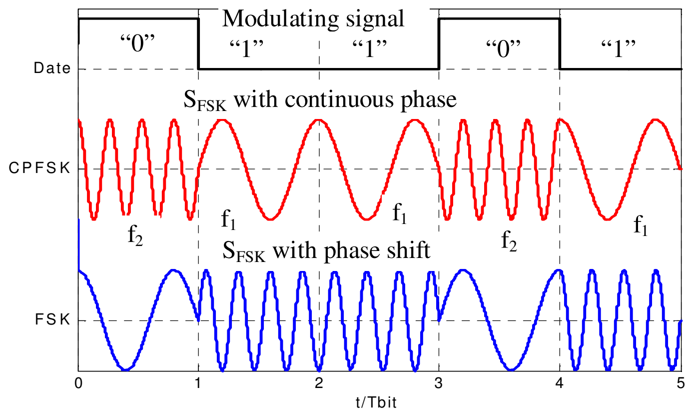
Frequency Shift Keying – FSK

- the FSK consists of frequency modulating a cosine (harmonic) channel carrier in terms of the logical value of the bit to be modulated.
- A cosine signal of frequency f_1 is transmitted for bit "1" and a frequency f_2 for bit "0".
- The frequency of the carrier is kept constant for the entire symbol period.
- Since a symbol carries 1 bit, the bit-period equals the symbol period

- This modulation is the binary version of the M-FSK modulation, where the momentary frequency has $M = 2^n$ values, corresponding to the M combinations of n bits.

Characteristics of FSK



- FSK is a frequency modulation with a continuous phase (CPFSK), i.e. the switching from one frequency to the other is performed with keeping the continuity of the carrier signal phase – see figure 1. "1" ↔ -1; "0" ↔ +1

Figure 1 B-FSK modulation rule

- 2-FSK is a non-coherent modulation, since it is not performed in the rhythm of a symbol-clock signal, but the temporary frequency is controlled by the level of the modulating signal. The symbol duration should not necessarily be an integer multiple of an elementary clock period.

Expression of the CP-FSK signal

- the FSK signal is expressed in terms of the frequencies f_k associated to the $M=2^n$ symbol values by:

$$s_{FSK}(t) = \sum_{k=0}^{\infty} A\cos\left(2\pi f_k t + \varphi_{k,k-1}\right) \cdot u_{T_s}(t - k \cdot T_s) \qquad (1)$$

- the momentary frequency of the M-FSK signal may take M distinct values (M=2 for B-FSK, fig. 1), while, for CPFSK, the $\varphi_{k,k-1}$ should take values which would ensure the phase continuity at the end of the symbol periods (when the modulating multibits are changed). T_s denotes the symbol period.
- for the phase-shift FSK, $\varphi_{k,k-1}$ is constant, depending on the difference of the initial phases of the two f_2 and f_1 oscillators, for each symbol period, see (9) below.
- taking into account that the FSK is a frequency modulated signal with a rectangular modulating signal, its Fourier-series decomposition is expressed by (2), where the parameters are defined in (3):

$$s(t) = A \sum_{n=-\infty}^{\infty} C_n \cos[(\omega_c + n\omega_N)t + \alpha]; \qquad (2)$$

$$\omega_c = \frac{\omega_1 + \omega_2}{2} \text{ a.}; \quad \omega_N = \frac{\pi}{T} = \frac{\omega_s}{2} \text{ b.}; \quad h = \frac{\omega_2 - \omega_1}{2\omega_N} = \frac{\Delta\omega_M}{\omega_s / 2} \text{ c.}; \ \omega_2 - \omega_1 = 2 \cdot \Delta\omega_M \qquad (3)$$

- for the "1010..." data sequence (also named the "1:1" sequence), which is the modulating signal that has the greatest fundamental frequency, the C_n coefficients have the values (4):

$$C_n = \frac{2 \cdot h}{\pi(h^2 - n^2)} \cdot \sin\frac{\pi}{2}(h + n); \qquad (4)$$

- the amplitudes of the spectral components decrease with the squared value of their index; therefore the energy of the modulated signal is concentrated close to the frequencies f_1 and f_2;
- the FSK signal may be considered as an FM signal. The f_c frequency would correspond to the FM carrier frequency, for $U_{mod} = 0V$.
- but, in FSK the modulating signal takes only bipolar values, +/-A (see (5)), so the modulated signal would have the momentary frequency equaling f_c only during the frequency shifts generated by the change of the bit's logical value.
- the FSK's signal momentary frequency may be expressed by (5), where $S_{m\,max} = |+/-A|$ and $S_m = (+/-A)$.

$$f_{ins} = f_c + \Delta f \cdot \frac{S_{modulator}}{S_{modulator\,max}}; \quad S_{modulator} = a_k \cdot A; \quad a_k = 1 - 2 \cdot b_k = \begin{cases} +1 & b_k = 0 \\ -1 & b_k = 1 \end{cases} \tag{5}$$

- due to the nature of the modulating signal, the frequency f_c is not regarded as the carrier frequency, but the central frequency of the spectrum of the modulated signal.
- for an arbitrary data sequence, the C_n coefficients have more complex expressions, the spectrum of the modulated signal becomes continuous, but its bandwidth is still infinite.
- the power spectral distribution of the FSK signal depends essentially of the modulating index, denoted here by h (6), where f_s is the symbol frequency:

$$h = \frac{f_2 - f_1}{f_s} = \frac{2\Delta f_M}{f_s} = \frac{\Delta f_M}{f_s/2}; \tag{6}$$

- recall that for an FM signal the modulation index is: $\beta = \Delta f_M / f_{mod\,max}$; (7)
- comparing (6) and (7) we get that *h* is a modulating index for a modulation where the maximum modulating frequency f_{mM} equals $f_s/2$, the Nyquist frequency, and the maximum frequency deviation Δf_M equals half of the difference between the f_1 and f_2 frequencies.
- therefore, the 2-FSK signal may also be expressed as an FM signal by (9); f_c is defined in (3).a; a_k in (5).

$$s_{B-FSK}(t) = A \cdot \cos\left(2\pi \cdot f_c \cdot t + 2\pi \cdot \Delta f_M \cdot \int_0^{t'} \left(\sum_{k=0}^{\infty} a_k u(\tau - kT_s) \right) \cdot d\tau \right) \tag{8}$$

- the carrier's phase variation $\Delta\Phi$ during one symbol period, i.e. for $t' \in [kT_s; (k+1)T_s]$, is:

$$s_{B-FSK}(t) = A \cdot \cos(2\pi \cdot f_c \cdot t + 2\pi \cdot \Delta f_M \cdot \int_0^{t'} a_k \cdot d\tau) = A \cdot \cos(2\pi \cdot f_c \cdot t + a_k \cdot 2\pi \cdot \Delta f_M \cdot t') \Rightarrow$$
$$\Rightarrow \Delta\Phi = 2\pi \cdot (f_c + a_k \cdot \Delta f_M) \cdot T_s; \quad \text{for} \quad t' \in [kT_s; (k+1)T_s) \tag{9}$$

- the computation of the power spectral density is complex and is performed in terms of the rated frequency F, defined by (10);

$$F = (f - f_c) / f_s; \tag{10}$$

- the PSD vs. F curve of the CPFSK, for a pseudorandom data sequence, is shown in figure 2, in its low-pass equivalent.
- for a modulated signal the spectrum should be translated on the central frequency f_c.
- a case of interest is h = 0,65:
- the power of the modulated signal is concentrated in a relatively narrow BW around the f_c (11). The width of the frequency band in (11) depends on the power percentage of the non-filtered FSK signal that is to be "retained" after filtering

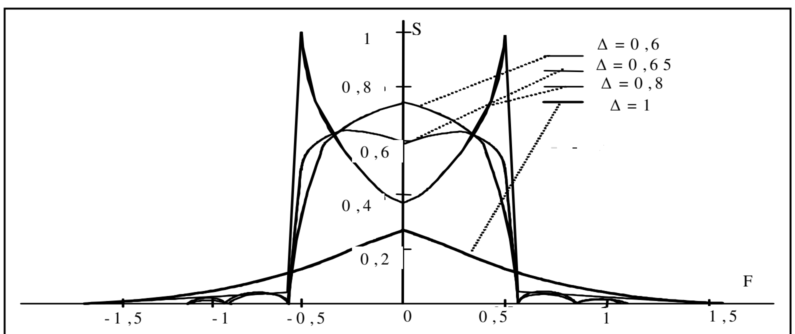


Figure 2. Power spectral density of the CP-FSK signal - schematical representation

$$F \in [-0{,}5\ +0{,}5] \leftrightarrow f \in [f_c - 0{,}5 f_s;\ f_c + 0{,}5 f_s] \text{ or } F \in [-0{,}6\ +0{,}6] \leftrightarrow f \in [f_c - 0{,}6 f_s;\ f_c + 0{,}6 f_s] \tag{11}$$

- the components outside this BW have small amplitudes and the sum of their powers is negligible;
- the distribution of the components inside this BW is relatively flat.

- for h ∈ (0,7;1) the spectrum exhibits the same concentration of the power distribution, but the components on f_1 and f_2 are very large, leading to complications in the demodulation process.
- for h ≥ 1, the spectral distribution is more flat, leading to the increase of the BW in which the signal power is concentrated and to a decreased immunity to noise.

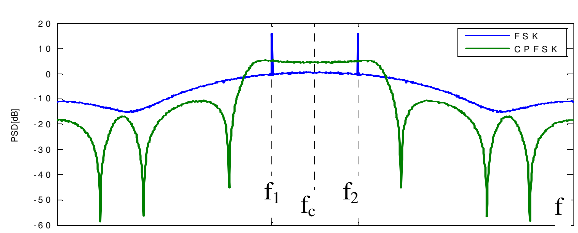


Figure 3. PSD of the FSK signals with phase-shifts

FSK with phase-shift

- the FSK with phase shift is shown in theory to exhibit a wider frequency spectrum that leads to significant amplitude variations after the modulated signal is filtered. Therefore, this type of FSK **should not be used**. Its power spectral distribution is represented with blue line in fig. 3.

Computation of the frequencies assigned to the two logical levels

- Considering (3).a and (6) we have the equation system:

$$\frac{f_2+f_1}{2}=f_c; \qquad \frac{f_2-f_1}{f_s}=h \tag{12}$$

- by choosing the f_c in the middle of the available FB and assuming h = 0,65, we get the frequencies f_1 and f_2 as solution of the system (12):

$$f_2=f_c+\frac{h\cdot f_s}{2}; \qquad f_1=f_c-\frac{h\cdot f_s}{2}; \tag{13}$$

- as an example, let's compute the f_1 and f_2 assigned in the 1200/600 bps modem for the vocal telephone channel, with the FB = [300, 3400] Hz, acc. to the ITU-T V.23 Recommendation.
- for the D=1200 bps, we choose f_c=1700 Hz and h=0,65. Using (13) we get f_2'= 2090 Hz and f_1'=1310 Hz. For a simpler implementation the two values are rounded to f_2(1200)=2100 Hz and f_1(1200)=1300 Hz, respectively. With these values the modulation index is h= 0.66, which generates about the same spectral distribution as h=0,65.
- for D=600 bps, a similar computation leads to f_2(600)=1900 Hz and f_1(600)=1500 Hz, for h=0,66.
- this would involve the implementation of a controlled oscillator that would switch between four frequencies, depending on the bit to be modulated and the bit rate to be transmitted. For a simpler implementation we impose f_1(600)=f_1(1200)=1300 Hz and imposing h=0,66 and solving the system (12), we get f_2(600)=1700 Hz and f_c(600) = 1500 Hz; so the oscillator switches between three frequencies
- Table 1 presents the frequency values assigned for the two logical levels, for the desired bit rate.

Data (Bit rate)	1 (600)	1 (1200)	0 (600)	0 (1200)
f_1 or f_2	1300 Hz	1300 Hz	1700 Hz	2100 Hz

- the standard allows for deviations of +/-10 Hz from the nominal values.
- the frequency BW of the FSK signal is computed using (11) and has the values:

$$B_{FSK}(1200)=[1100\text{ Hz};2300\text{ Hz}];\ B_{FSK}(600)=[1200\text{ Hz};1800\text{ Hz}] \tag{14}$$

- if the BW of the modulated signal is expressed by (11), the spectral efficiency, see (15), of this modulation is β_w = 1 bit/s/ Hz, indicating a rather inefficient employment of the frequency bandwidth.

$$\beta_w=\frac{D}{BW}[\frac{bps}{Hz}] \tag{15}$$

Generation of the FSK signal

- the FSK modulation can be generated both by analog and digital methods
- the analog FSK modulator are proven to be less accurate and stable and therefore they will not be discussed hereinafter

Digital generation of FSK by recursive computing of the carrier's momentary phase

- assuming a sampling frequency $f_e = p\cdot f_s$, p integer, and a discreet time $t = nT_e$, then the momentary phase of the B-FSK signal, see (9), at $t = (n+1)T_e$ could be written as:

$$\text{for } \Delta\omega_M=\pi\cdot h\cdot f_b \Rightarrow \Phi((n+1)T_e)=[\omega_c\cdot(n+1)\cdot Te+a_k\cdot\Delta\omega_M\cdot(n+1)\cdot T_e]_{mod\,2\pi}=$$
$$=[\omega_c\cdot n\cdot Te+a_k\cdot\Delta\omega_M\cdot n\cdot T_e+(\omega_c+a_k\cdot\Delta\omega_M)\cdot T_e]_{mod\,2\pi}=[\Phi(n\cdot T_e)+\Delta\Phi_p(a_k)]_{mod\,2\pi}; \tag{16}$$

- (16) shows that every sampling period the corresponding phase increment $\Delta\Phi_p(a_k)$ (17) is computed and added to the carrier phase of the previous sampling period.

$$\Delta\Phi_p(a_k)=(\omega_c\pm a_k\cdot\Delta\omega_M)\cdot T_e \tag{17}$$

- then the cosine function is applied to the resulted samples and the result is multiplied by the amplitude A to provide (8). The cosine function might be applied as such, if the hosting platform allows it, or might be computed by using numerical methods.
- the variation vs. time of the momentary phase of a 2-FSK signal is shown in figure 4

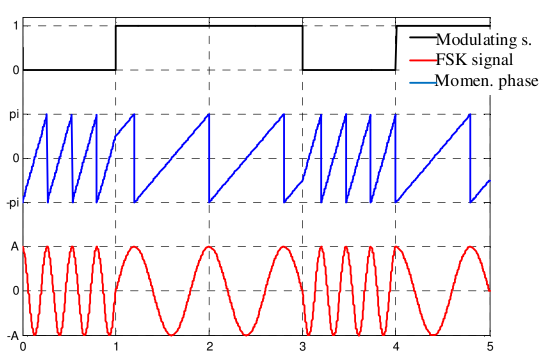


Fig. 4. Time variation of the momentary phase of 2-FSK

- other digital methods do not require the on-line computation of the cosine function, *see notes on blackboard*

Digital FSK modulator that employs the carrier generation with the Walsh functions

- this variant of digital implementation of the FSK modulation is based on the generation of the sinusoidal carrier signal using the Walsh functions. Then the frequency of this signal is controlled according to the input data and desired bit rate.
- a periodical function of period T and amplitude A, may be expanded in Walsh series:

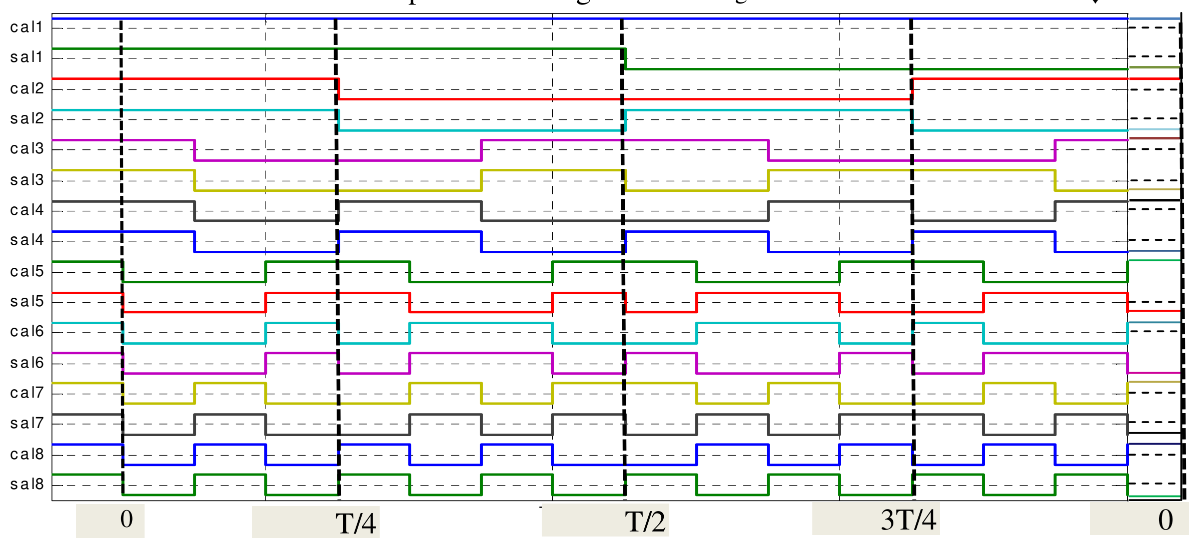
$$s(t) = A\sum_{i=1}^{\infty} s_i sal_i(2\pi t/T) + A\sum_{i=1}^{\infty} c_i cal_i(2\pi t/T) + w_0 \cdot A \cdot wal_0(2\pi t/T); \qquad (18)$$

- the coefficients s_i, c_i and w_0 are computed using:

$$s_i = \int_0^1 s(t/T)\cdot sal_i(2\pi t/T)\cdot d(t/T)\,;\; c_i = \int_0^1 s(t/T)\cdot cal_i(2\pi t/T)\cdot d(t/T)\,;\; w_0 = \int_0^1 s(t/T)\cdot wal(2\pi t/T)\cdot d(t/T)\,; \qquad (19)$$

- the first 16 Walsh functions are represented in figure 5.

Figure 5. The first 16 Walsh functions ↓



- since the sine function is odd and the cal_i functions of period T are even, the coefficients c_i, (19), will equal zero.
- the w_0 coefficient, which represents the d.c. component of the expanded signal, is also zero.
- the sine function can be represented as a sum of sal_i, functions, weighted by the s_i coefficients.
- but due to the symmetries exhibited by the sine function, the coefficients of the sal_i functions of even order would be zero. Therefore only sal_i functions of odd order would be employed. Retaining only four Walsh functions, the sine frequency is approximated by:

$$\sin\frac{2\pi t}{T} \approx s_1\cdot sal_1(2\pi t/T) + s_3\cdot sal_3(2\pi t/T) + s_5\cdot sal_5(2\pi t/T) + s_7\cdot sal_7(2\pi t/T); \qquad (20)$$

- if the amplitude of the sal_i functions is A, and that the computed values of the first four coefficients s_i are shown in (19) we get the amplitude of the resulting staircase sinusoid to be of about 0.977A

$$s_1 = 0{,}636; \quad s_3 = -0{,}265; \quad s_5 = -0{,}052; \quad s_7 = -0{,}128; \qquad (21)$$

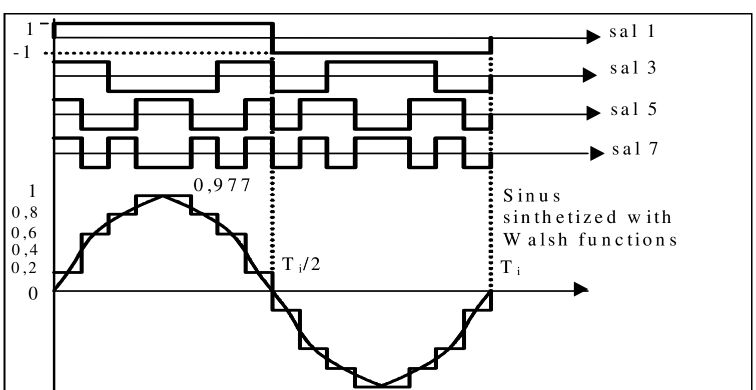


Figure 6. Sinusoid synthesis with 4 Walsh functions

- the signal generated by (20) and (21) is presented in figure 6.
- the truncation of the infinite sum (18) to the finite sum (20) leads to the occurrence of harmonic distortions, which are not very high. The additional spectral components are placed at multiples of $16f_i$ and are easy to filter. The filtering gets simpler with the increase of the number of terms of the truncated sum.
- the generation of the cosine function is similar, but employs the cal_i functions, because the cosine function is even;
– the cal_i functions are obtained by shifting the sal_i functions with T/4

Generation of the Sal_i functions

- the first four odd sal_i functions may be generated noting that the shortest level has a duration of $T_l/16$.
- considering a signal of frequency $8f_l$ and a counter with three ranges that generates signals with frequencies $4f_l$, $2f_l$ and f_l, the Walsh functions have the expressions:

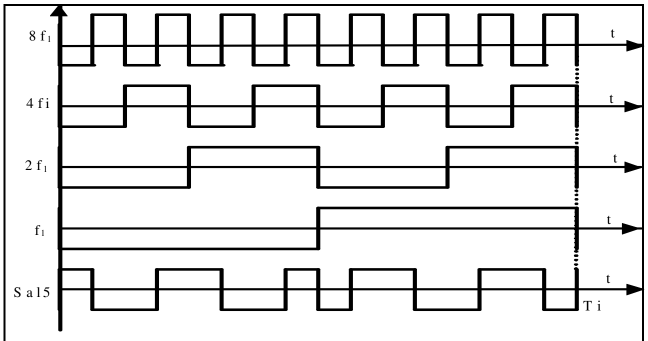


Figure 7. Synthesis of the sal_5 function

$$sal_1 = \overline{f_l}; \quad sal_3 = 4f_l \oplus 2f_l \oplus \overline{f_l}$$
$$sal_5 = 8f_l \oplus 4f_l \oplus \overline{f_l}; \qquad (22)$$
$$sal_7 = 8f_l \oplus 2f_l \oplus \overline{f_l};$$

- the synthesis of the sal_5 function is shown in figure 7.
- the diagram of the circuit that generates the sine function, using the first four odd-order sal functions is presented in figure 8.
- the weighted summation of (20) can be accomplished in two ways:

a. if the Walsh functions are generated as bipolar functions of amplitude A, then the addition is performed by means of an Op Amp, configured as an inverting adder, see fig. 8.

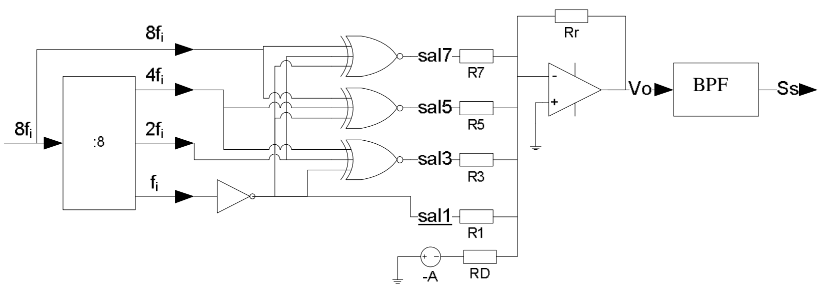


Figure 8.
Diagram of the sine generation using the synthesis with Walsh functions

The output voltage of the Op Amp is:

$$V_o = -\sum_{i=1}^{7} \frac{R_r}{R_i} sal_i; \qquad (23)$$

- identifying the coefficients between (21) and (23), we may compute the values of the R_i resistors by using (24) , provided that the value of the feedback resistor R_r is chosen.

$$R_i = \frac{R_r}{|s_i|}; \qquad (24)$$

- noting that the s_1 coefficient is positive (21), to obtain a negative value for the product of $s_1 \cdot sal_1$, we need the sal_1 function to be negative; this is accomplished by the inversion of the sal_1 function (a bipolar function).

b. if the Walsh functions are unipolar functions, generated by TTL circuits with levels of 0 V and +A V, the generated sinusoid would be negative, being centered on a d.c. component that equals -A/2;
- to obtain a sinusoid centered on a null d.c. component, the addition of a positive d.c. voltage equaling +A/2 is required.
- this is accomplished by adding a voltage – A by means of the resistor $R_d = R_r/2$; the R_d resistor, represented in figure 8 is required only if the Walsh functions are unipolar.

FSK modulator using the Walsh functions

- the FSK modulation is accomplished by changing the value of the $8f_i$ frequency, between $8f_1$ and $8f_2$, by means of a frequency divider controlled by the input data signal.
- this divider would change the division factor, by which it divides the frequency of the pilot oscillator with f_{osc} = l.c.m. ($8f_1$, $8f_2$), between n_1 and n_2, in terms of the data bit.
- the resulted sinusoidal signal would have the frequency f_1 or f_2. If the modulator of a V.23 modem is to be implemented, then f_{osc} = l.c.m. ($8f_1$, $8f_2(600)$, $8f_2(1200)$), and the dividing factor would be changed according to the bit rate control signal and to the data bit value.
- the change of the frequency division factor should be performed in such a manner that it would not affect the phase of the $8f_l$ signal; this ensures the maintenance of the phase continuity of the resulted sinusoidal signal.
- the block diagram of the FSK transmitter based on this approach is shown in figure 9.
- the transmission control block can enable/disable the modulator, depending on the RTS signal state, by enabling/disabling the controlled divider.

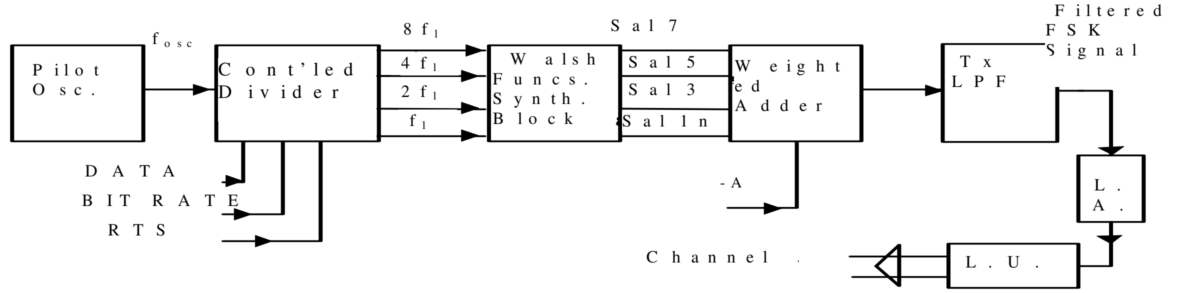


Figure 9. Block diagram of the FSK transmitter with Walsh synthesis of the carrier signal

- the generated FSK signal has a main spectrum centered around f_c, see figure 2, and similar spectra centered around the $16f_c$ and around its harmonics; this is because the sinusoid has a period equaling T_l and is generated in 16 steps.
- the values of the spectral components placed outside the desired spectrum are small, containing a negligible part of the modulated signal's power, so their filtering would not induce significant harmonic distortion; due to their position at a large "distance" from the desired spectrum, the filtering can be accomplished with a simple LP filter.
- the telegraphic distortion, see next paragraph, of the modulated signal generated by this method, is smaller than the one generated by the analog modulators, if the same receiver is employed.
- some other digital methods of generating the FSK signals are described in literature.

Filtering the FSK signals and its effects upon the modulated signals

- filtering is required to fit the spectrum of the FSK signal, an infinite one, into the channel's limited BW
- the frequency band that should be retained out of the FSK modulated signal, without affecting significantly its quality, is given by (11)
- the fulfillment of (11) is equivalent to the retaining of the components with indexes -1, 0, 1 out of the modulated signal (1) (for a 1:1 modulating data pattern);
- if $C_{-1} = -C_1$, (4) and neglecting the constant phase α, the FSK filtered signal has the expression:

$$s_{f\text{-}FSK}(t) = C_0 \cos\omega_c t + C_{-1}\cos(\omega_c - \omega_N)t + C_1\cos(\omega_c + \omega_N)t = R(t)\cos\phi_r(t);$$

$$R(t) = \sqrt{C_0^2 + 4C_1^2 \sin^2\omega_N t}\;;\; \phi_r(t) = \omega_c t + \operatorname{arctg}\left(\frac{2C_1}{C_0}\sin\omega_N t\right); \tag{25}$$

- (25) shows that the momentary phase of the filtered signal φ_r varies around the linear phase of the central frequency, with the Nyquist frequency $\omega_N/(2\pi)$, i.e. half of the fundamental frequency of the modulating signal.
- performing the derivative of the momentary phase of the modulated signal, with respect of time, we get the expression of the momentary pulsation:

$$\omega_{in}(t) = \frac{d\varphi_r(t)}{dt} = \omega_c + \frac{2 \cdot C_1 \cdot C_0 \cdot \omega_N \cdot \cos\omega_N t}{C_0^2 + 4C_1^2 \sin^2\omega_N t}; \tag{26}$$

- the fundamental of the modulating signal, the variation of the momentary frequency of the non-filtered and filtered 2-FSK signals are presented in figure 10.

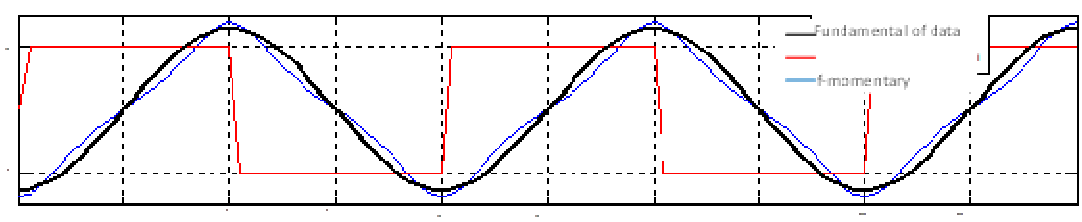


Figure 10 Momentary frequencies of 2-FSK before and after filtering and the fundamental of the modulating signal

- the momentary pulsation has a continuous variation in time and has at the middle of the bit period the value $\omega_{in} = \omega_c$; it is no longer a discreet variation as the one of the non-filtered modulated signal;
- the pulsation maximum deviation around the central carrier equals $(2C_1/C_0)\cdot\omega_N \approx 0.46\omega_{bit}$, value attained at the beginning of each bit period, $t = k\cdot T_s$.
- recalling that this expression is obtained for the 1:1 modulating data pattern, then ω_N is the fundamental of the modulating signal, regarded as a periodical rectangular signal of period $2T_s$.

- concluding, the filtered FSK signal is similar to a FM signal modulated with the fundamental of the data signal, i.e. *the deviation of the momentary frequency around the central frequency of the filtered FSK is approximately proportional to the level of the fundamental component of the data signal*
- relation (25) also shows that the filtered FSK signal is amplitude-modulated;
- the envelope also varies in the rhythm of the fundamental of the modulating signal, as shown by the expression (27) of the envelope R(t) of (25),:

$$R(t) = \sqrt{C_0^2 + 4 \cdot C_1^2 \cdot \sin^2 \omega_N t} = \sqrt{C_0^2 + 2 \cdot C_1^2 - 2 \cdot C_1^2 \cdot \cos 2\omega_N t} = \sqrt{C_0^2 + 2 \cdot C_1^2 - 2 \cdot C_1^2 \cdot \cos \omega_s t}\ ; \qquad (27)$$

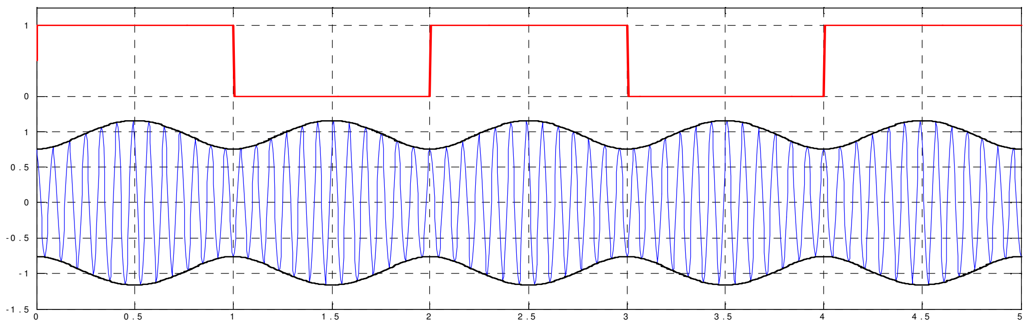


Figure 11. Filtered B-FSK's envelope variation - blue; modulating 1:1 data - red

- the post-filtering envelope has its maxima at the middle of the bit-period and minima at the margins of the bit-period.

Summarizing, the effects of the FSK signals filtering are:

- continuous variation of the momentary phase of the FSK filtered signal;
- continuous variation of the momentary frequency of the FSK filtered signal; it reaches its extreme values at the bit-period margins;
- the deviation of momentary frequency, around the central frequency, is approximately proportional to the level of the fundamental frequency of the modulating data signal;
- the occurrence of a "parasitic" amplitude modulation, proportional to the bit frequency that has minima at margins of the bit periods and maxima at their middle;

Telegraphic Distortion

- another consequence of the attenuation of some spectral components of the FSK signal is the occurrence of the telegraphic distortion;
- it consists in the variation of the durations of the demodulated data bits, around the nominal value T_s, see figure 12.

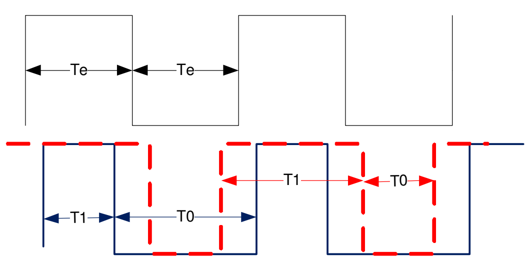


Figure 12. Schematic representation of the telegraphic distortion

- denoting by ΔT_s the maximum change of demodulated bit's duration and by BW the bandwidth of the receiver input filter, the telegraphic distortion is defined by (28).a, and its approximate value may be computed by (28).b;
- considering the bandwidth given by (10), this distortion is $\delta \approx 1\%$. The measurement of the telegraphic distortion is performed using a „1010…" data pattern.

$$\delta[\%] = \frac{2|\Delta T_s|}{T_s}; \text{a.} \qquad \delta[\%] = \frac{\pi - 2\sin\frac{BW}{2}\cdot T_s}{\frac{BW}{2}\cdot T_s - \sin\frac{BW}{2}\cdot T_s}; \text{b.} \qquad (28)$$

- the value of δ given by (28).b is an ideal one; the real value depends both of the harmonic distortions inserted by the modulator and of the demodulation method employed.
- note that the telegraphic distortion should not be confused with the jitter; it is generated by the signal processing in the modulator and demodulator, while the jitter is generated by the channel distortions and the clock synchronization system.