7. kapitola: Amplitudová modulace, demodulace
(rozšířená osnova)

Čas ke studiu: 4 hodiny

Cíl: Po prostudování této kapitoly budete mít:

- definovat princip analogové modulace
- definovat princip analogové demodulace
- popsát základní obvodové realizace modulátoru
- popsát základní obvodové realizace demodulátoru

Výklad

1. Amplitude modulation circuits (AM), demodulation

Amplitude modulation of a sine or cosine carrier results in a variation of the carrier amplitude that is proportional to the amplitude of the modulating signal. A modulating signal should produce an AM wave of the form

\[ S(t) = A_0 [1 + m(t)] \cos(2\pi f_c t) \]

\( m(t) \) - modulating signal

\( A \cos(2\pi f_c t) \) - Carrier signal (\( A \) – amplitude; \( f_c \) – frequency; \( \varphi \) – phase)

![Diagram of FET amplitude modulator](image)
A square law n-channel FET (Field Effect Transistor) will pass a drain-source current

\[ I_{ds} = K(V_g + V_p)^2 \]

where \( V_p \) is the FET's pinch-off voltage and \( V_g \) is the gate voltage.

It is evident that (superposition theorem)

\[ V_x = \frac{m[t]}{2} + A \cos(2\pi f_c t) \]

thus

\[ I_{ds} = \frac{K}{4} \left( m[t] + A \cos(2\pi f_c t) + V_p \right)^2 \]

This produces a drain voltage (output) of

\[ V_D = V_b - R_d I_{ds} \]

We know that \( \cos^2 \alpha = \frac{1 + \cos 2\alpha}{2} \). Hence the last term in expression for \( I_{ds} \) is a combination of a steady current and a fluctuation at the frequency \( -2f_c \). For simplicity we can arrange that the frequencies with which \( m(t) \) fluctuate are all \( \ll f_c \). This means that the first part of the expression consists of a steady current plus some fluctuations at frequencies well below \( f_c \). We can now use a bandpass filter, designed to only pass frequencies \( f_c \) to strip away low and high frequencies and obtain an output

\[ V_{out} = -\frac{R_d k}{4} \left[ 2m[t]A \cos(2\pi f_c t) + 2V_p A \cos(2\pi f_c t) \right] = -\frac{R_d k A V_p}{2} \left[ m[t]/V_p + 1 \right] \cos(2\pi f_c t) \]

which we can re-write in the form

\[ V_{out} = A_0' \cdot \left[ 1 + m'(t) \right] \cdot \cos(2\pi f_c t) \]

where

\[ m'(t) = m[t]/V_p; \quad A_0' = -R_d k V_p A / 2 \]

i.e. the output is a wave whose unmodulated amplitude is \( A_0' \) and is amplitude modulated by an amount, \( m'(t) \), proportional to the input modulating signal, \( m(t) \). The circuit therefore behaves as an amplitude modulator.

If we have now:

\[ m[t] = \sum_{n=1}^{N} a_m \cos(2\pi f_m t + \phi_m) \]
than

\[ m'i(t) = \sum_{n=1}^{N} a'_n \cos(2\pi f_m t + \varphi_m), \quad a'_n = a_n / V_p. \]

We easy get ( \( \sin \omega t \cdot \sin \omega t = \frac{1}{2} \cos(\omega t - \omega t) - \cos(\omega t + \omega t) \) )

\[ V_{out} = A'_{0} \cdot \left[ 1 + \sum_{n=1}^{N} a'_n \cos(2\pi f_m t + \varphi_m) \right] \cdot \cos(2\pi f_c t) = \]

\[ = A'_{0} \cdot \cos(2\pi f_c t) + \frac{A'_{0}}{2} \sum_{n=1}^{N} a'_n \cos\left[ 2\pi (f_c + f_m) t + \varphi_m \right] + \frac{A'_{0}}{2} \sum_{n=1}^{N} a'_n \cos\left[ 2\pi (f_c - f_m) t - \varphi_m \right] \]

For 100% modulation (m’ = 1.0 – the degree of modulation), the amplitude of each sideband will be just one-half of the carrier amplitude (voltage). We must keep \( 1 + m'i(t) > 1 \), always. If \( 1 + m'i(t) < 1 \) the signal is overmodulated. This signal cannot be recovered well in most detection systems.
There are various ways to measure or detect the amplitude. We'll consider one of the simplest, used by most portable radios, the *Envelope Detector*.

This is just a halfwave rectifier which charges a capacitor to a voltage ≈ to the peak voltage of the incoming AM waveform. When the input wave's amplitude increases, the capacitor voltage is increased via the rectifying diode. When the input's amplitude falls, the capacitor voltage is reduced by being discharged by a ‘bleed’ resistor, $R$. The main advantage of this form of AM *Demodulator* is that it is very simple and cheap! It contains just one diode, and one capacitor, and one resistor. That's why it is used so often. However, it does suffer from some practical problems.

All real diodes are non-linear – the current they pass varies with the applied voltage – as a result, the demodulated signal is slightly distorted. This simple type of AM demodulator isn't any good if we want the recovered waveform to be an accurate representation of the original modulating waveform (it is not Hi-Fi!!).

This circuit charges well the capacitor if the input voltage is greater than the capacitor voltage – it is the behavior of the diode. But this circuit blocks any current when the input voltage is below the capacitor voltage. The capacitor is discharged only via the resistor $R$ - *Ripple* and *Negative Peak Clipping*. The ripple effect happens because the capacitor will be discharged a small amount in between successive peaks of the input AM wave.

The illustration shows what happens in the worst possible situation where the modulating signal is a squarewave whose frequency isn't much lower than the carrier frequency.
The detector time constant (for discharging of capacitor) is $\tau = RC$. The time between successive peaks of the carrier will be $T = 1/f_c$. If we have $\tau >> T$ we have discharging current between each peak and the next almost constant $\approx V_{\text{max}}/R$. Now we can determine a change of charge: $\Delta Q \approx T V_{\text{max}}/R$. Then the change of voltage between successive peaks (Ripple) is

$$\Delta V \approx \Delta Q/C = T V_{\text{max}}/(RC) = V_{\text{max}} T/\tau = V_{\text{max}}/(\tau f_c)$$

A sudden, large reduction in the amplitude of the input AM wave means that capacitor charge isn’t being ‘topped up’ by each cycle peak. The capacitor voltage therefore falls exponentially until it reaches the new, smaller, peak value. To assess this effect, consider what happens when the AM wave’s amplitude suddenly reduces from $V_{\text{max}}$ to a much smaller value. The capacitor voltage then declines according to

$$V(t) = V_{\text{max}} \cdot \exp(-t/\tau)$$

This produces the negative peak clipping effect where any swift reductions in the AM wave’s amplitude are ‘rounded off’ and the output is distorted. Here we’ve chosen the worst possible case of squarewave modulation. In practice the modulating signal is normally restricted to a specific frequency range. This limits the maximum rate of fall of the AM wave's amplitude. We can therefore hope to avoid negative peak clipping by arranging that the detector’s time constant $\tau << t_m$ where

$$t_m = 1/f_m$$

and $f_m$ is the highest modulation frequency used in a given situation.

The above implies that we can avoid negative peak clipping by choosing a small value of $\tau$. However, to minimize ripple we want to make $\tau$ as large as possible. In practice we should therefore choose a value

$$1/f_m >> \tau >> 1/f_c$$

to minimize the signal distortions caused by these effects. This is clearly only possible if the modulation frequency $f_m << f_c$. Envelope detectors only work satisfactorily when we ensure this inequality is true.

An example of connection with transistor is in fig. below.
Simple diode modulator - principle

Diode modulation consists of a mixing network, a diode rectifier, and an LC tuned circuit, often. One diode is used as nonlinear element – it “creates” needed frequency spectrum.
**Balanced (lattice) modulator - principle**

A balanced modulator is a circuit that generates a DSB (double sideband) signal, suppressing the carrier and leaving only the sum and difference frequencies at the output. The output of a balanced modulator can be further processed by filters or phase-shifting circuitry to eliminate one of the sidebands, resulting in a SSB (single sideband) signal.

The carrier sine wave is used as a source of forward and reverse bias for the diodes.
- The carrier turns the diodes off and on at a high rate of speed.
- The diodes act like switches that connect the modulating signal at the secondary of T1 to the primary of T2.
- The carrier sine wave is considerably higher in frequency and amplitude than the modulating signal.

**Amplitude modulator with an analog multiplier**

The AD633 can be used as a linear amplitude modulator with no external components. Figure below shows the circuit. The carrier and modulation inputs to the AD633 are multiplied to produce a double-sideband signal. The carrier signal is fed forward to the AD633’s Z input where it is summed with the double-sideband signal to produce a double-sideband with carrier output.
It is evident that
\[ \sin \omega t \cdot \sin \omega_M t = \left[ \cos(\omega - \omega_M) t - \cos(\omega + \omega_M) t \right] / 2 \]

Thus all is clear, now.

Amplitudově modulovaný signál signal je definován vztahem
\[ u_{AM}(t) = U_n \left[ 1 + mf_m(t) \right] \cos(\omega_n t). \]  \hspace{1cm} (7.62)

kde \( m \) \((0 < m \leq 1)\) je modulační index AM nebo také hloubka modulace. \( U_n \) je amplituda nosné bez modulace, \( \omega_n \) je její úhlový kmítočet a \( f_m(t) \) je normovaný modulační signál. Pro případ \( f_m(t) = \cos(\Omega t) \) dostaneme časový průběh podle Obr. 7.37.

Obr. 7.37: Grafické znázornění AM v časové a kmítočtové oblasti
Text k prostudování


Další studijní texty

Otázky
Pro ověření, že jste dobře a úplně látku kapitoly zvládli, máte k dispozici několik teoretických otázek.

1. Princip amplitudové modulace.
2. Jaký je nutný minimální řád aproximace polynomu pro amplitudovou modulaci?
3. Lze linearizovat nelineární prvek v modulátoru?
4. Princip amplitudového demodulátoru.
5. Co je to hloubka modulace?
6. Základní zapojení amplitudových modulátorů.
7. Základní zapojení demodulátoru.

Odpovědi naleznete v uvedené literatuře.

Úlohy k řešení

Klék k řešení

Autokontrola
Pokud vyřešíte správně více než 2/3 problémů a otázek, můžete přejít ke studiu dalšího tématu.