Does 16-QAM Provide an Alternative to a Half-Rate GSM Speech Codec?

W. Webb, L. Hanzo, R.A. Salami, R. Steele

Dept. of Electr. & Comp. Sc., Univ. of Southampton
U.K. SO9 5NH

Abstract

The computational complexity, speech-quality, spectral efficiency and robustness trade-offs of speech transmission schemes for personal communications networks*(PCNs)* are addressed. Differentially encoded star *16-QAM* arrangements with and without oversampling and diversity are compared in terms of robustness to the parallel frequency division multiplex*(FDM) 16-QAM* system to find the most appropriate modem scheme. The *13.4 kbps* regular pulse excited *RPE-LTP* speech codec is subjected to rigorous bit-sensitivity analysis in terms of both cepstral distance*(CD)* and segmental signal to ratio*(SEGSNR)* degradations. A twin-class embedded binary Bose-Chaudhuri-Hocqunghem*(BCH)* bit protection scheme is devised to improve robustness. The *6.2 kBd* system proposed guarantees low-complexity, high-quality speech transmissions in a bandwidth of *12 kHz* for vehicular speeds of *30 mph* and channel signal to ratios*(SNRs)* in excess of *25 dB*, a value readily maintained in microcellular PCNs. When using diversity, our system has similar robustness to the *11.4 kbps* quadruple-complexity, lower speech-quality code-excited linear predictive speech codec/convolutional coding/minimum shift keying*(CELP/CC/MSK)* benchmarker utilised. Yes, *16-QAM* does provide an attractive alternative to half-rate speech coding.

1 Introduction

Since the standardisation of the *TDMA* Pan-European digital mobile radio system known as *GSM* was finalised, a number of new European initiatives have emerged *(DECT, CT-2)* with the incentive of proposing more cost-, bandwidth- and power-efficient solutions for personal communications networks *(PCNs)*. Simultaneously, in the U.K. three companies have been allocated operating licences for an up-converted *GSM*-type system in the *1.8 GHz* band, while the European speech-coding community is feverishly working on a half-rate *GSM* speech codec *[1]*. The exceptionally stringent specifications of the half-rate speech codec tolerate no speech quality degradation when compared to the regular pulse excited *(RPE)* full-rate system, operating at *13 kb/s* speech-coding rate and *22.8 kb/s* channel-coded rate. Accordingly, the half-rate system has to function at *11.4 kb/s* gross-rate (speech- plus channel-coding) and less than *90 ms* overall delay, while keeping the complexity and power consumption to less than four times that of the full-rate codec.

In this contribution we embark upon the investigation of an alternative solution to doubling the traffic capacity and bandwidth efficiency of the full-rate *GSM* system at significantly lower complexity than in case of a half-rate speech codec. Our proposed system is based on the *13 kb/s* full-rate *RPE-GSM* speech codec with carefully embedded low complexity binary Bose-Chaudhuri-Hocquenghem(*BCH*) error correction coding and *4 bits/symbol 16-QAM* modulation, made sufficiently robust for fading mobile channels.

2 Multi-Level Modulation for Rayleigh-Fading Channels

Until quite recently multi-level modulation schemes such as *16-QAM* had been considered unsuited for hostile fading channels, however desirable their superior bandwidth efficiency over inherently more robust constant envelope schemes such as *GMSK* might be. It is the violent phase and envelope changes of the *Rayleigh* channel which render robustness so difficult to achieve. Although there is plenty of evidence that small microcells of the emerging *PCN* have more friendly *Rician* characteristics, to achieve worst-case estimates we use in our studies the *Rayleigh* model.

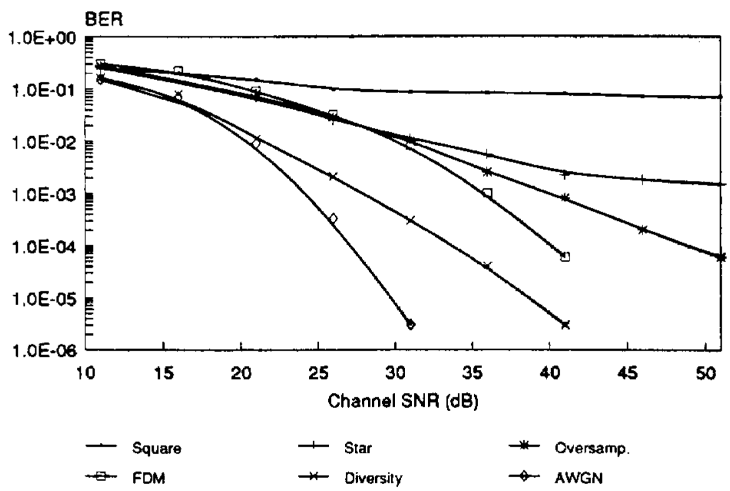
The ubiquitous 'maximum-distance' square-shaped *16-QAM* phasor constellation is optimum in terms of 'noise protection distances' amongst constellation points via *AWGN* channels but has a high residual bit error rate in the *Rayleigh*-environment as shown in *Figure 1 [2]*. For fading channels we have found that a twin circular constellation with a radii of one and three, respectively, coupled with differential encoding ensures considerably lower bit error rates, and removes the requirement of an accurate phase reference for demodulation *[3]*. The first bit of a *16-QAM* symbol encodes the phasor amplitude, i.e.,

CH2944-7/91/0000/0511 $1.00 © 1991 IEEE

pin-points which of the two circles is the corresponding constellation point assigned to, while the three remaining bits are differentially *Gray*-coded onto the phase angle on these circles. This differential 'star *QAM*' system greatly improves the *BER* performance, although there is still a persistant residual *BER* of approximately *0.1%* as seen in *Figure 1*. This is essentially removed by introducing the following oversampling-type broken-line approximation to the fading envelope. During any signalling interval several received signal observations are made, amongst which the last one is presumably the nearest to the first one in the forthcoming interval. Based on these observations a linear envelope-slope is inferred and used to linearly interpolate and remove the effect of fading envelope and erratic phase. Although very efficient in reducing the residual *BER* at high *SNR*s where the noise does not falsify this interpolation process, there is an infinitesimal *BER* degradation at low *SNR*s as demonstrated by *Figure 1*. Finally, *Gaussian*-like performance is achieved when using second-order switch-diversity. Clearly, in this systematic optimisation of our *16-QAM* system we achieved sufficiently low *BER*s at channel *SNR*s readily available in the microcells of the emerging *PCN*, while doubled the spectral efficiency of conventional constant-amplitude modems without quadrupling the speech-codec's complexity and power consumption. The major unknown is now how to embed the forward error correction coding *(FEC)* to best protect the speech-codec's bits against channel errors.

An alternative approach to the conventional serial data transmission using *QAM* is to transmit data in a parallel manner using frequency division multiplexing *(FDM)*. In its simplest conceptual form a serial data stream is converted to an *N* bit wide parallel stream. This is then transmitted over *N* subchannels each having a bandwidth of the order of *1/N* times the bandwidth of the serial channel. *FDM* has the advantage of dispersing the effects of a fade across the whole block of *N* symbols so that, whereas in serial transmission a few bits may be completely obliterated, in *FDM* each bit only suffers slight distortion and may still be recognizable. *FDM* as discussed above is complex to implement due to the need for *N* subchannel pulse-shaping filters, subchannel modems, etc. However, it can be shown *[5]* that if the serial stream is passed through an *FFT*, transmitted serially and then passed through an *IFFT* at the receiver, precisely the same effect is achieved for a substantial decrease in complexity. Using computer simulations the performance of *FDM/FFT* systems using *N=128* was evaluated for transmission over *Rayleigh*-fading channels. A single pilot tone *10 dB* higher in magnitude than the rest of the signal was transmitted at one end of the spectrum to allow the receiver to track channel attenuation and phase shift, which is essential if the *FDM* signal is to be accurately decoded.The results are shown in *Figure 1*. It can be seen that the *FDM* system is inferior to the serial system for channel *SNR*s below *30 dB*, but above this value quickly gains substantial improvements over the serial system which exhibits a residual *BER*.

Figure 1.: Performance of 16-QAM
modems over Rayleigh-fading channels



3 FEC for RPE-16-QAM Speech Transmissions

The most intelligent way of deploying *FEC* in multi-level modem schemes is to expand the symbol-set to accomodate additional bits for trellis-coded modulation *(TCM)* at unaltered signalling rate, which tacitly assumes convolutional coding, although trellis decoding of block codes is also possible. In an unfortunate analogy to the square-QAM constellation optimal for *AWGN* channels, *TCM* is optimised solely for *Gaussian* channels, and any practical extension of the symbol-set can accomodate only too low a number of additional bits for the *FEC* to cope with Rayleigh-fading channels.

Even for lower coding rates, i.e., high redundancy rates convolutional codes *(CC)* tend to be less robust against bursty channel error statistics than block codes. Furthermore, *CC*s heavily depend on long channel interleavers to randomise bursty errors and on soft decision channel measurement information as well as on high constraint lengths to successfully combat fading channels. Regretfully, the decoding complexity increases exponentially with the constraint length and with the number of constellation points supplying soft decision information. Lastly, they possess no error detection capabilities.

Due to their relatively high blocklengths block codes have higher freedom in the distribution of channel errors, particularly, when combined with interleavers. They also exhibit reliable error detection, and nicely curtail error propagation across speech-coding frames. For all these plausible advantages we favour block codes, however, for a rigorous comparison see *Reference [4]*. The channel interleaver has to disperse errors concentrated around deep fades so that the *FEC* codec can have approximately equal number of errors in each block. Naturally, long block codes inherently possess this randomising property, but are complex to implement. A very good compromise in terms of implementational complexity and error correcting power is constituted by the family of binary *BCH* codes of *63 bits* length. The *13 kb/s RPE-GSM* codec has a framelength of *20 ms*, whence we limit the interleaving depth, i.e., the transmitter's delay to one linear predictive coding *(LPC)* frame of *20 ms*. As an example, let us relate this to a vehicular speed of *30 mph* or *13.3 m/s* giving a travelling distance of *26.6 cm/20 ms*. For propagation frequencies of *1.8 GHz*, as in the planned *PCN*, the wavelength is around *15 cm*, and therefore interleaving over an interval of *20 ms* or approximately two wavelengths ensures efficient randomisation. However, for slowly walking pedestrians there is a danger of idling in deep fades, which is detrimental as regards to reception quality. In this situation the switch-diversity scheme described earlier is essential, unless frequency hopping is used, as in the *GSM* system.

4 Choice of Speech Codec

In our previous studies we proposed a number of speech transmission schemes for *PCN*, where we used either *16-QAM* and a *16 kb/s* subband codec *(SBC) [2]* to yield a *5 kBd* system, or *64-QAM* and a *4.8 kb/s* code-excited *(CELP)* codec *[6]* resulting in a *1.1 kBd* arrangement. In both of these cases the subjective speech quality of the codecs was inferior to that of the *13 kb/s GSM* codec. The *SBC* is essentially an open-loop codec giving near-toll quality at relatively low complexity. The *4.8 kb/s CELP* codec is an intelligent analysis-by-synthesis scheme, but inspite of perceptual error weighting and post-filtering its speech quality suffers from the exceptionally low bitrate. Indeed, the *CELP* is the potential candidate half-rate *GSM* codec we use to compare with our high-quality, low-complexity full-rate *RPE* codec.

Details of the *RPE-LTP GSM* codec are readily available in the literature *[1], [7]*, therefore only details necessary for embedding *FEC* are discussed here. The *LPC* analysis framelength is *20 ms* covering *160* input speech samples. We found using line spectrum frequencies *(LSFs)* rather than logarithmic area ratios *(LARs)* slightly more advantageous in terms of robustness due to their ordering properties, as well as in terms of subjective speech quality, which is a slight deviation from the standardised *RPE* codec. Instead of utilising *8 LAR*s quantised with a total of *36 bits* we use *10 LSF*s encoded similarly with a total of *36 bits*. The other subjectively and objectively important variation is the deployment of *4 bits* to quantise the long term predictor gains *(LTPG)* instead of the standardised *2 bits*. This increases the number of bits to *268/20 ms*, resulting in a bitrate of *13.4 kb/s*, but the reward in terms of speech quality makes it worthwhile. The final bit allocation scheme is summarised in *Table 1*.

Parameter	Bit No.	Bitpos. in Frame
8 LFSs	36	1-36
RPE gridpos.	2	37,38
Block max.	6	39-44
RPE exc. pulses	13x3=39	45-83
LTP delay	7	84-90
LTP gain	4	91-94
Per Subsegment	58	
Total bitrate: 36+4x58=268 / 20 ms = 13.4 kb/s		

Table 1: 13.4 kb/s RPE codec bit allocation

5 Bit-Sensitivity Analysis and Embedded Bit-Protection

With the aim of providing appropriate bit-protection we systematically corrupted each bit of the *RPE-LTP* codec in the *268-bit* frame, and evaluated the speech quality degradation in terms of both segmental signal-to-noise ratio *(SEG-SNR)* and cepstral distance *(CD)*, as shown in *Figure 2* and *Figure 3* respectively, for the first *94 bits* representing the *36 LSF bits* as well as the *58 bits* of the first subsegment. Impairments caused by the second, third and fourth subsegments for the bits *95-268* in the frame are identical to those of the first one, whence they are not depicted here.

Our basic observation is that corruption of the *LSF*s and *LTP* parameters as seen at beginning and end of both of the *SEG-SNR* and *CD* figures results in more serious degradations than that of the excitation bits. The *LSF*s determine the spectral envelope, while the *LTP* parameters the spectral fine-structure. However, the *SEG-SNR* sensitivities seem to emphasize the low-frequency *LSF*s and the *LTP* parameters, while the *CD* suggests that almost all *LSF*s are equally important, but underestimates the role of the *LTP* parameters. This is plausible, since the *CD* excels in measuring the envelope distortion, but not in estimating waveform fidelity. This is why the excitation bits hardly inflict *CD* degradation, although their role is clearly recognised down to least *(LSB)*, medium and most significant bit *(MSB)* level in terms of *SEG-SNR*s in *Figure 2*. Yet, *CD* objective measures are regarded to have the highest correlation with subjective listening tests.

Figure 2.: Bit sensitivities in SEGSNR degradation for the 13.4 kbps RPE codec

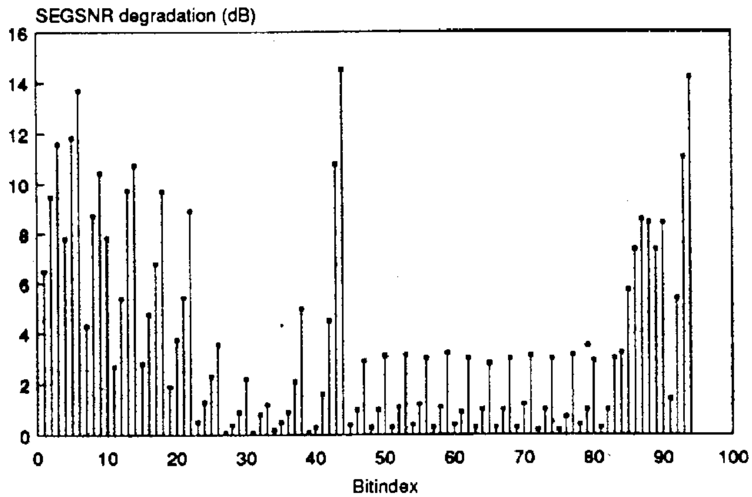
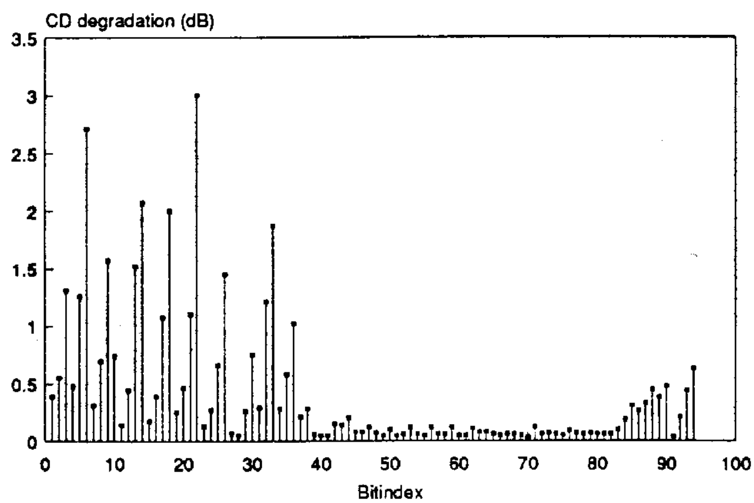


Figure 3.: Bit sensitivities in CD degradation for the 13.4 kbps RPE codec



To form a balanced view of objective bit-sensitivities and take proper account of subjective quality degradations we decided to amalgamate the *SEG-SNR* and *CD* measures into a combined sensitivity figure *S*. Therefore the bits were assigned a sensitivity rank in terms of both *CD* and *SEG-SNR*, where the most sensitive bit was allocated the sensitivity figure *1* and the last sensitive the figure of *94*. Then for each bit between *1* and *94* the *SEG-SNR* and *CD* sensitivity indices were added to form the combined sensitivity figure *S*. Following this the bits were sorted in ascending order of *S*, as tabulated in *Table 2*, where the constituent *SEG-SNR* and *CD* sensitivities are shown as well.

Armed with the sensitivity figures *S*, an arbitrarily complex embedded *FEC* scheme can be implemented. For practical reasons we opted for two protection classes. Observe that assigning any subsegment bit in the *RPE* frame with an index $95>X>36$ to a specific category implies the inclusion of the corresponding bits $(X+58)$, $(X+116)$ and $(X+174)$ of the other subsegments in the same class.

After careful consideration we accomodated *116 bits* in the high protection class and assigned a shortened binary *BCH(62,29,6)* code, while the remaining *152* less significant speech bits are protected by a less powerful shortened *BCH(62,38,4)* code. The codes are appropriately shortened to embrace the *116* and *152* class one and class two speech bits by four rectangularly interleaved *BCH* codewords in each category, which nicely restricts error propagation between adjacent *LPC* frames. Then in case of serious speech impairments the reliable error detection capability of the high protection *BCH* code can be invoked to activate the correlative speech post-processing algorithm proposed by *D.J. Goodman et al. [8]*, as portrayed in *[2]*.

The signalling rate of our *RPE/BCH/16-QAM* system is computed as follows. The *116* highly protected bits are coded into four consecutive *BCH(62,29,6)* codewords giving a total of *248 bits*. Similarly, the *152* low-protection bits are encoded by four *BCH(62,38,4)* codewords, yielding equally *248 bits*.Then a total of *124 16-QAM* symbols per each *20 ms LPC* frame are transmitted, and so the signalling rate is *6.2 kBd*, which requires a transmission bandwidth of around *12 kHz*.

6 Results and Discussion

The serial and *FDM 16-QAM* modems, the *BCH* codecs as well as the *13.4 kb/s RPE-LTP* codec described in earlier sections were simulated, and the *SEG-SNR* objective speech quality measure was evaluated via *Rayleigh*-fading channels with

No. in Frame	Index in Frame	SEGSNR Sensit.	CD Sensit.	S
1	6	3	2	5
2	14	8	3	11
3	22	13	1	14
4	3	5	9	14
5	5	4	10	14
6	9	9	6	15
7	18	11	4	15
8	13	10	7	17
9	94	2	19	21
10	8	14	17	31
11	2	12	21	33
12	93	6	27	33
13	10	18	16	34
14	17	22	13	35
15	21	25	12	37
16	90	17	22	39
17	26	33	8	41
18	88	16	25	41
19	4	19	23	42
20	44	1	43	44
21	87	15	31	46
22	1	23	28	51
23	89	21	30	51
24	12	26	26	52
25	43	7	47	54
26	20	32	24	56
27	85	24	33	57
28	16	29	29	58
29	86	20	38	58
30	38	28	35	63
31	7	31	32	63
32	33	58	5	63
33	30	51	15	66
34	25	50	18	68
35	92	27	42	69
36	42	30	46	76
37	84	34	44	78
38	32	72	11	83
39	36	69	14	83
40	59	35	50	85
41	71	38	51	89
42	53	37	53	90
43	15	48	45	93
44	19	53	40	93
45	37	52	41	93
46	24	56	37	93
47	35	74	20	94
48	50	39	56	95
49	56	41	54	95
50	83	40	57	97
51	47	45	52	97
52	11	49	48	97
53	62	44	55	99
54	77	36	67	103
55	29	71	39	110
56	80	46	68	114
57	68	43	74	117
58	46	62	60	122
59	74	42	80	122
60	65	47	77	124
61	23	75	49	124
62	34	90	36	126
63	64	65	61	126
64	31	93	34	127
65	76	73	58	131
66	58	61	70	131
67	73	66	65	131
68	79	68	64	132
69	67	64	71	135
70	52	60	76	136
71	82	67	69	136
72	41	54	83	137
73	45	78	59	137
74	55	59	84	143
75	48	81	63	144
76	63	84	62	146
77	91	55	93	148
78	78	76	73	149
79	70	57	94	151
80	54	77	75	152
81	49	63	89	152
82	61	70	86	156
83	27	94	66	160
84	57	82	79	161
85	39	92	72	164
86	69	87	82	169
87	72	91	78	169
88	81	88	81	169
89	28	80	90	170
90	40	85	85	170
91	60	79	92	171
92	51	83	88	171
93	75	89	87	176
94	66	86	91	177

Table 2: Combined Bit-Sensitivity Figures for the 13.4 kb/s RPE-LTP Codec

propagation frequency of *1.8 GHz* for a vehicular speed of *30 mph*. We portray our findings for the serial *16-QAM* modem with and without diversity, as well as for the *FDM* modem in *Figure 4*. The *FDM* modem is outperformed by the conventional serial *16-QAM* modem, which guarantees virtually error-free speech transmissions for channel *SNR*s in excess of about *25 dB*, a value easily maintained in microcellular *PCN*s. When second order diversity is used the system's robustness is further improved, approaching a *Gaussian*-like performance. As a benchmarker we also included in *Figure 4* the *SEG-SNR* performance curve of a half-rate type *CELP* codec, convolutional *(CC) FEC* codec and minimum shift keying *(MSK)* system *[9]*, requiring a channel *SNR* of about *20 dB* for near-toll-quality transmissions. The error-free subjective speech quality was rated formally by *GSM* to have a mean opinion score *(MOS)* of four over a range of operating conditions.

Yes, *16-QAM* does provide an attractive long-standing, low-complexity, high-quality alternative to 'any' quadruple complexity low-bitrate speech codec in microcellular *PCN*s, where the extra *5 dB* channel *SNR* reserve required by the *QAM* modem is provided by the link-budget designer. If diversity is deployed, our *RPE/BCH/16-QAM* system has similar robustness to that of the more complex, lower speech-quality *CELP/CC/MSK* system.

References

[1] H.J. Braun, J.E. Natvig: European DMR - The Standardisation Procedure on the Way from Full-Rate Coding to the Half-Rate System, Proc. of EURASIP Workshop on Medium-to-Low Rate Speech Coding, Hersbruck, W-Germany, Sept. 1989

[2] L. Hanzo, R. Steele, P.M. Fortune: A Subband Coding, BCH Coding and 16-QAM System for Mobile Radio Speech Communications, IEEE-VT, Nov. 1990, Vol 39, pp 327-340

[3] W.T. Webb, L. Hanzo, R. Steele: Bandwidth Efficient QAM Schemes for Rayleigh Channels, Proc. of the 5th Int. Conf. on Radio Receivers, Cambridge, U.K., 1990, No 325, pp 139-143

[4] K.H.H. Wong, L. Hanzo, R. Steele: Channel Coding for Satellite Mobile Channels, Int. Journal on Satellite Communications, Vol 7, No 2, 1989, pp 143-163

[5] S.B. Weinstein: Data Transmission by Frequency-Division Multiplexing Using Discrete Fourier Transform, IEEE Tr. on Communications, COM-19, No 5, Oct. 1971

[6] L. Hanzo, R. Salami, R. Steele, P.M. Fortune: Transmission of Digitally Encoded Speech at 1.2 KBd, submitted to IEEE Vehicular Technology, 1990;

[7] P. Vary et al.: RPE-LPC Codec - The Candidate for the GSM Radio Communication System, Int. Conf. on Digital Land Mobile Radio, 1987, Venice, Italy, pp 507-516

[8] O.J. Wasem, D.J. Goodman et al.: The Effect of Waveform Substitution on the Quality of PCM Packet Communications, IEEE Tr. ASSP, Vol 36, No 3, March 1988, pp 342-348

[9] R.A. Salami, K.H.H. Wong, R.Steele, D. Appleby: Performance of Error-Protected Binary Pulse Excitation Coders at 1.4 Kb/s over Mobile Radio Channels, Proc. of ICASSP'90, pp 473-476

Figure 4.: Speech quality of proposed RPE/BCH/16-QAM systems