On the Theoretical and Practical Limits of Digital QSOs

Klaus von der Heide, DJ5HG

1. Introduction

A usual digital communication system consists of hardware and software. In order to reach maximum sensitivity, every hardware component must carefully be tuned to optimal power or lowest noise. This paper deals with the part of the system which is implemented in software on the PC. The theoretical lower bound of digital information transfer is analysed under the special constraints of radio amateur QSOs.

Weak-signal communication systems usually are evaluated by the relation of the required energy per information bit $\mathbf{E_b}$ to the noise power per Hz bandwidth $\mathbf{N_0} = \boldsymbol{k}_\mathbf{B}\mathbf{T}$ with the Boltzmann-constant $\boldsymbol{k}_\mathbf{B}$ and the equivalent noise-temperature $\mathbf{T}$. In a spaceprobe, the value $\mathbf{E_b/N_0}$ determines the number of information bits that can be communicated with a limited battery. In an optimized BPSK-sytem, $\mathbf{E_b/N_0}$ is the SNR at the input of the decoder. On the other hand, radio amateurs use the **SNR** at the audio output of an SSB-transceiver, i.e. in a bandwidth of 2500 Hz. Unfortunately, digital modes with different periods cannot be compared directly in this case. The transformation between both scales is easily performed by

$$\mathbf{SNR} = \mathbf{E_b/N_0} + 10*\log_{10}(\text{ number\_infobits / period / bandwidth })$$
$$\mathbf{E_b/N_0} = \mathbf{SNR} - 10*\log_{10}(\text{ number\_infobits / period / bandwidth })$$

For the threshold sensitivity SNR=−24 dB of JT65 in bandwidth 2500 Hz we get

$$\mathbf{E_b/N_0} = -24 - 10*\log_{10}(72 / 47.8 / 2500) = +8.2 \text{ dB}$$

C.E.Shannon found by mathematical treatment [1] that confident communication only is possible if

$$\mathbf{E_b/N_0} > \ln(2) \quad \text{or in dB:} \quad \mathbf{E_b/N_0} > 10*\log_{10}(\ln(2)) = -1.6 \text{ dB}$$

Applied to the usual EME-case of transmission of about 70 bits within about 50 seconds we get the minimum possible SNR in bandwidth 2500 Hz (for communication at 100% correct decode):

$$\mathbf{SNR} > -1.6 + 10*\log_{10}(70 / 50 / 2500) = -34.1 \text{ dB.}$$

Therefore, at least theoretically, a gain of 10 dB over JT65 could be possible.

In 1959 Shannon refined the lower bound of communication as a function of the number $\mathbf{k}$ of information bits encoded as transmitted blocks, and of the code rate $\mathbf{r}$ which is the relation of the number of information bits $\mathbf{k}$ in a block and the number of resulting bits $\mathbf{n}$ of the transmitted block, and finally of the required block error rate $\mathbf{P_W}$ [2]. This lower bound is called the *sphere-packing lower bound*.

Figure 1 shows the sphere-packing bound (for $\mathbf{r} = 0$) over the number of information bits encoded in blocks, and the $\mathbf{E_b/N_0}$ at which interplanetary communication systems reach the low block error rate $\mathbf{P_W = 10^{-4}}$ [3]. Figure 1a adds the sphere-packing bound for lower code rates, and the Plotkin-bound [4].

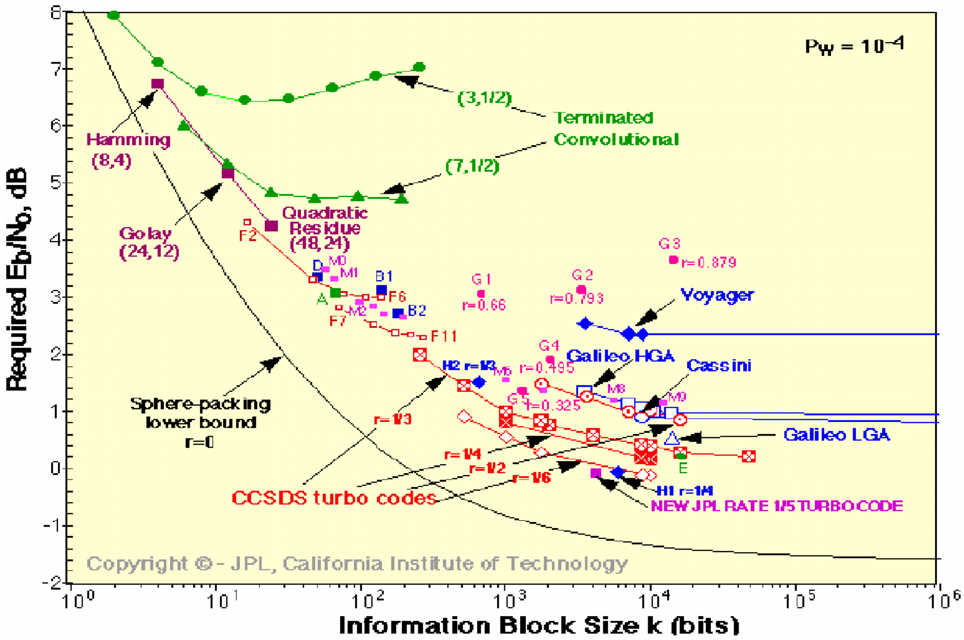

Figure 1. Evaluation of some codes and spacecrafts by the JPL for the block error rate $\mathbf{P_W=10^{-4}}$.

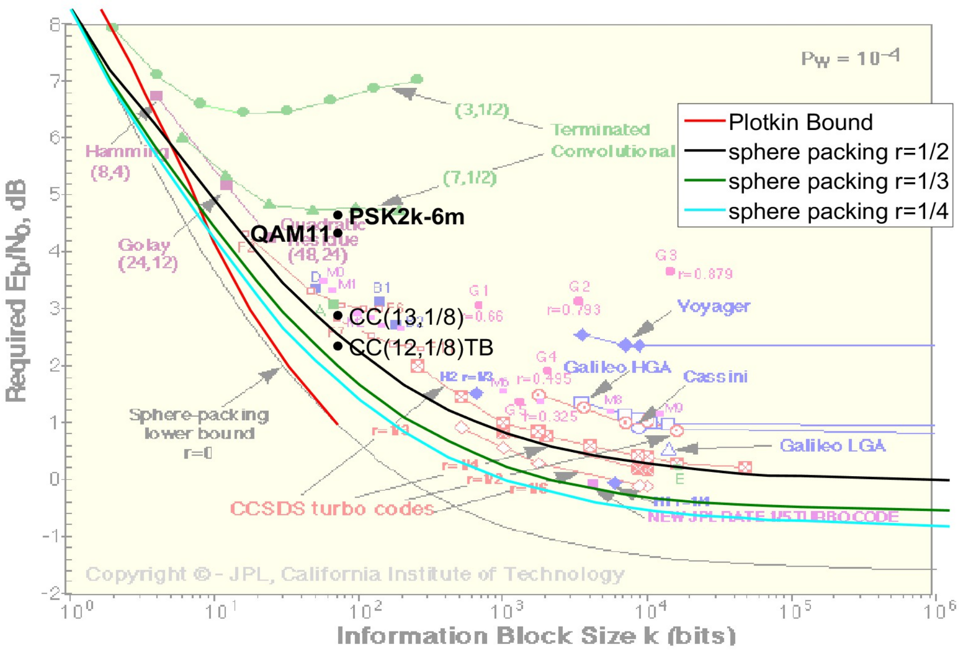

Figure 1a. This figure adds to Figure 1: the sphere-packing bound for **r** = 1/2, 1/3, and 1/4 [6], and the Plotkin bound (for any **r**). There is obviously some distance between the interplanetary communication and the sphere-packing bound for **r** = 0, but they nearly reach the limits set by the lower code rates. Also added are two codes found by the author and used in PSK2k and in the experimental mode QAM11. The required $\mathbf{E_b/N_0}$ of the modes PSK2k and QAM11 are considerably worse compared the the codes. This is caused by the necessary synchronization and phase-recovery (see Chapter 4 of this paper).

In contrast to commercial and scientific communication, we are interested to transmit small blocks of about **k**=60 bits of information, and we are happy if the rate of successful decodes is just **P**$_W$=50%. So the question is: Where are our weak-signal systems in a diagram for **P**$_W$**=0.5**? This is shown in Figure 2. The modes used by radio amateurs obviously are more or less far away from the theoretical limit. It is profitable to know the reasons, because the goal of a design of a weak-signal mode should be to approach the limit. We will now discuss the main three topics.

2. Codes for the Radio-Amateur Weak-Signal Modes

Figure 2 shows a gap in the required **E_b/N_0** of 6 dB between uncoded transmission and lower bound for optimally encoded transmission at the typical number **k** = 60 of information bits in a weak-signal QSO. A good code should approach the lower bound as near as possible, and decoding on a usual PC must be possible in about a second or less, but it's gain cannot be larger than 6 dB.

Convolutional Codes (**CC**) are a large class with selectable code rate **r** = 1/2, 1/3, ... , 1/16, and smaller. The distance between the lower bounds for codes with **r** = 1/2 and those with **r** = 0 is 1.6 dB at **k** = 60 . We therefore should choose a code with low rate (**r** = 1/8 ... 1/16).

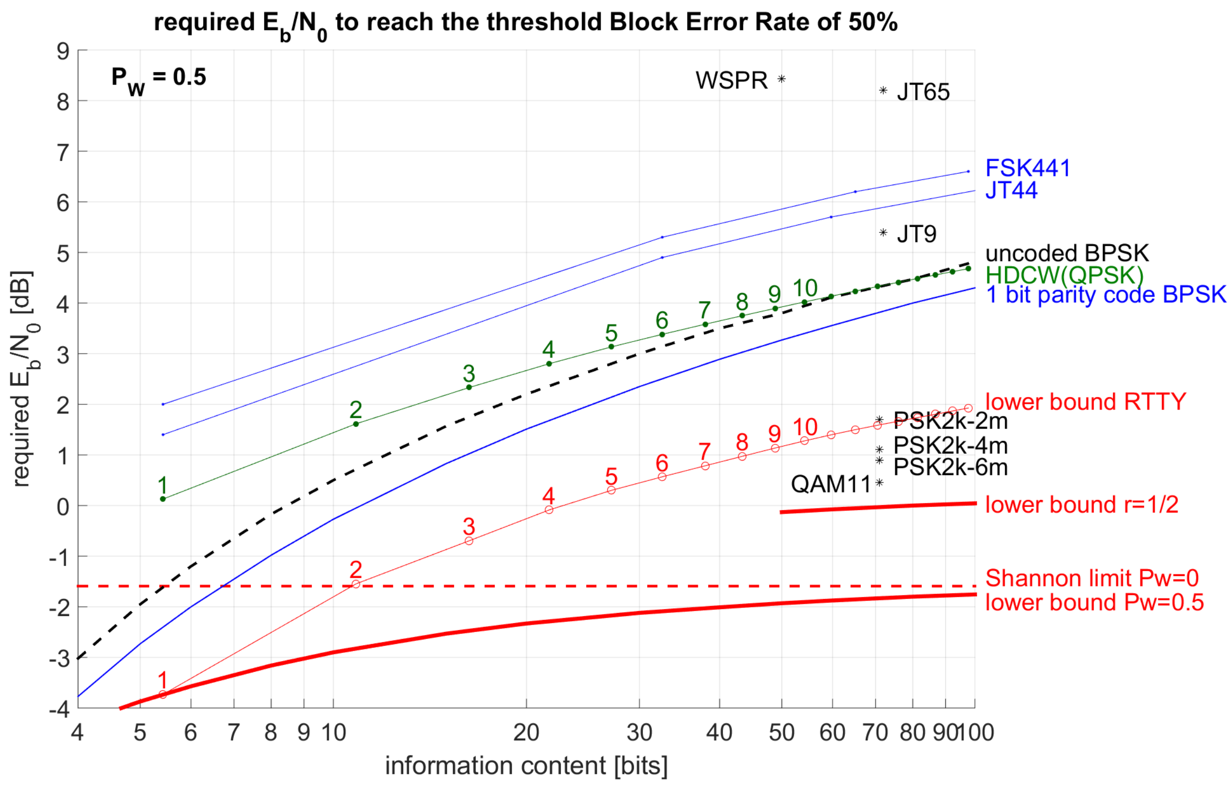

Figure 2. A diagram corresponding to Figure 1, but for **P**$_W$**= 0.5** looks quite different. The reason is that there is a considerable probability for receiving a correct block out of pure noise. Especially 1-bit-blocks are correct by 50% even if there is no signal. The horizontal scale of the diagram therefore is limited to the interesting region k = 4 ... 100. The green line named "lower bound RTTY" means an alphabet of 43 characters encoded by the optimal Hadamard43-code. JT44 and HDCW also belong to this class of character-oriented transmissions. The synchronization of single character-blocks causes a loss of about 3 dB (see Chapter 4).

A second parameter of convolutional codes is the so-called constraint length **cl** . Large constraint length makes the code better, but increasing **cl** by 1 nearly doubles the decoding time of the usual Viterbi-decoder. This limits **cl** to less than 16 or even lower. In addition, **cl** should not be larger than **k**/4. Otherwise the overhead of the long tail decreases the code performance.

A very effective method to increase the performance of convolutional codes is tail-biting. Given the constraint length **cl** (and **r** < 1/3), tailbiting makes the code much better if **k** is within the interval 4***cl** ... 20***cl**. Figures 3 and 4 show results of the author's simulations. The encoding of tail-ended and tail-biting convolutional codes is explained in the documentation of PSK2k [5]. The author decodes tail-biting blocks by applying the normal Viterbi-decoder to the concatenation of five times the received block enclosed by (1/**r**) * (**cl** − 1) zeros at the beginning and at the end. The result is an array of **cl** − 1 + 5***k** bits. Bits **cl**+2***k** ... **cl**+3***k**−1 are taken as the **k** decoded information bits. This procedure surely could considerably be improved. The results presented here were obtained with this simple algorithm.

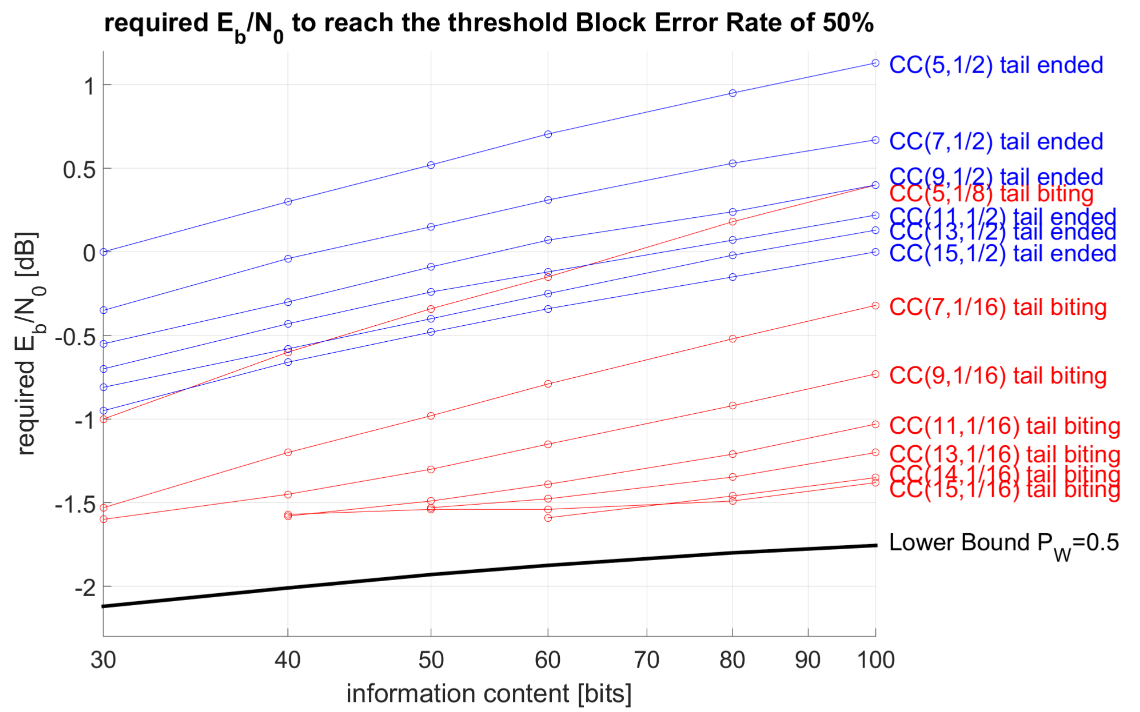

Figure 3. Required E_b/N_0 to reach the threshold block error rate of 50% by selected convolutional codes. The codes are named by CC(cl, r). All codes are taken from the author's web page [7]. In the essential region of **k** = 60 ... 70, some codes approach the theoretical lower bound to less than 0.4 dB distance.

We can conclude from Figures 3 and 4 that the convolutional codes with r = 1/16 and cl = 12, 13, and 14 are very good codes for the case of weak-signal applications in amateur radio. If the application does not allow such low code rates we must accept a loss. In the case of PSK2k-2m, the SSB-bandwidth does not allow a Baud-rate larger than 2000. This limits the number of bits of the codeword which can be sent within a short meteorscatter ping to about 250 including all additional pilote bits sent for synchronization and phase recovery. Therefore, the code rate cannot be lower than r = 1/2. This causes a loss of 1.6 dB. The 4m- and 6m-modes use r = 1/4 and r = 1/8 with a much lower loss. 1/3 of the transmitted bits are pilote bits. This overhead causes an additional loss of $10*\log_{10}(3/2)$ = 1.8 dB. The sensitivity of PSK2k-2m therefore should be

E_b/N_0 of the CC(13,1/2) at k = 71 taken from Figure 3: −0.1 dB

loss by adding the pilote (1 pilote bit per 2 codeword bits): 1.8 dB

the sum is the theoretical E_b/N_0 of PSK2k-2m: 1.7 dB

in scale with bandwidth 2500 Hz: **SNR** = 1.7 + 10*$\log_{10}$(71 / 0.129 / 2500) = −4.9 dB

This fits well to the experiments [8] with the existing PSK2k-receiver where −6 dB and −4 dB resulted in block error rates 90% and 0% resp. We can conclude this chapter 2 with the satisfactory message that there exist well usable codes for weak-signal applications which approach the theoretical lower bound better than 0.4 dB. The next chapter will show that this unfortunately only is true for PSK.

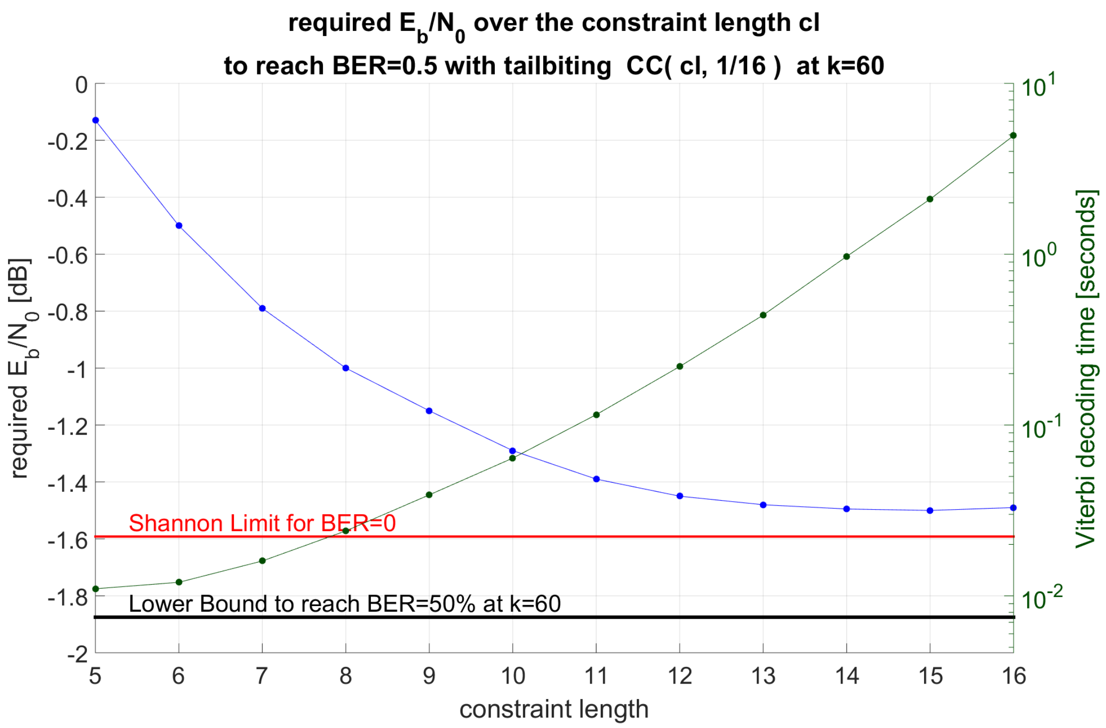

Figure 4. Required E_b/N_0 to reach the threshold block error rate of 50% by convolutional codes with different constraint length **cl** and fixed **r** = 1/16 (blue line), and decoding time on a PC (green line). A useful code obviously is the tail-biting CC(13, 1/16).

3. Modulation and Demodulation

Figure 5 compares the bit error rates at very low signal levels for some modulation types. It is obvious that PSK is much better than all other types. The difference of the required E_b/N_0 gets even larger with decreasing signal levels. At bit error rate 0.1, FSK and PSK differ by 6 dB (horizontal distance in Figure 5), at bit error rate 0.4, the difference already approaches 12 dB. Such large error rates (or corresponding small E_b/N_0 values) are quite normal if low-rate codes are used. Let for example be **r** = 1/16, that means the number of transmitted bits is 16 times larger than the number **k** of information bits. If the codeword is received with E_b/N_0 = −1.5 dB (the threshold sensitivity of CC(13,1/16)), then the energy per codeword bit is 12 dB less: −13.5 dB. We get the corresponding bit error rate of PSK from Figure 5. It is 0.38.

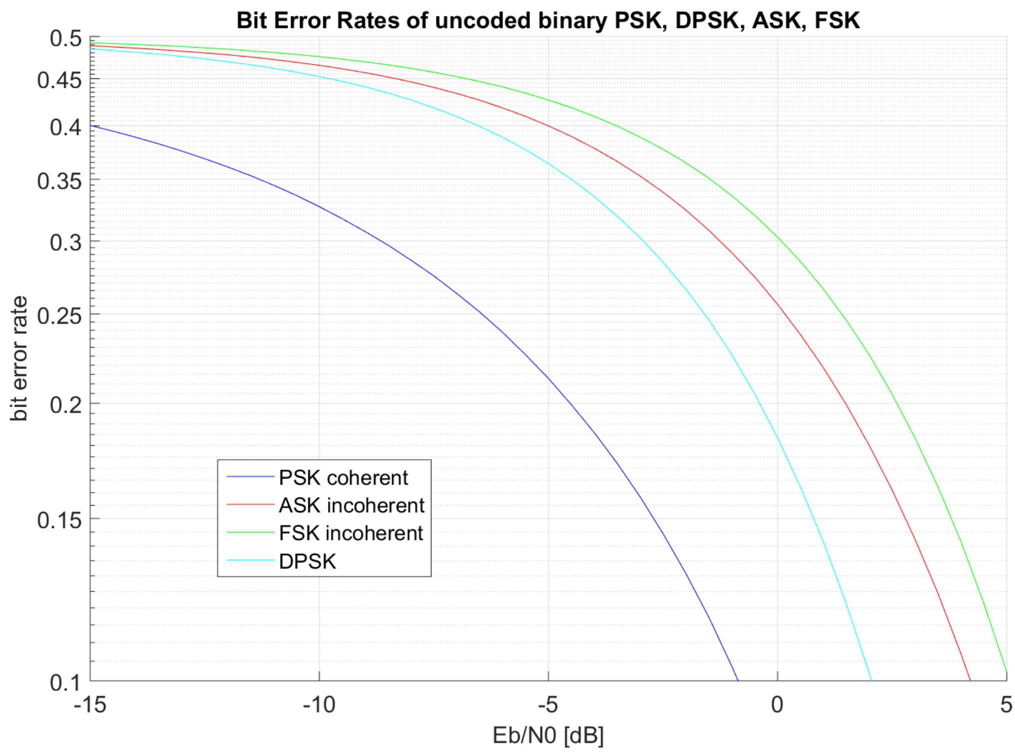

Figure 5. The bit error rates of modulation types differ very much at the same energy $\mathbf{E_b/N_0}$ in case of very low signal levels. The (horizontal) difference between PSK and FSK at the bit error rate 0.4 is nearly 12 dB, while it is 6 dB at the bit error rate 0.1.

The $\mathbf{E_b/N_0}$ to get the same bit error rate with FSK is 11 dB worse. I.e. if the same code is used combined with FSK instead of PSK, the transmitter must use more than 10 times the power needed by PSK. This shows that low-rate codes cannot be used conveniently with FSK.

A corresponding calculation with the higher-rate code CC(15,1/2) yields a difference of 6 dB between PSK and FSK. But if PSK with the low-rate code is compared to FSK with the high-rate code, then the difference is 8 dB.

ASK is about 1 dB more sensitive than FSK in the weak-signal area. It has the additional advantage of insensitivity to spreaded radio channels (FAI, aurora). On the other hand, there are the well known disadvantages of ASK, mainly the electromagnetic interference.

DPSK has nearly the same sensitivity as BPSK at large signal levels. It is even better than BPSK down to about $\mathbf{E_b/N_0} = 5$ dB if phase recovery of BPSK has to be supported by an additional pilote.

m-FSK is a good choice for uncoded communication. Figure 6 shows the symbol error rates of m-FSK for different values of m. The error rates are considerably reduced for $\mathbf{E_b/N_0} > 2$ dB at large m. But, following the argumentation of above, error correcting codes are nearly unusable.
An example for m-FSK is JT44. It uses a 43-FSK. A transmitted block consists of 22 symbols which are sent 3 times in sequence within 25 seconds. The 50% block error rate (all 22 symbols correct) is reached at the symbol error rate $1 - 0.5^{(1/22)} = 0.031$. It follows from figure 6 that $\mathbf{E_b/N_0} =$ +3.3 dB yields this symbol error rate in a 43-FSK. The JT44-block contains 66 data symbols interleaved by 69 synchronization symbols. The synchronization therefore causes a loss of 10*log10(135/66) = 3.1 dB. The transformation to SNR in 2500 Hz yields the 50%-threshold to get all 22 characters of the transmitted text fully correct:

$$\textbf{SNR} = 3.3 + 3.1 + 10*\log_{10}(22*\log2(43) / 25 / 2500) = -20.8 \text{ dB}$$

This theoretical lower bound fits well to the practical observation. It follows directly from Figure 2 that JT44 seems to be better than JT65. This is true only on non-fading channels, where JT44 needs only half the energy of JT65 to communicate an information bit. In contrast to PSK, coding cannot lead to considerable gain in sensitivity if FSK is used. But it does a good job in case of fading.

Our conclusion of this chapter is: BPSK is the only modulation for communication with very weak signals. In case of uncoded transmission, m-FSK may be a good alternative.

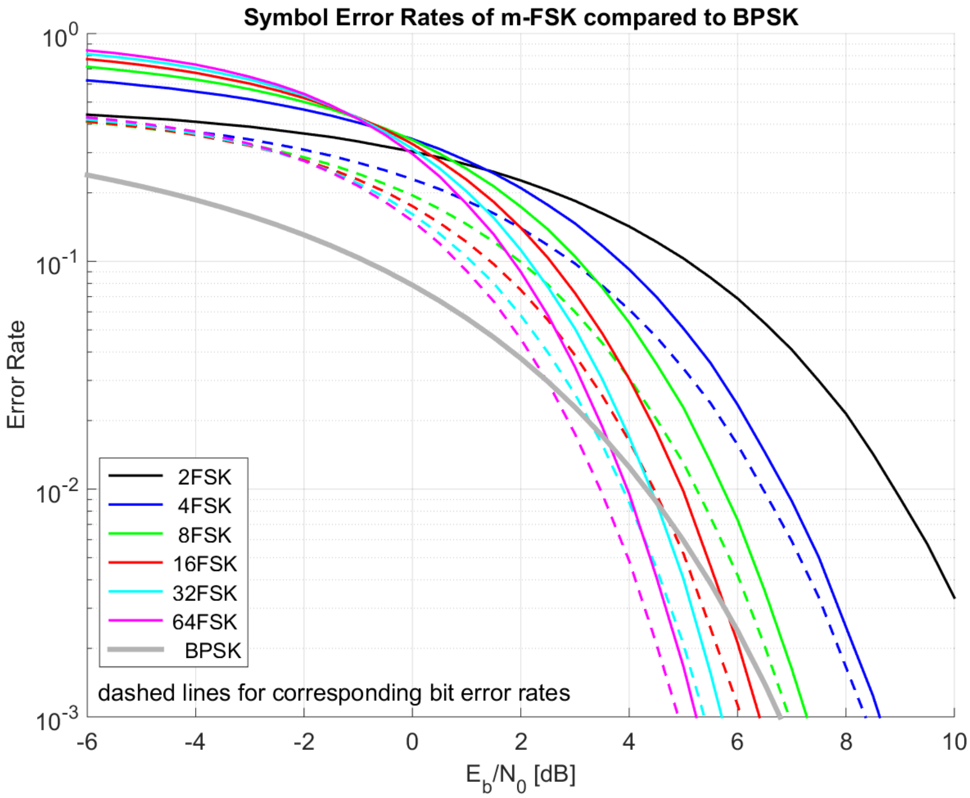

Figure 6. Symbol Error Rates of m-FSK and corresponding bit error rates (dashed lines)

4. Synchronization and Phase Recovery

If small blocks are used in a weak-signal communication, it is very difficult to find the block in time-domain and in frequency-domain. This problem is exhaustively discussed in [9], and the author made a contrution to it on the last EME-conference [10].

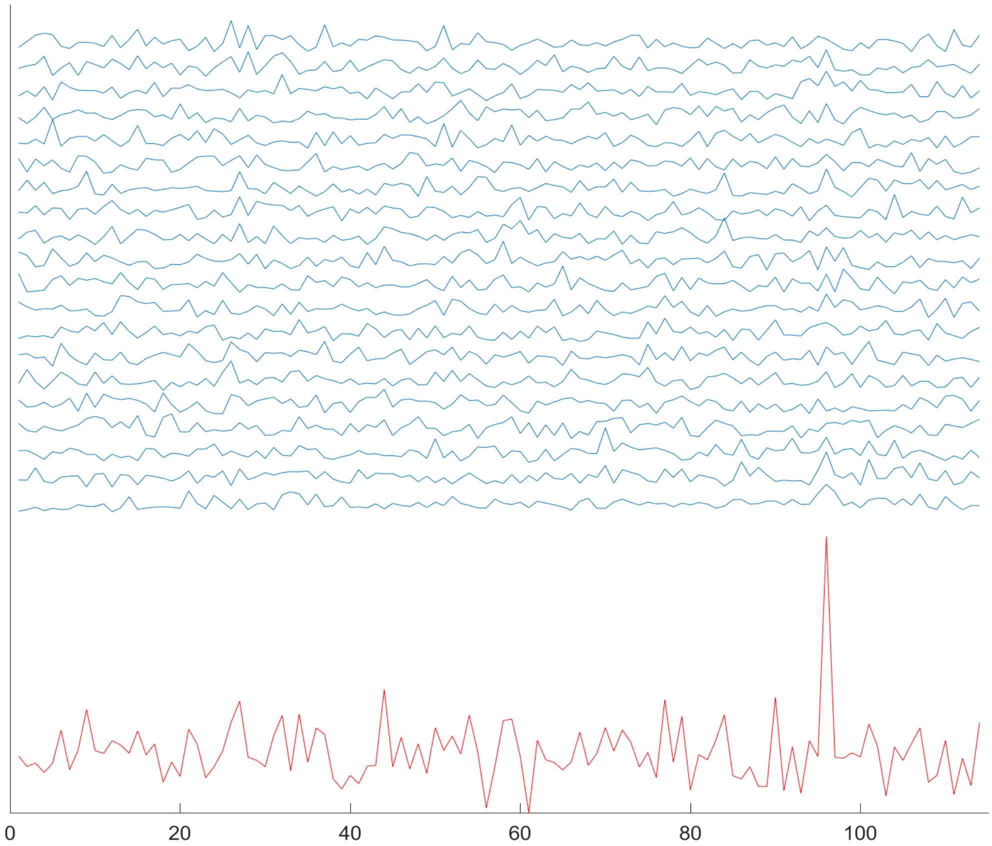

Figure 7. Example for the search of the pilote pattern of QAM11 (length 283 bits) at SNR = −29 dB and phase distortion of coherence length 5 s (corresponding to 28 bits). The pattern was divided into 20 overlapping segments of length 28. The complex input signal is filtered by all segments. The absolute-values of all 20 outputs are shown above. None of them has a significant indication of a hit, but the sum (the lower line) clearly has.

The search for a weak signal in frequency domain may be supported by a waterfall-diagram. But this does not work in case of PSK, and it fails completely for signals with SNR lower than about -28 dB in 2500 Hz bandwidth. Therefore the only solution is the exhaustive search on all parallel channels within a given bandwidth. The spacing of the channels is proportional to the Baud-rate. A high Baud-rate reduces the search in frequency domain. This fits well to the demand for low-rate codes.

The search in time-domain can easily be realized by interleaving the data-bits with a fixed binary pattern of low autocorrelation. The author prefers Hadamard-codes. In order to profit from the sensitivity of coherent PSK, the pattern must be divided into segments of length less than the coherence time. The correlation is performed with all segments and the results are absolutely added at correct delays. Figure 7 gives an example for the Hadamardcode of 283 bits and coherence length corresponding to 28 bits. The reverse pattern was divided into 20 overlapping segments of length 28 which where taken as the filter-coefficients of 20 filters. The figure shows the absolut-values of the individual filter outputs (already correctly shifted) and the sum of these outputs. The phases of the filteroutputs at the hit are good estimates of the carrierphase. A simple interpolation yields the development of the carrierphase over the whole length of the inputsignal. At this length of the pilote pattern, BPSK is 12 dB more sensitive than FSK, and the sensitivity even increases with longer patterns. The author uses this method of synchronization and phase-recovery in PSK2k with the interleaving sequence $d_1\, p_1\, d_2\, d_3\, p_2\, d_4\, d_5\, p_3\, d_6\, d_7\, p_4\, d_8$ ... where d_i and p_i denote data resp. pilote bits with index *i*.

The very simple experimental modes QAM11 and QAM66 use a tail-biting CC(12,1/8) to generate 568 data bits out of 71 information bits. The pilote also has 568 bits (see the appendix). But there is no interleaving of pilote and data. The data are modulated on the sinewave carrier (full amplitude), and the pilote bits are shifted by a halve bit and modulated on the cosinewave (halve amplitude). This Offset-QAM reduces the principal loss caused by sending the pilote to

$$10*\log_{10}(\text{ (energy of data) / (energy of pilote and data) }) = -1.0 \text{ dB}.$$

The reduction of the power for the pilote leads to an increase of the rate of failure of the synchronization at very low signal levels. Nevertheless, this tradeoff between synchronization failure and decoding failure leads to a 50%-threshold-SNR of −32 dB in 2500 Hz bandwidth on the AWGN-channel ($\mathbf{E_b/N_0}$ = 0.46 dB). That is a gain of 8 dB compared to JT65. And it is about 2.5 dB above the corresponding sphere packing bound (see Figure 1a). If we subtract the known losses, 1 dB loss by the pilote transmission and 0.5 dB loss by the used code, the remaining distance from the theoretical lower bound only is 1 dB. It is caused by different reasons, synchronization failures for example.

On a fading channel, the phase must be recovered as a function of time over the whole receiving period. If the coherence time for example is 1 second then there are only 11 noisy pilote bits within this time segment to support the phase recovery. The recovered phase therefore may be very noisy which reduces the sensitivity considerably. Figure 8 shows the block error rates of QAM11 and QAM66 for different coherence times. The phase recovery for PSK or QAM generally fails if the coherence time is less than 1 / Baudrate.

5. FSK vs. PSK

As long as the phase recovery is successful, PSK and QAM are more sensitive than FSK. But if the phase distortion by the radio channel is too large, only FSK and ASK are usable. It is interesting to get a closer look to the transition region. Let the coherence time of phase distortion be 0.4 s. From Figure 8 we get a block error rate of 90% at -25 dB. If the transmission period now is compressed to 1/6 (QAM66), and the power correspondingly increased by 7.8 dB, then QAM66 will run with a block error rate of only 6% at a 6 times larger bit rate, but with the same $\mathbf{E_b/N_0}$. The reason is the increase of sensitivity of PSK with lower (i.e. more stretched) phase distortion. See Figure 8 for the dependency of sensitivity on the coherence time in case of QAM11/66. Thus PSK sets a lower bound for the rate *bits* / (*coherence time*). In other words: if the coherence time is low, you have to use a high bitrate. To communicate the high bit rate on a weak radio channel you are forced to transmit with appropriate high power. The energy per bit remains the same as in Figure 1. On the other hand, the transmitting power is limited for radio amateurs. Therefore the demands for PSK cannot be satisfied in some cases. This especially concerns EME-contacts with strong libration fading on GHz-bands. If PSK does not work we must accept the general loss of other modulations plus the loss of coding gain which adds to at least 5 dB. This is not *weak-signal* communication, but *bad-channel* communication.

A further problem arises on channels with both, frequency-spreading and time-spreading. Time spreading leads to intersymbol interference (ISI), if the Baudrate is higher than 1 / (time spreading). PSK cannot be used on such over-spreaded channels. An example is aurora, but not EME.

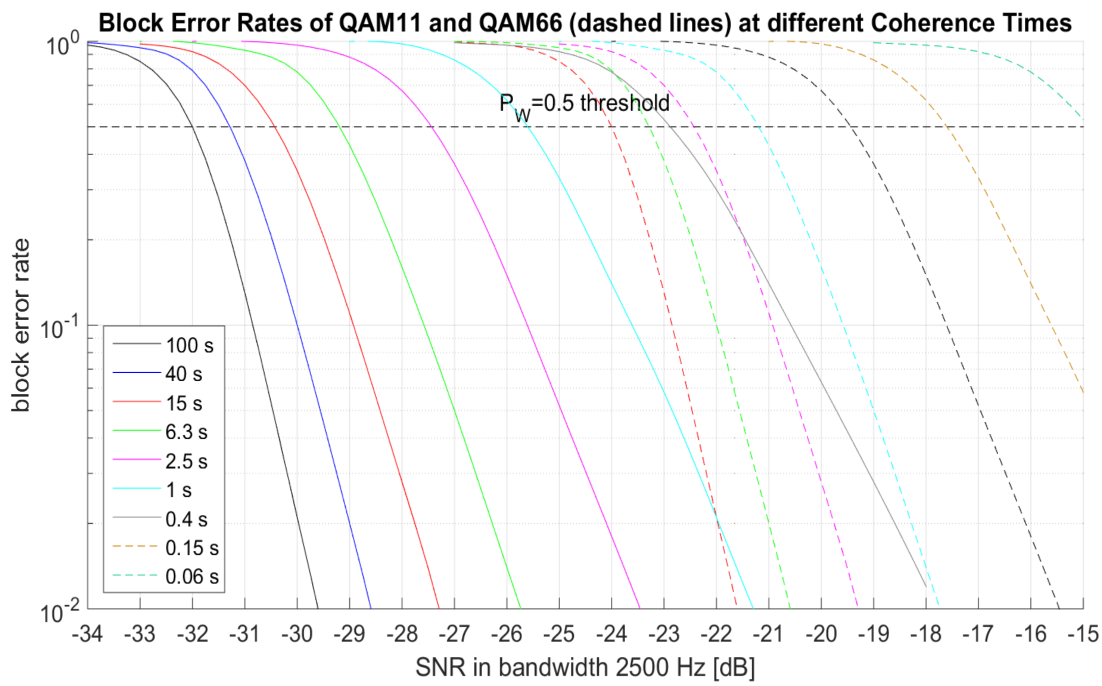

Figure 8. Block error rates of QAM11 and QAM66 for different coherence times.

References

[1] C.E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp.379-423, 623-656, (1948)

[2] C.E. Shannon, Probability of Error for Optimal Codes in a Gaussian Channel, The Bell System Technical Journal, Vol. 38, pp. 611-656, (1959)

[3] original URL http://www331.jpl.nasa.gov/public/AllCodesVsSize.GIF

[4] M. Plotkin, Binary codes with specified minimum distance, IRE Transactions on Information Theory, 6:445-450, (1960)

[5] http://www.dj5hg.de/digitalmodes/psk2k_documentation.pdf (chapter 5)

[6] S. Dolinar, D. Divsalar, and F. Pollara, Code Performance as a Function of Block Size, TMO Progress Report 42-133, (1998)

[7] http://www.dj5hg.de/code/codes.html

[8] K. von der Heide, FSK441 vs. PSK2k by Simulation Experiments, DUBUS vol. 42/2, pp. 67-77, (2012)

[9] K. von der Heide, Modulation, Demodulation, and Block Localization for Digital Weak-Signal Modes, DUBUS vol.42/3, pp.63-75, (2013)

[10] K. von der Heide, A Coherent Digital Mode for EME, EME 2014 conference, pp. 129-134

Appendix: QAM11 (QAM66)

Source Code:	56 bits of information plus 15 check bits for final error detection (71 bits)
Channel Code:	CC (12, 1/8) tail-biting and interleaved, results in 8*71 = 568 bits
Pilote Sequence:	Hadamard-283 interleaved with the encoded receiver-address of 284 bits
Receiver Address:	54 address bits plus 17 check bits = 71 bits encoded with CC (12, 1/4) TB
Modulation:	QAM, data on sine with amplitude 0.89, pilote on cosine with amplitude 0.45
Baud-rate:	8000 / 720 = 11.1111 (66.6666)
Total Bandwidth:	12 Hz (72 Hz)

Length of Transmission: 568.5 * 720 / 8000 = 51.165 s, period 60 s (8.5275 s no fixed period)