Demodulation of FSK signals

- the demodulation of FSK signals is a particular case of FM demodulation. Therefore, the general FM demodulation methods, as well as some specific demodulation methods that take advantage of the particularities of the FSK signals, may be employed.

The frequency-discriminator demodulator, see figure 13

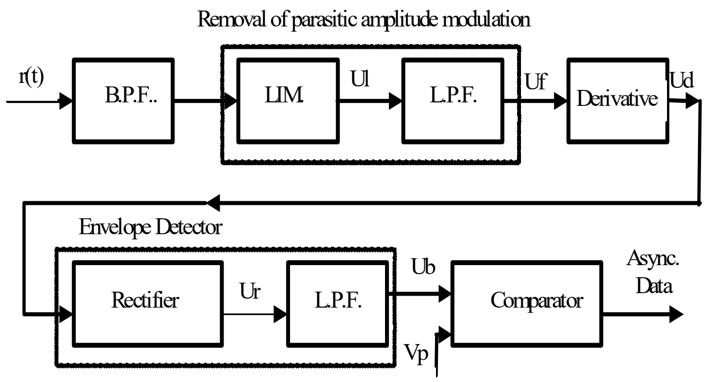


Figure 13. Block diagram of the FSK demodulator using a frequency discriminator

- before the actual demodulation, the received signal is BP-filtered to improve its SNR
- then, the "parasitic" AM is removed in a manner similar to the FM receiver, see FM lectures;
- the received signal, $r(t) = R(t)\cdot\cos\varphi_r(t)$, see (24), is performed according to:

$$U_l(t) = \begin{cases} +V; \cos\varphi_r(t) > 0 \\ -V; \cos\varphi_r(t) < 0 \end{cases} \qquad (28)$$

- since this signal is periodical, in terms of the $\varphi_r(t)$ variable, it may be expanded in Fourier series:

$$U_l(t) = \frac{4V}{\pi}[\sin\varphi(t) - \frac{1}{3}\sin 3\varphi(t) + \frac{1}{5}\sin 5\varphi(t)...]; \qquad (29)$$

- the higher odd harmonics are attenuated by an LPF, if the approximate separation condition (29) is observed.

$$3\cdot f_c - 2.2\cdot f_s > f_c + 1.2\cdot f_s \rightarrow f_c > 1.8\cdot f_s\,; \qquad (30)$$

- condition (30) is computed considering that the modulation index of the 3rd harmonic is $3\cdot h$, see (29) and the FM lectures, and the bandwidth of the modulated signal around this harmonic is approximately $4.4\cdot f_s$.
- the limited and filtered signal would be:

$$s_{lf}(t) = 4V\sin(\omega_c t + \varphi_r(t))/\pi; \qquad (31$$

- the actual demodulation consists of the two steps presented in the FM lectures:

- The derivative of the received FSK signal, generating an AM modulation proportional to the momentary frequency;
- The envelope detection that generates a low frequency signal whose amplitude is proportional to deviation of the momentary frequency around the central frequency ;

- the signal after the derivative is:

$$U_d(t) = [4V(\omega_c + \omega_{in}(t))\cos(\omega_c t + \varphi_r(t))]/\pi; \qquad (32)$$

- it is AM modulated with a signal that is proportional to the momentary frequency of the FSK, see (25) in the previous lecture on FSK.
- the envelope detection extracts a low-frequency signal whose amplitude is proportional to the amplitude of the fundamental of the modulating signal. This is accomplished by rectifying and LP-filtering the U_d signal, as shown in the Annex 2 of the second lecture on LM; K_a is the constant of the envelope detector:

$$U_b(t) = 4VK_a(\omega_c + \omega_{in}(t))/\pi; \qquad (33)$$

- other variants of performing the derivative and envelope detection are discussed in Annex 1 of the FM lecture; further details in can be found in reference [Ed. Nicolau].
- the $U_b(t)$ voltage is compared to a threshold voltage $V_p = 4VK_a/\pi$, which equals the d.c. component of the demodulated signal; this delivers two logical levels, which correspond to the demodulated asynchronous data:

$$+V \rightarrow \text{"0" and } -V \rightarrow \text{"1"}; \qquad (34)$$

- though it has good performance in the presence of noise, the method involves a rather complicated demodulation; it is an adapted version of a demodulator specific for the FM signals with analog modulating signals.

"Zero-Crossings" FSK demodulator

- this demodulator transforms the FSK signal into a Pulse-Frequency Modulation (PFM) and extracts by LP-filtering the fundamental of this signal; then a comparator generates the demodulated data

- the operation of this demodulator is based on two properties:

1) the deviation of the momentary frequency around the central frequency is approximately proportional to the level of the fundamental component of the modulating signal
2) the level of the fundamental component of the PFM signal is proportional to the deviation, around the f_c, of the momentary frequency of the received FSK signal

- the first property is justified by (26); the second will be proven below
- applying transitivity upon the two properties above, one may conclude that the level of the fundamental component of the PFM signal is approximately proportional to the level of the fundamental component of the modulating data signal
- by comparing the proportional to the level of the fundamental component of the modulating signal to a threshold U_p = 0V, which is actually the decision step, we get the estimated modulating data signal.
- the main steps of the ZC-demodulation are:

- transform the FSK into PFM
- extract the fundamental component of PFM
- perform the decision

- the block diagram of this demodulator is shown in figure 14 and the signal diagram in figure 15

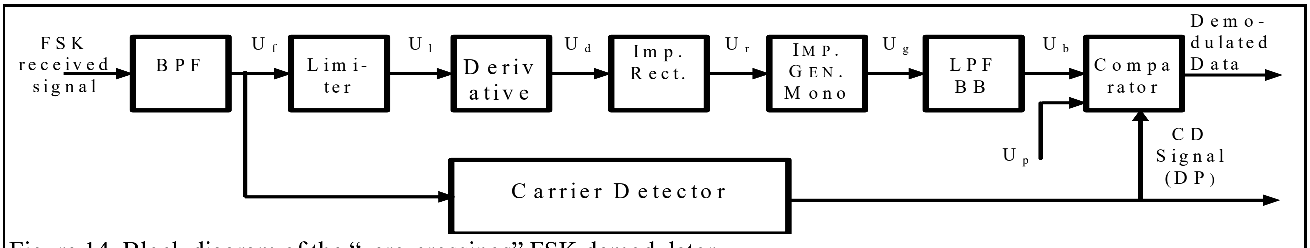


Figure 14. Block diagram of the “zero-crossings” FSK demodulator

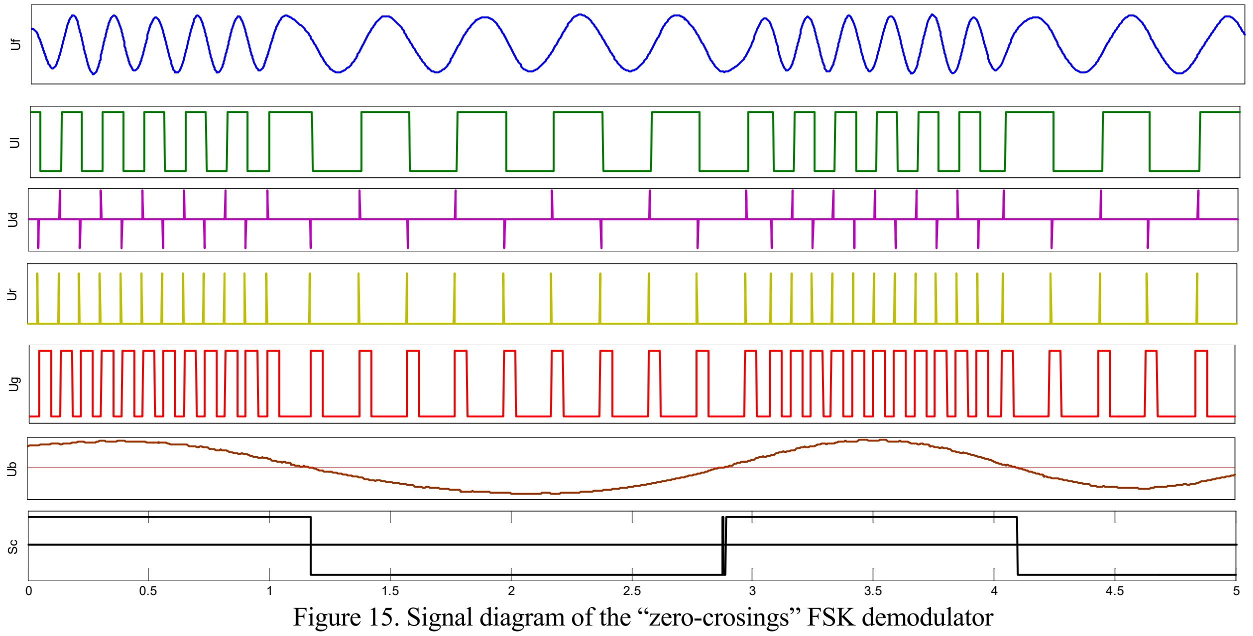


Figure 15. Signal diagram of the “zero-crosings” FSK demodulator

- the limitation of the received signal ensures, besides the removal of the parasitic AM, a larger variation of the signal near the zero-crossing moments; the derivative of the limited signal marks the zero-crossing moments by generating an impulse at every zero-crossing.
- these impulses are rectified and used to trigger a monostable circuit, which gives impulses of constant amplitude +U and duration τ, which outputs bipolar voltages.
- since the amplitude and duration of the impulse are constant, but the moments when the impulses start, due to zero-crossings, are variable, the processing including the derivative and impulse generation transforms the FSK modulation into a pulse-frequency modulation PFM.
- the fundamental of the PFM signal is obtained by BB-LPF, which has its cut-off frequency placed between the fundamental frequency and second harmonic of this signal.
- the second property mentioned above is based on the approximation of the LP filtering by the computation of the average values of the signal between two consecutive zero-crossings, followed by the accumulation of these values; this approximation is valid if the time-interval between two zero-crossings

is much smaller than the period of the signal which is extracted by filtering, namely. the fundamental of the PFM signal. This requirement is better observed by modulated signals that have more transitions through zero during a symbol period.

- to compute the average value of the U_g signal, between two zero-crossings t_n, t_{n+1}, we will consider that the monostable signal has the +U level for a duration equaling τ and –U for the rest of the interval between the two zero-crossings, i.e.:

$$U_G = \begin{cases} U\,; t_n < t \le t_n + \tau\,; \\ -U\,; t_n + \tau < t \le t_{n+1}\,; \end{cases} \tag{35}$$

- the momentary phase of the cosine carrier of the FSK signal observes at the zero-crossing moments t_n and t_{n+1}, the relations (36), where $\Psi_v(t)$ denotes the variable phase of the FSK signal generated by the frequency deviation.

$$\omega_c \cdot t_n + \psi_v(t_n) = \frac{(2n+1)\pi}{2}; \quad \omega_c \cdot t_{n+1} + \psi_v(t_{n+1}) = \frac{(2n+3)\pi}{2};$$

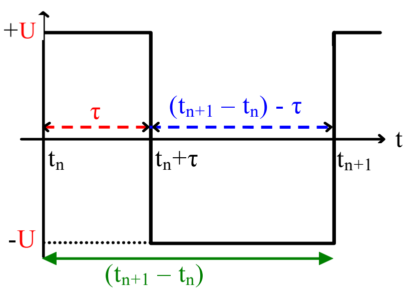
- subtracting the two equalities of (36) and knowing that the phase difference between two zero-crossings of a cosine signal equals π, we get:

$$\omega_c + \frac{\psi_v(t_{n+1} - \psi_v(t_n)}{t_{n+1} - t_n} = \frac{\pi}{t_{n+1} - t_n}; \tag{37}$$

- since the time interval between two zero-crossings is assumed to be extremely short, the second term of the sum in (37) approximates the derivative vs. time of the momentary phase, which equals the variable pulsation, so (37) becomes:

$$\omega_c + \psi_v^{'}(t_n) = \omega_c + \omega_v(t_n); \tag{38}$$

- replacing (37) in (38) and simplifying π, we get: $2f_c + 2f_v(t) = 1/(t_{n+1}-t_n)$; (39)



- on the other hand, the average value of $U_g(t)$ during the elementary interval $[t_n, t_{n+1}]$, is expressed by(40), see figure 16.

Figure 16. Computation of the average value R_0

$$R_0 = \frac{U \cdot (t_n + \tau - t_n) - U \cdot (t_{n+1} - t_n - \tau)}{t_{n+1} - t_n} = \frac{2U\tau}{t_{n+1} - t_n} - U\,; \tag{40}$$

- replacing (39) in (40) we get the average value of the signal which is extracted by LP filtering:

$$R_0(t) = U(4 \cdot f_c \cdot \tau - 1) + 4U \cdot \tau \cdot f_v(t) \tag{41}$$

- the average signal is composed of two voltages:

- a d.c. component, first term of (41), depending only of transmission parameters f_c, τ and U;
- a variable component, proportional to the frequency deviation, around f_c, of the received signal

- the d.c. component can be removed if the monostable time constant equals $\tau = 1/4f_c$; so the filtered signal becomes:

$$U_c = 0 \text{ if } \tau = \frac{1}{4f_c} \Rightarrow R_0 = \frac{U \cdot f_v(t)}{f_c}; \qquad (42$$

- relation (42) justifies the property 2) presented above.
- the variable pulsation can be computed for the “1010…” data sequence as:

$$\omega_v(t) = 2\pi \cdot f_{VM} \cdot \sin(\frac{\omega_s}{2} t); \tag{43}$$

while its maximum frequency deviation f_{VM} (= Δf_{max}) is computed using relations (6) and (7) and is expressed by:

$$f_{VM} = h \cdot f_s / 2; \tag{44}$$

- this deviation equals Δf_M from the FM transmissions and is imposed by the modulation index;
- it occurrs when the modulating signal takes the maximum admitted level and it is smaller than the maximum deviation existing in the spectrum of the modulated signal, see (25) in the previous FSK lecture and the considerations presented close to it.

- the variable component would have positive values for momentary frequencies > f_c and negative values for momentary frequencies < f_c. Therefore, the $R_0(t)$ signal would have negative values during the "1"-bits and positive values during the "0"-bit, because $f_2 > f_c > f_1$.
- comparing the $R_0(t)$ signal with a threshold voltage $U_p = U_c = 0$, we get the demodulated data that have the +A level for "0" and –A for "1", as shown in figures 14 and 15.
- since the synchronized receive clock was not employed in the demodulation process, the demodulated data are asynchronous.

Effects of the frequency-shifts upon the performances of the ZC-demodulator

- if the frequencies f_1 and f_2 are modified, at the transmission end, with $df_1 = df_2$, or the spectrum is frequency-shifted by the channel with df, then the central frequency f_c would be shifted by df_c:

$$df_c = (df_1 + df_2)/2 \; ; \qquad (45)$$

- if the frequency-shifts suffered by the f_1 and f_2 frequencies are small, compared to their nominal values, then the frequency deviation inserted by the modulating signal is not modified significantly; therefore the variable component f_v is approximately constant.
- using these assumptions, the signal after the BB-LPF $R_0(t)$ would be:

$$R_0(t) = U[4(f_c \pm df_c)\tau - 1] + 4Uf_v(t)\tau \; ; \qquad (46)$$

- using the value imposed to the time-constant τ, (42), the value of the filtered signal is:

$$R_0(t) = \pm\frac{df_c}{f_c}\cdot U + U\cdot\frac{f_v(t)}{f_c}; \qquad (47)$$

- compared to its ideal value, (42), the filtered signal has now an additional d.c. component

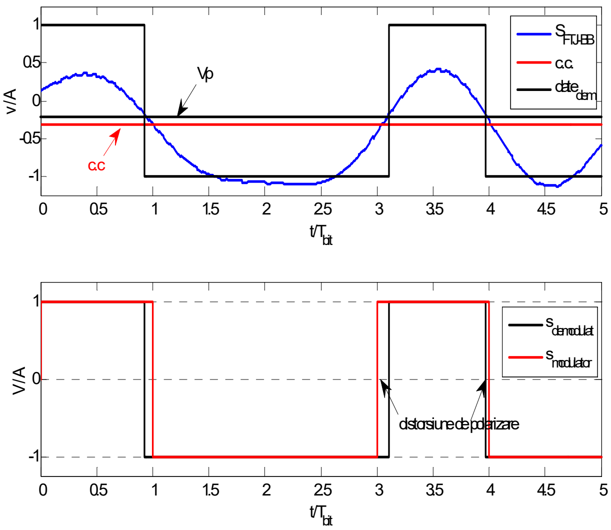


Figure 17. Representation of the bias distortion

- so the variable demodulated signal is shifted on a non-zero d.c. component, thus changing the durations of the two types of demodulated bits (after the comparator) and inserting the *bias distortion,* see figure 17. If df_c is significantly smaller than f_c, the effect of this d.c. component is negligible.
- as example, the ITU-T V.23 Recommendation allows tolerances $\Delta f_i = \pm 10$ Hz for the frequencies f_1 and f_2, which, associated with the maximum allowed frequency shift induced by the channel which equals ±6 Hz, lead to maximum frequency shifts of ±16 Hz of the frequency spectrum of the received signal. Then, the maximum value of the additional d.c. component that appears after the BB-LPF is negligible, see (48).

$$\frac{\pm df_c}{f_c}\cdot U = \frac{\pm 16}{1700}\cdot U < \pm 0{,}01\cdot U \; ; \qquad (48)$$

- the monostable circuit employed should operate with bipolar voltages.

HOMEWORK: Compute the expression of the $R_0(t)$ signal if the monostable operates between 0V and U and show the effects upon the demodulated data considering the noise margin as well.

- the SNR performances of this demodulator improve if the frequency of the modulated signal, compared to the symbol frequency, increases, i.e. the number of zero-crossings during a symbol period increases. With the increase of the frequency of the modulated signal, the SNR performances of this demodulator get closer to the ones of the frequency-discriminator demodulator.
- if the values of the frequencies employed for f_1 and f_2 decrease, compared to the symbol frequency, i.e. the number of zero-crossings during a symbol period gets small, the SNR performances of this demodulator decrease significantly.

FSK demodulator with delay-derivative and coherent detection

- this demodulator is also known as the differential detector and represents an adaptation of the same demodulator presented in the Annex 2 of the FM lecture. Its block diagram is shown in figure 18.
- after the removal of the parasitic amplitude modulation the received FSK signal is:

$$s_{FSK}'(t) = A\sin[\omega_c t + \Psi_v(t)]; \qquad (49)$$

where $\Psi_v(t)$ denotes the variable part of the momentary phase.

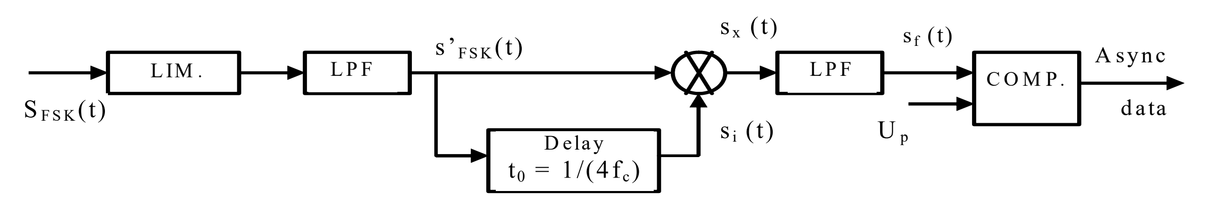


Figure 18. Block diagram of the FSK demodulator with delay-derivative and coherent detection

- the product between the signal (49) and its delayed version is expressed by(50), where K denotes the multiplier's constant.

$$s_x(t) = A\sin(\omega_c t + \psi_v(t)) \cdot A\sin(\omega_c(t - t_0) + \psi_v(t - t_0)) = \\ = \frac{A^2}{2K}\cos(\omega_c t_0 + \psi_v(t) - \psi_v(t - t_0)) - \frac{A^2}{2K}\cos(2\omega_c t + \omega_c t_o + \psi_v(t) + \psi_v(t - t_0)); \quad (50)$$

- after the LP filtering, which removes the spectral components around $2f_c$, the signal is:

$$s_f(t) = \frac{A^2}{2K}\cos\left[\omega_c t_0 + \psi_v(t) - \psi_v(t - t_0)\right] \quad (51)$$

- if the delay-time t_0 takes a value that corresponds to a $\pi/2$ phase-shift of the central pulsation ω_c,(52).a, the value of this delay-time t_0 being expressed in terms of f_c by(52).b, then the expression of the filtered signal is(52).c; this expression is obtained by using the approximation(52).d of the derivative of the variable part of the momentary phase.

$$\omega_c t_0 = \frac{\pi}{2};\ (a) \Rightarrow t_0 = \frac{1}{4f_c};\ (b) \Rightarrow$$

$$s_f(t) = -\frac{A^2}{2K}\cdot\sin(\psi_v(t) - \psi_v(t - t_0)) = -\frac{A^2}{2K}\cdot\sin(t_0\frac{\psi_v(t) - \psi_v(t - t_0)}{t_0}) \approx \quad (52)$$

$$\approx -\frac{A^2}{2K}\cdot\sin(\frac{2\pi\cdot f_v(t)}{4f_c}) = -\frac{A^2}{2K}\cdot\sin(\frac{\pi\cdot f_v(t)}{2f_c});\ (c);\quad \frac{\psi_v(t) - \psi_v(t - t_0)}{t_0} \approx \frac{d\psi_v(t)}{dt} = 2\pi\cdot f_v(t);\ (d)$$

- knowing that the frequency deviation around f_c is smaller than half of f_c, the sign of the sine function in (52).c would keep the sign of the frequency deviation, so the sign of the filtered signal $s_f(t)$ would be the same as the one of the frequency deviation.
- by comparing the filtered signal to 0V threshold voltage we get the demodulated data, the positive value corresponding to "0" and the negative one to "1".
- because it doesn't employ a symbol (bit) clock synchronized to the received signal, this demodulator is an asynchronous one, delivering asynchronous data at its output.
- literature shows that this demodulator ensure the best SNR performances, out of the three demodulators presented here, on channels that insert significant group-delay distortions.

Synchronization of the receiver's symbol (bit) clock
- the FSK demodulators deliver asynchronous data
- in synchronous transmissions, the modem should also provide to the destination computer, the receive clock RxCk of frequency f_{bit} (or f_{symbol} for transmissions mapping more than one bit/symbol), which should be synchronized to the demodulated data; this means that the negative edges of the synchronized clock should coincide to the bit-changing moments, if the signals are represented in a positive logic.
- to accomplish this, the demodulated asynchronous data are employed as a reference signal in the dynamic and fast synchronization circuits of a synchronization block, as the one described in the BB transmissions chapter.
- the synchronization block consists of a dynamic synchro circuit and sometimes of a fast synchro one; the resynchronization circuit is not required, because a phase incertitude of 180º of the reference signal may not occur. The synchronization block also delivers data that are synchronous with the RxCk.
- the dynamic synchro, synchronizes a local clock with frequency f_{bit}, because the binary FSK carries 1 bit/symbol, using a $f_{atack} = 2^n\cdot f_{bit}$.
- as a phase reference signal, the phase comparator employs the asynchronous demodulated data.
- during the RTS/CTS interval, the data pattern transmitted the 1:1 pattern provided by the modem, because it ensures a maximum number of transitions of the phase reference signal.

Block diagram of a modem that employs FSK – individual study

- the block diagram of such a modem is depicted in figure 19.

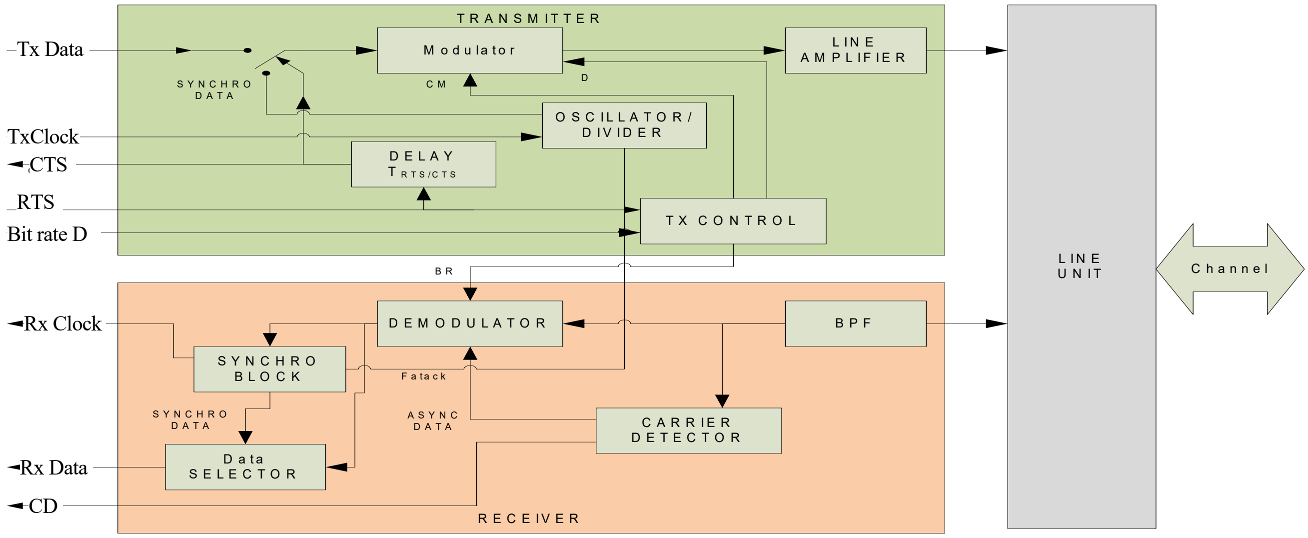


Figure 19. Block diagram of the FSK modem – type V.23 ITU-T

- the **transmitter** is composed of :

- transmission control block, which, enables or disables the modulator (CM), depending of the state of the RTS (Request to Send) signal, and, for the half-duplex connection, it enables/disables the receiver (BR); for this type of connection the transmission has priority. For a full-duplex transmission, the control of the reception by the transmission control block is disabled. This block also performs the selection of the set of frequencies assigned to the logical levels, depending of the bit rate employed, by means of the control signal D.
- the Delay RTS/CTS block controls the state of the CTS (Clear to Send) signal, depending on the RTS state and commands, and, depending of the CTS state, the position of the K switch (CK) that inserts the data into the modulator. During the synchronization interval T_0, it inserts the synchronization data (a 1:1 pattern) from the oscillator-divider circuit; then, after the CTS goes active (high), it inserts the computer data, TxD.
- the oscillator-divider block that, employing a crystal-controlled oscillator, generates the transmission clock TxCk, the f_{atack} clock signal and the synchronization data pattern.
- the line-amplifier (AL) is an amplifier whose gain can be modified in steps; it establishes the level of the transmitted signal.

- the **receiver** is composed of:

- the input BP filter that limits the BW of the received signal to the one of the FSK signal, improving the SNR;
- the carrier detector block (CD-DP), which compares the level of the received signal with a reference level, enabling the receiver only if the received signal has the highest level.
- the demodulator that delivers the asynchronous demodulated data, RxD;
- the local clock synchronization block, which synchronizes a locally generated clock, $f_{local} = f_{bit}$, to the demodulated data, providing the synchronized receive data and the receive clock RxCk. The type of the data (asynchronous or synchronous) delivered to the computer is selected by the user, by selecting the desired output, depending of the type of connection employed.
- the line unit, which ensures the connection to the transmission channel and the type of connection, i.e. half or full-duplex

Error performances of the FSK modulation

- the BER vs. SNR is the main metric for performance evaluation
- the theoretical evaluation of BER vs. SNR should consider some particularities of the FSK, namely:

- FSK filtering generates a non-linear variation of the momentary frequency of the filtered signal, see (24) of the previous FSK lecture;
- the received signal is affected by noise in different manners for different modulating data sequences, due to the non-linearity of the demodulation process;
- a Gaussian distributed noise at the input of the demodulator generates a non-Gaussian distributed noise

at the demodulator's output.

- the computation of the average BER vs. SNR of the FSK is complex and should also consider the operating principle of the demodulator and the filter employed.
- the value of BER provided by the coherent demodulator can be computed approximately by (53), where SNR is expressed in dB.

$$\mathrm{BER} = Q\left(\sqrt{\rho\cdot\left(1-\frac{\sin(2\pi\cdot h)}{2\pi\cdot h}\right)}\right) \approx \frac{e^{-\frac{\rho}{2}\cdot(1-\frac{\sin(2\pi\cdot h)}{2\pi\cdot h})}}{\sqrt{2\pi\cdot\rho\cdot\left(1-\frac{\sin(2\pi\cdot h)}{2\pi\cdot h}\right)}};\quad \rho = 10^{(\mathrm{SNR}/10)} \tag{53}$$

- the BER ensured by the ZC (non-coherent) demodulator can be approximated by:

$$\mathrm{BER} = 0.5\exp(-\rho/2);\ \ \rho = 10^{(\mathrm{SNR}/10)} \tag{54}$$

- figure 20 presents the BER vs. SNR (at the input of the demodulator) provided by the coherent and by the non-coherent demodulators, for modulating data sequences that ensure equal occurrence-probabilities for the "0" and "1" values.

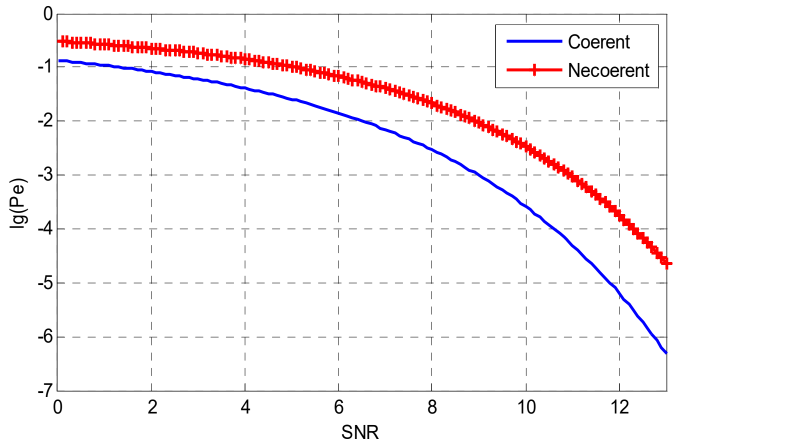


Figure 20. BER vs. SNR of the FSK –coherent and non-coherent demodulation

- comparing the two curves one might notice that the non-coherent demodulator provides greater BER values than the coherent demodulator, at the same SNR and for the same input filter.
- the SNR needed by 2-FSK to ensure a target BER value is with greater than the one needed by 2-PSK; for BER=$1\cdot10^{-5}$, 2-FSK requires an SNR greater with approximately 3 dB.
- the telegraphic distortion of the demodulated asynchronous data depends of the passing band of the filter. A larger BW, compared to the one defined by (10), decreases this distortion, but increases the power of noise, decreasing the SNR.
- the measured values of this distortion also depend of the modulation method, i.e.:

• for a zero crossing demodulator and a flat filter, the telegraphic distortion is 5-6%, if an analog modulator that modulates on a triangular carrier is employed;
• for the same demodulator and filter, but for a Walsh-synthesis modulator, the telegraphic distortion is about 2-3%.

- for the synchronous demodulated data, the receive clock is synchronized to the data, so the telegraphic distortion is removed; it is "replaced" by the phase jitter of the local clock, which equals the phase step of the dynamic synchronization.
- the frequency shifts have reduced effects upon the performances of the FSK, as long as they are within the limits imposed by the pertinent standards.
- the FSK is almost insensitive to the group delay distortion of the channel, especially when the coherent demodulator is employed.
- concluding, the FSK is recommendable on channels that exhibit significant phase and frequency distortions, inserted either by propagation causes, or by processing imperfections or by the mobility of transmitters/receivers.