_z^2 and thus maximizes the cut–off–frequency f_c. For a first order spectra the SOSH metric Γ_{SOSH} for the VA is formally

$$\begin{aligned}\Gamma_{SOSH}(\mu) &= E_{RDS}(\mu) \\ &= \sum_{\nu=-\infty}^{\mu} RDS_\nu^2 \\ &= \sum_{\nu=-\infty}^{\mu} \Big| \sum_{\kappa=-\infty}^{\nu} x_\kappa \Big|^2 \\ &= \sum_{\nu=-\infty}^{\mu} \Big| \sum_{\kappa=-\infty}^{\nu} M\big(b_\kappa + (m'_\kappa \oplus h_\kappa)\big) \Big|^2 .\end{aligned} \tag{16}$$

With Γ_{SOSH} the VA chooses the output sequence with lowest RDS energy, among all possible output sequences. The distribution of the RDS is concentrated around zero and hence Γ_{SOSH} produces a spectral null, at least for stochastic input sequences. The RDS is constrained to a certain interval as it is necessary to eliminate the DC–component.

However, for many shaping codes there seem to exist deterministic sequences that lead to a steadily growing RDS. But this can easily be avoided by scrambling the shaping input sequence s. Moreover, the behavior of the shaping decoder is significantly affected by the choice of the mapping function, which may also be adapted to handle deterministic sequences.

3.3 Combined spectral and distribution shaping

The major shortcoming of the method introduced in the preceding subsection is a power penalty of (almost) 3 dB due to the constellation expansion. The basic idea to overcome this problem is a linear combination of the RDS–metric (16) introduced in the last subsection and the energy–minimizing metric (5) mentioned in section 2.

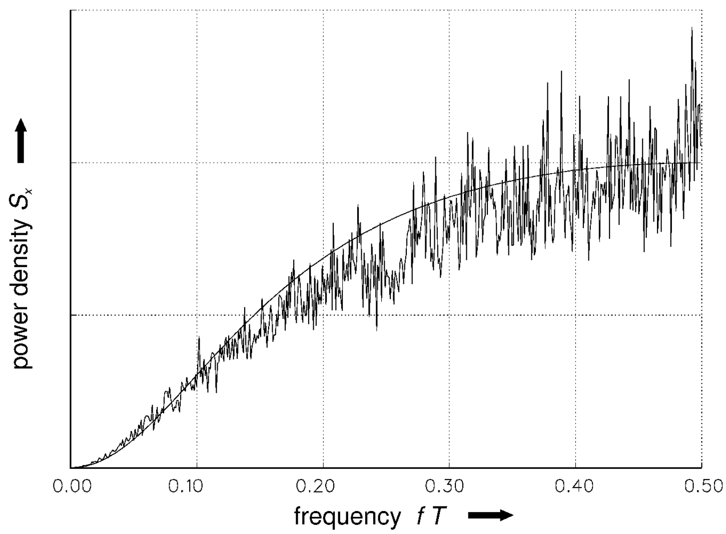

Figure 2. Power spectral density for pure spectral shaping. (code C_4, $f_c\,T = 0.135$).

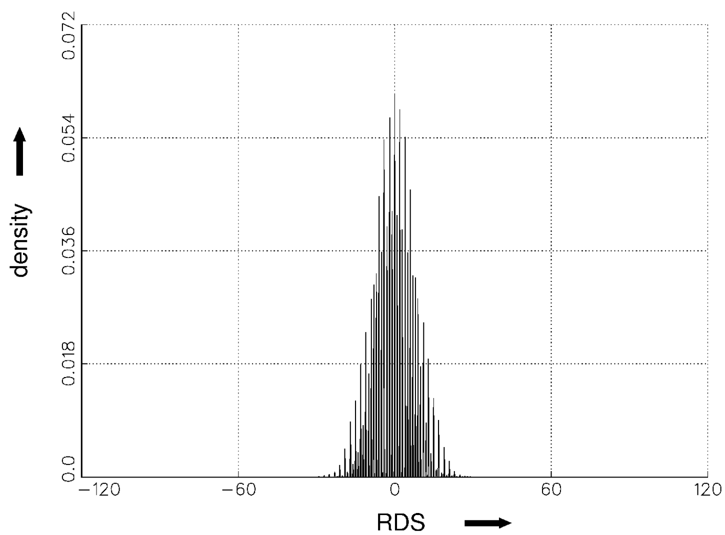

Figure 3. Histogram function of the running digital sum (RDS) for pure spectral shaping (16). (same parameters as in figure 2).

$$\begin{aligned}\Gamma(\mu) &= \Gamma_{POSH}(\mu) + k \cdot \Gamma_{SOSH}(\mu) \\ &= \sum_{\nu=-\infty}^{\mu} \left(\left| x_\nu \right|^2 + k \cdot \left| \sum_{\kappa=-\infty}^{\nu} x_\kappa \right|^2 \right) \\ &= \sum_{\nu=-\infty}^{\mu} \left(\left| M\big(b_\nu + (m'_\nu \oplus h_\nu)\big) \right|^2 + k \left| \sum_{\kappa=-\infty}^{\nu} M\big(b_\kappa + (m'_\kappa \oplus h_\kappa)\big) \right|^2 \right)\end{aligned} \tag{17}$$

A VA minimizing this metric operates as follows: As long as the RDS energy doesn't become too large (depending on k), it selects a sequence h to minimize the average energy. In general this leads to a steadily growing RDS, and after some time the decoder starts minimizing the DC–component. This again causes the path energy to increase, and so on. Compared to an usual implementation of the VA this special decoder requires an additional register per path to keep track of the RDS sequence. The multiplication by k could, for faster operation, be replaced by a shift operation if k is a power of two. An important detail is that the energy metric must be reset from time to time to prevent it from overflow, while the RDS register must not.

The constant k determines the weight of the RDS compared to the path energy. In other words, a large k forces the decoder to keep the variance of the RDS low, and the energy of the transmitted sequence increases. This leads to a loss in average energy compared to the unshaped constellation. On the other hand, a small k gives priority to the path energy metric preserving most of the shaping gain, while the variance of the RDS increases. It nevertheless forces a spectral null.

Using the complex RDS (12) as part of the metric (17) produces a spectral null at any desired frequency, cf. section 4 for examples. Similarly, a second order spectral null is obtained by controlling the RDSS in the same manner.

4. Numerical results

In this section we concentrate on the effect of spectral shaping and the trade–off between SOSH and POSH by changing the parameter k in (17). Since the decoding complexity increases with the number of states of the shaping code C_S we present results for three simple codes C_2, C_4 and C_8 with their generator matrices given in table 1.

Figure 2 shows the effect of spectral shaping using C_4 and the RDS energy as metric (cf. (16)). Compared to the unshaped constellation this leads to an increase in the average power of (i.e. a loss of) approximately -2.17dB. This is better than the theoretical loss of -3dB since SOSH influences also the amplitude distribution. The distribution of the RDS (figure 3) is concentrated around zero and Gaussian–like with the probability of growing large tending to zero.

Table 1. Trellis codes used for simulations

code	states	generator matrix
C_2	2	$\mathbf{G_2} = [1, D+1]$
C_4	4	$\mathbf{G_4} = [D^2+1, D^2+D+1]$
C_8	8	$\mathbf{G_8} = [D^3+D+1, D^3+D^2+1]$

Using the combined metric (17) with $k = 11$ instead of the pure RDS (16) realizes a shaping gain of approximately 0.44 dB compared to the unshaped constellation *and* preserves the spectral null at zero frequency (figure 4). Due to the greater variance of the RDS the cut–off frequency (half power frequency) is lower than that in figure 2.

The trade-off between shaping gain (a negative shaping gain means a higher average energy compared to the unshaped constellation) and cut–off frequency for the three codes mentioned above is depicted in figure 5. If the cut–off frequency is chosen lower than $f_c T \approx 0.05$ a positive shaping gain is attained with codes with four or more states.

Figure 6 shows an example for a spectral zero at $f_0\, T = 0.32$. Besides the complex path metric (12) in the VA this case is similar to a spectral null at DC.

The effect of a second order spectral null obtained by a RDSS–controlled shaping decoder using a combined metric of RDSS energy and signal energy is depicted in figure 7. The shaping gain in this case is about 0 dB.

The spectral shaping scheme can also be expanded to multidimensional constellations, with 1 bit redundancy per N dimensions. This is demonstrated in figure 8. The width of the spectral null decreases with increasing dimensionality as expected, since an increase in dimension corresponds to a decrease in redundancy ($R = 1/N$ bit).

In figure 9 SOSH with the code C_8 is compared to the bound for decision feedback equalization (DFE) given in [1] and also to some multilevel trellis codes with spectral nulls designed by Calderbank and Mazo [2]. In any case the combination of SOSH and POSH is at least 1 dB better then DFE and 1.5 to 2 dB better than the trellis codes, which are among the best the author has found. It also can be seen that the gain compared to DFE increases as f_c becomes larger.

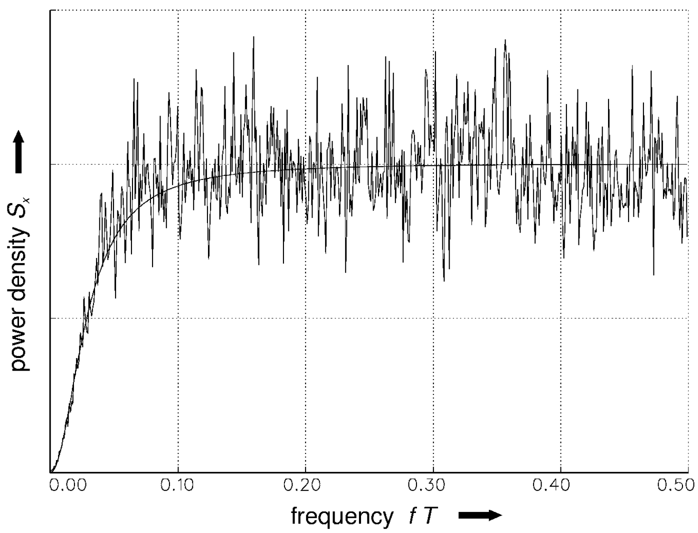

Figure 4. Power spectral density for the combined metric (17) with $k = 11$. (code C_4, $f_c\,T = 0.028$, shaping gain= 0.44dB).

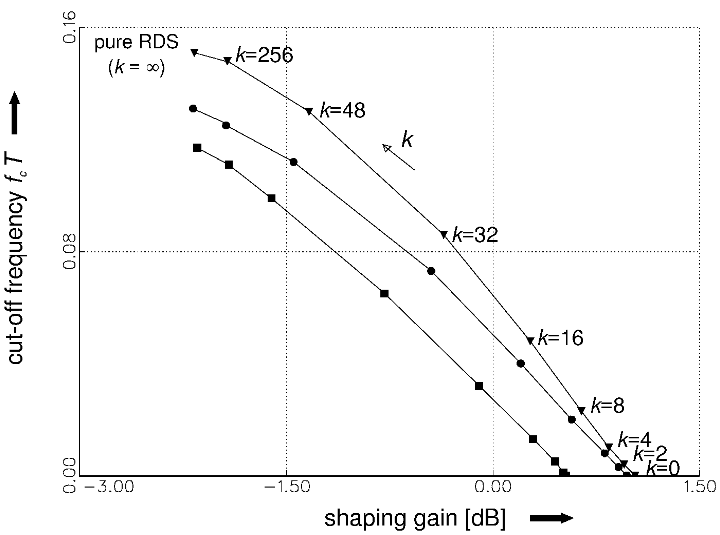

Figure 5. Trade–off between cut–off frequency f_c and shaping gain for C_2 (■), C_4 (•) and C_8 (▼).

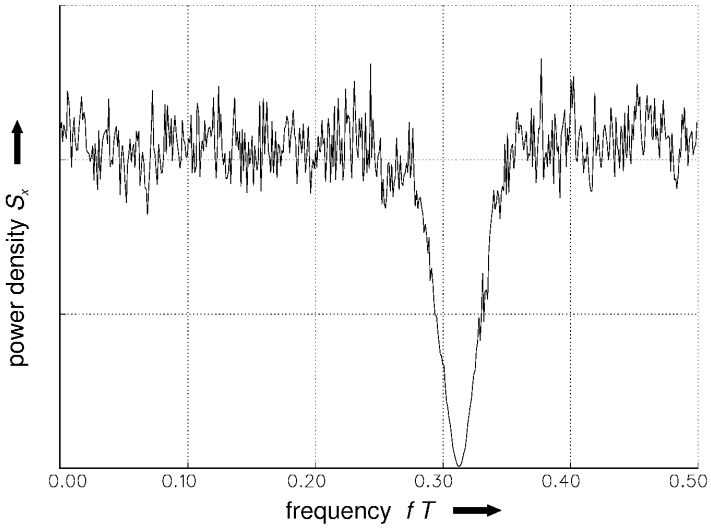

Figure 6. Spectral null at $f_0\,T = 0.32$. (code C_4, $f_c\,T = 0.018$, shaping gain=0.26dB).

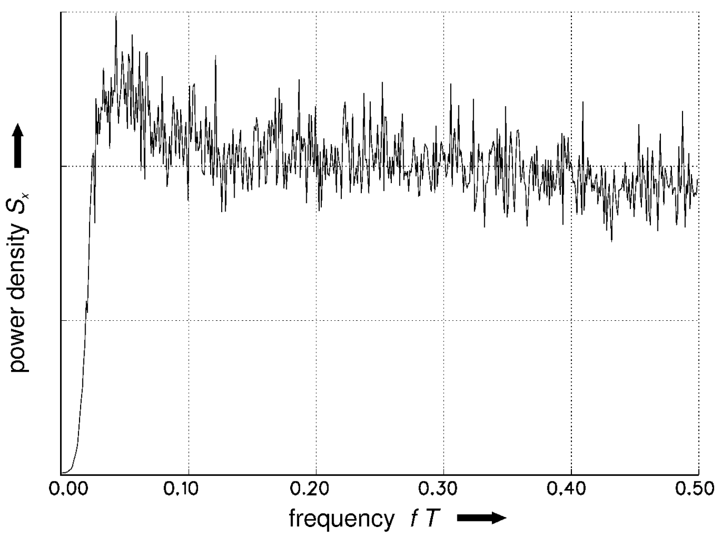

Figure 7. A second order spectral null with a combined RDSS/energy minimizing metric. (code C_4, $f_c\,T = 0.019$, shaping gain $= -0.16$ dB).

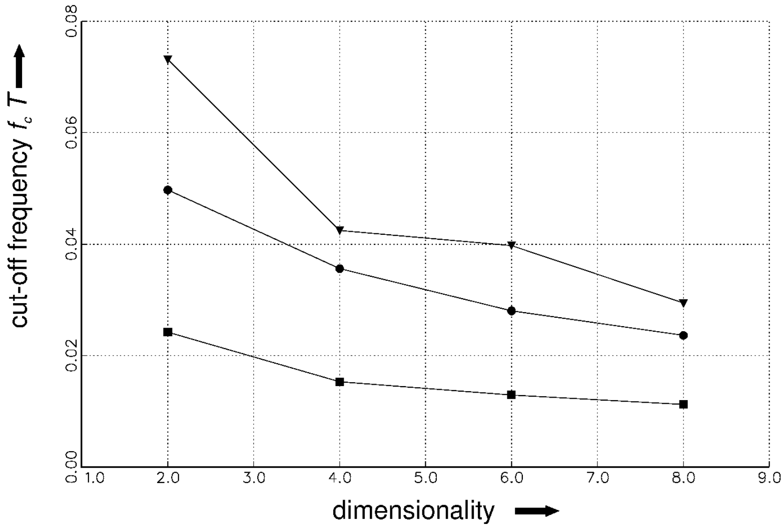

Figure 8. Trade–off between cut–off frequency f_c and dimensions with shaping gain 0.5 dB (■), 0.0 dB (●) and -0.5 dB (▼).

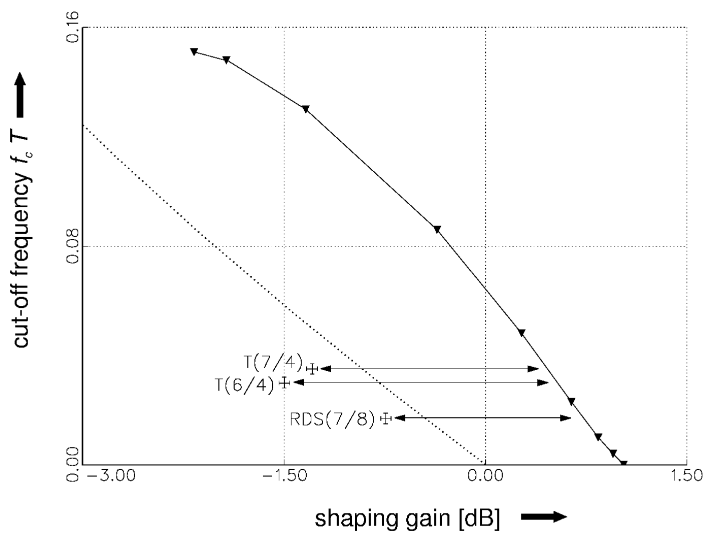

Figure 9. Comparison of SOSH (∇) with C_8 to a bound for DFE (···) given in [1] and trellis codes with spectral nulls given in [3].

5. Conclusion

A method has been introduced to produce a spectral zero at an arbitrary frequency in the spectrum. The basic idea is to combine Forney's trellis shaping algorithm with a controlled running digital sum. This allows preserving a significant part of the possible shaping gain while a true spectral null is obtained, as long as the width of the spectral null is not chosen too large. The complexity is that of a four to eight state Viterbi algorithm with double the expense of a minimum distance decoder for the metric calculation.

Acknowledgement

The author wish to thank Prof. K. Tröndle for his support and his detailed suggestions to improve the clarity of presentations.

REFERENCES

1. J. Huber: *Loss of signal/noise ratio caused by highpass–filtering in digital baseband transmission systems with decision feedback equalization.* Frequenz, vol. 41, pp. 112–114, 1987.
2. A.R. Calderbank, Ting–Ann Lee, J.E. Mazo: *Baseband trellis codes with a Spectral Null at Zero.* IEEE Trans. Inform. Theory, vol. IT–34, pp. 425–434, 1988.
3. A.R. Calderbank, J.E. Mazo: *Baseband line codes via spectral factorization.* IEEE J. Select. Areas Commun., vol. SAC–7, no. 6, pp. 914–928, Aug. 1989.
4. A.R. Calderbank, J.E. Mazo: *Spectral nulls and coding with large alphabets.* IEEE Commun. Mag., vol. 29, no. 12, pp. 58–67, December 1991.
5. G.D. Forney, Jr., A.R. Calderbank: *Coset codes for partial response channels; or coset codes with spectral nulls.* IEEE Trans. Inform. Theory, vol. IT–35, pp. 925–943, 1989.
6. G.D. Forney, Jr.: *Trellis shaping.* IEEE Trans. Inform. Theory, vol. IT–38, no. 2, pp. 281–300, Mar. 1992.
7. G.D. Forney, Jr.: *Convolutional codes I: Algebraic structure.* IEEE Trans. Inform. Theory, vol. IT–16, pp. 720–738, Nov. 1970.
8. G.D. Forney, Jr.: *Structural analysis of convolutional codes via dual codes.* IEEE Trans. Inform. Theory, vol. IT–19, pp. 512–518, July 1973.
9. G.D. Forney, Jr.: *The Viterbi algorithm.* Proc. IEEE, vol. 61, pp. 268–278, Mar. 1973.
10. J.K. Forster: *A Lower Bound to the Maximal Entropy of Line Codes.* Int. Symp. on Information Theory & Its Applications, Sydney Australia, Australia Proceedings of Papers, vol. 94/5, Nov. 1994, pp. 611–616.
11. J. Justesen: *Information rates and power spectra of digital codes.* IEEE Trans. Inform. Theory, vol. IT–28, no. 3, pp. 457–472, May 1982.
12. G.L. Pierobon: *Codes for zero spectral density at zero frequency.* IEEE Trans. Inform. Theory, vol. IT–30, no. 2, pp. 436–439, March 1984.

13. C.M. Monti, G.L. Pierobon: *Codes with a multiple spectral null at zero frequency.* IEEE Trans. Inform. Theory, vol. IT–35, pp. 463–471, 1989.

14. K. Tröndle, R. Matzner, J. Forster: *Trellis–Shaping und Vorcodierung.* Universität der Bundeswehr München, Abschlußbericht, Institut für Nachrichtentechnik, Dez. 1993, (unpublished).