FM FROM THE BEGINNING

John Linsley Hood continues his short series on FM transmission and reception with a look at the requirements of FM receiver design.

Last month I looked at the way in which the left and right channels are encoded before transmission as a stereo FM signal and at the performance limitations this process involves. In this article I will be looking at the design of VHF/FM receivers and the influence the receiver circuitry has on the quality of the final audio signal.

Figure 1 shows a typical FM receiver circuit in block diagram form. The RF, mixer and oscillator stages (the 'front end') are usually grouped on a small PCB mounted in an individually screened metal box, and this tuner unit is often made by some outside specialist manufacturer who supplies several different hi-fi companies.

Tuning System

In the earlier days of FM, such head units were usually tuned by two or three-gang air-spaced variable capacitors. Nowadays varicap tuning diodes are almost always employed, except where the unit combines AM and FM in which case a variable capacitor may still be employed. This is because the large values of tuning capacitance required for use on MW and LW are difficult to obtain with varicaps.

Varicap diodes are basically reverse-biased silicon diodes having a fairly large, heavily doped, PN junction area. This provides a useful junction capacitance which can be varied by an external control voltage. Apart from cheapness, the advantages of varicap diodes are that they allow a greater number of tuned stages, they allow the tuning capacitors to be placed nearer to the coils, and they allow remote (possibly computer controlled) tuning.

The snags are that varicap diodes have a greater RF loss factor than air spaced capacitors and this spoils the Q of the coils (though this probably doesn't matter very much at VHF where pre-mixer selectivity isn't very important). They also suffer from a temperature drift of capcitance — even for a constant tuning voltage — and the tuned frequency may be modulated by the RF voltage appearing across the tuned circuit if the signal is large enough. Twin varicap units of the type shown in Fig. 2 cost a bit more but are much better in this respect, and are used in better quality units.

RF Devices

In inexpensive tuner heads, the amplifying devices are usually bipolar junction transistors. These are still the best choice at frequencies over 1GHz or so, although gallium arsenide MOSFETs are challenging this position.

The main snag with bipolar devices is that they do not provide a good match with the high impedances of the tuned circuits, and this loss of power gain tends to degrade the noise performance of the stage. They are also very non-linear, which makes the head unit more susceptible to cross-modulation in the presence of strong input signals.

The best device in terms of noise figure and linearity is the junction FET, but these also suffer from too high a value of internal (drain/gate) feedback capacitance, and

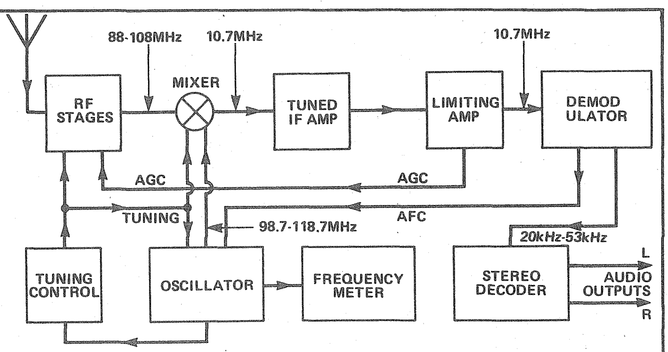


Fig. 1 Block diagram of a typical modern FM tuner.

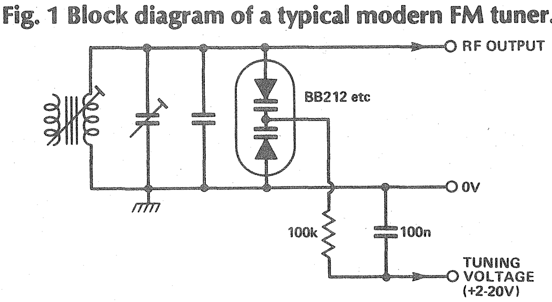


Fig. 2 Tuned stage using a twin varicap diode.

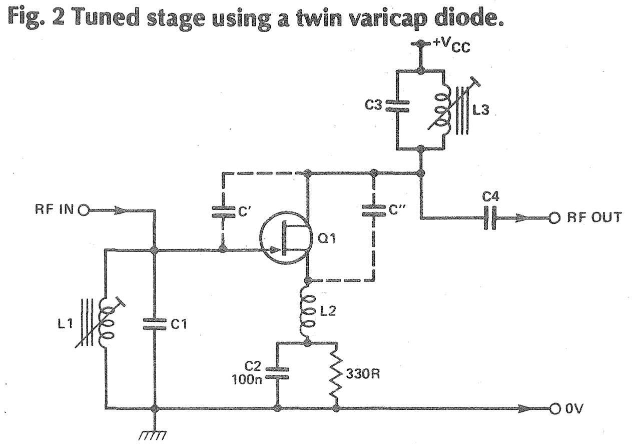


Fig. 3 The use of a small inductor in the supply lead to neutralise an FET stage.

if this is not neutralised it will lead to poor tuned circuit performance and RF instability.

A very neat way of neutralising an RF FET stage is shown in Fig. 3. In this, a small inductance is formed in the PCB track leading to the source of the FET, and if the inductance is chosen so that

$$C'' \times L2 = C' \times L1$$

(where C′ is the internal feedback capacitance, and C″ is the drain source capacitance) then the effect of the internal capacitative feedback will be neutralised and the stage will be stable.

In typical junction FETs, the value of C″ is about 10x C′, so the inductance of L2 will not need to be very big for amplifiers working in the 100MHz range.

Although they are not the quietest or the most linear of the available transistor types, dual gate MOSFETs are the easiest devices to use because they have very low internal capacitance between the drain and gate 1, permitting stable operation at RF. They also have a second (screening) gate which can be used for automatic gain control (AGC) purposes.

For mixer use, the second gate on a dual gate MOSFET provides a useful point for local oscillator injection, as shown in the circuit layout of Fig. 4a. More up-market tuners might use a double balanced mixer circuit based on a pair of junction FETs, as shown in Fig. 4b.

The main concern at this stage is freedom from cross modulation by strong unwanted adjacent channel signals — an important feature for those who live near to transmitters — and good sensitivity and signal-to-noise ratio, which is important for those who live further away from the programme source.

Having two or three tuned circuits between the RF amplifier and the mixer input also helps to reduce the problems of cross-modulation. A good quality tuner head based on dual gate MOSFETs might have the circuit layout shown in Fig. 5.

Bipolar transistors make quite good local oscillators, though it is good practice to buffer their outputs with a pair of junction FETs — one to the mixer, the other in the line to the frequency meter.

In presenting these comments on the design of tuner front ends I am not aiming to encourage DIY enthusiasts to design and build their own. Although I do not think there is any particular magic involved in the design of such units, the layout is critical if complete stability is to be achieved. For this reason it is usually better to purchase a commercial unit ready-built. These notes should help you choose a good one, particularly if you can examine the circuit diagram of a tuner unit before purchasing.

Selectivity

FM tuners, like almost all other modern radio receivers, employ the superhet principle. The incoming signals are converted by a frequency changer stage down to a fixed intermediate frequency (IF) at which the bulk of the RF amplification is obtained, along with all of the adjacent channel selectivity.

The convention for VHF/FM tuners is that the IF is 10.7MHz, and the most common method of obtaining the required selectivity is by way of ceramic surface acoustic wave (SAW) filters.

There isn't room here to describe in detail how these devices work (see ETI November 1985 — Ed.) but basically they consist of two transducer elements deposited on the surface of a piezo-electric material. Applying an input signal to one of the transducers causes a physical ripple to be launched across the surface of the chip where it is received by the second transducer and converted back into electrical energy. By careful design

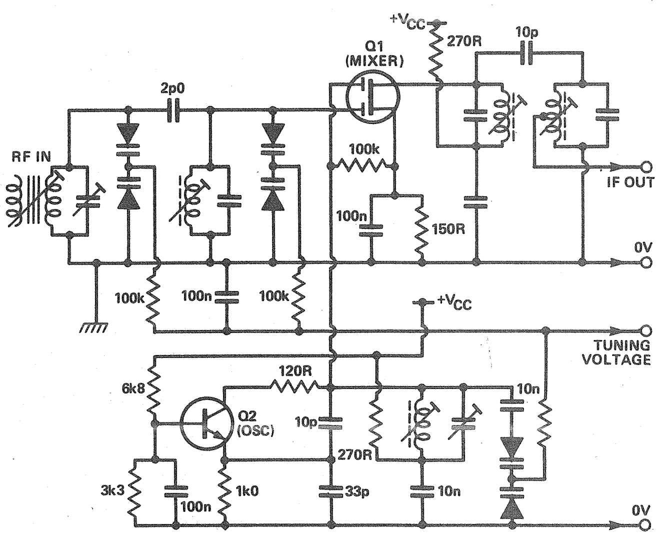
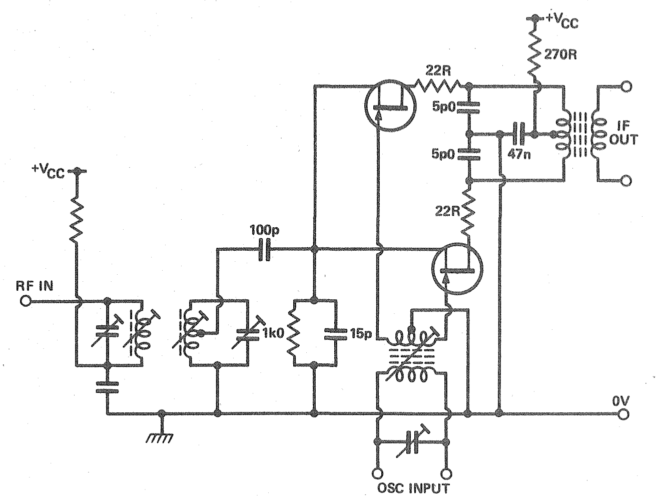


Fig. 4a (above) an FM mixer stage using a dual-gate MOSFET, and Fig. 4b (Below) a balanced mixer system using junction FETs.



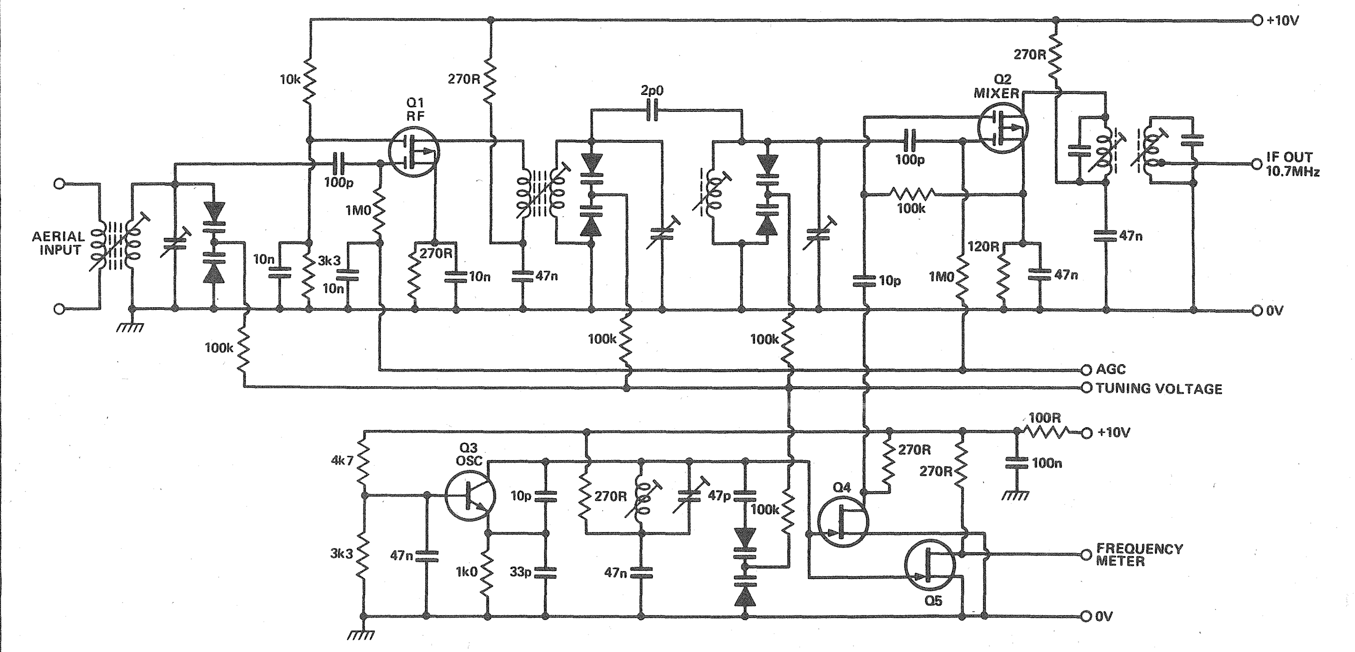
of the physical shape of the transducer elements, it is possible to define frequencies at which the surface waves will combine constructively and others at which they will cancel. In this way a near-ideal filter characteristic can be produced. A typical SAW filter pass-band response is shown in Fig. 6.

An important advantage of SAW filters is that they allow the relative phase shift of the signals to be determined independently of the RF pass-band shape. This relative phase shift is important for low distortion, and is usually called group delay.

The snag with SAW filters is that they invariably introduced a substantial loss of signal strength, typically 10x, and they must be driven and terminated with a low impedance, usually 300 ohms. Often the tuner head will have an output impedance of the right order to drive a filter, but either an RF IC or a transistor will be needed between stages to make up for the signal loss.

A typical circuit for an input IF stage is shown in Fig. 7. A high degree of selectivity is essential in order to avoid multiple signals causing cross-modulation in the limiting IF stage. For this reason, two pairs of SAW filters are often used in series. However, too much selectivity will spoil the stereo image separation in the recovered audio signal, so a degree of compromise is required.

Fig. 5 Circuit diagram of a typical high quality receiver 'front end'.



The Limiting IF Stage

In the FM transmission system, the amplitude modulation characteristics of the signal are not of interest — indeed, the idea is to get rid of them, to suppress impulse type noise. One of the benefits of this is that the final IF stage can be designed to have a high gain so that it clips the signal to a fixed and constant output level.

In contemporary practice this is always done by a specialised type of IC which also contains the FM demodulator, and probably a range of other useful bits and pieces of circuitry as well.

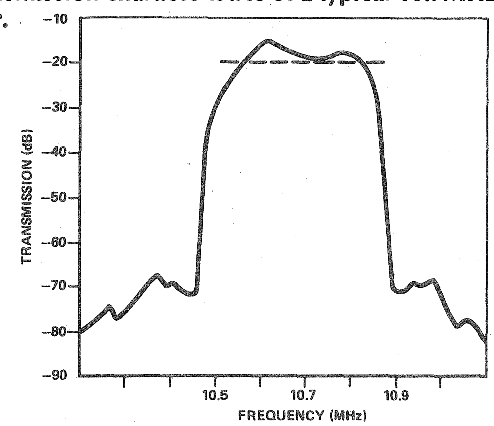
In the earlier ICs of this type designers took advantage of the ease with which bipolar transistor amplifier stages could be incorporated. In the Texas SN76660N, for example, no less than six push-pull gain stages were used before the demodulator, as shown in Fig. 8. Since the collector-emitter voltage on each of the amplifying stages was held to the same value as the base emitter voltage of the following stage (about 0.6V), this caused the amplitude of the RF signal to be limited to a peak level of the same value.

The penalty is that the gain from each stage is low, and a lot of stages will be needed if a clipping level input sensitivity of 20uV or so is to be achieved. In consequence, the cumulative transistor noise is quite substantial.

Later designs, such as the Motorola MC1351 or the TAA930, interposed an emitter follower between the succeeding stages, as shown in Fig. 9. This raised the load impedance for each amplifying stage, increasing the individual stage gain and allowing the number of amplifying stages to be reduced to three for the same input sensitivity. The result was a substantial improvement in the overall signal-to noise ratio.

The latest step in the development process came with the introduction of the RCA CA3089E/3189E ICs. These use a fully symmetrical push-pull amplifier system (as used in the TI SN76660N IC but dropped in the 1351 and its derivatives), an input series-connected transistor cascode arrangement which provides the best possible input stage gain and noise figure, emitter-follower load buffering in each stage for high gain, and a pair of back-to-back diodes at the output to give a clean, symmétrical, limiting chracteristic. This is a rather better way of limiting than simply letting the amplifying stages run up against their supply voltage limits.

Fig. 6 Transmission characteristics of a typical 10.7MHz SAW filter.



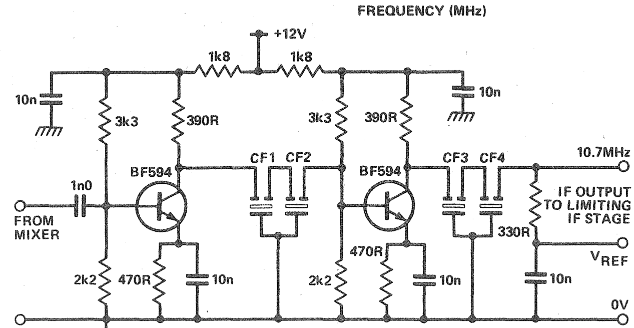


Fig. 7 An FM IF stage using ceramic filters.

The 3189E is very similar in circuit layout to the 3089E, but a few small detailed design improvements have been included, mainly aimed at reducing still further the background noise level. This IC (and copies of it made by other manufacturers), is deservedly the most popular device for use in FM limiting amplifier stages. A slightly simplified circuit diagram of the limiting amplifier chain is shown in Fig. 10.

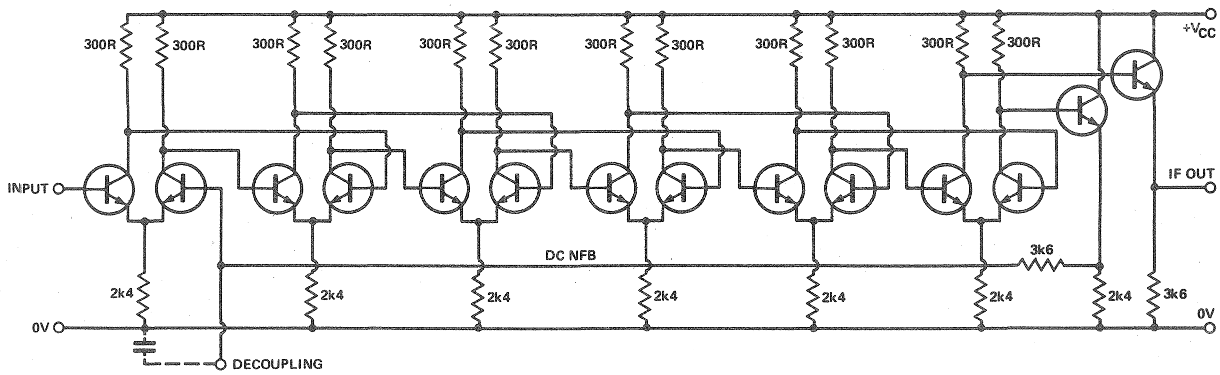


Fig. 8 The limiting IF amplifier circuit used in the Texas SN76660N.

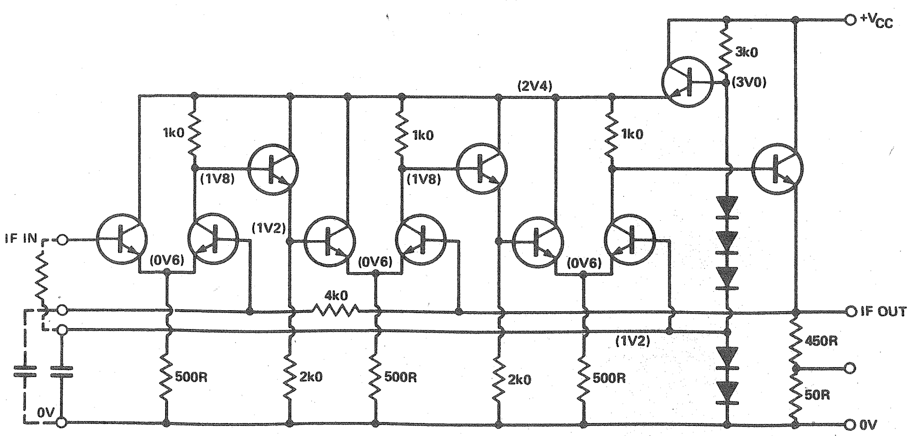


Fig. 9 The limiting IF amplifier circuit used in the MC1351 and TAA930 ICs.

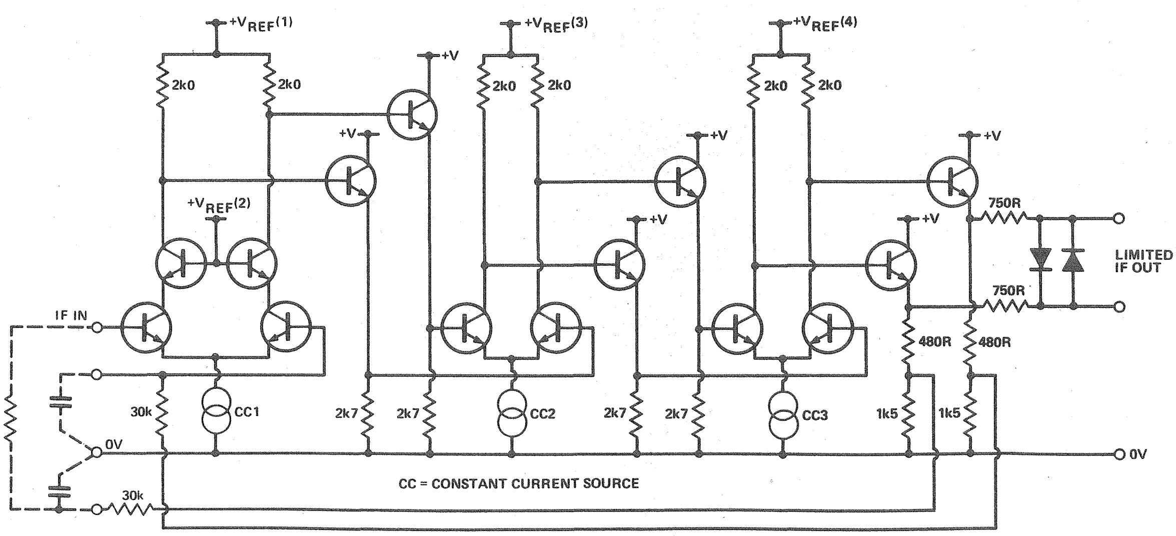


Fig. 10 The limiting IF amplifier circuit used in the RCA CA3089/3189 (simplified for clarity).

The Demodulator Stage

This is the key stage in the FM tuner. It determines the freedom from distortion of the recovered audio signal, the rejection of intruding AM-type noise (motor car ignition interference and the like) and the capture ratio of the receiver.

This last term refers to the ability of the receiver to reject an interfering FM signal that it is less strong than the wanted one, even if it is on the same frequency. This rejection is expressed in dB: a value of 1dB is excellent, while 3dB or greater would nowadays be thought to be a very poor rejection ratio indeed.

The earliest of the specifically-designed FM demodulator circuits was the 'Round-Travis' circuit shown in Fig. 11a. This consists basically of a pair of tuned circuits with associated diode detectors, one tuned above the centre frequency and arranged to give a positive-going output while the other is tuned below the centre frequency and has a negative-going output. When these two signals are combined, the result is as shown in Fig. 11b.

This layout gives no AM rejection except for signals occuring precisely at F_o, and its linearity depends entirely on the shapes of the response curves of L2C1/

L3C2 and the skill with which one tunes these two circuits.

The early circuits were gradually replaced with demodulators of the phase detector type. These fall under three headings, the Foster-Seeley, the Ratio Detector and the Gate Coincidence Demodulator type. All rely on the fact that if an RF signal is applied to a tuned circuit at a frequency which is below the resonant frequency of that circuit (F_o), the tuned circuit presents an inductive load and the phase of the voltage leads that of the current. On the other hand, if the RF signal has a frequency above F_o, the tuned circuit looks like a capacitance and the phase of the voltage lags behind that of the current.

In the case of the ratio detector layout shown in Fig. 12a (and the Foster-Seeley circuit operates in a very similar manner) a small third, untuned, winding on the RF transformer is arranged to inject a signal into the centre tap of the secondary.

Since the two rectifier diodes are connected in opposition, there will normally be no AF output at this centre tap point at any frequency. However, if the input RF frequency is either above or below F_o, the signal from L3 will add to that in the upper half of L2 and subtract from that in the lower, or vice-versa, as the phase of the voltage across L2 changes. The result will be an output response of the form shown in Fig. 12b.

This was the most popular type of FM demodulator circuit until about fifteen years ago, since when it has been almost entirely replaced by IC-based systems using gate coincidence techniques.

Gate Coincidence Demodulators

Gate coincidence demodulators use the transistor layout shown in Fig. 13. The current from the constant current source CC1 is shared between the transistors Q1 and Q2, then divided again between transistors Q3-Q6 and finally recombined in the load resistors R1 and R2.

Obviously, if all the transistors are indentical and V_{in} (1) is the same as V_{ref} (1) and V_{in} (2) is the same as V_{ref} (2), the output current through R1 will be the same as that through R2. Moreover, because of the cross-connection of Q3-Q5 and Q4-Q6, a change in the potential of V_{in} (1) or V_{in} (2) will not alter this situation, provided that it does not occur at both inputs simultaneously.

However, if both V_{in} (1) and V_{in} (2) were to go positive simultaneously, the current through R1 would increase and that through R2 would decrease, causing a positive-going output change at point B. On the other hand, if V_{in} (1) were to go positive while V_{in} (2) were to go negative, the output voltage at B would be negative going.

This allows the circuit to be used as an FM demodulator in an arrangement of the kind shown in Fig. 14. If the IF signal from the limiting amplifier LA1 is taken to point E and a small amount of this signal is used to drive a tuned circuit connected across points C and D, then, as the input frequency varies with respect to the natural resonant frequency of L1/L3, so the phase of the signal applied to point C will alter and the output at B will change.

The linearity of the audio output from this type of demodualtor can be very good, but the AF output and signal-to-noise ratio decrease if the quadrature input from L1/C3 gets smaller. For this reason, the coupling capacitor C2 must not be small, but then again it must not be so big that it swamps the input at C. Also, the performance of the circuit is better if the Q of L1/C3 is high, but it most not be so high that this circuit cannot ade-

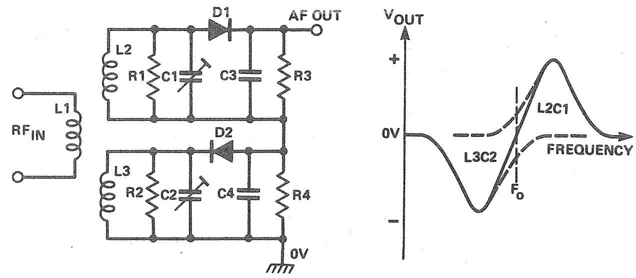


Fig. 11 The circuit and output voltage/frequency characteristics of a Round-Travis FM detector.

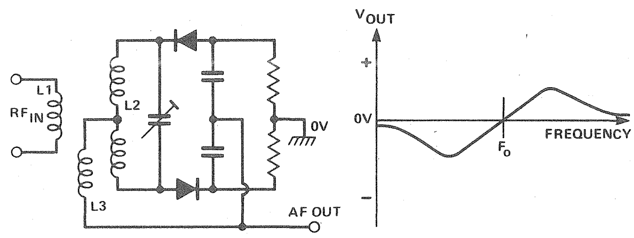


Fig. 12 The circuit and output voltage/frequency characteristics of a ratio detector.

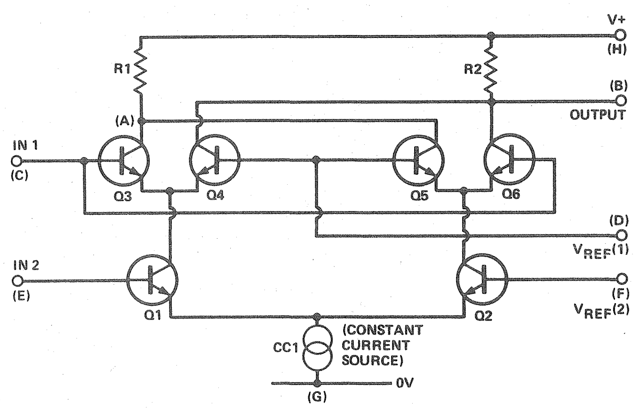


Fig. 13 A transistor gate-coincidence array.

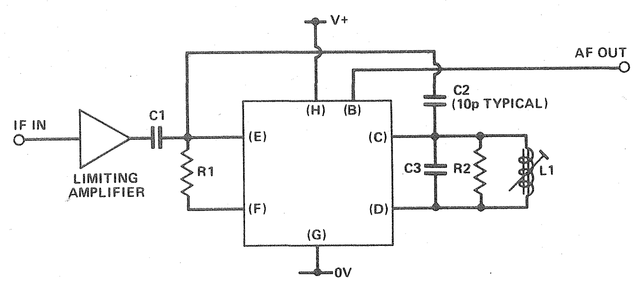


Fig. 14 Using a gate-coincidence detector for FM demodulation.

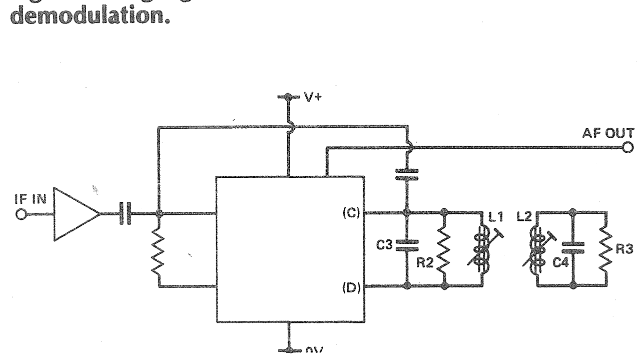


Fig. 15 Improving the linearity of a gate-coincidence demodulator by means of a band-pass quadrature circuit.

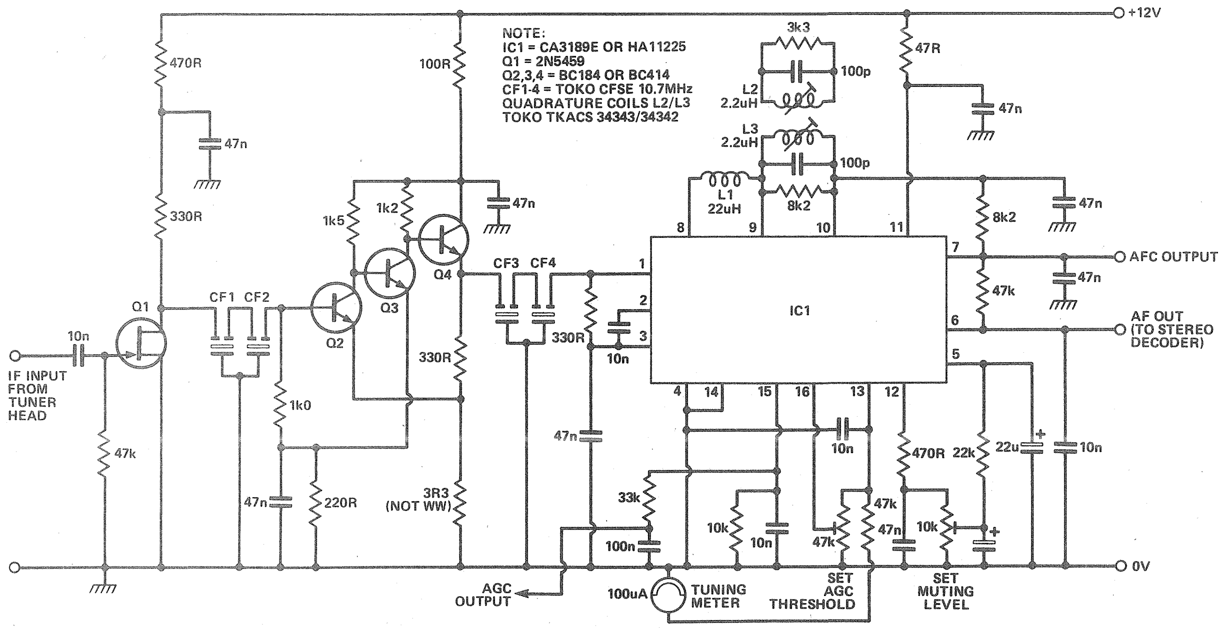


Fig. 16 A complete FM IF and demodulator circuit using the RCA CA3189E or the Hitachi HA11225.

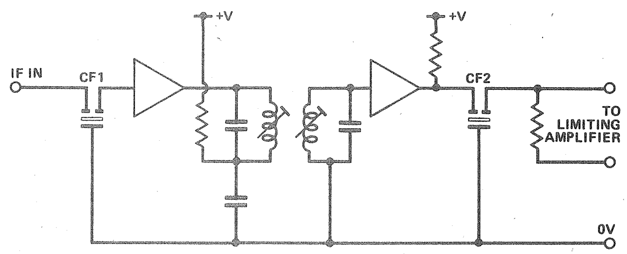


Fig. 17 Arrangement for group delay compensation in an IF amplifier.

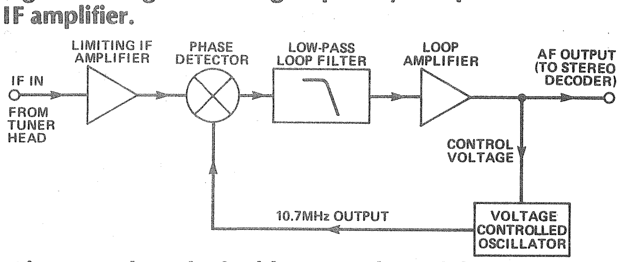


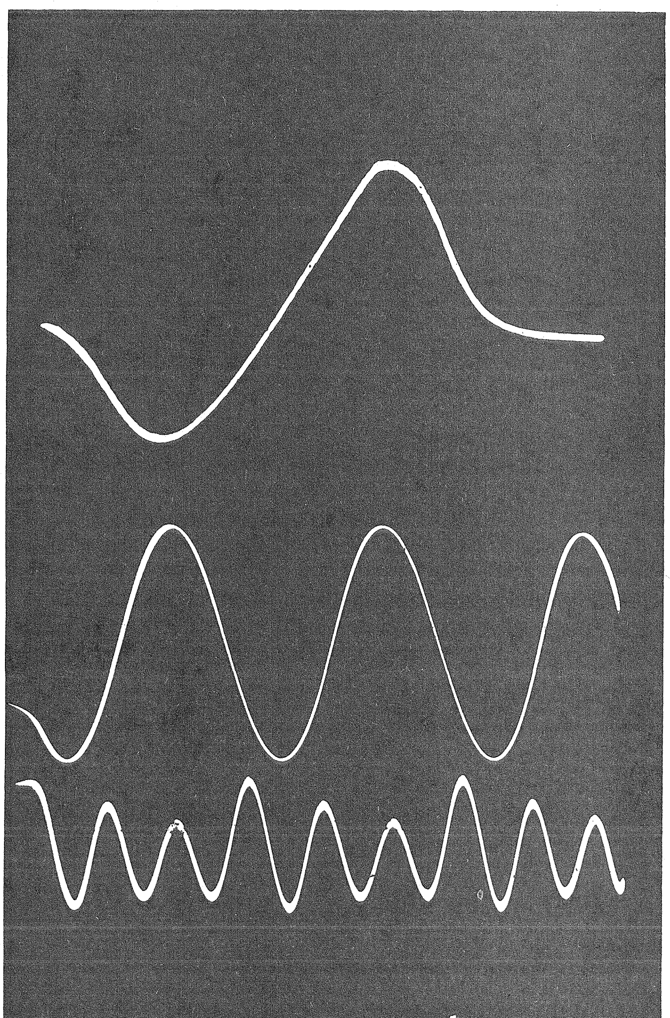
Fig. 18 A phase-locked-loop FM demodulator system.

quately cover the required FM IF bandwidth (±150kHz or so).

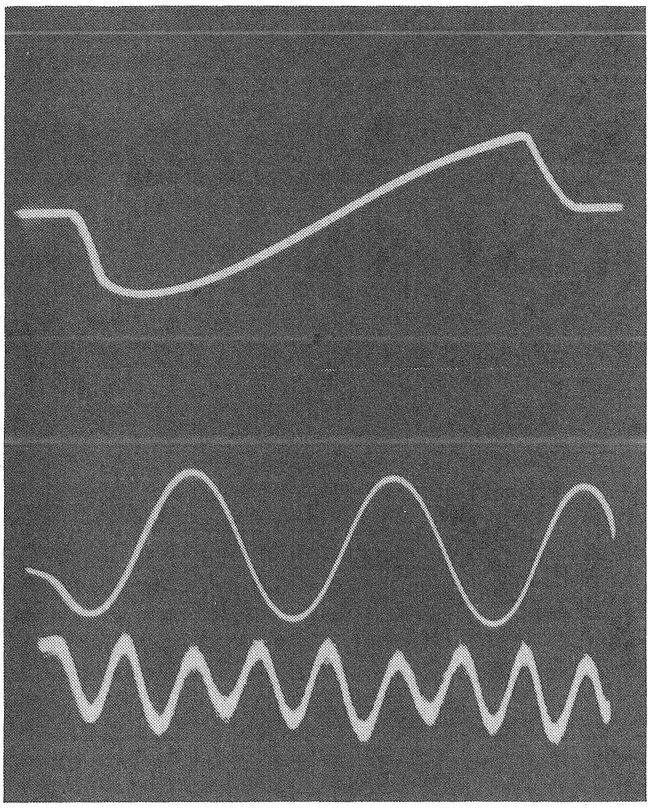
A useful improvement in performance is given if L1/C3 is elaborated to a bandpass coupled pair of circuits, as shown in Fig. 15. This allows both L1 and L2 to have higher values of Q without impairing the usable bandwidth. A complete FM IF layout based on a CA3189E and a double-tuned quadrature coil system is shown in Fig. 16.

THD values down to 0.1% have been claimed for this particular demodulator configuration, as compared with 0.3-0.5% for normal, single coil systems.

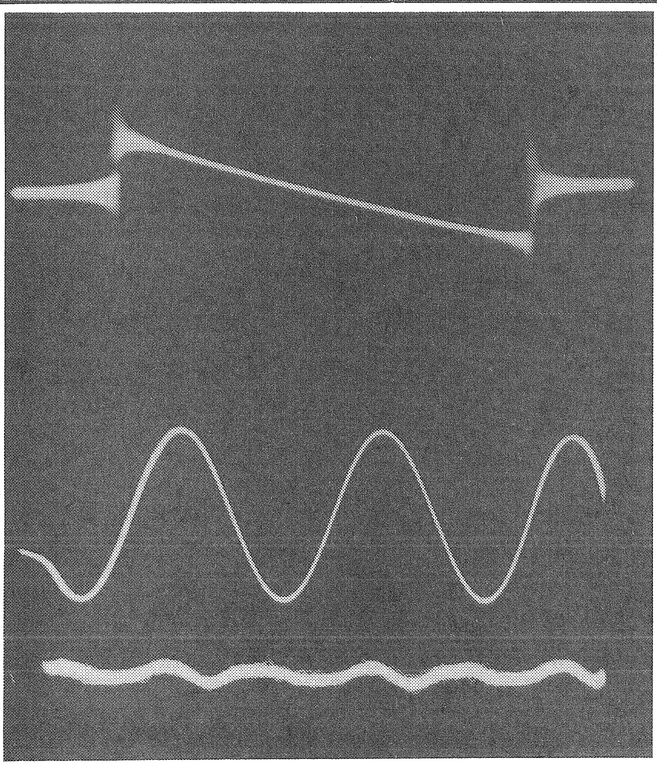
However, although the demodulator system on its own might give values of this kind, any phase-sensitive demodulator system (including the Foster-Seeley, the ratio detector, and the gate-coincidence system) will be influenced by the frequency/phase relationships imposed by preceding amplifier stages, RF coils, ceramic filters and like. So, in practice, the final result will never be quite as good as the figures claimed by IC manufacturers.



The voltage/frequency characteristics of an FM ratio detector and the resultant 1kHz demodulator harmonic distortion (about 1.2%).



The voltage/frequency characteristics of an FM gate-coincidence detector circuit (single coil type) and the resultant 1kHz demodulator harmonic distortion (about 0.6%).



The voltage/frequency characteristics of a phase-locked-loop FM demodulator and the resultant 1kHz harmonic distortion (about 0.15%).

A simple circuit arrangement which is used commercially to compensate for the errors in the group delay characteristics of ceramic filters, and thereby to improve the overall THD figure of the receiver, is shown in Fig. 17. In this, advantage is taken of the fact that the phase characteristics of bandpass-coupled tuned circuits tend to be similar to those of ceramic filters but opposite in sign. The combination of these two kinds of elements therefore tends to cancel the phase error.

The influence of the frequency/phase characteristics of preceding stages also explains why it is very important that incipient instability in the RF or IF tuned circuits must be avoided. It can have a drastic effect on the phase linearity and the final THD of the whole receiver system.

Other Demodulator Systems

In view of the dependence of phase detector systems on the phase linearity of the preceding stages, it is surprising that so little commercial interest has been shown in the alternatives, of which the two most promising are the phase-locked loop (PLL), and the pulse counting system.

In both of these cases the output signal voltage is related directly to the incoming signal frequency, and the demodulator linearity is inherent in itself rather than dependent on the performance of the preceding parts of the system.

The phase-locked loop in particular is a very elegant way of demodulating an FM signal, but is seldom if ever used even though a PLL arrangement is commonly employed in the stereo decoder section.

The basic layout of a PLL FM demodulator is shown in Fig. 18. In this the incoming 10.7MHz IF signal from the limiting amplifier is fed to a phase detector along with the output from a linear voltage controlled oscillator. The phase detector can have the kind of circuit layout shown in Fig. 13. Its output (which will consist of the sum and difference frequencies of the two input components) is passed through a low-pass filter and an amplifier (if needed) to form a control voltage for the VCO.

If the VCO pulls into lock, this control voltage will swing up and down as required to force the VCO to stay in frequency synchronisation with the incoming signal. The control voltage, as fed to the VCO, will be as linear an equivalent of the signal modulation as the characteristics of the VCO allows.

The benefits of such a demodulator system are clearly shown in the three photographs which were taken directly off the screen of an oscilloscope. They show various types of FM receivers driven at their aerial inputs by a frequency modulated oscillator. The actual harmonic distortion of the recovered signal for a 1kHz sinewave-modulated input is also shown.

In the case of the two phase detector systems, the linearity of the demodulator itself would have been somewhat better than is shown: a good part of the final distortion is due to the inadequacies of the preceding stages.

In the next part of this series I will describe a practical PLL FM demodulator system having a very high linearity, and I will also look at the problems presented by the stereo decoder stage of the receiver and some possible solutions.

ETI