# Experiment No. 4. Pre-Lab

## Linear and Non-Linear Systems, Determining Diode Parameters, Amplitude Modulation, Envelope and Square-Law Diode Detectors

**By:** Prof. Gabriel M. Rebeiz
The University of Michigan
EECS Dept.
Ann Arbor, Michigan

### Linear and Non-Linear Systems:

Any system (amplifier, filter, circuit, etc.) which has a transfer function:

$$V_o = A\,V_i \quad \text{where } A \equiv H(\omega)$$

is a *linear* system. The transfer function, A, can have a different value and phase for different frequencies as we have measured in the lab (20 Hz - 20 kHz audio amplifier, low-Q and high-Q filters, etc.). Linear systems have several basic properties which make them desirable for use in electric circuits. Some of these properties are:

1. Linearity: If the input is a signal $a_1$ and the output signal is $b_1(t)$, then if the input is $na_1(t)$, the output will be exactly $nb_1(t)$ (n = constant). (At any frequency, $V_o/V_i$ is a straight line!).

2. Superposition: If the input is a signal $a_1$ and the output signal is $b_1$, and if the input signal is $a_2$ and the output signal is $b_2$, then if the input is $a_1 + a_2$, the output will be *exactly* $b_1 + b_2$.

Talking in frequency domain, the input signal contains a frequency f, then the output will have the same *exact* frequency f, and the only difference between the input and output is an amplitude and phase change. If the input signal contains several frequencies ($f_1$, $f_2$, $f_3$), then the output signal will have *exactly* the same frequencies ($f_1$, $f_2$, $f_3$) each changed differently in its amplitude and phase, depending on the transfer function. However, *no* new frequencies are generated in linear systems.

A *non-linear* transfer function is represented by:

$$V_o = A\,V_i + B\,V_i^2 + C\,V_i^3 + \ldots$$

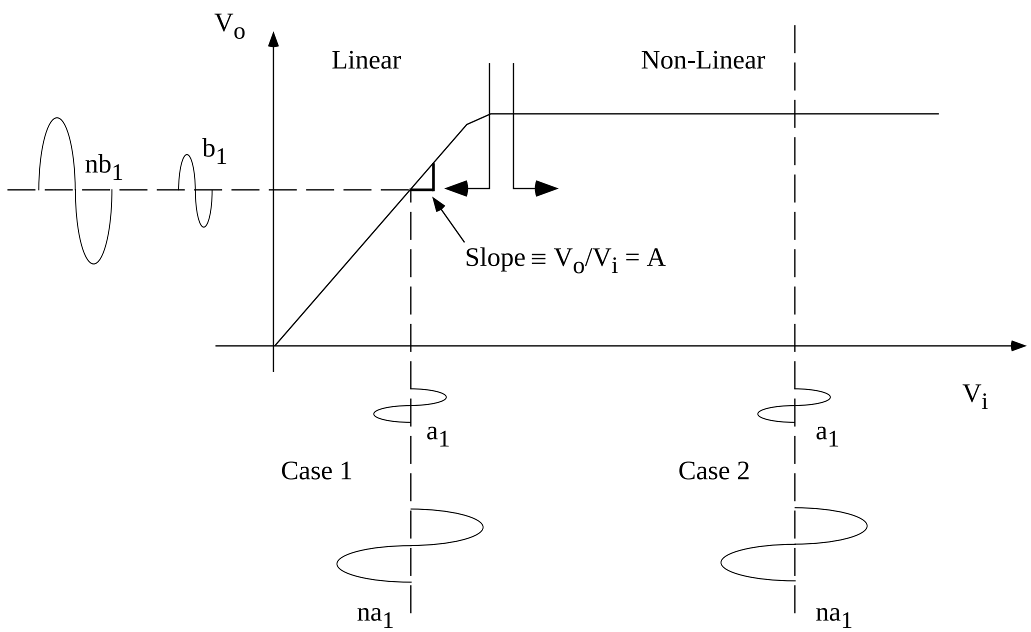
where A,B,C are dependent on frequency ($A \equiv H(\omega)$, $B \equiv B(\omega)$, $C \equiv C(\omega)$ ...)

The above mentioned properties of linearity and superposition do not apply in non-linear systems. In non-linear systems, if the input signal is $a_1(t)$, and the output signal is $b_1(t)$, then if the input is $na_1(t)$, the output is not necessarily $nb_1(t)$! Figure 1 shows an amplifier operating in the linear and non-linear regime. In the non-linear region, the output is constant *independent* of the input! As is evident, in a non-linear system, the output does not have the same form as the input in time domain. This means that if the input signal contains a frequency f, then the output signal will not only contain f but also 2f, 3f, ..., each at a different amplitude depending on B(ω), C(ω), etc. The 2f, 3f, ... components are called *harmonics* of the signal. They are generally non-desirable components in amplifiers, filters, etc. (systems which should act linearly). Remember in EECS 210 how an amplifier generates a lot of harmonics when driven into clipping (non-linear behavior)!

Real-Life Amplifiers:

In real life, linear amplifiers have some very small non-linear components (that is, the straight line is not perfectly straight). In audio amplifiers, it is important to keep the non-linear components (the harmonics) below -40 dB of the fundamental component. This translates to a total harmonic distortion of:

$$THD \leq \frac{\sqrt{\Sigma P_{harmonics}}}{\sqrt{P_{signal}}} \leq \frac{\sqrt{0.0001}}{\sqrt{1}} \leq 1\%.$$




Since most amplifiers clip asymmetrically, they generate only $V_i^3$, $V_i^5$, etc. components. The transfer function is therefore $V_o = AV_i + BV_i^3 + CV_i^5 + \ldots$. If the $V_i^3$ is the largest component, then for a THD of 1%, B = 0.01 ≡ - 40 dB (for A = 1, or B/A = 0.01 for A ≠ 1). Some people can hear this distortion level and hi-fi audio amplifiers are designed to give a THD of 0.1% and even 0.01%. For a THD of 0.01%, B = 0.0001 ≡ - 80 dB (for A = 1, or B/A = 0.0001 for A ≠ 1), which is a -80 dB third harmonic component compared to the fundamental!

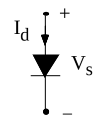
In communication systems, especially those radiating KWs of power such as radio stations, TV stations, radars, etc., it is important to keep the system very linear and to generate very low (-60 to -80 dB is very common) harmonic levels so as not to interfere with other stations at the harmonic frequencies. This means if a TV station is radiating 50 KW at 200 MHz with a -60 dB harmonic content, it will not radiate more than 50 mW(!) at 400 MHz, 600 MHz, etc. For low power applications (0.2-1W) such as hand-held analog telephones at 50 MHz and digital phones at 800-900 MHz, the harmonic content is about -30 dB.

Operation of Diodes in Small-Signal (Linear) Regime:

A diode is a very non-linear device. The 1-V curve is *exponential* and is given by $I = I_s\left(e^{V/n\,V_T} - 1\right)$. However, as discussed in class, if a small-signal, $V_S$, is applied across the diode around a DC bias condition of ($I_D$, $V_D$), then the diode equation can be written as:

$$I_d = I_D\left(\frac{V_s}{nV_T} + \frac{1}{2!}\left(\frac{V_s}{nV_T}\right)^2 + \frac{1}{3!}\left(\frac{V_s}{nV_T}\right)^3 + \ldots\right).$$

The small signal (linear component) is $i_d = I_D\frac{V_s}{nV_T}$, which is, of course, at the same frequency as $V_S$. The other components, $V_S^2$, $V_S^3$, ... are non-linear components and generate higher order harmonics. In class, we said that $V_{spk}$ < 5 mV for "small-signal" or "linear" operation. Let us now calculate the harmonic content for different values of $V_{spk}$.




$$V_s = A\cos(\omega t) \qquad (V_{spk} = A)$$

$$I_d = I_D\left(\frac{A\cos(\omega t)}{nV_T} + \frac{1}{2}\frac{A^2\cos^2(\omega t)}{n^2 V_T^{\ 2}} + \frac{1}{6}\frac{A^3\cos^3(\omega t)}{n^3 V_T^{\ 3}} + \ldots\right)$$

Fundamental Component:

$$I_{d(f)} \approx \frac{I_D A}{nV_T}\cos(\omega t) + \frac{1}{8}\frac{I_D A^3}{n^3 V_T^{\ 3}}\cos(\omega t)$$

negligible for $A << nV_T$

Second Harmonic (and DC!) Components:

$$_{d(2f)} \approx \frac{I_D A^2}{4n^2 V_T 2}\cos(2\omega t)$$

$$I_{d(DC)} = \frac{1}{4}\frac{I_D A^2}{n^2 V_T^{\ 2}}$$

Look: a DC component!

Third Harmonic Component: $I_{d(3f)} \approx \frac{I_D A^3}{24 n^3 V_T^{\ 3}}\cos(3\omega t)$

Dividing, we have:

$$\frac{\left|I_{(2f)}\right|}{\left|I_{(f)}\right|} \approx \frac{1}{4}\frac{A}{(nV_T)} \quad \text{and} \quad \frac{\left|I_{(3f)}\right|}{\left|I_{(f)}\right|} \approx \frac{1}{24}\left(\frac{A}{nV_T}\right)^2$$

and

| $nV_T = 30\,mV$ | | | |
| --- | --- | --- | --- |
| $V_{spk} = A$ (mV) | $\frac{I_{2f}}{I_f}$ (dB) | $\frac{I_{3f}}{I_f}$ (dB) | THD |
| 1 | –42 | –87 | < 1% |
| 5 | –28 | –59 | ~ 4% |
| 10 | –22 | –47 | ~ 8% |

with Vspk being the voltage across the *diode junction*.

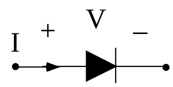
It is seen that the second and third harmonic levels are strongly dependent on the peak input voltage (A). So, by limiting $V_{spk}$ (= A) $\leq$ 5 mV, we ensure that all harmonic are less than -28 dB and that the THD is less than 4%. This may simply not be enough in many applications, and we may limit $V_{spk} \leq 1$ mV if we want the harmonics to be less than -40 dB.

Non-Linear Systems Put To Good Use:

Do not think that non-linear systems are all bad! In many industrial problems, the process is non-linear and it is best to use a non-linear system to control it. Also, most biological sensors are non-linear (your eye, ear, pain sensors immediately go non-linear and saturate if too much light, sound, or pain is applied). Finally, in communication systems, many non-linear devices are expressly used to translate frequencies from 800-900 MHz to 10-20 MHz (and vice versa). These components are called "mixers" and "multipliers" and are used in every communication system today. You will study them in detail in EECS 411 and 522.

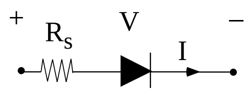
**Determining Diode Parameters:**

$$I = I_s \left( e^{V/nV_T} - 1 \right)$$

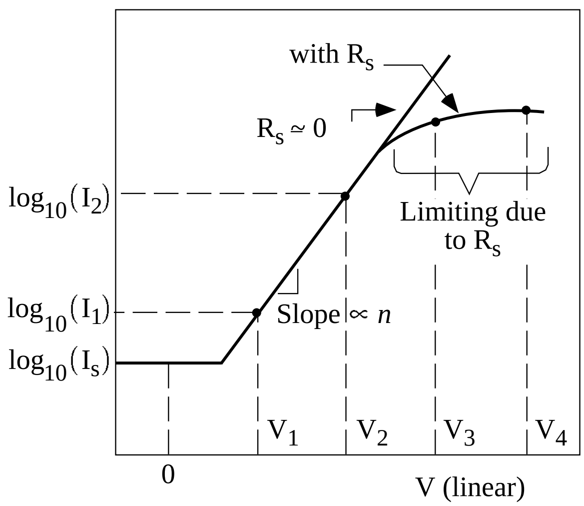



$\left(\text{with } R_s \sim 0\right)$

$$I = I_s \left( e^{\frac{V - IR_s}{nV_T}} - 1 \right)$$




$\left(\text{with } R_s\right)$




$\begin{cases} V < 0 \\ V << V_T \end{cases} \quad I = I_s$

$V >> V_T$ ,

$$I = I_s e^{V/nV_T}$$

$$\ell n \frac{I}{I_s} = \frac{V}{nV_T}$$

$$\boxed{V = nV_T \left( \ell n I - \ell n I_s \right)}$$

$\left(\text{Do a } \ell n(I) \text{ to } \log_{10}(I) \textit{ change}\right)$

Slope of the curve $\propto \underline{\underline{n}} V_T$

Get $n \Rightarrow V_2 - V_1 = 2.3\, nV_T \left[ \log_{10}(I_2) - \log_{10}(I_1) \right]$

At large currents, voltage drop is series resistance limited.

$$\frac{I}{I_s} = \left( e^{\frac{V - IR_s}{nV_T}} \right) \quad \text{and} \quad \ell n \frac{I}{I_s} = \frac{V - IR_s}{nV_T}$$

$$\Rightarrow \quad V = \left( nV_T \ \ell n \frac{I}{I_s} \right) + IR_s$$

$$V_4 - V_3 = nV_T \left[ \ell n (I_4) - \ell n (I_3) \right] + (I_4 - I_3)\, R_s$$

if $I_4 \approx I_s$ then $\dfrac{V_4 - V_3}{I_4 - I_3} \simeq R_s$

$\Rightarrow$ The "flatter" the curve the better the approximation. Good at large currents.

# Amplitude Modulation, Envelope and Square-Law Diode Detectors

## 1.0 Amplitude Modulation:

An amplitude modulated (AM) wave is given by:

$$V(t) = \{A + a(t)\}\sin(\omega_c t)$$

where a(t) ≡ audio signal (commonly called baseband signal, or modulation signal)
$\omega_C$ ≡ carrier frequency
A ≡ Amplitude of carrier frequency for no modulation signal

For the AM radio broadcast, the carrier frequency is between 550 KHz and 1550 KHz, and the bandwidth of the audio signal is 5 KHz. The audio signal is composed of a range of frequencies from 20 Hz to 20 KHz and each frequency is associated with a certain amplitude, resulting in the audio spectrum for a specific instrument or human voice. Most of the power of the human voice is below 5 KHz, so AM is excellent for talk-shows and not for music.

Let us assume that the audio signal (or baseband signal, or modulation signal) is composed of a single frequency, $\omega_1$ (for example, $\omega_1$ = 1 KHz). The AM signal is then:

$$\begin{aligned} V(t) &= \{A + a_1 \sin(\omega_1 t)\}\sin(\omega_c t) \\ &= A\sin(\omega_c t) + a_1 \sin(\omega_1 t)\sin(\omega_c t) \\ &= A\sin(\omega_c)t - \frac{a_1}{2}\cos(\omega_c + \omega_1)t + \frac{a_1}{2}\cos(\omega_c - \omega_1)t \end{aligned}$$

We see that the AM signal is composed of three frequencies: the carrier frequency ($\omega_C$) with amplitude A, the upper sideband frequency ($\omega_C+\omega_1$) and the lower sideband frequency ($\omega_C-\omega_1$), each with amplitude $a_1/2$. An AM signal with $a_1$=0.9A is shown in Fig. 1. The peak of the signal is 1.9A and the minimum is 0.1A. The frequency spectrum of the signal is also shown for $f_C$= 1 MHz and $\omega_1$= 5 KHz. Notice that the bandwidth of an AM signal (centered around the carrier) is twice the bandwidth of the baseband signal and in this case is 10 KHz.

An often used term is the modulation index, m, defined by m = $a_1$/A. Using this definition, the AM signal can be written as:

$$\begin{aligned} V(t) &= A(1 + m\sin(\omega_1 t))\sin(\omega_c t) \\ &= A\left[\sin(\omega_c t) - \frac{m}{2}\cos(\omega_c + \omega_1)t + \frac{m}{2}\cos(\omega_c - \omega_1)t\right] \end{aligned}$$

The total power contained in an AM signal is:

$$P_s = \left(\frac{A^2}{2R_e\{Z\}}\right)(1 + m^2/2)$$

where Z is the impedance seen by the AM signal, and $P_C$ is the power in the carrier frequency ($P_C=A^2/2R_e\{Z\}$). The sidebands contains only $m^2/2$ of the power relative to the carrier signal power. For m=1 (100% modulation), $P_S=P_C(1+1/2)$ and the sidebands contain half of the carrier power, and one third of the total AM signal power. Amplitude modulation is therefore not an efficient method for transmitting information, but it is so easy to generate and to detect that it is the predominant modulation technique in low-cost low-bandwidth systems (in other words, no one uses AM modulation for satellite and wireless communications). Most AM broadcast stations maintain a modulation index, m, between 85 and 95% as a compromise between spectral efficiency (power in baseband vs. power in carrier) and the risk of drifting into overmodulation and therefore severe distortion (m > 100%).

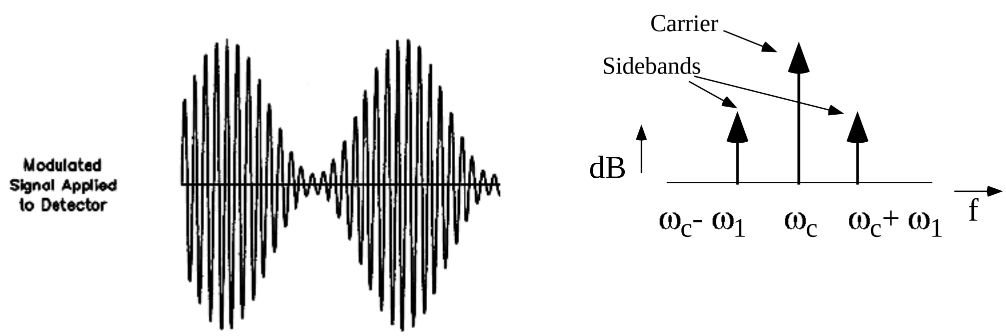



**Figure 1**: An amplitude modulated (AM) signal in time and frequency domain.

AM modulation is not limited to radio broadcasts and to audio signals. Actually, when you use a TV remote control, or an infrared transmitter on a calculator, or an infrared data link between your computer and a hard disk, you are actually AM modulating the infrared diode with a digital (on, off) modulation. In this case, the carrier is at an extremely high frequency (~300 THz for a red diode) and the baseband signal is a digital signal with a frequency of 100's of Hertz for the case of the TV remote control to several MHz for the case of a hard disk data link. We will build an AM infrared link in Experiment #5.

## 2.0 A Square-Law Detector:

If an AM signal (or any type of signal) passes by a linear system with $v_o = Av_i$ (A=constant), then the output is a replica of the input (in time and frequency domain). The output differs from the input only by its amplitude and phase but does not contain any new frequency components. The amplifiers are called linear amplifiers (such as audio amplifiers, communication amplifiers, etc....) and generate very little total harmonic distortion (i.e., no harmonics, no new frequency components). This is the essential characteristic of linear systems which I tried hard to get across in EECS 210. Therefore, linear systems cannot detect modulation of any type!

If one needs to detect a signal, a frequency translation must occur. Basically, the information in the sidebands at ($\omega_c + \omega_1$) and at ($\omega_c - \omega_1$) must be translated to $\omega_1$! For example, the information in the 1005 KHz and 995 KHz sidebands of Figure 1 must be translated into 5 KHz so as to be heard as an audio signal. In order to do this, the detector must have a non-linear transfer function. The simplest non-linear transfer function is the square-law detector. Let us assume that a detector exhibits a response given by:

$V_o = kV_i^2$ k = constant, and $V_i(t) = A\left(1 + m\sin(\omega_1 t)\sin(\omega_c t)\right)$ (the AM signal)

The output voltage is:

$$
\begin{aligned}
V_o(t) = {} & \frac{kA^2}{2} + \frac{kA^2 m^2}{4} && \text{DC Component} \\
& + kA^2 m \sin(\omega_1 t) && \text{Audio Component} \\
& + \frac{kA^2 m^2}{4} \sin(2\omega_1 t) && \text{Distorted Audio Component} \\
& + \frac{kA^2}{2} \cos(2\omega_c t)\left[1 + m\sin(\omega_1 t)\right]^2 && \text{Components at } 2\omega_c \text{ with respective sidebands}
\end{aligned}
$$

Notice that we have frequency translation! An input frequency spectrum of $\omega_C$, $\omega_C+\omega_1$ and $\omega_C-\omega_1$ (the AM signal) results in an output frequency spectrum containing a DC component, a baseband (or modulation or audio) component at at $\omega_1$, a component at $2\omega_1$ (called distortion), and frequency components $2\omega_C$ with respective sidebands. If you remember EECS 210, we called these components harmonics and intermodulation products and they occurred whenever an amplifier was driven into the non-linear region (clipping). Well, here we are using the intermodulation components of a square-law detector to our advantage so as to translate the sidebands at $\omega_C + \omega_1$ to an audio frequency of $\omega_1$.

### 3.0 AM Signal Detection (or Demodulation):

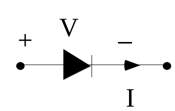
The question is then: which component do we use to get a good square-law response? An excellent and very inexpensive electronic component is the pn junction diode (or the Schottky diode for high frequency applications, f > 100 MHz). The diode IV relationship is given by:

$$I = I_s\left(e^{\alpha V} - 1\right), \quad \alpha = {}^{1}\!/_{nV_T}$$

with Is= Diode Saturation Current, $V_T$= 26 mV and n=1.2–1.8 for silicon diodes and germanium diodes.

If we expand the exponential equation into its Taylor series, we get:

$$
\begin{aligned}
I &= I_s\left(1 + \alpha V + \frac{\alpha^2 V^2}{2} + \ldots - 1\right) \\
&= \alpha I_s V + \frac{\alpha^2 I_s V^2}{2} + \ldots
\end{aligned}
$$




and we can put it in the form:

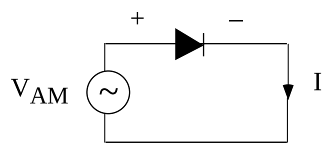
$$I = k_1 V + k_2 V^2 + \ldots \quad \text{with} \quad k_1 = \alpha I_s \quad \text{and} \quad k_2 = {}^{\alpha^2 I_s}\!/_{2}$$

Now, let us assume that an AM voltage, $V_i(t) = V_{AM}(t) = A\left(1 + m\sin(\omega_1 t)\right)\sin(\omega_c t)$, is impressed across the diode, it does not take a genius to recognize that the V2 component in the diode current equation will demodulate the AM signal! The resulting current in the diode is:

$$I = k_1 V_{AM}(t) + k_2 V_{AM}^2(t) + \ldots$$

Expanding the sine/cosine terms and grouping them together, we get the diode current I to be:

$$
\begin{aligned}
I = {} & \frac{k_2 A^2}{2} + k_2 A^2 \frac{m^2}{4} && \text{DC Component} \\
& + k_2 A^2 m \sin(\omega_1 t) && \text{Audio Component} \\
& + k_2 \frac{A^2 m^2}{4} \sin(2\omega_1 t) && \text{Distorted Audio Component} \\
& + k_1 A\left(1 + m\sin(\omega_1 t)\right)\sin(\omega_c t) && \text{Components at } \omega_C \text{ and sidebands} \\
& + \frac{k_2 A^2}{2}\left(1 + m\sin(\omega_1 t)\right)^2 \cos(2\omega_c t) && \text{Components at } 2\omega_C \text{ and sidebands}
\end{aligned}
$$




Notice that the diode current has a DC component and therefore the diode self-biases at $\left(k_2 A^2/2\right)\left(1 + m^2/2\right)$ (Amperes). This is the result of the non-linear (demodulation) process.

The diode current is then passed by a parallel RC network (a low-pass filter). This network presents an resistance R at low frequencies (ω<<1/RC) and a short circuit a high frequencies (ω >>1/RC).

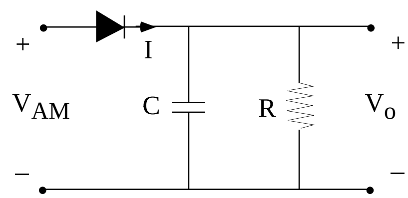



**Figure 2**: AM demodulation circuit (envelope detector).

The resulting output voltage is therefore:

$$V_o = \left(\frac{k_2 A^2}{2} + \frac{k_2 A^2 m^2}{4}\right)[R||C]_{\omega=0} \quad \leftarrow = R$$

$$+\ k_2\, A^2\, m\ sin(\omega_1 t)\ [R||C]_{\omega_1} \quad \leftarrow \sim R$$

$$+\ k_2 \frac{A^2 m^2}{4} sin(2\omega_1 t)\ [R||C]_{2\omega_1} \quad \leftarrow \sim R$$

$$+ k_1 A\left(1 + m\, sin(\omega_c t)\right) sin(\omega_c t)\ [R||C]_{\omega_c} \quad \leftarrow \sim 0! \quad \text{(negligible component in } V_o\text{)}$$

$$+ (\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ )\, cos(2\omega_c t)\,[R||C]_{2\omega_c} \quad \leftarrow \sim 0! \quad \text{(negligible component in } V_o\text{)}$$

Notice that the output voltage has a DC component, a baseband components at $\omega_1$ and distortion components $2\omega_1$. The DC component can be removed with a simple series capacitor and voila! we have the baseband (audio) signal. The AM demodulation circuit is commonly called an "envelope" detector since it follows the envelope of the AM signal.

The distortion signal at $2\omega_1$ is dependent on $m^2/4$, and the audio signal is dependent on m. For high fidelity AM, m should be kept around 0.5 which results in a distortion component of around -20 dB with respect to the audio component. In this case, the sidebands contain only 12% of the power of the carrier which is not an efficient use of the radio power spectrum. Therefore, it is common to use m ≈ 0.9 for AM broadcast of human voice which results in a distortion component around -13 dB relative to the signal.

For audio links using AM modulated infrared systems, spectral power is of no concern since the distances are generally short (for ex: TV room, conference room). In this case, m is kept around 0.2-0.3 for low distortion response in analog systems.

Another Point of View:

A simple way to explain the AM detector circuit which is common in radio amateur circles is presented below: Consider the circuit of Figure 2 with an impressed AM voltage of $V_i(t) = A(1+m\sin(\omega_1 t))\sin(\omega_c t)$. The capacitor is chosen to be a short-circuit at the carrier frequency and therefore $v_i(t)$ is imposed across the diode. The diode rectifies the signal, much like a half-wave bridge rectifier, and the capacitor charges to the peak of the signal. The resistor is chosen so as to discharge the capacitor slowly, and therefore the output waveform follows the envelope of the AM modulated signal. The output waveform has a DC component, and a series capacitor is used before the baseband (audio) amplifier to block this DC voltage. This is illustrated in Figure 3 (from the Radio Amateur's Handbook).

AM detectors are generally driven from low impedance sources since the non-linear "action" is due to the diode current. A low source impedance ensures a large current in the diode for a given source voltage. This is the reason why the AM detector is placed after an op-amp (low output impedance), after the emitter of a transistor (very low impedance) or after a transformer which ensures a low impedance drive.

Problems with AM detection:

The simplicity of AM detection is also its downfall. Anything which exhibits a non-linear response can demodulate an AM signal! In an AM receiver, one must be careful not to drive any of the pre-amplifiers into saturation (clipping), otherwise, "free and unwanted" AM demodulation occurs. The non-linear response is not limited to electronic circuits. Speakers can demodulate AM signals and I have heard many a radio amateur voice popping out on my speakers (speaker coils are non-linear)! One interesting true story is about this person who had two metals in his tooth (gold and silver). The junction between these metals with the conducting alkaline fluid (saliva) in his mouth created a non-linear diode and therefore demodulated AM stations. Since this happened in his mouth, the demodulation process did not need to be efficient, and he could hear music and talk shows whenever he drove by an AM station!

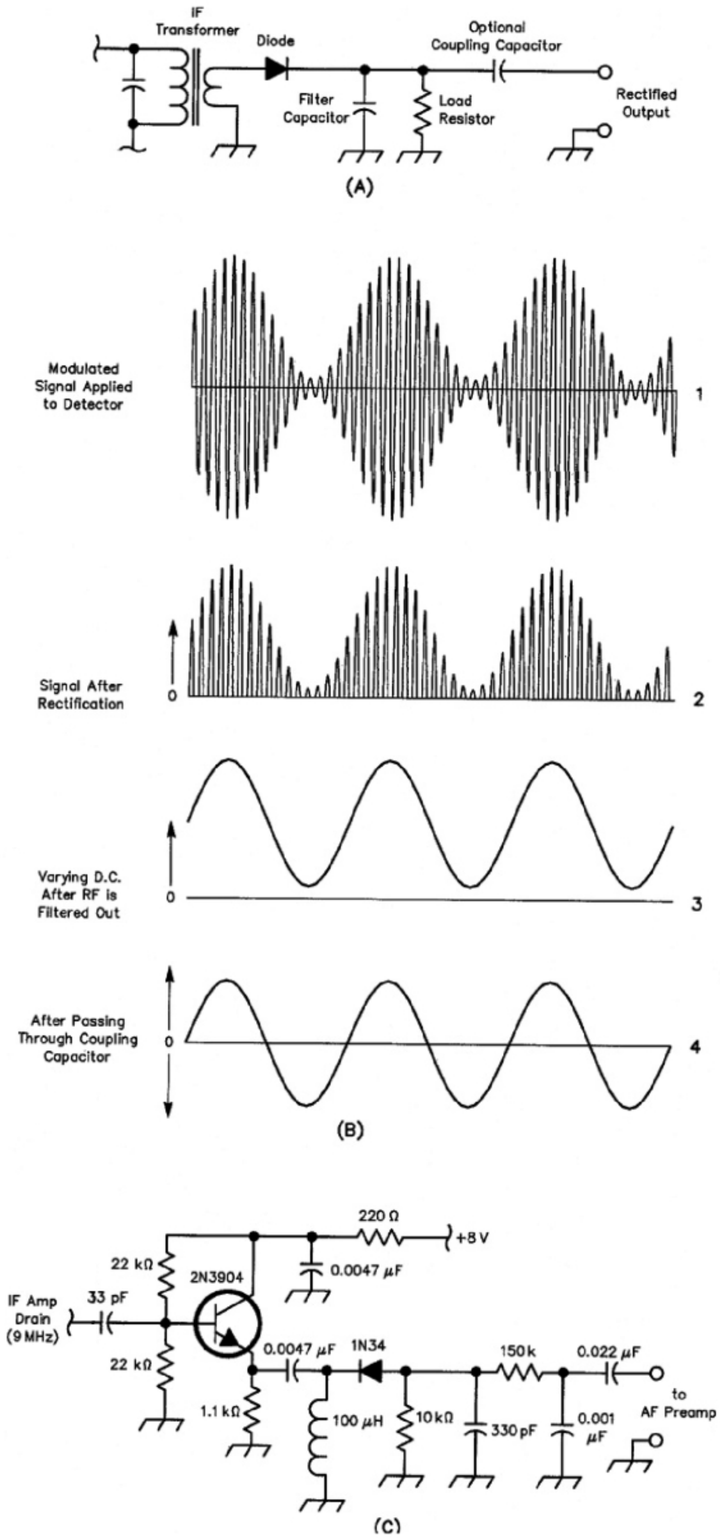



**Fig. 3:** Radio's simplest demodulator, the diode rectifier (A), demodulates an AM signal by multiplying its carrier and sidebands to produce frequency sums and differences, two of which sum into a replica of the original modulation (B). Modern receivers often use an emitter follower to provide low-impedance drive for their diode detectors (C).