RF INTEGRATED CIRCUITS
(15A04804)

LECTURE NOTES

B.TECH

(IV YEAR–II SEM)

(2019-20)

PREPARED BY:

MR.G.SIVA KOTESWARA RAO, ASSISTANT PROFESSOR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

VEMU INSTITUTE OF TECHNOLOGY

(Approved by AICTE, New Delhi and affiliated to JNTUA, Ananthapuramu)

NEAR PAKALA, P.KOTHAKOTA, CHITTOOR- TIRUPATHI HIGHWAY

CHITTOOR-517512 Andhra Pradesh

WEB SITE: www.Vemu.Org

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR

B.Tech IV-II Sem (E.C.E)	**T**	**Tu**	**C**
	3	**1**	**3**

(15A04804) RF INTEGRATED CIRCUITS

C422_1: Explain the Basic architectures of RF Transceivers.
C422_2: Explain the MOS device Review, High frequency Amplifiers and Bandwidth Estimation Techniques.
C422_3: Explain the Noise present in the Active and Passive Elements, LNA and Mixers.
C422_4: Design the RF power amplifiers, Negative Resistance Oscillators and PLL.
C422_5: Explain the Frequency Synthesizers and Radio architectures.

UNIT – I

Introduction RF systems – basic architectures, Transmission media and reflections, Maximum power transfer , Passive RLC Networks, Parallel RLC tank, Q, Series RLC networks, matching, Pi match, T match, Passive IC Components Interconnects and skin effect, Resistors, capacitors Inductors

UNIT – II

Review of MOS Device Physics - MOS device review, Distributed Systems, Transmission lines, reflection coefficient, the wave equation, examples, Lossy transmission lines, Smith charts – plotting Gamma, High Frequency Amplifier Design, Bandwidth estimation using open-circuit time constants, Bandwidth estimation, using short-circuit time constants, Rise time, delay and bandwidth, Zeros to enhance bandwidth, Shunt-series amplifiers, tuned amplifiers, Cascaded amplifiers

UNIT - III

Noise - Thermal noise, flicker noise review, Noise figure, LNA Design, Intrinsic MOS noise parameters, Power match versus, noise match, large signal performance, design examples & Multiplier based mixers. Mixer Design, Sub sampling mixers.

UNIT – IV

RF Power Amplifiers, Class A, AB, B, C amplifiers, Class D, E, F amplifiers, RF Power amplifier design examples, Voltage controlled oscillators, Resonators, Negative resistance oscillators, Phase locked loops, Linearized PLL models, Phase detectors, charge pumps, Loop filters, and PLL design examples

UNIT - V

Frequency synthesis and oscillators, Frequency division, integer-N synthesis, Fractional frequency, synthesis, Phase noise, General considerations, and Circuit examples, Radio architectures, GSM radio architectures, CDMA, UMTS radio architectures

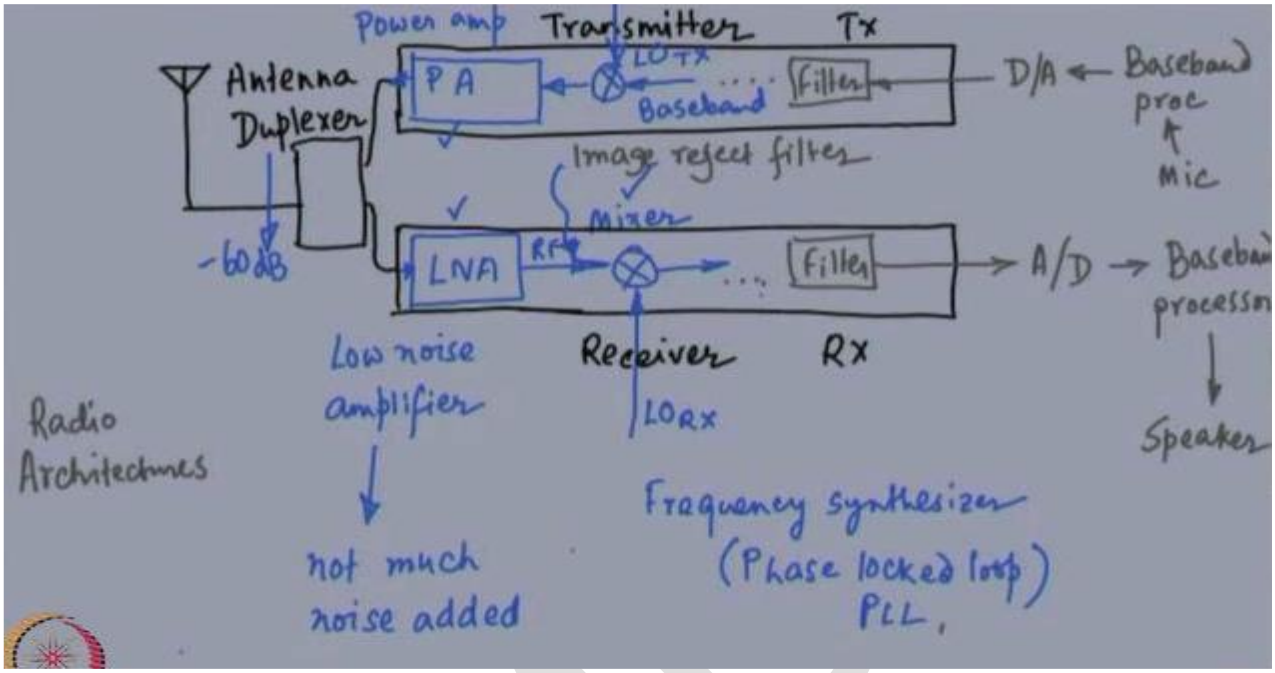
Text books:

1. The design of CMOS Radio frequency integrated circuits by Thomas H. Lee Cambridge university press, 2004.
2. RF Micro Electronics by Behzad Razavi, Prentice Hall, 1997.

UNIT-I

Introduction RF systems

1. INTRODUCTION RF SYSTEMS – BASIC ARCHITECTURE:

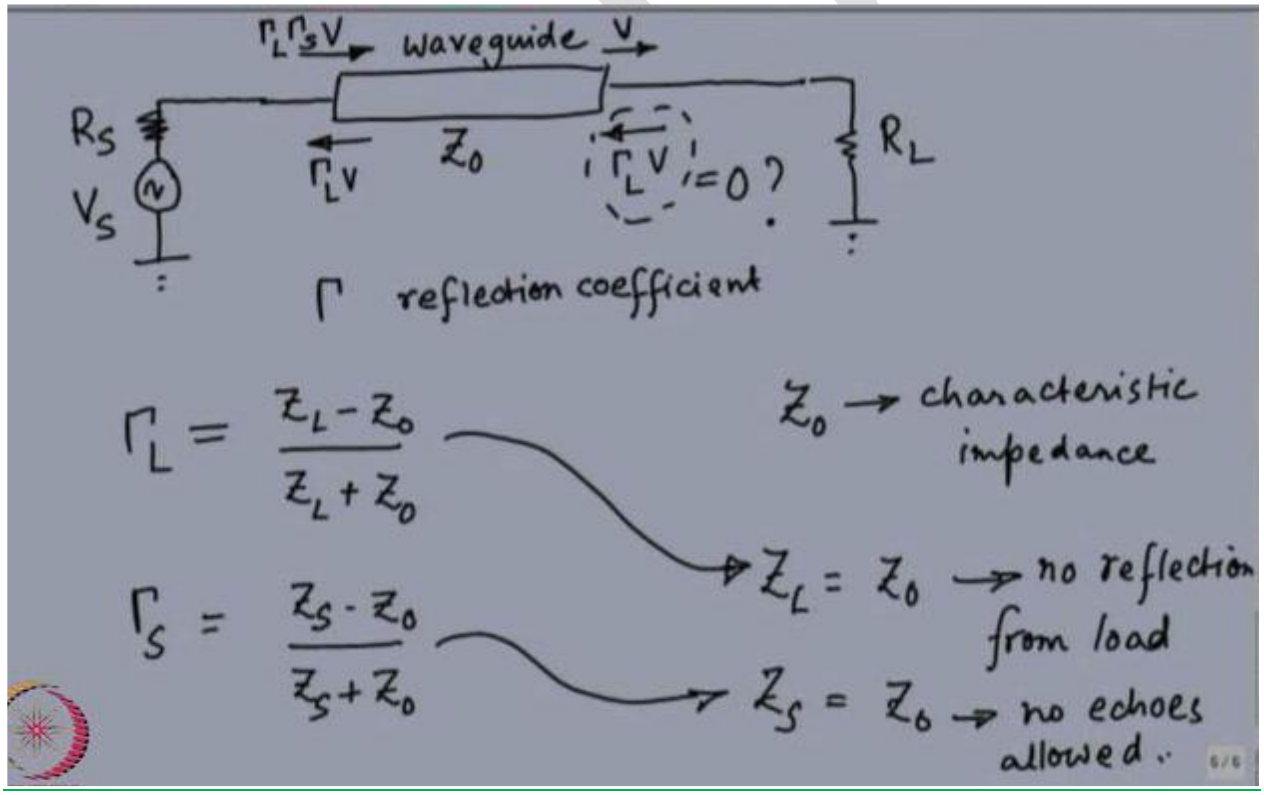


The first and for most thing that it has is something called an antenna. This is the symbol for the antenna. The cell phone has inside it an antenna. Now, there is a transmitter and there is a receiver. So, there is a transmitter and there is a receiver; both are going to be using the same antenna. How is it possible? How can both the transmitter and the receiver use the same antenna? Now, for this, there is something called it is like a switch; it is called a duplexer. Now, this switch separates the transmit path from the receive path. Duplexer is a surface acoustic wave kind of component, that is why this duplexer has to be a very good duplexer – something that separates the transmit chain from the receive chain; it could be a filter; in which case, it has to have an extremely good isolation. It could be a switch maybe when the transmitter is working, the receiver is not working. In that case also, it has to have very low attenuation; typically, it is a semi-mechanical switch.

Let us look at the receiver side now, The first thing that you need is something called a low noise amplifier, receiving a tiny signal from the atmosphere, this tiny signal has to be amplified, so that you can make sense of what was spoken on the other side. So, it has got to be an amplifier, Second thing is it has to be low noise. Low noise here means that, it does not add too much noise on its own, A low noise amplifier does not add too much noise, Therefore, a low noise amplifier cannot throw out the noise and keep the signal; it has to handle both the noise and the signal that it has already received. Every system unless it is a passive lossless system, every other system adds noise. If it burns power, it adds noise.

So, a low noise amplifier is most probably going to burn power; Just before throwing out the signal to the antenna to the duplexer, what we need is something called a power amplifier, We want to blast as much power as possible into the atmosphere, so that the base station can here me clearly. So, that is a power amplifier. That is the last block on the transmit chain, let us try to understand that, a cell phone; when you are receiving signal, it is probably going to use 800 megahertz or a 1600 megahertz or some extremely high frequency. If it is an extremely high frequency, we do not like these extremely high frequencies, because it is hard to work with them. So, the first thing that we need to do is to bring it down to a lower frequency. So, how do we bring it down to a lower frequency? We use something called a multiplier. Or, in other words, it is called a mixer. it down converts the high frequency that you received to something that is more manageable, LO stands for local oscillator. The local oscillator for the transmitter is typically different from the local oscillator of the receiver, you do not want transmit and receive to be working at the same frequency band. So, these two frequencies are generated on chip; they are different. And these two frequencies mix with the RF signal or with the baseband signal and create the low frequency or the high frequency whatever you want depending on Rx or Tx. the transmit side local oscillator will be oscillating at a frequency different from the receive side local oscillator, Why cannot I transmit and receive at the same frequency? Because if I do not have them different, then mostly, what I am going to be hearing on the receive is an echo of what I transmitted.

2.TRANSMISSION MEDIA AND REFLECTIONS:-



an electromagnetic waveguide could be a wire; it could be a real waveguide; it could be the atmosphere; could be anything, Gamma is the reflection coefficient. So, if there is a wave – it is a waveguide; there is a wave hitting a certain object; a portion of the wave gets absorbed into the object; a portion of the wave reflects back from the object.

at the load, this reflection coefficient is Z L minus Z 0 divided by Z L plus Z 0; where, Z L is basically R L in the case that I have drawn;
and Z 0 is the characteristic impedance of the waveguide in question. So, this is how we are going to understand the reflections.

So, the reflection coefficient happens to be equal to this. At the source side, it is going to be something very similar – Z S minus Z 0 divided by Z S plus Z 0; Z S in this case is the source resistance of the voltage that has been applied, the antenna is passing over the signal to the low noise amplifier. If the low noise amplifier has an input impedance of Z L; and if the antenna – receive antenna has the characteristic impedance of Z naught; and if Z naught equal to Z l, then all the signal that hits the low noise amplifier is going to be absorbed by the low noise amplifier; nothing of that signal is going to be reflected back, the characteristics impedance of the antenna is chosen to be 50 ohms.

3. PASSIVE *RLC* NETWORKS:

3.1.INTRODUCTION:- One characteristic of R F circuits is the relatively large ratio of passive to active components. In stark contrast with digital VLSI circuits (or even with other analog circuits. such as op-amps), many of those passive components may be inductors or even transformers, This chapter hopes to convey some underlying intuition that it is useful in the design of *RLC* networks. As we build up that intuition, we'll begin to understand the many good reasons for the preponderance of *RLC* networks in RFcircuits. Among the mOSIcompelling of these are that they can be used to match or

Otherwise modify impedances (important for efficient power transfer. for example), cancel transistor parasitic to provide high gain at high frequencies and filler out unwanted signals. To understand how *RLC* networks may confer these and other benefits. let's revisit some simple second-order examples from undergraduate introductory network theory. By looking a: how these networks behave from a couple of different viewpoints .well build up intuition that will prove useful in understanding networks of much higher order.

3.2.PARALLEL *RLC* TANK: - Let's just jump right into the study of a parallel *R LC* circuit. As you probably know. This circuit exhibits resonant behavior: we'll see what this implies momentarily. This circuit is also often called a *tank circuit* \ (or simply *tank),* We begin by studying its complex impedance. or more directly. its admittance (more convenient for a parallel network: see Figure 1.1. For this network, we know that the admittance is simply

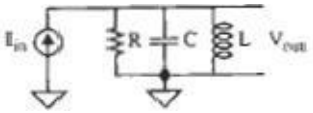


Figure 1.1 Parallel RLC network

$$Y = G + j\omega C + \frac{1}{j\omega L} = G + j\left(\omega C - \frac{1}{\omega L}\right).$$

Therefore say that. at very low frequencies. the network's admittance is essentially that of the inductor (since its admittance dominates the combination) and is also that of the capacitor at very high frequencies.

What divides "low" from "high" is the frequency at which the inductive and capacitive admittances cancel. Known as the resonant frequency. this is given by

$$\left(\omega_0 C - \frac{1}{\omega_0 L}\right) = 0 \implies \omega_0 = \frac{1}{\sqrt{LC}}.$$

3.3. QUALITY FACTOR:-

$$Q \equiv \omega \frac{\text{energy stored}}{\text{average power dissipated}}.$$

specific about what stores or dissipates the energy. So. as we'll see later on, it applies perfectly well even to distributed systems, such as microwave resonant cavities, where it is not possible to identify individual inductances, capacitances, and resistances. It should also be clear that the notion of *Q* applies both to resonant and nonresonant systems, so one may talk of the *Q* of an *RC* circuit. A high-order system may exhibit multiple resonances, each with its own peak *Q* value. From the fundamental definition, we also sec that the value we compute depends on whether or not

we include external loading, and perhaps also on how that load connects to the network in question. If we neglect the loading then we refer to the computed value as the unloaded *Q*, and if we include it then we call it the loaded *Q*. whenever the context is ambiguous and the distinction mailers. it is important to identify explicitly the type of *Q* under discussion.

Let's now use this definition to derive expressions for the Q of our parallel *RLC*circuit at resonance. At the resonant frequency, which we'll denote by *Wo* the voltage across the net• work is simply *linR*. Recall that energy in such a network sloshes back and forth between the inductor and capacitor, with a constant sum at resonance. As a consequence then network energy and power is given by

$$E_{\text{tot}} = \tfrac{1}{2} C (I_{\text{pk}} R)^2.$$

$$P_{\text{avg}} = \tfrac{1}{2} I_{\text{pk}}^2 R.$$

Then the Q of the network is given by

$$Q = \omega_0 \frac{E_{\text{tot}}}{P_{\text{avg}}} = \frac{1}{\sqrt{LC}} \frac{\tfrac{1}{2} C (I_{\text{pk}} R)^2}{\tfrac{1}{2} I_{\text{pk}}^2 R} = \frac{R}{\sqrt{L/C}}.$$

Where The quantity of √L/C is called characteristic impedance

3.4. SERIES *RLC* NETWORKS:- We may follow an exactly analogous dual approach to deduce the properties of scries *R LC* circuits. The details of the derivations arc relatively uninteresting. so here we simply present the relevant observations and equations.The resonant condition corresponds again to the frequency where the capacitance and inductance cancel.

Rather than resulting in an admittance minimum, though. resonance here results in an impedance minimum, with a value of *R*. The equation for *Q* involves the same terms as for the parallel case. but in reciprocal form:

$$Q = \frac{\sqrt{L/C}}{R}.$$

At resonance. the voltage across either the inductor or capacitor is *Q* times as great as that across the resistor. Thus. if a series *RLC* network with a *Q* of 1000 is driven at resonance with a one- volt source. then the resistor will have that one volt across it yet a thrilling one thousand volts will appear across the inductor and capacitor.

3.5. OTHER RESONANT *RLC* NETWORKS :- Purely parallel or series *R LC* networks rarely exist in practice. so it's important to take *a* look at configurations that might be more realistically representative. Consider. for example. the case sketched in Figure 1.2. Because inductors tend to be significantly lossier than capacitors. the model shown in the figure is often a more realistic approximation to typical parallel *RLC* circuits. Since we've already analyzed the purely parallel *RLC* network in detail. it would be nice if we could re-use as much of this work as possible. So, let's convert the

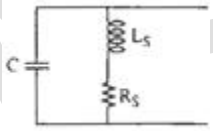


Figure 1.2 not a quite parallel RLC circuits

circuit of Figure 3.2 to a purely parallel *RLC* network by replacing the series *LR* section with ,1 parallel one. Clearly. such a substitution cannot be valid in general. but over a suitably restricted frequency range (e.g .. near resonance) the equivalence is pretty reasonable. To show this formally. let 's equate the impedances of the series and parallel *L R* sections:

$$j\omega_0 L_S + R_S = [(j\omega_0 L_P) \parallel R_P] = \frac{(\omega_0 L_P)^2 R_P + j\omega_0 L_P R_P^2}{R_P^2 + (\omega_0 L_P)^2}.$$

If we equate real parts and note that $Q = R_P/\omega_0 L_P = \omega_0 L_S/R_S$,[7] we obtain

$$R_P = R_S(Q^2 + 1).$$

Similarly, equating imaginary parts yields

$$L_P = L_S\left(\frac{Q^2 + 1}{Q^2}\right).$$

we may also derive a similar set of equations for computing series and parallel *RC* equivalents:

$$R_P = R_S(Q^2 + 1).$$
$$C_P = C_S\left(\frac{Q^2}{Q^2 + 1}\right).$$

Let's pause for *a* moment and look at these transformation formulas, Upon closer examination. it's clear that we *may* express them in a universal form that applies to both *RC* and *LR* networks:

$$R_P = R_S(Q^2 + 1)$$
$$X_P = X_S\left(\frac{Q^2 + 1}{Q^2}\right).$$

where X is the imaginary part of the impedance. This way, one need only remember a single pair of "universal" formulas in order to convert any "impure" *RLC* network into a purely parallel (or series) one that is straightforward to analyze. However. One must bear in mind that/he equivalences hold only over a narrow range of frequencies centered about *W0.*

3.6. RLC NETWORKS AS IMPEDANCE TRANSFORMERS:- The relative abundance of power gain at low frequencies allows designers to treat it essentially as an infinite resource. Design specifications are thus often expressed simply in terms of *a* voltage gain. for example, without any explicit reference to or concern for power gain. Hence. circuit design at low frequencies usually proceeds in blissful ignorance of the maximum power transfer theorem derived in every undergraduate network theory course. In striking contrast with that insouciance. RF circuit design is frequently *preoccupied* with power gain because of its relative scarcity. Irnpedancc-rransforming networks thus play a prominent role in the radio frequency domain. figure1.3 maximum power transfer

3.6.1 THE MAXIMUM POWER TRANSFER THEOREM :- To understand more explicitly the value of impedance transformers. we now review the maximum power transfer theorem as shown in Figure 1.3. The problem is this: Given a *fixed source* impedance *Zs.* what load impedance *ZL* maximizes the power delivered to the load? The power delivered to the load impedance is entirely due to *RL.* since reactive elements do not dissipate power. Hence. the power delivered is simply

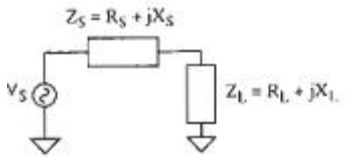
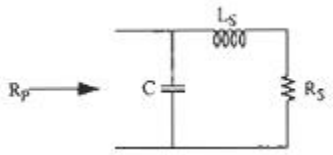


figure1.3 maximum power transfer

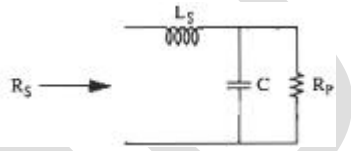
$$\frac{|V_R|^2}{R_L} = \frac{R_L|V_S|^2}{(R_L + R_S)^2 + (X_L + X_S)^2}.$$

where *VR* and *Vs* are the rrnx voltages across the load resistance and source respectively. To maximize the power delivered to *RL*. it's clear that *XL* and *Xs* should be inverses so that they sum to zero. In addition. Maximizing above Eqn. under that condition leads to the result that *RL* should equal *Rs*. Hence. the maximum power transfer from a fixed source impedance to a load occurs when the load and source impedances are complex conjugates. Having established mathematically the condition for maximum power transfer, we now consider practical methods for achieving it .

3.6.2 THE L-MATCH :- The multiplication by *Q* of voltages or currents in resonant *RLe* networks hints at their impedance-modifying potential. Indeed. the series-parallel *RClLR* network conversion formulas developed in the previous section actually show this property explicitly. To make this clearer. Consider once again the circuit of Figure 1.2. Redrawn slightly as Figure 1.4. Here we treat *Rs* as a load resistance for the network. When this resistance is viewed across the capacitor. it is transformed to an equivalent



1.4 Upward impedance transformer



1.5 Downward impedance transformer

R>>1 according to the formulas developed in the previous section. From inspection of those "universal" equations. il is clear that *Rp* will always be larger them *Rs*. so the network of Figure

1.4 transforms resistances upward. To get a downward impedance conversion. just interchange ports as shown in the Figure 3.5.

This circuit is known as an *L-match* because of its shape (perhaps you have to be lying on your side and dyslexic to see this). and it docs have the attribute of simplicity. However, there are only two degrees of freedom (one can choose only *L* and *e).*Hence. once the impedance transformation ratio and resonant frequency have been specified, network *Q* is automatically determined. If you want a different value of *Q* then you must use a network that offers additional degrees of freedom: we'll study some of these shortly. As a final note on the L-match the "universal" equations can be simplified if *Q′* >>l. If this inequality is satisfied. then the following approximate equations hold is

$$R_P \approx R_S Q^2 = R_S\left(\frac{1}{\omega_0 R_S C}\right)^2 = \frac{1}{R_S}\frac{L_S}{C}.$$

$$R_P R_S \approx \frac{L_S}{C} = Z_0^2.$$

Which may be written as

where Zo is the characteristic impedance of the network. One may also deduce that *Q* is approximately the square root of the transformation ratio is given by

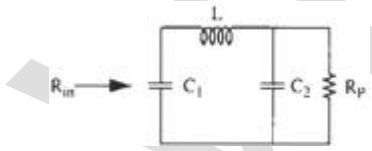
$$Q \approx \sqrt{\frac{R_P}{R_S}}.$$

Finally. the reactance's don't vary much in undergoing the transformation is

$$X_P \approx X_S.$$

As long as *Q* is greater than about 3 or4. the error incurred will be under about 10%.If *Q* is greater than 10. the maximum error will be in the neighborhood of 1% or 50.Hence. for quick. Back-of-the-envelope calculations. these simplified equations are adequate. Final design values can be computed using the full "universal" equations.

3.6.3 THE Pi-MATCH :- one limitation of the L-match is that one can specify only two of center frequency. Impedance transformation ratio. and *Q*. To acquire a third degree of freedom. one can employ the network shown in Figure 1.6.This circuit is known as a *Pi* -match, again because of its shape. The most expedient way to understand this matching network is to view it as two L matches connected in cascade. one that transforms down and one that transforms up; see Figure 1.7. Here. the load resistance *Rp* is transformed to a lower resistance (known as the *image* or *intermediate resistance,* here denoted *R,)* at the junction of the two inductances. The image resistance is then transformed up to a value Rin by a second L-match section.



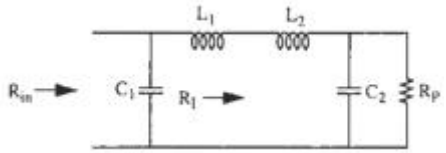


figure 1.6 The Pi match figure 1.7 The Pi match as cascaded of two L-matches

In order to derive the design equations. first transform the parallel *RC* subnetwork of the right- hand L-section into its series equivalent. as shown in Figure 1.8, When we replace the output parallel LC network with its series equivalent. the series resistance is. of course. *R,*. Hence. the *Q* of the right-hand L-scction may be written as

$$\frac{\omega_0 L_2}{R_I} = \sqrt{\frac{R_P}{R_I} - 1} = Q_{right}.$$

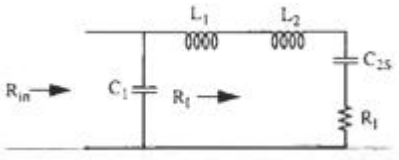


Figure 1.8Pi-match with transformed right hand L-section.

At the same Time. recognize that the left-hand L section about series a resistance of *R*, at the center frequency. Therefore. its *Q* is given by

$$\frac{\omega_0 L_1}{R_I} = \sqrt{\frac{R_{in}}{R_I} - 1} = Q_{left}.$$

The overall network of Q is given by

$$Q = \frac{\omega_0 (L_1 + L_2)}{R_I} = \sqrt{\frac{R_{in}}{R_I} - 1} + \sqrt{\frac{R_P}{R_I} - 1}.$$

Above equation is allows us to find the image resistance. given *Q* and the transformation resistances. Once *R*, is computed. the total inductance is quickly found is

$$L_1 + L_2 = \frac{QR_I}{\omega_0}.$$

The values of capacitances are given by

$$C_1 = \frac{Q_{left}}{\omega_0 R_{in}}.$$

$$C_2 = \frac{Q_{right}}{\omega_0 R_P}.$$

As a practical matter. note that finding *Ri* generally requires iteration. A good starting value can be obtained by assuming that *Q* is large . In that case. *Ri* is approximately given by

$$R_I \approx \frac{\left(\sqrt{R_{in}} + \sqrt{R_P}\right)^2}{Q^2}.$$

If *Q* is very large, or if you're just doing some preliminary "cocktail napkin" calculations. then iteration may not even be necessary. And that's all there is to it .As a parting note. one final bit of trivial deserves mention. An additional reason that the Pi -match is popular is that the parasitic capacitances of whatever connects to it can be absorbed into the network design. This property is particularly valuable because capacitance is the dominant parasitic clement in many practical cases.

3.6.4 THE T·MATCH :- The zr-march results from cascading two L-sections in one particular way. Connecting up the L-sections another way leads to the dual of the Pi-match as shown in Figure 1.9. Here what would be a single capacitor in a practical implementation has been decomposed explicitly into two separate ones. The (parallel) image resistance is seen across these capacitors. Either looking to the right or looking to the left as in the *Pi-match*.

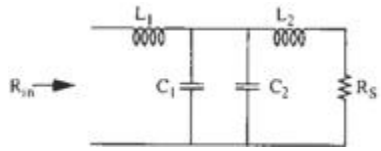


Figure 1.9 T-match

The design equations are readily derived by following an approach analogous to that used for the Pi-match. The overall network *Q* is simply

$$Q = \omega_0 R_I (C_1 + C_2) = \sqrt{\frac{R_I}{R_{in}} - 1} + \sqrt{\frac{R_I}{R_S} - 1},$$

From which image resistance may be found then

$$C_1 + C_2 = \frac{Q}{\omega_0 R_I},$$

$$L_1 = \frac{Q_{left} R_{in}}{\omega_0},$$

$$L_2 = \frac{Q_{right} R_S}{\omega_0}.$$

the T-match is particularly useful when the source and termination parasitic." are primarily inductive in nature. Allowing them to be absorbed into the network.

4. PASSIVE IC COMPONENTS:-

4.1 INTRODUCTION:- We've seen that RF circuits generally have many passive components. Successful design therefore depends critically on a detailed understanding of their characteristics .Since main~treall1integrated circuit (IC) processes have evolved largely to satisfy the demands of digital electronics, the RF IC designer has been left with a limited palette of passive devices, For example, inductors larger than about 10 nH consume significant die area and have relatively poor *Q* (typically below 10) and low self-resonant frequency. Capacitors with high *Q* and low temperature coefficient are available. But tolerances are relatively loose (e.g.. order of 20% or worse). Additionally. the most area-efficient capacitors also tend to have high loss and poor voltage coefficients. Resistors with low self-capacitance and temperature coefficient are hard to come by and one must also occasionally contend with high voltage coefficients, loose tolerances. and a limited range of values .

4.2 INTERCONNECT AT RADIO FREQUENCIES: SKIN EFFECT

At low frequencies, the properties of interconnect we care about most are resistivity, current-handling ability, and perhaps capacitance. As frequency increases. we find that inductance might become important. Furthermore, we invariably discover that the resistance increases owing to a phenomenon known as the *skill effect.*

Skin effect is usually described as the tendency of current to now primarily on the surface (skill) of a conductor as frequency increases. Because the inner regions of the conductor are thus less effective at carrying current than at low frequencies, the useful cross-sectional area of a conductor is reduced. thereby producing a corresponding increase in resistance.

To develop a deeper understanding of the phenomenon. we need to appreciate explicitly the role of the magnetic field in producing the skin effect. To do so qualitatively. let's consider a solid cylindrical conductor carrying a time-varying current, as shown in Figure 4.1. Assume for now that the return current (there must always be one in any real system) is far enough away that its influence may be neglected. A time varying current *I* generate a time-varying magnetic field *H.* That lime-varying field induces a voltage around the rectangular path shown. in accordance with Faraday's law. Ohm's law then tells us that the induced voltage in turn produces a current flow along that same rectangular path. as indicated by the arrows. Now here's the key observation: The direction of the induced current along path A is *opposite* that along8. The induced current thus adds to the current flowing along one side of the rectangle and subtracts from the other. Taking care to keep track of algebraic signs. We see that the current along the surface is the one that is augmented whereas the current below the surface is diminished. In other words. current flow is strongest near the surface; that's the skin effect

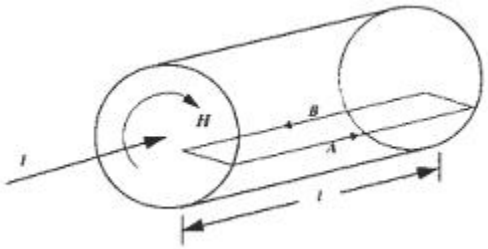


figure 4.1. Illustration 01 skin effect with isolated cylindrical conductor

To develop this idea a little more quantitatively. let's apply Kirchhoff's voltage law (with proper accounting for the induced voltage term. both in magnitude and sign) around the rectangular path to obtain

$$J_B \rho l - J_A \rho l + \frac{d\phi}{dt} = 0,$$

where J is the current density. p is the resistivity. and ¢. the flux. is perpendicular to the rectangle shown.We see that. as deduced earlier. the current density along path A is indeed larger than along 8 by an amount that increases as either the depth, frequency. or magnetic field strength increases and also as the resistivity decreases. Any of these mechanisms acts to exacerbate the skin effect. Furthermore. the presence of the derivative tells us that the current undergoes more than a simple decrease with increasing depth; there is a phase shift as well.

if we now increase the radius of curvature to infinity. we may convert the cylinder into the rectangular structure that is more commonly analyzed to introduce skin effect; sec Figure 4:2. We will provide only {he barest outline of how to set up the problem. and then Simply present the solution. Computing the voltage induced by H around the rectangular contour proceeds th Kirchhoff's voltage law is given by

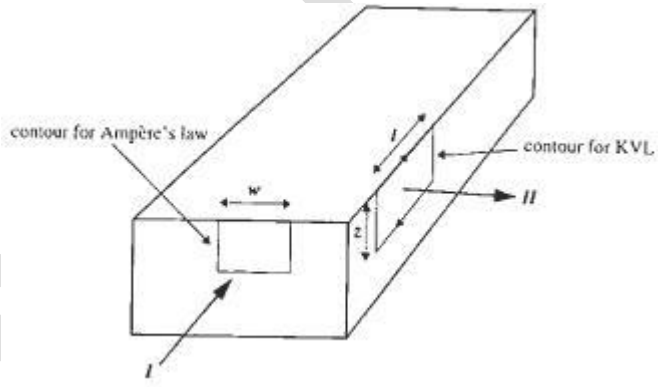


Figure 4.2. Subsection of semi 'infinite conductive block.

$$J\rho l - J_0 \rho l = \frac{d\phi}{dt} = -\frac{d}{dt}\int_0^z Bl\, dz,$$

here the subscript 0 denotes the value at the surface of the conducting block... Now express land *H* (and thus *B)* explicitly as sinusoidally time-varying quantities. Based on this we can states the equation of skin effects is given by

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} = \sqrt{\frac{2}{\omega\mu\sigma}}$$

Notice that the current density decays exponentially from its surface value, Notice also (from the second exponential factor) that there is indeed a phase shift. as argued earlier, with a l-rad lag at a depth equal to δ.

For this case of an infinitely wide. Infinitely long. and infinitely deep conductive block. the skin depth is the distance below the surface at which the current density has dropped by a factor of *e*. For copper at I GHz. the skin depth is approximately211m. For aluminum. that number increases a little bit. to about 2.5 urn. What this exponential decay implies is that making a conductor much thicker than a skin depth provides negligible resistance reduction because the added material carries very little current. Furthermore. we may compute the effective resistance as that of a conductor of thickness li in which the current density is uniform. This fact is often used to simplify computation of the AC resistance of conductors. To make sure that the result is valid. however. the boundary conditions must match those used in deriving our system of equations: The return currents must be infinitely far away. and the conductor must resemble a semi-infinite block. The latter criterion is satisfied reasonably well if all radii of curvature, and all thicknesses. are at least 3-4 skin depths.

4.3 RESISTORS:- are relatively few good resistor options in standard CMOS (complementary metal-oxide silicon) processes. One possibility is to use polysilicon poly interconnect material. Since it is more resistive than metal. However, most poly these days is solicited specifically to reduce resistance. Resistivity's tend to be in the vicinity of roughly *S-IO* ohms per square (within a factor of about 2-4. usually). so poly is appropriate mainly for moderately small-valued resistors. Its tolerance is often poor (e.g., *35%),* and the temperature coefficient, defined as

$$\mathrm{TC} = \frac{1}{R}\frac{\partial R}{\partial T},$$

depends on doping and composition and is typically in the neighborhood of 1000ppm/°C. Unsolicited poly has a higher resistivity (by approximately an order of magnitude, depending on doping), and the TC can vary widely (even to zero, in cer• tain cases) as a function of processing details. It is usually not tightly controlled. So unsolicited poly, if available as an option at all. Frequently Possesses very loose tol• erances (e.g., 50%). Advanced bipolar technologies use self- aligned poly emitters, so poly resistors are an option there, too. In addition to their moderate TC.

poly resistors have a reasonably low parasitic ca• pacitance per unit area and the lowest voltage coefficient of all the resistor materials available in a standard CMOS technology, Resistors made from source-drain diffusions are also an option. The resistivity's and temperature coefficients are generally similar (within a factor of 2, typically) to those of solicited poly silicon, with lower TC associated with heavier doping. There is also significant parasitic (junction) capacitance as well as a noticeable voltage coefficient. The former limits the useful frequency range of the resistor, while the latter limits the dynamic range of voltages that may be applied without introducing sig• nificant distortion. Additionally, care must be taken to avoid forward-biasing either end of the resistor. These characteristics usually limit the use of diffused resistors to noncritical circuits.In modern VLSI (very large-scale integration) technologies. source-drain "diffusions" arc defined by ion implantation. The source-drain regions formed in this way are quite shallow (usually no deeper than about 200-300 nrn, scaling roughly with channel length), quite heavily doped, and almost universally silicided, leading to moderately low temperature coefficients (order of SOO-IOOO ppm/oC).Wells may be used for high-value resistors. since resistivities are typically in the range of 1-10 kQ per square. Unfortunately, the parasitic capacitance is substantial because of the large-area junction formed between the well and the substrate; the resulting resistor has poor initial tolerance (±SO -80%), large temperature coefficient (typically about 3000-S000 ppm/oC. owing to the light doping). and large voltage coefficient. Well resistors must therefore be used with care.Sometimes, a MOS transistor is used as a resistor, even a variable one. With a suitable gate-to-source voltage. a compact resistor can be formed. From first-ordertheory, recall that the incremental resistance of a long-channel MOS transistor in the triode region is

$$r_{ds} \approx \left[\mu C_{ox}\frac{W}{L}[(V_{GS} - V_T) - V_{DS}]\right]^{-1}.$$

Unfortunately, implicit in this equation is that a MOS resistor has loose tolerance (because it depends on the mobility and threshold), high temperature coefficient (because of mobility and threshold variation with temperature) and is quite nonlinear (because it depends on *VDS)*. These characteristics frequently limit its use to noncritical circuits outside of the signal path. An exception is use of such a resistor in certain gain control applications in which the gate drive is derived from a feedback loop so that variations in device characteristics are automatically compensated.One other option that is occasionally useful, particularly to prevent thermal runway in bipolar power stages with paralleled devices. is to use metal interconnect asa small resistor. In most interconnect technologies, metal resistivities are usually onthe order 01*SO*mQ /square, so resistances up to around 10 Q are practical.Aluminum is most commonly used in interconnect and has a temperature coefficient of about 3900 ppm/°C. The TC varies little with temperature and the resistance may be considered PTAT (proportional 10 absolute temperature) over the military temperature range *(-SSOC* to 125°C) to a reasonable approximation is

$$R(T) \approx R_0 \frac{T}{T_0}.$$

where one data point, the resistance Roat temperature *To,* is known.Some processes offer one or more layers of interconnect made of some silicide(mainly for its superior electrornigrution properties). The resistivity is about an order of magnitude larger than that of pure aluminum or copper, while the TC is about the same.A few companies that specialize in analog circuits have modified their processes to provide excellent resistors, such as those made ofNiCr(nichrome) or SiCr (sichrome). These resistors possess low TC (order of 100 ppm/oC or Jess). and thin-film versions arc easily trimmed with a laser to absolute accuracies better than a percent. Unfortunately, these processes are not universally available. and the additional process steps increase die cost significantly.

4.4 CAPACITORS:- the interconnect layers may be used to make traditional parallel plate capacitors as shown in figure 4.8 However ordinary interlevel dielectric tends to be rather thick (order of 0.5-1 *lim).* precisely 10 reduce the capacitance between layers. so the capacitance per unit area is small (a typical value is 5 x 10-5 pF/J..I.m2). Additionally .one must be aware of the capacitance formed by the bottom plate and any conductors (especially the substrate) beneath it.

This parasitic bottom plate capacitance is frequently as large as *1O-30o/c* (or more) of the main capacitance and often severely limits circuit performance.

Figure 4.8 parallel plate capacitor The standard capacitance

formula is given by

$$C \approx \varepsilon \frac{A}{H} = \varepsilon \frac{W \cdot L}{H}.$$

Somewhat underestimates capacitance because it does not take fringing into account. but it is accurate as long as the plate dimensions are much larger than the plate separation *H.* In cases where this inequality is not well satisfied. a rough first-order correction for the fringing may be provided by adding between *H and 2H to each* of W and *L* in computing the area of the plates, Choosing the maximum yields

$$C \approx \varepsilon \frac{(W+2H)\cdot(L+2H)}{H} \approx \varepsilon\left[\frac{WL}{H}+2W+2L\right].$$

One of the few bits of good news in IC passive components is that the TC metal capacitors are quite low. Usually in the range of approximately 30-50 ppm/f'C.and is dominated by the TC of the oxide's dielectric constant itself. as dimensional variations with temperature are negligible.

A simple structure that illustrates the general idea is shown in Figure 4.9, where th two terminals of the capacitor are distinguished by different shadings. As can be seen. the "top" and "bottom" plates. constructed out of the same metal layer. Alternate to exploit the lateral flux. Ordinary vertical flux may also be exploited by arranging the segments of a different metal layer in a complementary pattern, as shown in figure 4.10.

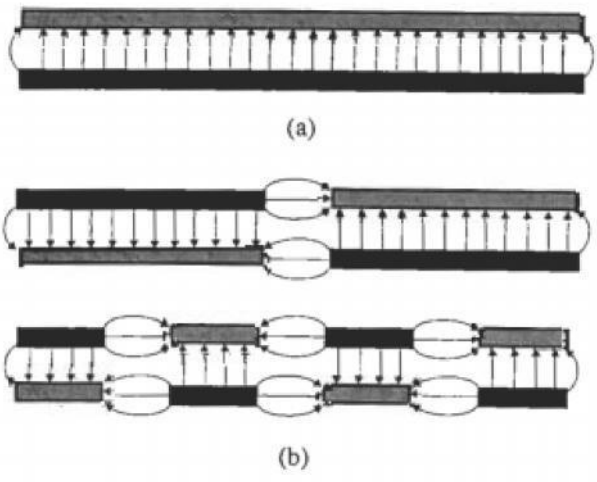


figure 4.9 example of lateral flux capacitor (top-view**)** figure 4.10 example of lateral flux capacitor (side-view)

An important attribute of a lateral flux capacitor is that the parasitic bottom plate capacitance IS much smaller than for an ordinary parallel plate structure, since it consumes less area for a given value of total capacitance. In addition, adjacent plates help steal flux away from the substrate, further reducing bottom plate parasitic capacitance, as seen in Figure 4.11.

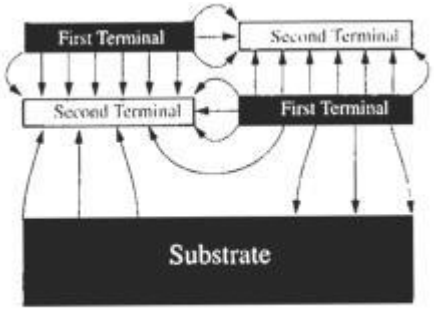


FIGURE 4.11. Illustration of flux stealing.

4.5 INDUCTORS:- From the point of view of R F circuits. the lack of a good inductor is by far the most conspicuous shortcoming of standard Ie processes. Although active circuits can sometimes synthesize the equivalent of an inductor. they always have higher noise. distortion, and power consumption than "real" inductors made with some number of turns of wire.

4.5.1. SPIRAL INDUCTORS:- The most widely used on-chip inductor is the planar spiral, which can assume many shapes as shown in Figure 412. The choice of shape is more often made on the basis of convenience (e.g., whether the layout tool accommodates non-Manhattan geometries) or habit than anything else. Despite stubborn lore 10 the contrary. the inductance and *Q* values attainable are very much second-order functions of shape, so engineers should feel free to usc their favorite shape with relative impunity. Octagonal or circular spirals are moderately better than squares (typically on the order of 10'70)and hence arc favored when layout tools permit their use - or when that modest difference represents the margin between success and failure, The most common realizations use the topmost metal layer for the main part of the inductor (occasionally with two or more levels strapped together to reduce resistance) and provide a connection to the center of the spiral with a cross under implemented with some lower level of meta\. These conventions arise from quite practical considerations: the topmost metal layers in an integrated circuit are usually the thickest and thus generally the lowest in resistance. Furthermore. maximizing the distance to the substrate minimizes the parasitic capacitance between the inductor and the substrate.

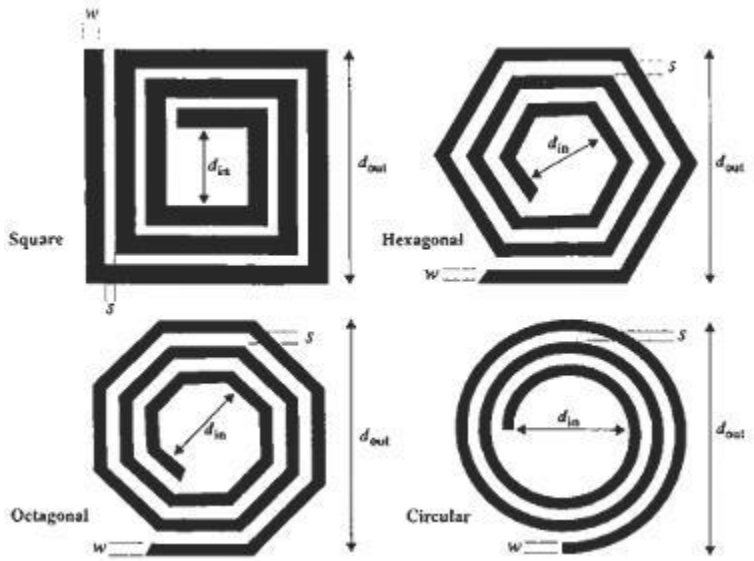


Figure 4.12 planar spiral inductors

the inductance of an arbitrary spiral is a complicated function of geometry, and accurate computations require the use of field solvers or Greenhouse's method." However, a (very) crude zeroth-order estimate. suitable for quick hand calculations, is

$$L \approx \mu_0 n^2 r = 4\pi \times 10^{-7} n^2 r \approx 1.2 \times 10^{-6} n^2 r.$$

where L is in henries, II is the number of turns. and r is the radius of the spiral in meters. This equation typically yields numbers on the high side, but generally within 30% of the correct value (and often better than that).For shapes other than square spirals, multiply the value given by the square spiral formula by the square root of the area ratio to obtain a crude estimate of the correct value Thus. for circular spirals, multiply the square-spiral value by $(\pi/4)^{0.5} \sim 0.89$, and by 0.91 for octagonal spirals. Perhaps more useful for the approximate *design* of a square spiral inductor is the following equation is

$$n \approx \left[\frac{PL}{\mu_0}\right]^{1/3} \approx \left[\frac{PL}{1.2 \times 10^{-6}}\right]^{1/3},$$

where P is the winding pitch ill turns/meter: wc have assumed that the permeability is that of free space. The first of these. which applies to a hollow square spiral inductor is shown in figure 4,22 is

$$L \approx \frac{9.375\mu_0 n^2 (d_{avg})^2}{11 d_{out} - 7 d_{avg}},$$

where dout is the outer diameter and davg is the arithmetic mean of the inner and outer diameters. Checks with a field solver reveal that this modified Wheeler formulae exhibits errors below 5% for typical IC inductors. The inductance of planar spirals of all regular shapes can be cast in a simple unified form if we base a derivation on the properties of a uniform current sheet is

$$L \approx \frac{\mu_0 n^2 d_{avg} c_1}{2}\left[\ln\left(\frac{c_2}{\rho}\right) + c_3\rho + c_4\rho^2\right].$$

here *p* is the *fill factor,* defined as

$$\rho \equiv \frac{d_{out} - d_{in}}{d_{out} + d_{in}}.$$

Figure 4.13 shows a relatively complete model for on-chip spirals. 17 The model is symmetrical. even though actual spirals are not. Fortunately. the error introduced is negligible in most instances. An estimate for the series resistance may be obtained from the following equation is

$$R_S \approx \frac{l}{w \cdot \sigma \cdot \delta(1 - e^{-t/\delta})}.$$

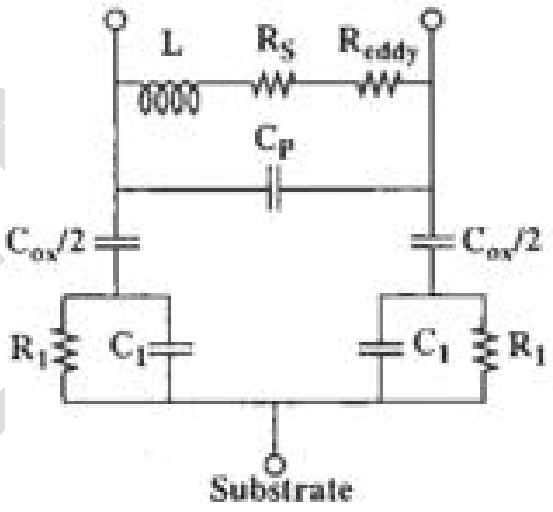


Figure 4.14 model for on-chip inductors

UNIT-2

REVIEW OF MOS DEVICE PHYSICS

Review of MOS Device Physics - MOS device review, Distributed Systems, Transmission lines, reflection coefficient, the wave equation, examples, Lossy transmission lines, Smith charts – plottingGamma, High Frequency Amplifier Design, Bandwidth estimation using open-circuit time constants, Bandwidth estimation, using short-circuit time constants, Rise time, delay and bandwidth, Zeros to enhance bandwidth, Shunt-series amplifiers, tuned amplifiers, Cascaded amplifiers

Basic structure of a MOS Transistor:-

The basic structure of a MOS transistor is given below. On a lightly doped substrate of silicon two islands of diffusion regions called as source and drain, of opposite polarity of that of the substrate, are created. Between these two regions, a thin insulating layer of silicon dioxide is formed and on top of this a conducting material made of poly-silicon or metal called gate is deposited.

Drain Current in Triode Region.

$$\mathrm{Id} = \mathrm{UnCoxW}/L\left[\left(\mathrm{Vgs}-\mathrm{Vth}\right)\mathrm{Vds}-1/2\,V^{2}\mathrm{ds}\right]$$

Maxwell's Equations for free space:-

$$\nabla . \varepsilon_o \vec{E} = \rho$$

$$\nabla . \mu_o \vec{H} = 0$$

$$\nabla \times \vec{E} = -\frac{\partial \mu_o \vec{H}}{\partial t}$$

$$\nabla \times \vec{H} = J + \frac{\partial \varepsilon_o \vec{E}}{\partial t}$$

Delay time of a CMOS:-

The delay time td is given by the expression

$$t_d = \left[\frac{L_n}{K_n W_n} + \frac{L_p}{K_p W_p}\right] \frac{C_L}{V_{dd}\left(1-\frac{V_t}{V_{dd}}\right)^2}$$

Where C is the load capacitance, V_{dd} is the supply voltage and V_t is the threshold voltages of the MOS transistors

Fall time:-

$$t_f \approx K_n \frac{C_L}{\beta_n V_{DD}}$$

Gauss Law:-

The total of the electric flux out of a closed surface is equal to the charge enclosed divided by the permittivity. The electric flux through an area is defined as the electric field multiplied by the area of the surface projected in a plane perpendicular to the field.

Characteristic Impedance:-

The characteristic impedance or surge impedance (usually written Z_0) of a uniform transmission line is the ratio of the amplitudes of voltage and current of a single wave propagating along the line; that is, a wave travelling in one direction in the absence of reflections in the other direction.

Smith chart:-

Smith chart is a graphical aid or monogram designed for electrical and electronics engineers specializing in radio frequency (RF) engineering to assist in solving problems with transmission lines and matching circuits.

Open circuit time constant:-

The open-circuit time constant method is an approximate analysis technique used in Electronic circuit design to determine the corner frequency of complex circuits. It also is known as the zero-value time constant technique.

Short circuit Time Constant:-

The short circuit time constant method to determine the low frequency band limit in case of Multi stage amplifier.

Propagation constant:-

The propagation constant of a sinusoidal electromagnetic wave is a measure of the change Undergone by the amplitude and phase of the wave as it propagates in a given direction. The Quantity being measured can be the voltage or current in a circuit or a field vector such as Electric field strength or flux density.

Cascaded Amplifiers:-

A cascade amplifier is any two-port network constructed from a series of amplifiers, where each amplifier sends its output to the input of the next amplifier in a daisy chain.

Tuned Amplifiers:-

A tuned amplifier is an electronic amplifier which includes band pass filtering components within the amplifier circuitry. They are widely used in all kinds of wireless applications. The response of tuned amplifier is maximum at resonant frequency and it falls sharply for frequencies below and above the resonant frequency

UNIT-III

NOISE

1. NOISE–INTRODUCTION:-

The sensitivity of communications systems is limited by noise. The broadest definition of noise as "everything except the desired signal" is most emphatically notwhat we will use herehowever, because it does not separate say, artificial noisesources (e.g .. 60-Hz power-line hum) from more fundamental (and therefore irreducible) sources of noise that we discuss in this.

That these fundamental noise sources exist was widely appreciated only after theinvention of the vacuum tube amplifier. when engineers finally had access to enoughgain to make these noise sources noticeable. It became obvious that simply cascadingmore amplifiers eventually produces no further improvement in sensitivity because a mysterious noise exists that is amplified along with the signal. In audio systems, thisnoise is recognizable as a continuous hiss while, in video, the noise manifests itselfas the characteristic "snow" of analog TV systems.

1.1 THERMAL NOISE:-

Johnson was the first to report careful measurements of noise in resistors, and his colleague Nyquist J explained them as a consequence of Brownian motion: thermally agitated charge carriers in a conductor constitute a randomly varying current that gives rise to a random voltage (via Ohm's law). In honor of these fellows. Thermal noise is often called Johnson noise or less frequently Nyquist noise.

Because the noise process is random. one cannot identify a specific value of voltage at a particular time (in fact. the amplitude has a Gaussian distribution) and the only recourse is to characterize the noise with statistical measures. such ax the mean square or root-mean-square values.

Because of the thermal origin, we would expect a dependence on the absolute temperature. It turns out that thermal noise power is exactly proportional to T (the astutemight even guess that it is proportional to kTJ. Specifically. a quantity called the *available noise power* is given by

$$P_{NA} = kT\Delta f, \quad (1)$$

where$k$ is Boltzmann's constant (about 1.38 x $10^{-23}$$J/K$). T is the absolute temperature in kelvins, and Δfthe noise bandwidth in hertz (equivalent brickwall bandwidth) over which the measurement is made.

We will clarify shortly what is meant by the terms "available noise power" and" noise bandwidth:' but for now simply note that the noise source is very broadband (infinitely so in fact in the simplified picture presented here"), so that the total noise power depends on the measurement bandwidth.

With Eqn. 1, we can compute that the available noise power over a 1-Hz bandwidth is about 4 x 10^{-21} W (or -174 dBm) s at room temperature. Further note that the constancy of the noise density implies that the thermal noise power is the same over any given *absolute* bandwidth. Therefore the noise power in the interval between 1 MHz and 2 MHz is the same as between 1 GHz and 1.001 GHz. Because of this constancy, thermal noise is often described as "white," by analogy with white light. However the analogy is not exact, since white light consists of constant energy per *wavelength* whereas white noise has constant energy per *hertz.*

The term "available noise power" is simply the maximum power that can be delivered to a load. Recall that the condition for maximum power transfer (for a resistive network) is equality of the load and source resistances. This suggests the use of the network shown in Figure 1.1 to compute the available noise power.

The model of the noisy resistor is enclosed within the dashed box and is here shown as a noise voltage generator in series with the resistor itself. The power delivered by this noisy resistor to another resistor of equal value is by definition theavailable noise power is

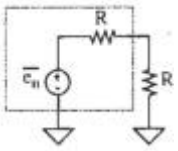


Figure 1.1 network for computing the thermal noise of resistor.

$$P_{NA} = kT\Delta f = \frac{\overline{e_n^2}}{4R}.$$

Where en- is the open-circuit rrns noise voltage generated by the resistor *R* over the bandwidth Δf at a given temperature, The mean-square open-circuit noise voltage is therefore

$$\overline{e_n^2} = 4kTR\Delta f.$$

The two noise models for a resistor are displayed in Figure 1.2. Note that the polarity indications on the noise voltage source and the arrow on the noise current.

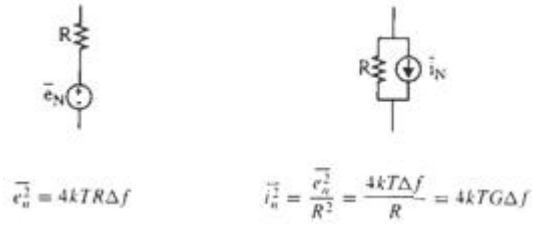


Figure 1. 2, Resistor thermal noise model

Source is simply references. They do not imply that the noise has a particular constant polarity (in fact, the noise has a zero mean).

The two noise models for a resistor are displayed in Figure 1.2. Note that thepolarity indications on the noise voltage source and the arrow on the noise current

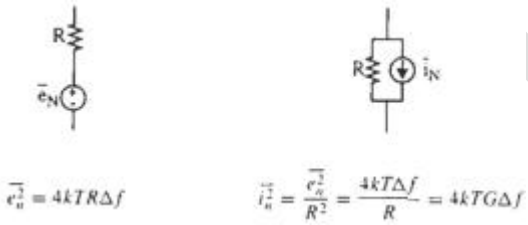


figure 1. 2, Resistor thermal noise model

Source is simply references. They do not imply that the noise has a particular constant polarity (in fact, the noise has a zero mean). Also note that, since the noise arises from the random thermal agitation of charge in the conductor, the only ways to reduce the noise of a given resistance are to keep the temperature as low as possible and to limit the bandwidth to the minimum useful value as well.

The distinction is made to underscore that the noise bandwidth Δf generally is not the same as the -3-dBbandwidth. Rather. the noise bandwidth is that of a perfect. brick wall (rectangular) filter that possesses the same area and peak value as the actual power gain-versus frequency characteristic of the system. including that of the measurement apparatus. The noise bandwidth is therefore

$$\Delta f \equiv \frac{1}{|H_{pk}|^2} \int_0^\infty |H(f)|^2 df.$$

Where *H*pk is the peak value of the magnitude of the filter voltage transfer function *H(f)*·

This normalization concept allows comparisons to be made on a standard basis. As a specific example, consider a single-pole *RC* low-pass filter. We know that the-3-dB bandwidth (in hertz) is simply 1/2piRC but the equivalent noise bandwidth is computed as

$$\Delta f \equiv \frac{1}{|1|^2}\int_0^\infty \left[\frac{1}{(2\pi fRC)^2+1}\right] df = \frac{1}{2\pi RC}\arctan 2\pi fRC\Big|_0^\infty$$
$$= \frac{\pi}{2} f_{3\,\mathrm{dB}} = \frac{1}{4RC}.$$

We see that a single-pole low-pass filter (LPF) has a noise bandwidth that is about1.57 times the -3-dB bandwidth. That the noise bandwidth exceeds the -3-dBbandwidth makes sense, since the lazy roll off of a single-pole filter allows spectral components of noise beyond the filter's -3-dB frequency to contribute significantly to the output energy of the filter.

This result follows from a more detailed derivation that takes into account the actual distribution of carrier energies modified by considerations related to the He is enberguncertainty principle." A more general expression for the thermal noise voltage is as follows

$$\overline{e_n^2} = \frac{h\omega R\Delta f}{\pi}\coth\left(\frac{h\omega}{4\pi kT}\right),$$

where h is Planck's constant, about 6.62×10^{-34} J-s.

1.2 .THERMAL NOISE IN MOSFETs:-

(a)Drain Current Noise:-Since FETs (both junction and MOS) are essentially voltage-controlled resistors,they exhibit thermal noise. In the triode region of operation particularly one would expect noise commensurate with the resistance value. Indeed detailed theoretical considerations" lead to the following expression for the drain current noise of FETs:

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f,$$

Where *gdD* is the drain-source conductance at zero *VDS*. The parameter *y* has a value of unity at zero *VDS* and, in long devices decreases toward a value of 2/3 in saturation. Note that the drain current noise at zero *VDS* is precisely that of an ordinary conductance of value *gdo*

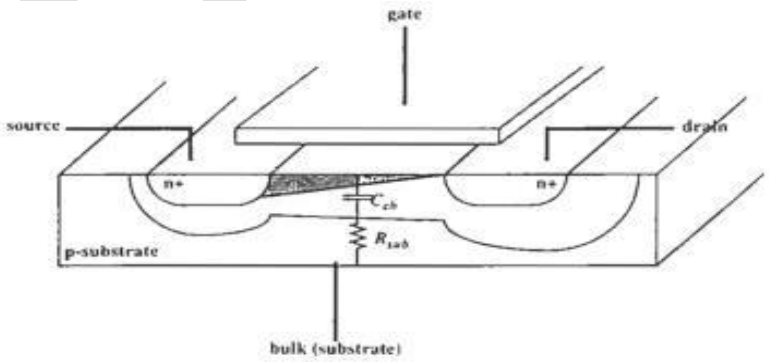


Figure1.3.Simplified illustration of substrate thermal noise

figure1.3 shows a simplified picture of how the thermal noise associated withthe substrate resistance can produce measurable effects at the main terminals of thedevice. At frequencies low enough that we may ignore *Ccb,* the thermal noise of *R*submodulates the potential of the back gate, contributing some noisy drain current is

$$\overline{i_{nd,sub}^2} = 4kTR_{sub}g_{mb}^2\Delta f. \tag{1}$$

Depending on bias conditions - and also on the magnitude of the effective substrate resistance and size of the back-gate transconductance - the noise generated by this mechanism (often called *epi noise,* whether or not epitaxial layers are present) may actually exceed the thermal noise contribution of the ordinary channel charge. In this regime layout strategies that reduce substrate resistance (e.g .. liberal use of substrate contacts tied together to ground) have a noticeable and beneficial effect on noise.

At frequencies well above the pole formed by *C cb* and R, ub however, the substrate thermal noise becomes unimportant, as is readily apparent from inspection of the physical structure and the corresponding frequency-dependent expression for the substrate noise contribution:

$$\overline{i_{nd,sub}^2} = \frac{4kTR_{sub}g_{mb}^2}{1+(\omega R_{sub}C_{cb})^2}\Delta f. \tag{2}$$

n addition to drain current noise, the thermal agitation of channel charge has another important consequence: gate noise. The fluctuating channel potential couplescapacitively into the gate terminal. leading to a noisy gate current (see Figure 1.4).Noisy gate current may also be produced by thermally noisy resistive gate material. Although this noise (whatever its source) is

Negligible at low frequencies. it can dominate at radio frequencies. Van der Ziel16 has shown that the gate noise may be expressed as

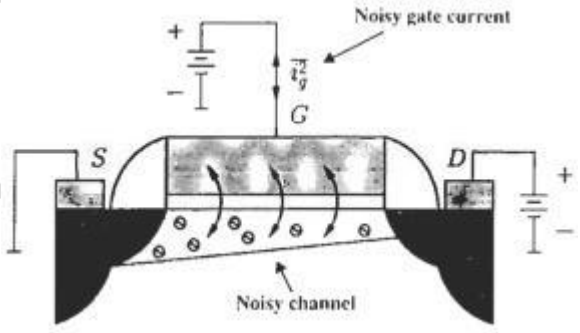


figure 1.4.Induced gate noise

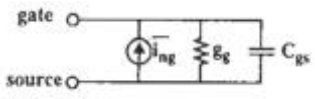


figure 1.5.gate noise circuit model

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f,$$

$$g_g = \frac{\omega^2 C_{gs}^2}{5g_{d0}}.$$

Van der Ziel gives a value of 4/3 (twice *y)* for the gate noise coefficient, 8. in long channel devices.

The circuit model for gate noise that follows directly from Eqn. 8 and Eqn. 9 is a conductance connected between gate and source, shunted by a noise current source(see Figure 11.5). The gate noise current clearly has a spectral density that is not constant. In fact it increases with frequency. so perhaps it ought to be called "blue noise" to continue the optical analogy.

For those who prefer not to analyze systems that have blue noise sources. It is possible to recast the model in a form with a noise *voltage* source that possesses a constant spectral density." To derive this alternative model first transforms the parallel *RC* network into an equivalent series *RC* network. If one assumes a reasonably high *Q*, then the capacitance stays roughly constant during the transformation. The parallel resistance becomes a series resistance whose value is

$$r_g = \frac{1}{g_g} \cdot \frac{1}{Q^2+1} \approx \frac{1}{g_g} \cdot \frac{1}{Q^2} = \frac{1}{5g_{d0}},$$

which is independent of frequencyFinally equate the short-circuit currents of the original network and the transformed version again with the assumption of high *Q*. The equivalent series noisevoltage source is then found to be

$$\overline{v_{ng}^2} = 4kT\delta r_g \Delta f,$$

which possesses a constant spectral density. Hence this alternative gate noise modelconsists of elements whose values arc independent of frequency; see Figure 1.6.

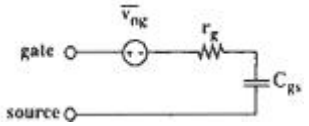


figure 1.6.alternate gate noise model

Although the noise behavior of long-channel devices is fairly well understood. Theprecise behavior of δ in the short-channel regime is unknown at present. Given thatboth the gate noise and drain noise share a common origin, however it is probablyreasonable as a crude approximation to assume that 8 continues to be about twice aslarge as *y*. Hence, just as *y* is typically 1-2 for short-channel NMOS devices δ maybe taken as 2-4.

1.3. SHOT NOISE:-

The fundamental basis for shot noise is the granular nature of the electronic charge, but how this granularity translates into noise is perhaps not as straightforward as one might think.

Two conditions must be satisfied for shot noise to occur. There must be a direct current flow and there must also be a potential barrier over which the charge carriers hop. The second condition tells us that ordinary, linear resistors do not generates hot noise despite the quantized nature of electronic charge.

The fact that charge comes in discrete bundles means that there arc discontinuous pulses of current every time electron hops energy barrier. It is the randomness of the arrival limes that gives rise to the whiteness of shot noise. If all the carrier shopped simultaneously shot noise would have a much more benign character. As we'll see in a later example.

We would expect the shot noise current to depend on the charge of the electron (since smaller charge would result in less lumpiness and therefore less noise). The total DC current Row (less

current also means fewer lumps) and the bandwidth (justas with thermal noise). In facts hot noise does depend on all of those quantities. As seen in the following equation:

$$\overline{i_n^2} = 2qI_{DC}\Delta f.$$

where**in**- is the rms noise current. *q*is the electronic charge (about 1.6 x 10^{-19} C), IDC is the DC current in amperes. and*/o"f*is again the noise bandwidth in hertz. Note that. like thermal noise. Shot noise (ideally) is white and has amplitude that possesses a Gaussian distribution. As a reference point. therrns current noise density is approximately 18 *pA/-/Hz* for a I-rnA value of *IDC.*

The requirement for a potential barrier implies that shot noise will only be associated with nonlinear devices, although not all nonlinear devices necessarily exhibit hot noise. For example whereas both the base and collector currents *are* sources of shot noise in a bipolar transistor because potential barriers arc definitely involved there (two junctions) only the DC gate leakage current of FETs (both MOS adjunction types of FETs) contributes shot noise. Because this gate current is normally very small, it is rarely a significant noise source (sadly. though. the same cannot be said of base current).

1.4. FLICKER NOISE:-

The most mysterious type of noise is flicker noise (also known as 1/f noise or pink-" noise). No universal mechanism for flicker noise has been identified, yet it is ubiquitous. Phenomena that have no obvious connection, such as cell membrane potentials, the earth's rotation rate, galactic radiation noise, and transistor noise all have fluctuations with a 1/f character.

As the term "1/f" suggests, the noise is characterized by a spectral density that increases, apparently without limit, as frequency decreases. Measurements have verified this behavior in electronic systems down to a small fraction of a micro hertz. One unfortunate implication of the increasing noise with decreasing frequency is the failure of averaging (band limiting) to improve measurement accuracy, since the noise power increases just as fast as the averaging interval. Because of the lack of a unifying theory, mathematical expressions for "1/f" noise invariably contain various empirical parameters (in contrast with the theoretical cleanliness of the equations for thermal and shot noise). as can be seen in the following equation is

$$\overline{N^2} = \frac{K}{f^n}\Delta f.$$

Here N is the rms noise (either voltage or current). K is an empirical parameter that is device- specific (and generally also bias-dependent), and n is an exponent that is usually (but not always) close to unity.

A question that often arises in connection with "1/f" noise concerns the infinity at DC implied by a "1/f" functional dependency. It's instructive to carry out a calculation with typical numbers to see why there is no problem, practically speaking.

First, let the parameter II have its commonly occurring value of unity. Then, integrate the density to find the total noise in a frequency band bounded by a lower frequency fl and an upper frequency fh.

$$\overline{N^2} = \int_{f_l}^{f_h} \frac{K}{f}\,df = K \ln\left(\frac{f_h}{f_l}\right).$$

This equation tells us that the total mean-square noise depends on the *log* of the frequency *ratio,* rather than simply on the frequency *difference* (as in thermal and shot noise). Hence, the mean- square value of l/f noise is the same for equal frequency *ratios;* there is thus a certain constant amount of mean-square noise per *decade* of frequency, say, or some specific amount of rms noise per *root* decade of frequency (units again for rrns quantities).

As a specific numerical example, suppose measurements on an amplifier reveal that its 1/f noise has a

density of 10*u*Vrrns per root decade. Thus, for the 16-decadefrequency interval below I Hz, the total 1/f noise would be just four times larger, or 40*u*Vrrns. Recognize that 16decades below 1 hertz is equal to one cycle about every 320 million years." and you have to concede that "DC" infinities are simply not a practical problem. The resolution of the apparent paradox thus lies in recognizing that true DC implies an infinitely long observation interval, and that humans' and the electronic age have been around for only a finite time. For any finite observation interval, the infinities simply don't materialize.

1.4.2. FLICKER NOISE IN RESISTORS:-

Flicker noise also shows up in ordinary resistors, where it is often called "excess noise," since this noise is in addition to what is expected from thermal noise considerations. It is found that a resistor exhibits 1/f noise only when there is DC current flowing through it, with the noise increasing with the current. In the discrete world, garden-variety carbon composition resistors are the most conspicuous offenders, while metal- film and wire wound resistors exhibit the smallest amounts of excess noise.

The current-dependent excess noise of carbon composition resistors has been explained by some as the result of the random formation and extinction of "micro-arcs" among neighboring carbon granules. *"carbon film "*resistors, which are made differently, exhibit much less excess noise than do carbon composition types. Whatever the explanation, it is certainly true that excess noise increases with the DC bias, so one should minimize the DC drop across a resistor.

The following approximate expression shows explicitly the dependency of thisnoise on various parameters is

$$\overline{e_n^2} = \frac{K}{f} \cdot \frac{R_\square^2}{A} \cdot V^2 \Delta f.$$

Where *A* is the area of the resistorRis the sheet resistivity, *V* is the voltage acrossthe resistor, and *K* is a material-specific parameter. For diffused and ion-implantedresistors, *K* has a value of roughly 5 x 10^{-28} S^2-m^2, whereas for thick-film resistors (not normally available in CMOS processes), *K* is about an order of magnitude larger.

1.4.2.FLICKER NOISE IN MOSFETS:-

In electronic devices 1/f noise arises from a number of different mechanisms and is most prominent in devices that are sensitive to surface phenomena. Hence, MOSFETs exhibit significantly more 1/f noise than do bipolar devices. One means of comparison is to specify a "corner frequency." where the 1/f and thermal or shot noise components are equal, All other things held equal. a lower 1/f corner implies less total noise. It is relatively trivial to build bipolar devices whose 1/f corners are below tens or hundreds of hertz, and many MOS devices routinely exhibit 1/f corners of tens of kilohertz to a megahertz or more.

Charge trapping phenomena are usually invoked to explain 1/f noise in transistors. Some types of defects and certain impurities (most plentiful at a surface orinterface of some kind) can randomly trap and release charge. The trapping times are distributed in a way that can lead to a 1/f noise spectrum in both MOS and bipolar transistors. Since MOSFETs are surface devices (at least in the way that they are conventionally fabricated), they exhibit this type of noise to a much greater degree than bipolar transistors (which are bulk devices). Larger MOSFETs exhibit less 1/f noise because their larger gate capacitance smooths the fluctuations in channel charge. Hence, if good 1/f noise performance is to be obtained from MOSFETs, thelargest practical device sizes must be used (for a given *gm)*

The mean-square 1/fdrain noise current is given by

$$\overline{i_n^2} = \frac{K}{f} \cdot \frac{g_m^2}{WLC_{ox}^2} \cdot \Delta f \approx \frac{K}{f} \cdot \omega_T^2 \cdot A \cdot \Delta f.$$

Where *A* is the area of the gate (= *WL)* and *K* is a device-specific constant. Thus for a fixed trans conductance. a larger gate area and a thinner dielectric reduce this noise term

For PMOS devices, *K* is typically about 10^{-28} C^2/m^2 whereas for NMOS devices it is about 50 times larger." One should keep in mind that these constants vary considerably from process to process, and even from run to run, so the values of *K* given here should be treated as crude estimates. In particular the superior *1/f* performance of PMOS devices may be a temporary situation, as it is due to the use of buried channels that may cease to be widely used in the future.

1.4.3. FLICKER NOISE IN JUNCTIONS:-

Forward-biased junctions also exhibit *1/f*noise. The noise is proportional to the biascurrent and inversely proportional to the junction area

$$\overline{i_f^2} = \frac{K}{f} \cdot \frac{I}{A_j} \cdot \Delta f.$$

where the constant K typically has a value of around 10^{-25}A.m^2. Once again however, considerable variation from process to process is not uncommon."Flicker noise in bipolar transistors is attributed entirely to the base-emitter junction (since it is the only one in forward bias). It has been established experimentally that only the base current exhibits *1/f* noise.

1.5. CLASSICAL TWO-PORT NOISE THEORY (NOISE FIGURE):-

1.5.1 NOISE FACTOR:-

A useful measure of the noise performance of a system is the noise factor. Usually denoted *F*. To define it and understand why it is useful, consider a noisy (but linear) two-port driven by a source that has an admittance *Y* and an equivalent shunt noise current as shown in Figure 1.1.

If we are concerned only with overall input-output behavior, it is an unnecessary complication to keep track of all of the internal noise sources. Fortunately, the net effect of all of those sources can be represented by just one pair of external sources, a noise voltage and a noise current. This huge simplification allows rapid evaluation ofhow the source admittance affects the overall noise performance. As a consequence,we can identify the criteria one must satisfy for optimum noise performance.

The noise factor is defined a $$F = \frac{\text{total output noise power}}{\text{output noise due to input source}}, \quad (1)$$

where by convention the source is at a temperature of 290 K. The noise factor isa measure of the degradation in signal-to-noise ratio that a system introduces. Thelarger the degradation the larger the noise factor. If a system adds no noise of itsown then the total output noise is due entirely to the source and the noise factor istherefore unity.

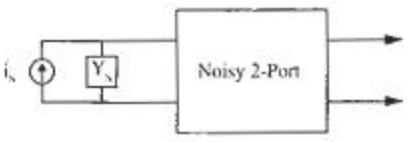


figure1.1 noisy two-port driven by noise source

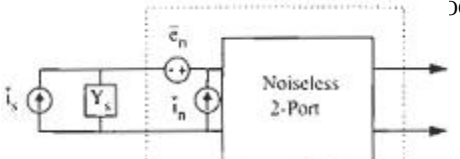


figure 1.2 equivalent noise model

In the model of Figure 11.8, all of the noise appears as inputs to the noiseless network, so we may compute the noise figure there. A calculation based directly on equation (1) requires the computation of the total power due to all of the sources and dividing that result by the power due to the input source. An equivalent (and simpler)method is to compute the total short-circuit mean-square noise current and then divide that total by the short-circuit mean-square noise current due to the input source. This alternative method is equivalent because the individual power contributions areproportional to the short-circuit mean-square current. with a proportionality constant(which involves the current division ratio between the source and two- port) that isthe same for all of the terms.

In carrying out this computation, one generally encounters the problem of combining noise sources that have varying degrees of correlation with one another. In thespecial case of zero correlation the individual *powers* superpose. For example, if weassume as seems reasonable. That the noise powers of the source and of the two-portare uncorrelated, then the expression for noise figure becomes

$$F = \frac{\overline{i_s^2} + \overline{|i_n + Y_s e_n|^2}}{\overline{i_s^2}}. \qquad (2)$$

Although we have assumed that the noise of the source is uncorrelated withthe two equivalent noise generators of the two-port equation 2 does *not*assume thatthe two-port 's generators are also uncorrelated with each other.

In order to accommodate the possibility of correlations between *en*and *In* express*In*as the sum of two components. One *ic*is correlated with *en* and the other*i*u is given by

$$i_n = i_c + i_u. \qquad (3)$$

Since *ic*is correlated with *en*it may be treated as proportional to it through a constant whose dimensions are those of admittance is given

$$i_c = Y_c e_n; \qquad (4)$$

the constant *Yc*is known as the *correlation admittance*

Submit the equations 2,3.4 in equation 1,the noise factor is then

$$F = \frac{\overline{i_s^2} + \overline{|i_u + (Y_c + Y_s)e_n|^2}}{i_s^2} = 1 + \frac{\overline{i_n^2} + |Y_c + Y_s|^2\overline{e_n^2}}{i_s^2}. \qquad (5)$$

The expression in Equation 5 contains three independent noise sources, each of whichmay be treated us thermal noise produced by an equivalent resistance or conductance(whether or not such a resistance or conductance actually is the source of the noise)

$$R_n \equiv \frac{\overline{e_n^2}}{4kT\Delta f},$$

$$G_n \equiv \frac{\overline{i_u^2}}{4kT\Delta f},$$

$$G_s \equiv \frac{\overline{i_s^2}}{4kT\Delta f}.$$

Using these equivalences, the expression for noise factor can be written purely interms of impedances and admittances is

$$F = 1 + \frac{G_u + |Y_c + Y_s|^2 R_n}{G_s}$$
$$= 1 + \frac{G_u + [(G_c + G_s)^2 + (B_c + B_s)^2]R_n}{G_s}. \qquad (6)$$

where we have explicitly decomposed each admittance into a sum of a conductanceC and a susceptance*B.*

1.5.2 OPTIMUM SOURCE ADMITTANCE:-

Once a given two-port's noise has been characterized with its four noise parameters *(Gr,Bc,Rn,Bs)*Eqn. 6 allows us to identify the general conditions forminimizing the noise factor. Taking the first derivative with respect to the source admittance and selling it equal to zero yields

$$B_s = -B_c = B_{opt},$$
$$G_s = \sqrt{\frac{G_u}{R_n} + G_c^2} = G_{opt}. \qquad (7),(8)$$

Hence.to minimize the noise factor. the source susceptance should be made equal tothe inverse of the correlation susceptance. while the source conductance should beset equal to the value in Eqn. 8

The noise factor corresponding to this choice is found by direct substitution ofEqn. 7 and Eqn. 8 into Eqn. 6

$$F_{min} = 1 + 2R_n[G_{opt} + G_c] = 1 + 2R_n\left[\sqrt{\frac{G_u}{R_n} + G_c^2} + G_c\right]. \quad (9)$$

We may also express the noise factor in terms of Fmin and the source admittance is

$$F = F_{min} + \frac{R_n}{G_s}[(G_s - G_{opt})^2 + (B_s - B_{opt})^2]. \quad (10)$$

Thus contours of constant noise factor are non-overlapping circles in the admittanceplane.

The ratio *Rn/Gs,* appears as a multiplier in front of the second term of Eqn. 10. Fora fixed source conductance *Rn* tells us something about the relative sensitivity of thenoise figure to departures from the optimum conditions. A large *Rn* implies a high sensitivity; circuits or devices with high *R"* obligate us to work harder to identifyachieve, and maintain optimum conditions. We will shortly see that operation at lowbias currents is associated with large *R",* in keeping with the general intuition thatachieving high performance only gets more difficult as the power budget tightens.

It is important to recognize thatalthough minimizing the noise factor has something of the flavor of maximizing power transfer. the source admittances leading tothese conditions are generally not the same - as is apparent by inspection of Eqn. 7 and Eqn. 8. For example there is no reason to expect the correlation susceptance to equal the input susceptance (except by coincidence). As a consequence one must generally accept less than maximum power gain if noise performance is to be optimized, and vice versa.

1.5.3 LIMITATIONS OF CLASSICALNOISE OPTIMIZATION:-

The classical theory just presented implicitly assumes that one is given a device with particular fixed characteristics and defines the source admittance that will yield theminimum noise figure given such a device.

Although one starts with fixed devices indiscrete RF design. The freedom to choose device dimensions in IC realizations points out a serious shortcoming of the classical approach: There are no specific guidelines about what device size will minimize noise. Further more power consumption is frequently a parameter of great interest (even an obsessive one in many portable applications) but power is simply not considered at all in classical noise optimization. We will return to these themes in great detail in the chapter on LNA design. But for now simply be aware of the incompleteness of the classical approach.

1.5.4 NOISE FIGURE AND NOISE TEMPERATURE:-

In addition to noise factor, other figures of merit that often crop up in the literature are noise figure and noise temperature. The noise figure (NF) is simply the noise factor expressed in decibels.

Noise temperature, *TN*, is an alternative way of expressing the effect of an amplifier's noise contribution and is defined as the increase in temperature required of the source resistance for it to account for all of the output noise at the reference temperature *Tref* (which is 290 K). It is related to the noise factor as follows

$$F = 1 + \frac{T_N}{T_{\text{ref}}} \implies T_N = T_{\text{ref}} \cdot (F - 1).$$

An amplifier that adds no noise of its own has a noise temperature of 0 K.Noise temperature is particularly useful for describing the performance of cascaded amplifiers and those whose noise factor isquite close to unity (or whose noise figure is very close to 0 dB), since the noise temperature offers a higher-resolution description of noise performance in such cases Noise figures in the range of 2-3 dB are generallyconsidered very good, with values around or below 1 dB considered outstanding.

2. LNA DESIGN:-

2.1 DERIVATION OF INTRINSIC MOSFETTWO-PORT NOISE PARAMETERS:-

The MOSFET noise model consists of two sources. The mean-square drain current noise is

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f; \qquad (1)$$

Gate current noise is

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f.$$

$$g_g = \frac{\omega^2 C_{gs}^2}{5g_{d0}}.$$

Further recall that the gate noise is correlated with the drain noise, with a correlationcoefficient defined formally as

$$c \equiv \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}}}.$$

then the reference direction for the gate noise is from the source 10 gate (as in Fig• ure 1.5)and that for the drain noise is from the drain to the source (as in Figure 1.4), then the long-channel value of *c* is theoretically - jO.395.

we will neglect the thermal noise due to the resistive gate materialalthough this source of noise can actually dominate the gate noise when operating device well below *fT*,where nonquasistatic effects (such as induced gate noise) willbe less prominent.' We will also neglect *Cgd* to simplify the derivation. While theachievable noise figure is little affected by *Cgd* the input impedance can be a strong function of *Cgd* and this effect must be taken into account when designing the inputmatching network.

To derive the four equivalent two-port noise parameters repeated here for convenience

$$R_n \equiv \frac{\overline{e_n^2}}{4kT\Delta f}.$$

$$G_u \equiv \frac{\overline{i_u^2}}{4kT\Delta f}.$$

$$Y_c \equiv \frac{i_c}{e_n} = G_c + jB_c.$$

we first reflect the two fundamental MOSFET noise sources back to the input port as a different pair of equivalent input generators (one voltage and one current source).The equivalent input noise voltage generator accounts for the output noise observed when the input port is short- circuited (incrementally speaking). To determineits value, reflect the drain current noise back to the input as a noise voltage and recognize that the ratio of these quantities is simply gm,. Thus,

$$\overline{e_n^2} = \frac{\overline{i_{nd}^2}}{g_m^2} = \frac{4kT\gamma g_{d0}\Delta f}{g_m^2}.$$

From which it is apparent that the equivalent input noise voltage is completely correlated, and in phase, with the drain current noise. Thus, we can immediately determinethat

$$R_n \cong \frac{\overline{e_n^2}}{4kT\Delta f} = \frac{\gamma g_{d0}}{g_m^2}.$$

he equivalent input noise voltage generator by itself does not fully account for thedrain current noise. however because a noisy drain current also flows even when theinput is open-circuited and induced gate current noise is ignored. Under this opencircuit condition. dividing the drain current noise by the transconductance yields anequivalent input voltage which. when multiplied in turn by the input admittance,gives us the value of an equivalent input current noise that completes the modelingof *imd*is given by

$$\overline{i_{n1}^2} = \frac{\overline{i_{nd}^2}(j\omega C_{gs})^2}{g_m^2} = \frac{4kT\gamma g_{d0}\Delta f(j\omega C_{gs})^2}{g_m^2} = \overline{e_n^2}(j\omega C_{gs})^2. \quad (2)$$

In this step of the derivation, we have assumed that the input admittance of a MOSFETis purely capacitive. This assumption is a good approximation for frequencies well below *Wr*, if appropriate high-frequency layout practice is observed to minimize gateresistance. Given this assumption, Eqn. 2 shows that the input noise current in1 isin quadrature and therefore completely correlated, with the equivalent input noise voltage *en*

The total equivalent input current noise is the sum of the reflected drain noisecontribution of Eqn. 2 and the induced gate current noise. The induced gate noisecurrent itself consists of two terms. One which we'll denote *ingc* is fully correlatedwith the drain current noise, while the other, *ingu*is completely uncorrelated with thedrain current noise. Hence we may express the correlation admittance as follows:

$$Y_c = \frac{i_{n1} + i_{ngc}}{e_n} = j\omega C_{gs} + \frac{i_{ngc}}{e_n}$$

$$= j\omega C_{gs} + \frac{g_m}{i_{nd}} \cdot i_{ngc} = j\omega C_{gs} + g_m \cdot \frac{i_{ngc}}{i_{nd}} \quad (3)$$

To express Yc in a more useful form we need to incorporate the gate noise correlationfactor explicitly. To do so, we must manipulate the last term ofEqn. 3 in waysthat will initially appear mysterious. First, we express it in terms of cross-correlationsby multiplying both numerator and denominator by the conjugate of the drain noise current and then averaging each one is

$$g_m \cdot \frac{i_{ngc}}{i_{nd}} = g_m \cdot \frac{\overline{i_{ngc} \cdot i_{nd}^*}}{\overline{i_{nd} \cdot i_{nd}^*}} = g_m \cdot \frac{\overline{i_{ngc} \cdot i_{nd}^*}}{\overline{i_{nd}^2}} = g_m \cdot \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\overline{i_{nd}^2}}. \tag{4}$$

The last equality, in which *ing* replaces *ings* is valid because the uncorrellatcd portionof the gate noise current necessarily contributes nothing to the cross-correlation.Using Eqn. 3, we may write the correlation admittance as

$$Y_c = j\omega C_{gs} + g_m \cdot \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\overline{i_{nd}^2}} = j\omega C_{gs} + g_m \cdot \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{nd}^2}}\sqrt{\overline{i_{nd}^2}}} \sqrt{\frac{\overline{i_{ng}^2}}{\overline{i_{ng}^2}}}, \tag{5}$$

which, in turn, may be expressed as

$$Y_c = j\omega C_{gs} + g_m \cdot \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}}} \sqrt{\frac{\overline{i_{ng}^2}}{\overline{i_{nd}^2}}} = j\omega C_{gs} + g_m \cdot c \sqrt{\frac{\overline{i_{ng}^2}}{\overline{i_{nd}^2}}}, \tag{6}$$

which explains all of the maneuvering, since the correlation coefficient has finallymade an explicit appearance.Substituting for the term under the radical yields

$$Y_c = j\omega C_{gs} + g_m \cdot c \sqrt{\frac{\delta \omega^2 C_{gs}^2}{5\gamma g_{d0}^2}} = j\omega C_{gs} + \frac{g_m}{g_{d0}} \cdot c \sqrt{\frac{\delta}{5\gamma}} \cdot \omega C_{gs}. \tag{7}$$

If we assume that e continues to be purely imaginary, even in the short-channelregime. We finally obtain a useful expression for the correlation admittance is

$$Y_c = j\omega C_{gs} - j\omega C_{gs} \frac{g_m}{g_{d0}} \cdot |c| \sqrt{\frac{\delta}{5\gamma}} = j\omega C_{gs} \left(1 - \alpha |c| \sqrt{\frac{\delta}{5\gamma}}\right). \tag{8}$$

where we have used the substitution

$$\alpha = \frac{g_m}{g_{d0}}. \tag{9}$$

Since α is unity for long-channel devices and progressively decreases as channellengths shrink. it is one measure of the departure from the long-channel regime.

We see from Eqn.8 that the correlation admittance is purely imaginary. so that G; =0More significant, however, is the fact that Yc does not equal the admittanceof Cgs although it is some multiple of it. Hence one cannot maximize power transferand minimize noise figure simultaneously. To investigate further the importantimplications of this impossibility though we need to derive the last remaining noise parameter Gh.

Using the definition of the correlation coefficient, we may express the induced gate noise as follows on

$$\overline{i_{ng}^2} = \overline{(i_{ngc} + i_{ngu})^2} = 4kT\Delta f\delta g_g|c|^2 + 4kT\Delta f\delta g_g(1 - |c|^2). \tag{10}$$

The very last term in Eqn. 10 is the uncorrelated portion of the gate noise current, sothat, finally,

$$G_u \equiv \frac{\overline{i_u^2}}{4kT\Delta f} = \frac{4kT\Delta f\delta g_g(1 - |c|^2)}{4kT\Delta f} = \frac{\delta\omega^2 C_{gs}^2(1 - |c|^2)}{5g_{d0}}. \tag{11}$$

we can determine both the source impedance that minimizes the noise figureas well as the minimum noise figure itself is

$$B_{opt} = -B_c = -\omega C_{gs}\left(1 - \alpha|c|\sqrt{\frac{\delta}{5\gamma}}\right). \tag{12}$$

From Eqn. 12 we see that the optimum source susceptance is essentially inductivein character except that it has the wrong frequency behavior. Hence achieving abroadband noise match is fundamentally difficult.Continuing the real part of the optimum source admittance is

$$G_{opt} = \sqrt{\frac{G_u}{R_n} + G_c^2} = \alpha\omega C_{gs}\sqrt{\frac{\delta}{5\gamma}(1 - |c|^2)}. \quad (13)$$

and the minimum noise figure is given by

$$F_{min} = 1 + 2R_n[G_{opt} + G_c] \approx 1 + \frac{2}{\sqrt{5}}\frac{\omega}{\omega_T}\sqrt{\gamma\delta(1 - |c|^2)}. \quad (14)$$

In Eqn, 14 the approximation is exact if one treats Wτ as simply the ratio of gmto CS" Note that if there were no gate current noise the minimum noise figure would be 0 dB. That unrealistic prediction alone should be enough to suspect that gate noise must indeed exist. Also note that, in principle, increasing the correlation between drain and gate current noise would improve noise figure, although correlation coefficients unrealistically near unity would be required to effect large reductions in noise figure.

Generalizing, the noise parameters for a device of width Ware

$$G_c = \frac{W}{W_0}G_{c0},$$
$$B_c = \frac{W}{W_0}B_{c0},$$
$$G_u = \frac{W}{W_0}G_{u0};$$
$$R_n = \frac{W_0}{W}R_{n0}. \quad (15)$$

2.2. LNA TOPOLOGIES: POWER MATCHVERSUS NOISE MATCH:-

One straightforward approach to providing a reasonably broadband 50-Q termination is simply to put a 50-Q resistor across the input terminals of a common-source amplifier; this is shown in Figure 1.

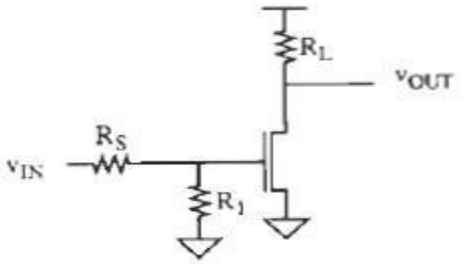


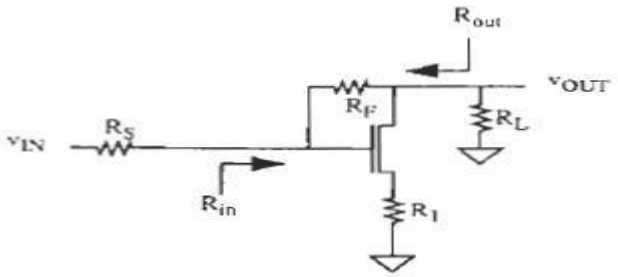
Figure1.1. Common-source amplifier with shunt input resistor (biasing not shown)

Unfortunately the resistor Rl adds thermal noise of its own and so attenuates the signal (by a factor of 2) ahead of the transistor. The combination of these two effects generally produces unacceptably high noise figures. More formally it is straight forward to establish the following lower bound on the noise figure of this circuit is

$$F \geq 2 + \frac{4\gamma}{\alpha} \cdot \frac{1}{g_m R}, \tag{1}$$

Where Rs = R1 = R. This bound applies only in the low-frequency limit and ignores gate current noise altogether. Naturally, the noise figure is worse at higher frequencies and when gate noise is taken into account.

The shunt-series amplifier, is another circuit that provide broadband real input Impedance Since it does not reduce the signal with a noisy attenuators before amplifying, we expect its noise figure to be substantially better than that of the circuit of Figure 1.2.



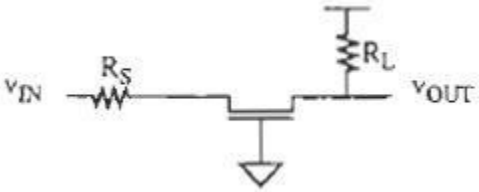


Figure 1.2 shunt-series amplifier as bias is not shown figure 1.3 common gate amplifier

The amplifier sketched in Figure 1.3 suffers from fewer problems than the previous circuit, yet the resistive feedback network continues to generate thermal noise of its own and also fails to present to the transistor an impedance that equals Zopts at all frequencies (.perhaps at any frequency). As a consequence, the overall amplifier's noise figure while usually much better than that of Figure 1.2 still generally exceeds the device F min by a considerable amount (typically a few decibels), Nonetheless the broad band capability of this circuit is frequently enough of a compensating advantage that the shunt-series amplifier is found in many LNA applications even though Its noise figure is not the minimum possible.

Another method for realizing a resistive input impedance is to use a common-sate configuration, Since the resistance looking into the source terminal is 1/gm a proper selection of device size and bias current can provide the desired 50ohm resistance' see Figure 1.4.

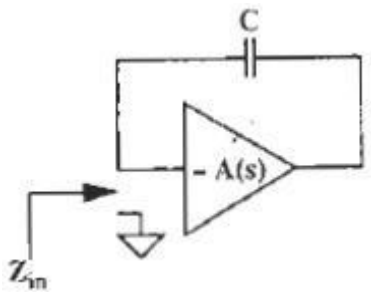


Figure 1.4 impedance transformation model

lower bound on the noise figure of the common-gate amplifier (again. at low frequencies and neglecting gate current noise) with the circuit shown in figure 1.2.

$$F \geq 1 + \gamma/\alpha. \quad (2)$$

This bound assumes that (he resistance looking into the source terminal is adjusted toequal the source resistance. and is about 2.2 dB in the long-channel limit and perhapsas high as 4.8 dB for short devices ($\gamma/\alpha.$ = 2). The noise figure will be significantlyworse at high frequencies and when gate current noise is taken into account.

The essential features of this mechanism may be understood by examining the abstractionof Figure 1.4. The amplifier is ideal in all respects except for a frequencydependentgain A(s),

Theinput terminal is analogous to the gate (or grid) terminaland the amplifier output is connected to the bottom plate of the gate capacitance. Theinput impedance of this circuit is

$$Z_{in} = \frac{1}{sC[1 + A(s)]} \tag{3}$$

Now let A(s) have gain and phase shift is

$$A(s) = A_0 e^{-j\phi};$$

then

$$Z_{in} = \frac{1}{j\omega C[1 + A_0 e^{-j\phi}]} = \frac{1}{j\omega C[1 + A_0(\cos\phi - j\sin\phi)]}.$$

Collecting terms and focusing on the dcnominator, we obtain

$$Y_{in} = j\omega C[1 + A_0\cos\phi] + A_0\omega C\sin\phi. \tag{4,5,6}$$

from which it is apparent that the input admittance indeed possesses a real part whosevalue depends on the phase lag ¢. With zero phase lug.the admittance is purely capacitive.as anticipated from quasistatic analyses. If the more realistic scenario of anonzero phase lag is considered. the equivalent shunt conductance is seen to increasewith frequency. Perhaps it is no surprise that measurements show that the phase lagitself grows with frequency, and the equivalent shunt conductance typically increasesas the square of frequency. to a good approximation.

Transit-lime effects also cause a resistive component of input impedance in vacuum tubes where the phenomenon was first observed. Because of the finite velocity of charge. Then a real term is an unavoidable reality in charge-controlled devicessuch as vacuum tubes and FETs.In the context of low-noise amplifiers we actually seek to enhance this effect, forit can be used to create a resistive input impedance without the noise of real resistors.From the foregoing. it is clear that one possible method is to modify the device (e.g ..elongate it) in order to enhance transit time effects directly. However, this approachhas the undesirable side effect of degrading high- frequency gain.

A better method is to employ inductive source degeneration. With such an inductance current flow lags behind an applied gate voltage behavior that is qualitativelysimilar to the mechanism described. An important advantage of this met bod is thatone then has control over the value of to be real part of the impedance through choiceof inductance, as is clear from computing the input resistance of the circuit shown in 1.5

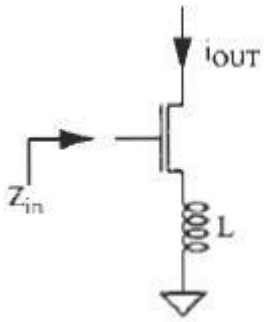


Figure 1.5.Induclivelydegeneratedcommon-source amplifier.

$$Z_{in} = sL + \frac{1}{sC_{gs}} + \frac{g_m}{C_{gs}}L \approx sL + \frac{1}{sC_{gs}} + \omega_T L. \tag{7}$$

Hence the input impedance is that of a series RLC network with a resistive ten that is directly proportional to the inductance value. More generally an arbitrary source degeneration impedance Z is modified by afactor equal to [B(jw+1)] when reflected to the gate circuit where B(jw) is the current gain is

$$\beta(j\omega) = \frac{\omega_T}{j\omega}. \tag{8}$$

The current gain magnitude goes to unity at WT (as it should), and it has a capacitivephase angle because of Cgs. Hence, for the general case.

$$Z_{in}(j\omega) = \frac{1}{j\omega C_{gs}} + [\beta(j\omega) + 1]Z = \frac{1}{j\omega C_{gs}} + Z + \left[\frac{\omega_T}{j\omega}\right]Z. \tag{9}$$

Note that capacitive degeneration contributes a negative resistance to the input impedance. Hence any source-to-substrate capacitance offsets the positive resistancefrom inductive degeneration. It is important to take this effect into account in anyactual design.

Whatever the value of this resistive term it is important to emphasize that itdoesnot bring with it the thermal noise of an ordinary resistor because a pure reactanceis noiseless. We may therefore exploit this property to provide specified input impedancewithout degrading the noise performance of the amplifier.

The form of Eqn. 7 clearly shows that the input impedance is purely resistiveat only one frequency (at resonance) however so this method can provide only a narrowband impedance match. Fortunately, there are numerous instances whennarrowband operation is not only acceptable but actually desirable. So inductive degenerationis certainly a valuable technique. The LNA topology we will examine forthe rest of this chapter is therefore as shown in Figure 1.6

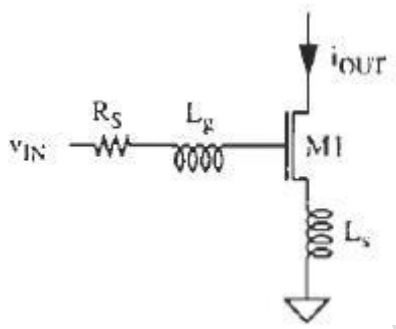


Figure 1.6Narrowband LNAwith inductive sourcedegeneration.

The inductance Ls

is chosen to provide the desired input resistance (equal to Rsthe source resistance). Since the input impedance is purely resistive only at resonance.an additional degree of freedom, provided by inductance Lg. is needed toguarantee this condition. Now, at resonance, the gate-to-source voltage is Q timesas large as the input voltage. The overall stage transconductanceGm, under this conditionis therefore

$$G_m = g_{m1} Q_{in} = \frac{g_{m1}}{\omega_0 C_{gs}(R_s + \omega_T L_s)} = \frac{\omega_T}{2\omega_0 R_s}, \quad (10)$$

where we have used the approximation that WT is the ratio of gml to Cgs

Let us continue to neglect Cgd,gmb and Csb. Then it is straightforward to showthat the input impedance of the circuit in Figure 1.5 is

$$Z_{in}(j\omega) = \frac{1}{j\omega C_{gs}} + j\omega L + g_m \frac{L}{C_{gs}} \left(\frac{r_0}{r_0 + j\omega L + Z_L} \right). \quad (11)$$

where ZL is the impedance attached to the drain. Comparing this result with our previous equation, we see that finite output resistance alters the third term in the impedanceequation.

2.3.POWER CONSTRAINED NOISE OPTIMIZATION:-

To develop the desired noise optimization technique, we must express noise figurein a way that takes power consumption explicitly into account. Given a specifiedbound on power consumption. the method should then yield the optimum device thatminimizes noise. Although the detailed derivations are somewhat complex. the endresults are remarkably simple. Readers interested primarily in applying the methodare invited to skip to the end of this section.

We start with the general expression for noise figure as given by classical noisetheory is

$$F = F_{min} + \frac{R_n}{G_s}[(G_s - G_{opt})^2 + (B_s - B_{opt})^2]. \quad (1)$$

The goal here is ultimately to reformulate the expression for noise figure in terms ofpower consumption. Once we derive such an equation,we'll minimize it subject tothe constraint of fixed power and then solve for the width of the transistor that correspondsto this optimum condition.

To simplify the development let us assume that the source susceptance Bs is chosen sufficiently close to Bopt that we may neglect the difference between the two. Wewill justify this step formally at a later time. Given this assumption, the expressionfor noise figure reduces to

$$F = F_{min} + \frac{R_n}{G_s}[(G_s - G_{opt})^2 + (B_s - B_{opt})^2]. \quad (2)$$

The goal here is ultimately to reformulate the expression for noise figure in terms ofpower consumption. Once we derive such an equation. we'll minimize it subject tothe constraint of fixed power and then solve for the width of the transistor that correspondsto this optimum condition.

To simplify the development. let us assume that the source susceptance Bs is chosensufficiently close to Bopt that we may neglect the difference between the two. Wewill justify this step formally at a later time. Given this assumption, the expressionfor noise figure reduces to

$$F = F_{\min} + \frac{R_n}{G_s}(G_s - G_{\text{opt}})^2. \tag{3}$$

Next rearrange the expression for Gopt to define a parameter with the dimensionsof a quality factor. This maneuver will help reduce clutter in the equations to come

$$\frac{G_{\text{opt}}}{\omega C_{gs}} = \alpha\sqrt{\frac{\delta}{5\gamma}(1 - |c|^2)} = Q_{\text{opt}}. \tag{4}$$

To accommodate the possibility of operation with source conductance's other than Gopt we also define a similar Q in which Gopt is replaced by Gs the actual source conductance is

$$Q_s \equiv \frac{1}{\omega C_{gs} R_s}. \tag{5}$$

Now re-express and the noise parameters of

$$\begin{aligned} F &= F_{\min} + \frac{(\gamma/\alpha)(1/g_m)}{Q_s \omega C_{gs}}(Q_s \omega C_{gs} - Q_{\text{opt}} \omega C_{gs})^2 \\ &= F_{\min} + \left[\frac{\gamma}{\alpha g_m R_s}\right]\left[1 - \frac{Q_{\text{opt}}}{Q_s}\right]^2. \end{aligned} \tag{6}$$

The parameters α ,gm, Qopt and Qs in Eqn. 6 are linked to power dissipation.We need to make the linkage explicit. However, and rewrite those terms directly interms of power. To do so, first recall that a simple expression for the drain current is

$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{gs} - V_t)[(V_{gs} - V_t) \parallel (LE_{sat})], \tag{7}$$

which may be rewritten as

$$I_D = WLC_{ox}v_{sat}E_{sat}\frac{\rho^2}{1+\rho},$$

where

$$v_{sat} = \frac{\mu_n}{2}E_{sat}$$

and

$$\rho = \frac{V_{gs} - V_t}{LE_{sat}} = \frac{V_{od}}{LE_{sat}}. \tag{8,9,10}$$

Given Eqn. 8 the power dissipation can be written as follows as

$$P_D = V_{DD}I_D = V_{DD}WLC_{ox}v_{sat}E_{sat}\frac{\rho^2}{1+\rho}. \tag{11}$$

Furthermore the transconductance gm can be found by differentiating Eqn, 9. After a little rearrangement this may be expressed as

$$g_m = \left[\frac{1+\rho/2}{(1+\rho)^2}\right]\left[\mu_n C_{ox}\frac{W}{L}V_{od}\right] = \alpha\left[\mu_n C_{ox}\frac{W}{L}V_{od}\right] = \alpha g_{d0}. \tag{12}$$

Another of the parameters of Eqn, 6 linked to power is Qs. Recall that Qs is a function of Cgs which in turn is a function of device width. Equation 11 may besolved for W. and the resulting expression substituted into the equation for Qs withthe following result is

$$Q_s = \frac{P_0}{P_D}\frac{\rho^2}{(1+\rho)},$$

$$P_0 = \frac{3}{2}\frac{V_{DD}v_{sat}E_{sat}}{\omega R_s}. \tag{13,14}$$

$$\rho^2 \approx \frac{P_D}{P_0}\sqrt{\frac{\delta}{5\gamma}(1-|c|^2)}\left[1+\sqrt{\frac{7}{4}}\right]. \tag{15}$$

$$Q_{sP} = |c|\sqrt{\frac{5\gamma}{\delta}}\left[1+\sqrt{1+\frac{3}{|c|^2}\left(1+\frac{\delta}{5\gamma}\right)}\right] \approx 4. \tag{16}$$

$$W_{opt\,P} = \frac{3}{2}\frac{1}{\omega L C_{ox} R_s Q_{sP}} \approx \frac{1}{3\omega L C_{ox} R_s}. \tag{17}$$

$$F_{min\,P} \approx 1 + 2.4\frac{\gamma}{\alpha}\left[\frac{\omega}{\omega_T}\right]. \tag{18}$$

$$F_{min} \approx 1 + 2.3\left[\frac{\omega}{\omega_T}\right]. \tag{19}$$

3. DESIGN EXAMPLES:-

3.1 SINGLE-ENDED LNA:-

To complete the design largely requires only the addition of bias and output circuitry. For narrowband applications it is advantageous to tune out the output capacitance to increase gain. Hence a typical single-ended LNA might appear as shown in Figure 1.

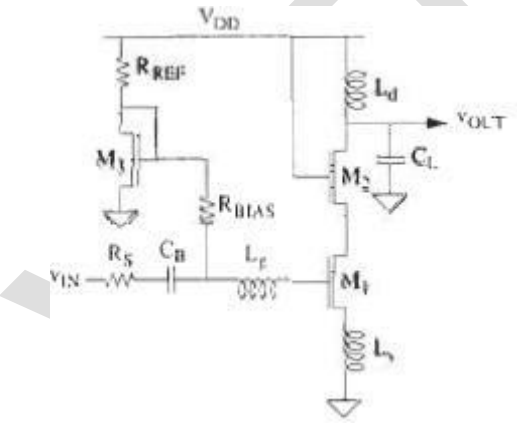


Figure 1 single ended LNA

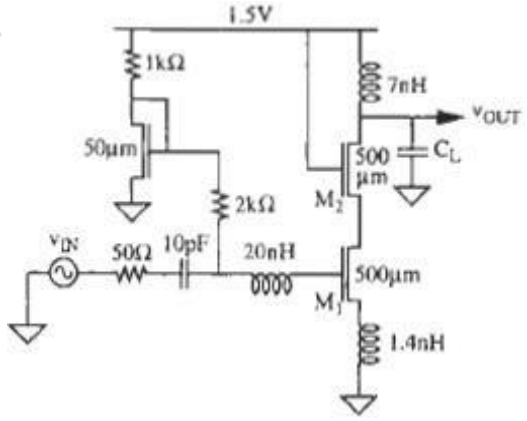


figure 2 Complete 1.5-GHz, 8-mW single-ended LNA

The cascoding transistor is chosen here to have the same width as the main device. This choice is common, but it is somewhat arbitrary and thus not necessarily ideal. Two competing considerations constrain the size of the cascoding transistor.The gate-drain overlap capacitance can reduce the impedance looking into the gate and drain of *MI* considerably, degrading both the noise performance and input match.It is a straightforward matter to show that, for equal-sized common-source and cascading devices, the resistive component at the input is given by

$$\mathrm{Re}[Z_{in}] = \frac{\omega_T L_s}{1 + 2C_{gd}/C_{gs}}. \tag{1}$$

Another potential source of NF degradation is the thermal noise of the substrate, as discussed in the previous chapter. Recall that this epi noise produces a drain noise component whose value is given by

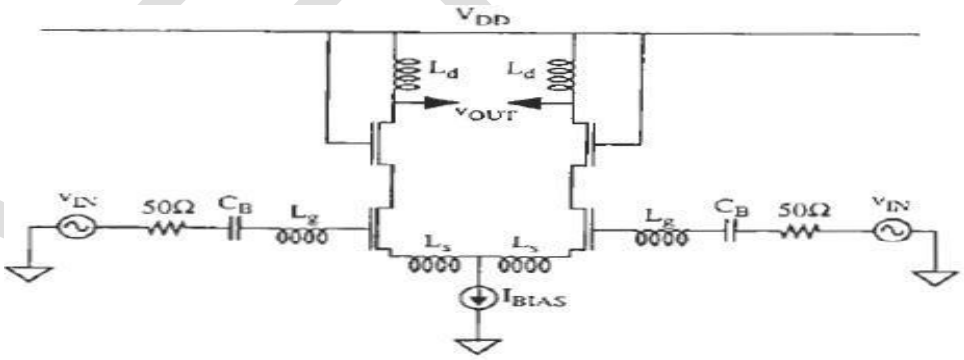
$$\overline{i_{nd,sub}^2} = \frac{4kTR_{sub}g_{mb}^2}{1 + (\omega R_{sub}C_{cb})^2}\Delta f. \tag{2}$$

At frequencies well below the pole frequency of 1/RsubCcb this extrinsic noise contribution can nonetheless remain negligible if model parameters satisfy the following inequality is

$$R_{sub} \leq 2R_s. \tag{3}$$

3.2 DIFFERENTIAIL LNA:-

fig 1.Differential LNA



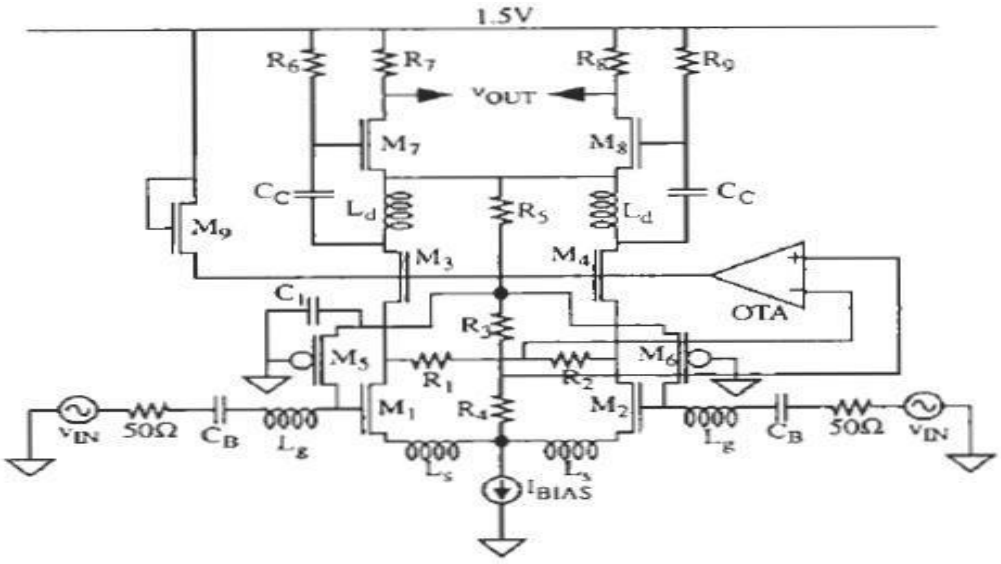


Fig.2.Complete 12-mW, 1.5·GHz differential LNA.

The incremental ground located at the symmetrical point of a differential structure. See Figure 1 where the source degenerating inductances return to a virtual ground (for differential signals). Any parasitic reactance in series with the bias current source is largely irrelevant, since a current source in series with impedance is still a current source. Hence, the real part of the input impedance is controlled only by *Ls* and is unaffected by parasitics in the current source's ground return path.

The circuit of Figure 1 ignores biasing details - in particular, it doesn't show how the DC gate potentials are established. Furthermore, the gate bias for the cascading transistors is the supply voltage, and such a choice may impress enough voltage across the bottom transistors to increase the values of *gamma* and *alpha* at higher supply voltages. To mitigate such hot-electron noise degradation, it is prudent to use only enough drain bias to keep transistors comfortably out of the triode region, and no more.

One of many possible methods for achieving the desired results is shown in Figure 2 In this circuit transistors *M* I through *M* ~ are the LNA core of the previous simplified schematic. The output of this first stage is AC-coupled through Cc to *M7*and *Ms,* which provide additional gain. To save power the current through the second gain stage is *re-used* to supply the four core

transistors. Hence, the output tuning inductors *Ld* return to the common source connection of *M7* and *M8* rather than to the positive supply.

To keep all of the transistors in the stack in saturation with a low supply voltage, a common- mode bias feedback loop sets the drain voltage of M1 and *M2* equal to a fixed fraction of *Vgs1* To the extent that Vdsat roughly tracks *Vgs1* this strategy guarantees that no more than the minimum supply voltage is consumed.

The bias loop works as follows. The voltage at the top of resistor *R)* is equal to the common- mode gate voltage of MI and *M2* because, with no current flowing through junction of *R*...and *R*~ is therefore some fraction of the common-mode gate-to-source voltage of the input pair. The operational amplifier+' compares this voltage with the common-mode drain voltage of the input pair (measured at the midpoint of *RI/R2)* and drives the gates of the cascoding transistors to make the two voltages equal. Hence we have

$$V_{d1,2} = \frac{R_4}{R_3 + R_4} V_{gs1,2}, \quad (1)$$

where the voltages are referenced to the common source connection of the input pair. Transistor *M9* guarantees start-up of the bias loop by providing some default gate voltage for the cascoding transistors until the op-arnp has a chance to act. Finally because of the modest performance required of the op-amp, negligibly low bias currents may be used there.

3. MIXER

CHAPTER THIRTEEN

MIXERS

13.1 INTRODUCTION

Most circuit analysis proceeds with the assumptions of linearity and time invariance. Violations of those assumptions, if considered at all, are usually treated as undesirable. However, the high performance of modern communications equipment actually depends critically on the presence of at least one element that fails to satisfy linear time invariance: the mixer. We will see shortly that mixers are still ideally linear but depend fundamentally on a purposeful violation of time invariance. As noted in Chapter 1, the superheterodyne[1] receiver uses a mixer to perform an important frequency translation of signals. Armstrong's invention has been the dominant architecture for 70 years because this frequency translation solves many problems in one fell swoop (see Figure 13.1).[2]

In this architecture, the mixer translates an incoming RF signal to a lower frequency,[3] known as the intermediate frequency (IF). Although Armstrong originally sought this frequency lowering simply to make it easier to obtain the requisite gain, other significant advantages accrue as well. As one example, tuning is now accomplished by varying the frequency of a local oscillator, rather than by varying the center frequency of a multipole bandpass filter. Thus, instead of adjusting several *LC* networks in tandem to tune to a desired signal, one simply varies a single *LC* combination to change the frequency of a local oscillator (LO). The intermediate frequency stages can then use *fixed* bandpass filters. Selectivity is therefore determined

[1] Why "super"heterodyne? The reason is that Fessenden had already invented something called the "heterodyne," and Armstrong had to distinguish his invention from Fessenden's.

[2] Proving once again that success has many fathers (while failure is an orphan, to complete J. F. Kennedy's saying), Lucien Lévy and Walter Schottky, among others, laid claim to the superheterodyne. While it is certainly true that Armstrong was not the first to conceive of the heterodyne principle, he *was* the first to recognize how neatly it solves so many thorny problems, and he was certainly the first to pursue its development vigorously.

[3] Actually, one may also translate to a *higher* frequency, but we will defer a discussion of that case to Chapter 19.

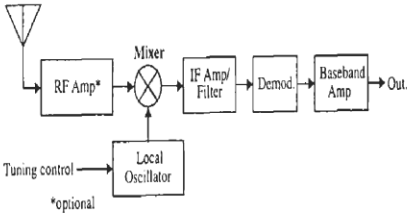


FIGURE 13.1. Superheterodyne receiver block diagram.

by these fixed-frequency IF filters, which are much easier to realize than variable-frequency filters. Additionally, the overall gain of the system is distributed over a number of different frequency bands (RF, IF, and baseband), so that the required receiver gain (typically 120–140 dB on a power basis) can be obtained without much worry about potential oscillations arising from parasitic feedback loops. These important attributes explain why the superheterodyne architecture still dominates, more than 70 years after its invention.

13.2 MIXER FUNDAMENTALS

Since linear, time-invariant systems cannot produce outputs with spectral components not present at the input, mixers must be either nonlinear or time-varying elements in order to provide frequency translation. Historically, many devices (e.g., electrolytic cells, magnetic ribbons, brain tissue, and rusty scissors – in addition to more traditional devices such as vacuum tubes and transistors) operating on a host of diverse principles have been used, demonstrating that virtually *any* nonlinear element can be used as a mixer.[4]

At the core of all mixers presently in use is a multiplication of two signals in the time domain. The fundamental usefulness of multiplication may be understood from examination of the following trigonometric identity:

$$(A\cos\omega_1 t)(B\cos\omega_2 t) = \frac{AB}{2}[\cos(\omega_1-\omega_2)t + \cos(\omega_1+\omega_2)t]. \qquad (1)$$

Multiplication thus results in output signals at the sum and difference frequencies of the input, signals whose amplitudes are proportional to the product of the RF and LO amplitudes. Hence, if the LO amplitude is constant (as it usually is), any amplitude modulation in the RF signal is transferred to the IF signal. By a similar mechanism, an undesired transfer of modulation from one signal to another can also occur through nonlinear interaction in both mixers and amplifiers. In that context the result is called

[4] Of course, some nonlinearities work better than others, so we will focus on the more practical types.

cross-modulation, as mentioned in Chapter 12, and its suppression through improved linearity is an important design consideration.

Having recognized the fundamental role of multiplication, we now enumerate and define the most significant characteristics of mixers.

13.2.1 CONVERSION GAIN

One important mixer characteristic is conversion gain (or loss), which is defined as the ratio of the desired IF output to the value of the RF input. For the multiplier described by Eqn. 1, the conversion gain is therefore the IF output, $AB/2$, divided by A (if that is the amplitude of the RF input). Hence, the conversion gain in this example is $B/2$, or half the LO amplitude.

Conversion gain, if expressed as a power ratio, can be greater than unity in active mixers; passive mixers are generally capable only of voltage or current gain at best.[5] Conversion gain in excess of unity is often convenient since the mixer then provides amplification along with the frequency translation. However, it does not necessarily follow that sensitivity improves, since noise figure must also be considered. For this reason, passive mixers may offer superior performance in some cases despite their conversion loss.

13.2.2 NOISE FIGURE: SSB VERSUS DSB

Noise figure is defined as one might expect: it's the signal-to-noise ratio (SNR) at the input (RF) port divided by the SNR at the output (IF) port. There's an important subtlety, however, that often trips up both the uninitiated and a substantial fraction of practicing engineers. To appreciate this difficulty, we first need to make an important observation: In a typical mixer, there are actually *two* input frequencies that will generate a given intermediate frequency. One is the desired RF signal, and the other is called the *image* signal. In the context of mixers, these two signals are frequently referred to collectively as *sidebands*.

The reason that two such frequencies exist is that the IF is simply the *magnitude* of the difference between the RF and LO frequencies. Hence, signals both above and below ω_{LO} by an amount equal to the IF will produce IF outputs of the same frequency. The two input frequencies are therefore separated by $2\omega_{IF}$. As a specific numerical example, suppose that our system's IF is 100 MHz and we wish to tune to a signal at 900 MHz by selecting an LO frequency of 1 GHz. Aside from the desired 900-MHz RF input, a 1.1-GHz image signal will also produce a difference frequency component at the IF of 100 MHz.

[5] An exception is a class of systems known as parametric converters or parametric amplifiers, in which power from the LO is transferred to the IF through reactive nonlinear interaction (typically with varactors), thus making power gain possible.

The existence of an image frequency complicates noise figure computations because noise originating in both the desired and image frequencies therefore becomes IF noise, yet there is generally no desired signal at the image frequency. In the usual case where the desired signal exists at only one frequency, the noise figure that one measures is called the *single-sideband* noise figure (SSB NF); the rarer case, where both the "main" RF and image signals contain useful information, leads to a double-sideband (DSB) noise figure.

Clearly, the SSB noise figure will be greater than for the DSB case, since both have the same IF noise but the former has signal power in only a single sideband. Hence, the SSB NF will normally be 3 dB higher than the DSB NF.[6] Unfortunately, DSB NF is reported much more often because it is numerically smaller and thus falsely conveys the impression of better performance, even though there are few communications systems for which DSB NF is an appropriate figure of merit.[7] Frequently, a noise figure is stated without any indication as to whether it is a DSB or SSB value. In such cases, one may usually assume that a DSB figure is being quoted.

Noise figures for mixers tend to be considerably higher than those for amplifiers because noise from frequencies other than at the desired RF can mix down to the IF. Representative values for SSB noise figures range from 10 dB to 15 dB or more. It is mainly because of this larger mixer noise that one uses LNAs in a receiver. If the LNA has sufficient gain then the signal will be amplified to levels well above the noise of the mixer and subsequent stages, so the overall receiver NF will be dominated by the LNA instead of the mixer. If mixers were not as noisy as they are, then the need for LNAs would diminish considerably. We will return to this theme in the chapter on receiver architectures (Chapter 19).

13.2.3 LINEARITY AND ISOLATION

Dynamic range requirements in modern, high-performance telecommunications systems are quite severe, frequently exceeding 80 dB and approaching 100 dB in many instances. As discussed in the previous chapter, the floor is established by the noise figure, which conveys something about how small a signal may be processed, whereas the ceiling is set by the onset of severe nonlinearities that accompany large input signals.

As with amplifiers, the compression point is one measure of this dynamic range ceiling and is defined the same way. Ideally, we would like the IF output to be proportional to the RF input signal amplitude; this is the sense in which we interpret the

[6] This 3-dB difference assumes that the conversion gain to two equal sidebands is the same. Although this assumption is usually well satisfied, it need not be.

[7] Two important exceptions in which both sidebands contain useful information are radio astronomy (as in the measurements of the echoes of the Big Bang) and direct-conversion receivers (see Chapter 19).

term "linearity" in the context of mixers. However, as with amplifiers (and virtually any other physical system), real mixers have some limit beyond which the output has a sublinear dependence on the input. The compression point is the value of RF signal[8] at which a calibrated departure from the ideal linear curve occurs. Usually, a 1-dB (or, more rarely, a 3-dB) compression value is specified. One may specify either the input or output signal strength at which this compression occurs, together with the conversion gain, to allow fair comparisons among different mixers.

The two-tone third-order intercept is also used to characterize mixer linearity. A two-tone intermodulation test is a relevant way to evaluate mixer performance because it mimics the real-world scenario in which both a desired signal and a potential interferer (perhaps at a frequency just one channel away) feed a mixer input. Ideally, each of two superposed RF inputs would be translated in frequency without interacting with each other. Of course, practical mixers will always exhibit some intermodulation effects, and the output of the mixer will thus contain frequency-translated versions of third-order IM components whose frequencies are $2\omega_{RF1} \pm \omega_{RF2}$ and $2\omega_{RF2} \pm \omega_{RF1}$. The difference frequency terms may heterodyne into components that lie within the IF passband and are therefore generally the troublesome ones, while the sum frequency signals can usually be filtered out.

As a measure of the degree of departure from linear mixing behavior, one can plot the desired output and the third-order IM output as a function of input RF level. The third-order intercept is the extrapolated intersection of these two curves. In general, the higher the intercept, the more linear the mixer. Again, one ought to specify whether the intercept is input- or output-referred, as well as the conversion gain, to permit fair comparisons among mixers. Additionally, it is customary to abbreviate the intercept as IP3, or perhaps IIP3 or OIP3 (for input and output third-order intercept point, respectively). These definitions are summarized in Figure 13.2.

Cubic nonlinearity can also cause trouble with a *single* RF input. As a specific example, consider building a low-cost AM radio. The standard IF for AM radios happens, unfortunately, to be 455 kHz (mainly for historical reasons). Tuning in a station at 910 kHz (a legitimate AM radio frequency) requires that the LO be set to 1365 kHz.[9] The cubic nonlinearity could generate a component at $2\omega_{RF} - \omega_{LO}$, which in this case happens to coincide with our IF of 455 kHz.

One might be tempted to assert that such a component is not a problem because it adds to the desired output. One therefore might even be tempted to consider this an asset. However, the third-order IM products have amplitudes that are no longer

[8] Some manufacturers (and authors) report an *output* compression point. If the conversion gain at that point is known then the figure can be reflected back to the input point. Sadly, many insist on burying that bit of information, making it extremely difficult to perform fair comparisons of mixer performance. We will always state explicitly whether the figure is an input or output parameter.

[9] A local oscillator frequency of 455 kHz also works, but it is a less practical choice because such "low-side injection" requires the local oscillator to tune over a larger range than if the LO frequencies were above the desired RF.

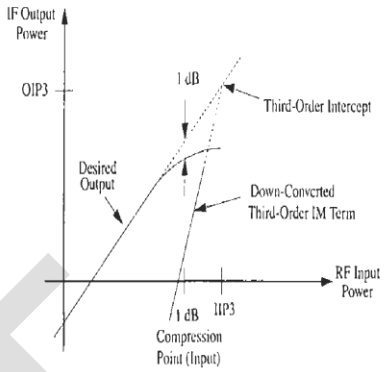


FIGURE 13.2. Definition of mixer linearity parameters.

proportional to the input signal amplitude. Hence, they represent amplitude distortion that can corrupt the "correct" output (we're talking about an amplitude-modulated signal, after all).

Even if the exact numerological coincidence of the foregoing example does not occur, various third-order IM terms can possess frequencies within the passband of the IF amplifier, ultimately degrading signal-to-noise or signal-to-distortion ratios.

Another parameter of great practical importance is isolation. It is generally desirable to minimize interaction among the RF, IF, and LO ports. For instance, since the LO signal power is generally quite large compared with that of the RF signal, any LO feedthrough to the IF output might cause problems at subsequent stages in the signal processing chain. This problem is exacerbated if the IF and LO frequencies are similar, so that filtering is ineffective. Even reverse isolation is important in many instances, since poor reverse isolation might permit the strong LO signal (or its harmonics) to work its way back to the antenna, where it can radiate and cause interference to other receivers.

13.2.4 SPURS

Mixers, by their nature, may heterodyne a variety of frequency components that you never intended to mix. For example, harmonics of some signal (desired or not) could lie (or be generated) within the passband of the mixer system and subsequently beat against the local oscillator signal (and its harmonics). Some of the resulting components may end up within the IF passband. The undesired signals that do ultimately emerge from the output of the mixer are known as spurious responses, or just *spurs*. Evaluation of mixer spurs is straightforward in principle but *highly* tedious in practice (so much so that a hazing ritual for newly minted RF engineers in days past often

included evaluation of mixer spurs).[10] The availability of software tools to take care of this task has eliminated the tedium, but it's instructive to describe the process, just the same.[11]

Let m and n be the harmonic numbers of the RF input and LO frequencies, respectively. Then the spur products present at the output of the mixer (prior to any filtering) are given by

$$f_{spur} = mf_{RF} + nf_{LO}. \tag{2}$$

The apparent simplicity of this equation is misleading: The calculation must be repeated for all combinations *and signs* of m and n, ranging up to the maximum harmonic order you care to consider. To make a laborious procedure even more so, one must actually consider RF signals of frequencies below the nominal input passband – at least down to the lower passband edge frequency divided by the maximum value of m. One may also have to consider input frequencies somewhat above the nominal upper passband frequency of the RF filter. Since no filter is perfect and since no LO is completely free of distortion, harmonics of the LO can still heterodyne with nominally out-of-band RF signals that leak through the filter. The resulting interaction can produce spurs at the mixer output that happen to lie within the IF passband. If the out-of-band interferer is strong enough, the spurious IF signal can severely degrade receiver performance.

For each (m, n) pair, examine the spur frequency and then determine whether it lies within the IF passband, or sufficiently close to it, to merit further consideration. For each spur that does, work backward to the implied RF input frequency and evaluate the likelihood that there will be a signal of sufficient strength at that frequency to be a source of trouble. Then make appropriate modifications either to the input filtering or to the choice (or quality) of LO or IF, if necessary, to avoid those troubles.

This exercise is sometimes performed with the worst-case assumption that there is no filtering of any kind at the RF input port. In that case, the number of calculations grows very large quite quickly. If one is patient enough, however, the information generated can be used to guide the design of the input filter (or the frequency plan and other architectural details of the receiver).

As a specific example, suppose we wish to design a mixer for an FM receiver, whose nominal input passband is to accommodate signals spanning 88.1 MHz to 108.1 MHz. With a 10.7-MHz IF (conventional for consumer FM receivers), the LO needs to tune from 77.4 MHz to 97.4 MHz (assuming low-side injection). To keep

[10] In vacuum-tube days, the initiation also usually included setting a neophyte off to the stockroom in search of a grid-leak drip pan; it's the geek equivalent of a snipe hunt.

[11] An excellent program that performs this calculation (and many others of great value to the RF/microwave designer) is *AppCAD*, originally from HP, now from Agilent. Versions for older and newer PC operating systems are currently available for free download from http://www.hp.woodshot.com.

Table 13.1. *Spur table for FM radio example*

m	f_{RF}, low	f_{RF}, high	n
−3	73.8	93.9	3
−2	72.0	92.1	2
1	88.0	108.2	−1
2	82.7	102.8	−2
3	80.9	101.0	−3

the numbers easy, assume a bit unrealistically that the IF system possesses a nominal bandwidth of approximately 200 kHz. Further assume that the LO is pure enough – and the input RF filtered enough – that we need not consider harmonic orders higher than 3.[12] With these assumptions, we can construct Table 13.1.

Examining the first entry in the table, we see that the third harmonic of RF signals in the 73.8–93.9-MHz frequency band may heterodyne with the third harmonic of the LO to produce signals within the 10.6–10.8-MHz IF passband. Notice that improved input filtering would be only partially effective at best, because much of the spurious input band overlaps the desired FM radio band. If there were indeed significant interferers within this spurious band, then our only choice would be to improve the spectral purity of the LO (specifically, we would need to minimize its third harmonic content). Such an improvement would also benefit the spur problems implied by the last row. Similarly, reduction in *second* harmonic LO content would be the only practical way to avoid the problems implied in the second and fourth rows of the spur table. The third row does not describe undesired components; because $m = n = 1$, it describes the intended mode of operation for the receiver and is included simply for completeness.

By carrying out this laborious procedure for any contemplated system, it is possible to assess the sensitivity to various imperfections and thus to evaluate the need for remediation.

13.3 NONLINEAR SYSTEMS AS LINEAR MIXERS

We now consider how to implement the multiplication that is the heart of mixing action. Some mixers directly implement a multiplication, while others provide it incidentally through a nonlinearity. We follow an historical path and first examine

[12] Remember that the spectrum of a signal rolls off approximately as $1/n$, where n is the number of derivatives needed to produce impulses from the time-domain representation of the signal. Most signals of practical interest have spectra that roll off fast enough that consideration of orders higher than about 5 or 7 is probably overkill for typical situations. Your mileage may vary, however.

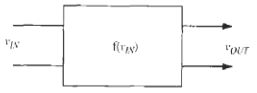


FIGURE 13.3. General two-port nonlinearity.

a general two-port nonlinearity,[13] since mixers of that type predate those designed specifically to behave as multipliers. See Figure 13.3.

If the nonlinearity is "well-behaved" (in the mathematical sense), we can describe the input–output relationship with a series expansion:

$$v_{OUT} = \sum_{n=0}^{N} c_n (v_{IN})^n. \tag{3}$$

Using an Nth-order nonlinearity as a mixer requires the signal v_{IN} to be the sum of the RF input and the local oscillator signals. In general, the output will consist of three types of products: DC terms, harmonics of the inputs, and intermodulation products of those harmonics.[14] Not all of these spectral components are desirable, so part of the challenge in mixer design is to devise topologies that inherently generate few undesired terms.

Even-order nonlinear factors in Eqn. 3 contribute DC terms; these are readily filtered out by AC coupling, if desired. Harmonic terms, at $m\omega_{LO}$ and $m\omega_{RF}$, extend from the fundamental ($m = 1$) all the way up to the Nth harmonic. As with the DC terms, they are also often relatively easy to filter out because their frequencies are usually well away from the desired IF.

The intermodulation (IM) products are the various sum and difference frequency terms. These have frequencies expressible as $p\omega_{RF} \pm q\omega_{LO}$, where integers p and q are greater than zero and sum to values up to N. Only the second-order intermodulation term ($p = q = 1$) is normally desired.[15] Unfortunately, other IM products might have frequencies close to the desired IF, making them difficult to remove, as we shall see. Since it is generally true that high-order nonlinearities (i.e., large values of N in the power series expansion) tend to generate more of these undesirable terms,[16] mixers should approximate square-law behavior (the lowest-order nonlinearity) if they have only one input port (as shown in Figure 13.3). We now consider the specific properties of a square-law mixer to identify its advantages over higher-order nonlinear mixers.

[13] We will shortly see the advantages of three-port mixers.

[14] Keep in mind that fundamentals are harmonics.

[15] The order of a given IM term is the sum of p and q, so a second-order IM product arises from the quadratic term in the series expansion.

[16] As with most sweeping generalities, there are exceptions to this one. In building frequency multipliers, for example, high-order harmonic nonlinearities are extremely useful. In mixer design, however, it is usually true that high-order nonlinearities are undesirable.

TWO-PORT EXAMPLE: SQUARE-LAW MIXER

To see explicitly where the desired multiplication arises in a square-law mixer, note that the only nonzero coefficients in the series expansion are the c_1 and c_2 terms.[17] If we then assume that the input signal v_{IN} is the sum of two sinusoids,

$$v_{IN} = v_{RF}\cos(\omega_{RF}t) + v_{LO}\cos(\omega_{LO}t), \tag{4}$$

then the output of this mixer may be expressed as the sum of three distinct components:

$$v_{OUT} = v_{fund} + v_{square} + v_{cross}, \tag{5}$$

where

$$v_{fund} = c_1[v_{RF}\cos(\omega_{RF}t) + v_{LO}\cos(\omega_{LO}t)], \tag{6}$$

$$v_{square} = c_2\{[v_{RF}\cos(\omega_{RF}t)]^2 + [v_{LO}\cos(\omega_{LO}t)]^2\}, \tag{7}$$

$$v_{cross} = 2c_2 v_{RF} v_{LO}[\cos(\omega_{RF}t)][\cos(\omega_{LO}t)]. \tag{8}$$

The fundamental terms are simply scaled versions of the original inputs and therefore represent no useful mixer output; they must be removed by filtering. The v_{square} components similarly represent no useful mixer output, as is evident from the following special case of Eqn. 1:

$$[\cos\omega t]^2 = \tfrac{1}{2}[1 + \cos 2\omega t]. \tag{9}$$

Thus, we see that the v_{square} components contribute a DC offset as well as second harmonics of the input signals. These also must generally be removed by filtering.

The useful output comes from the v_{cross} components because of the multiplication evident in Eqn. 8. Using Eqn. 1, we may rewrite v_{cross} in a form that shows the mixing action more clearly:

$$v_{cross} = c_2 v_{RF} v_{LO}[\cos(\omega_{RF} - \omega_{LO})t + \cos(\omega_{RF} + \omega_{LO})t]. \tag{10}$$

For a fixed LO amplitude, the IF output amplitude is linearly proportional to the RF input amplitude. That is, this nonlinearity implements a linear mixing, since the output is proportional to the input.

The conversion gain for this nonlinearity is readily found from Eqn. 10:

$$G_c = \frac{c_2 v_{RF} v_{LO}}{v_{RF}} = c_2 v_{LO}. \tag{11}$$

[17] There may also be a nonzero DC term (i.e., c_0 may be nonzero), but this component is easily removed by filtering, so we will ignore it at the outset to reduce equation clutter.

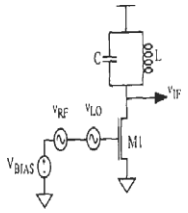


FIGURE 13.4. Square-law MOSFET mixer (simplified).

Just as any other gain parameter, conversion gain may be a dimensionless quantity (or a transconductance, transresistance, etc.). It is customary in discrete designs to express conversion gain as a power ratio (or its decibel equivalent), but the unequal input and output impedance levels in typical IC mixers makes a voltage or current conversion gain appropriate also. To avoid confusion, of course, it is essential to state explicitly the type of gain.[18]

As asserted earlier, the square-law mixer's advantages are that the undesired spectral components are usually at a frequency quite different from the intermediate frequency, and thus readily removed. For this reason, two-port mixers are often designed to conform to square-law behavior to the maximum practical extent.

Excellent square-law mixers may be realized with long-channel MOSFETs, or approximated by virtually any other type of nonlinearity in which the quadratic term dominates; see Figure 13.4. In this simplified schematic, the bias, RF, and LO terms are shown as driving the gate in series. The summation of RF and LO signals can be accomplished in practical circuits with resistive or reactive summers. Because the RF and LO signals are in series, there is poor isolation between them.

An alternative (but functionally equivalent) arrangement that reduces the effect of the relatively large LO signal on the RF port is shown in Figure 13.5. The RF signal drives the gate directly (through a DC-blocking capacitor), while the LO drives the source terminal. This way, the gate-to-source voltage is the sum of ground-referenced LO and RF signals. The bias current is set directly with a current source, and the DC gate voltage is determined by the value of V_{BIAS}. Resistor R_{BIAS} is chosen large enough to avoid excessive loading, and also to minimize its noise contribution.

In deriving an expression for conversion gain, we use our assumption that the device is long enough (and biased appropriately) to allow us to express the drain current as follows:

[18] All too frequently, published "power" gain figures are essentially voltage gain measurements and are therefore grossly in error if the input and output impedance levels differ significantly, as they often do. It seems necessary to emphasize that watts and volts are not the same.

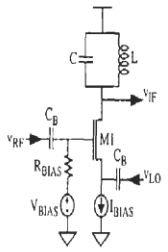


FIGURE 13.5. Square-law MOSFET mixer (alternative configuration).

$$i_D = \frac{\mu C_{ox} W}{2L}(V_{gs} - V_T)^2. \tag{12}$$

Short-channel (high-field) devices are more linear as a result of velocity saturation, and thus are generally inferior to long devices as mixers.[19]

If the gate–source voltage V_{gs} is the sum of RF, LO, and bias terms, then we may write

$$i_D = \frac{\mu C_{ox} W}{2L}\{V_{BIAS}[v_{RF}\cos(\omega_{RF}t) + v_{LO}\cos(\omega_{LO}t)] - V_T\}^2. \tag{13}$$

from which one may readily find that the conversion gain (here, a transconductance) is simply

$$G_c = \frac{\mu C_{ox} W}{2L} \cdot v_{LO}. \tag{14}$$

This square-law device thus has a conversion transconductance that is independent of bias.[20] It is still dependent on temperature (through mobility variation) and LO drive amplitude, however.

Because perfect square-law behavior is not necessary to obtain mixing action, M_1 can be a bipolar transistor, for example, because the quadratic factor in the series expansion for the exponential i_C–v_{BE} relationship dominates over a limited range of input amplitudes. Precisely because many nonlinearities are well approximated by a square-law shape over some suitably restricted interval, one can estimate the conversion gain for other nonlinear devices used as mixers once the value of the quadratic

[19] The reader is reminded once again that "short-channel" actually means "high-field." Hence, even "short" devices may still behave quadratically for suitably small drain-source voltages.

[20] This independence of bias holds only in the square-law regime. Enough bias must therefore be supplied to guarantee this condition. Hence, V_{BIAS} is not permitted to equal zero. In fact, it must be chosen large enough to guarantee that the gate–source voltage always exceeds the threshold voltage, since a MOSFET behaves exponentially in weak inversion.

coefficient (c_2) is found. To underscore this point, let's estimate the conversion gain for one more nonlinear element, a bipolar transistor.

Conversion Gain of a Single Bipolar Transistor Mixer

To simplify the calculation, let us continue to ignore dynamic effects. Then we can use the exponential v_{BE} law:

$$i_C \approx I_S e^{v_{BE}/V_T}. \tag{15}$$

Expansion of this familiar relationship up to the second-order term yields[21]

$$i_C \approx I_C\left[1 + \frac{v_{IN}}{V_T} + \frac{1}{2}\left(\frac{v_{IN}}{V_T}\right)^2\right]. \tag{16}$$

By inspection (well, almost),

$$c_2 = \frac{g_m}{2V_T}, \tag{17}$$

so that an estimate of the conversion gain is:

$$G_c = c_2 v_{LO} = g_m \cdot \frac{v_{LO}}{2V_T}. \tag{18}$$

The conversion gain here is a transconductance proportional both to the standard incremental transconductance and to the ratio of the local oscillator drive amplitude to the thermal voltage. The conversion gain for a bipolar transistor is therefore dependent on bias current, LO amplitude, and temperature.

As in the corresponding derivation for a MOSFET, the foregoing computation ignores parasitic series base and emitter resistances. These resistances can linearize the transistor and therefore weaken mixer action. Thoughtful device layout is thus mandatory to minimize this effect.

13.4 MULTIPLIER-BASED MIXERS

We have seen that nonlinearities produce mixing incidentally through the multiplications they provide. Precisely because the multiplication is only incidental, these nonlinearities usually generate a host of undesired spectral components. Additionally, since two-port mixers have only one input port, the RF and LO signals are generally not well isolated from each other. This lack of isolation can cause the problems mentioned earlier, such as overloading of IF amplifiers, as well as radiation of the LO signal (or its harmonics) back out through the antenna.

Mixers based directly on multiplication generally exhibit superior performance because they ideally generate only the desired intermodulation product. Furthermore, because the inputs to a multiplier enter at separate ports, there can be a high

[21] We have implicitly assumed that the base-emitter drive contains a DC component as well as the RF and LO components, so that I_C is nonzero.

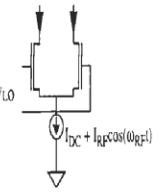


FIGURE 13.6. Single-balanced mixer.

degree of isolation among all three signals (RF, LO, and IF). Finally, CMOS technology provides excellent switches, and one can implement outstanding multipliers with switches.

13.4.1 SINGLE-BALANCED MIXER

One extremely common family of multipliers first converts the incoming RF voltage into a current and then performs a multiplication in the current domain. The simplest multiplier cell of this type is sketched in Figure 13.6.[22] In this mixer, v_{LO} is chosen large enough so that the transistors alternately switch (commutate) all of the tail current from one side to the other at the LO frequency.[23] The tail current is therefore effectively multiplied by a square wave whose frequency is that of the local oscillator:

$$i_{out}(t) = \text{sgn}[\cos\omega_{LO}t]\{I_{BIAS} + I_{RF}\cos\omega_{RF}t\}. \tag{19}$$

Because a square wave consists of odd harmonics of the fundamental, multiplication of the tail current by the square wave results in an output spectrum that appears as shown in Figure 13.7 (ω_{RF} is here chosen atypically low compared with ω_{LO} to reduce clutter in the graph).

The output thus consists of sum and difference components, each the result of an odd harmonic of the LO mixing with the RF signal. In addition, odd harmonics of the LO appear directly in the output as a consequence of the DC bias current multiplying with the LO signal. Because of the presence of the LO in the output spectrum, this type of mixer is known as a *single-balanced* mixer. Double-balanced mixers, which we'll study shortly, exploit symmetry to remove the undesired output LO component through cancellation.

[22] Mixers of this general kind are often lumped together and called Gilbert mixers, but only some actually are. True Gilbert multipliers function entirely in the current domain, deferring the problem of *V–I* conversion by assuming that all variables are already available in the form of currents. See Barrie Gilbert's landmark paper, "A Precise Four-Quadrant Multiplier with Subnanosecond Response," *IEEE J. Solid-State Circuits*, December 1968, pp. 365–73.

[23] One may also interchange the roles of LO and RF input, but the resulting mixer has lower conversion gain and worse noise performance, among other deficiencies. A more detailed discussion of this issue is deferred to a later section.

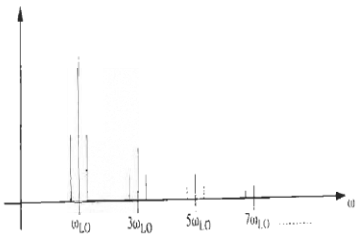


FIGURE 13.7. Representative output spectrum of single-balanced mixer.

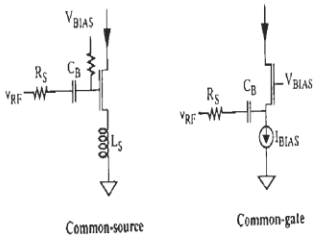


FIGURE 13.8. RF transconductors for mixers.

Although the current source of Figure 13.6 includes a component that is perfectly proportional to the RF input signal, *V–I* converters of all real mixers are imperfect. Hence, an important design challenge is to maximize the linearity of the RF transconductance. Linearity is most commonly enhanced through some type of source degeneration, in both common-gate and common-source transconductors; see Figure 13.8. The common-gate circuit uses the source resistance R_s to linearize the transfer characteristic. This linearization is most effective if the admittance looking into the source terminal of the transistor is much larger than the conductance of R_s. In that case, the transconductance of the stage approaches $1/R_s$.

Inductive degeneration is usually preferred over resistive degeneration for several reasons.[24] An inductance has neither thermal noise to degrade noise figure nor DC voltage drop to diminish supply headroom. This last consideration is particularly relevant for low-voltage–low-power applications. Finally, the increasing reactance of an inductor with increasing frequency helps to attenuate high frequency harmonic and intermodulation components.

[24] Capacitive degeneration has also been tried but is markedly inferior to inductive degeneration because it increases noise and distortion at high frequencies.

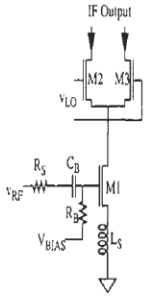


FIGURE 13.9. Single-balanced mixer with linearized transconductance.

A more complete single-balanced mixer that incorporates a linearized transconductance is shown in Figure 13.9. The value of V_{BIAS} establishes the bias current of the cell, while R_B is chosen large enough not to load down the gate circuit (and also to reduce its noise contribution). The RF signal is applied to the gate through a DC blocking capacitor C_B. In practice, a filter would be used to remove the LO and other undesired spectral components from the output.

The conversion transconductance of this mixer can be estimated by assuming that the LO-driven transistors behave as perfect switches. Then the differential output current may be regarded as the result of multiplying the drain current of M_1 by a unit-amplitude square wave. Since the amplitude of the fundamental component of a square wave is $4/\pi$ times the amplitude of the square wave, we may write:

$$G_c = \frac{2}{\pi} g_{m1} \tag{20}$$

where g_{m1} is the transconductance of the *V–I* converter and G_c is itself a transconductance. The coefficient is $2/\pi$ rather than $4/\pi$ because the IF signal is divided evenly between sum and difference components.

13.4.2 ACTIVE DOUBLE-BALANCED MIXER

To prevent the LO products from getting to the output in the first place, two single-balanced circuits may be combined to produce a double-balanced mixer; see Figure 13.10. We assume once again that the LO drive is large enough to make the differential pairs act like current-steering switches. Note that the two single-balanced mixers are connected in antiparallel as far as the LO is concerned but in parallel for the RF signal. Therefore, the LO terms sum to zero in the output, whereas the converted RF signal is doubled in the output. This mixer thus provides a high degree of

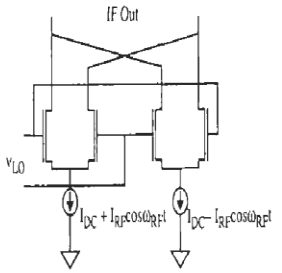


FIGURE 13.10. Active double-balanced mixer.

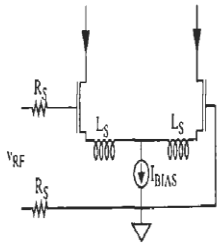


FIGURE 13.11. Linearized differential RF transconductor for double-balanced mixer.

LO–IF isolation, easing filtering requirements at the output. If care is taken in layout, IC realizations of this circuit routinely provide 40 dB of LO–IF isolation, with values in excess of 60 dB possible.

As in the single-balanced active mixer, the dynamic range is limited in part by the linearity of the *V–I* converter in the RF port of the mixer. So, most of the design effort is spent attempting to find better ways of providing this *V–I* conversion. The basic linearizing techniques used in the single-balanced mixer may be adapted to the double-balanced case, as shown in Figure 13.11.

In low-voltage applications, the DC current source can be replaced by a parallel *LC* tank to create a zero-headroom AC current source. The resonant frequency of the tank should be chosen to provide rejection of whatever common-mode component is most objectionable. If several such components exist, one may use series combinations of parallel *LC* tanks. With such a choice, a complete double-balanced mixer appears as shown in Figure 13.12. The expression for the conversion transconductance is the same as for the single-balanced case.

Noise Figure of Gilbert-Type Mixers

Computing the noise figure of mixers is difficult because of the cyclostationary nature of the noise sources. One technique involves characterization of the time-varying

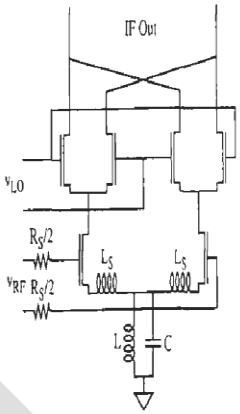


FIGURE 13.12. Minimum supply-headroom double-balanced mixer.

impulse response, arguing that a mixer is at least linear, if not time-invariant.[25] Although the method is accurate and quite suitable for analysis, its complexity does inhibit acquisition of design insight. Nonetheless, we can identify several important noise sources and make general recommendations about how to minimize noise figure.

One noise source is certainly the transconductor itself, so that its noise figure establishes a lower bound on the mixer noise figure. The same approach used in computing LNA noise figure may be used to compute the transconductor noise figure. The low-headroom mixer of Figure 13.12 may also be modified to act as a low *noise* mixer simply by adding suitable gate inductances to the inductively degenerated pair that receives the RF input. By following a prescription essentially identical to that for stand-alone LNAs (see Chapter 12), it is possible to construct a low-headroom, low-noise mixer that may obviate the need for a separate LNA in some applications. Adjustment of the tuning of the input loop allows a variable tradeoff among conversion gain, noise figure, and distortion.

The differential pair also degrades noise performance in a number of ways. One noise figure contribution arises from imperfect switching, which causes attenuation of the signal current. Hence, one challenge in such mixers is to design the switches (and associated LO drive) to provide as little attenuation as possible.

Another NF contribution of the switching transistors arises from the interval of time in which both transistors conduct current and hence generate noise. Additionally, any noise in the LO is also magnified during this active gain interval. Minimizing the simultaneous conduction interval reduces this degradation, so sufficient LO drive must

[25] C. D. Hull and R. G. Meyer, "A Systematic Approach to the Analysis of Noise in Mixers," *IEEE Trans. Circuits and Systems I*, v. 40, no. 12, December 1993, pp. 909–19.

be supplied to make the differential pair approximate ideal, infinitely fast switches to the maximum practical extent. Finally, the 3-dB attenuation inherent in ignoring either the sum or difference signal automatically degrades noise figure (by 3 dB) since the noise cannot be discarded so readily. As a result, practical current-mode mixers typically exhibit SSB noise figures of at least 10 dB, with values more frequently in the neighborhood of 15 dB.

Linearity of Gilbert-Type Mixers

The IP3 of this type of mixer is bounded by that of the transconductor, so the three-point method used to estimate the IP3 of ordinary amplifiers may also be used here to estimate the IP3 of the transconductor. If the LO-driven transistors act as good switches then the overall mixer IP3 generally differs little from that of the transconductor. To guarantee good switching, it is important to note that – although sufficient LO drive is necessary – excessive LO drive is to be avoided. To understand the reason that excessive LO drive is a liability rather than an asset, consider the effect of ever-present capacitive parasitic loading on the common-source connection of a differential pair. As each gate is driven far beyond what's necessary for good switching, the common-source voltage is similarly overdriven. A spike in current results. In extreme cases, this spike can cause transistors to leave the saturation region. Even if that does not occur, the output spectrum can become dominated by the components arising from the spikes, rather than the downconverted RF. Hence, one should use only enough LO drive to guarantee reliable switching, and no more.

A Short Note on Simulation of Mixer IP3 with Time-Domain Simulators

Just as we noted with simulations of intermodular distortion in amplifiers, common circuit simulators such as SPICE provide accurate mixer simulations only reluctantly, if at all. The problem stems from two fundamental sources: The wide dynamic range of signals in a mixer forces the use of far tighter numerical tolerances than are adequate for "normal" circuit simulations; and the large span of frequencies of important spectral components forces long simulation times. Hence, obtaining an accurate value for IP3 from a transient simulation, for example, is usually quite challenging. Furthermore, a correct noise figure simulation for CMOS mixers is not possible at all with commercially available tools, because the device noise models presently in use are incorrect. The reader is therefore cautioned to treat mixer simulation results with great skepticism.

Because even the "accurate" options available in some simulation tools are orders of magnitude too loose to be useful for IP3 simulations, one specific action that mitigates some of these problems is to tighten tolerances progressively until the simulation results stop changing significantly. In particular, the behavior of the IM3 component in an IP3 simulation is an extremely sensitive indicator of whether the tolerances are sufficiently tight. If the IM3 terms do not exhibit a +3 slope (on a dB scale), chances are high that the tolerances are too loose. One must also make sure

that the amplitudes of the two input tones are chosen small enough (i.e., well below either the compression or intercept points) to guarantee quasilinear operation of the mixer; otherwise, higher-order terms in the nonlinearity will contribute significantly to the output and confound the results. In the early phases of design, the three-point method may be applied to the transconductor to estimate its IP3 without having to suffer the agony of a transient simulation.

Another subtle consideration is to guarantee equal time spacing in the transient simulation, since FFT algorithms generally assume uniform sampling. Because some simulators use adaptive time stepping to speed up convergence, significant spectral artifacts can arise when computing the FFT. One may set the time step to a tiny fraction of the fastest time interval of interest to assure convergence without resort to adaptive time stepping. As an example, one might have to use a time step (parameter "delmax" in HSPICE[26]) that is three orders of magnitude smaller than the period of the RF signal. Hence, for a 1-GHz RF input, one might need to use a 1-ps time step. It is this combination of iteration, tight time step, and numerical tolerance problems that causes IP3 simulations to execute so slowly.[27] Again, as with the amplifier case, alternatives to time-domain simulators have evolved in response to these problems.

Additional Linearization Techniques

Because the linearity of these current-mode mixers is controlled primarily by the quality of the transconductance, it is worthwhile to consider additional ways to extend linearity. Philosophically, there are four methods for doing so: predistortion, feedback, feedforward, and piecewise approximation. These techniques can be used alone or in combination. What follows is a representative (but hardly exhaustive) set of examples of these methods.

Predistortion cascades two nonlinearities that are inverses of each other, and it shares with feedforward the need for careful matching. Predistortion is actually nearly ubiquitous, as it is the principle underlying the operation of current mirrors. In a mirror, an input current is converted to a gate-to-source voltage through some nonlinear function that is then undone to produce an output current exactly proportional to the input. Predistortion is also fundamental to the operation of true Gilbert mixers, where a pair of junctions computes the inverse hyperbolic tangent of an input differential current, and a differential pair subsequently undoes that nonlinearity.

Negative feedback computes an estimate of error, inverts it, and adds it back to the input, thereby helping to cancel the errors that distortion represents. The reduction in distortion is large as long as the loop transmission magnitude is large. Because a negative feedback system computes the error a posteriori, the overall closed-loop

[26] HSPICE is a trademark of the Meta-Software Corporation.

[27] Measuring these quantities in the laboratory also requires some care. As with the simulation, the amplitudes of the two input tones must be low enough to avoid excitation of higher-order nonlinearities (which would cause a slope of other than +3) yet sufficiently larger than the noise floor.

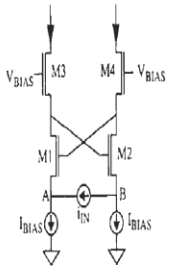


FIGURE 13.13. MOSFET cross-quad.

bandwidth must be kept a small fraction of the inherent bandwidth capabilities of the elements comprising the system; otherwise, the a posteriori estimate will be irrelevant at best and destabilizing at worst. The series feedback examples of this chapter are popular methods for linearizing high-frequency transconductors.

Contrary to common prejudice, positive feedback cannot be precluded as a linearizing technique. Furthermore, since loop transmission magnitudes must be less than unity to guarantee stability, the bandwidth penalty is much less severe than for negative feedback. As an illustrative example, the *cross-quad*, adapted from its bipolar progenitor, uses positive feedback to synthesize a virtual short-circuit; see Figure 13.13.

To show that this connection presents a short circuit to an applied current, i_{in}, consider how the voltages at the sources of M_1 and M_2 change as i_{in} changes. As i_{in} increases, the gate–source voltages of M_2 and M_4 increase by an equal amount, while those of M_1 and M_3 similarly decrease. The voltage at node A is:

$$V_{BIAS} - (V_{gs4} + V_{gs1}). \tag{21}$$

Similarly, the voltage at node B is:

$$V_{BIAS} - (V_{gs3} + V_{gs2}). \tag{22}$$

That is, the voltage at each source terminal is below V_{BIAS} by an amount equal to the sum of a high V_{gs} and a low V_{gs}. Hence, the two source voltages are always equal; the circuit synthesizes a virtual short circuit.[28]

Such a short circuit can be used to shift the burden of linearity away from active elements to a passive element, such as a resistance; this is shown in Figure 13.14. Because nodes A and B are at the same potential, the current injected into A is equal to v_{in}/R_s. This injected current is thus perfectly proportional to the input voltage and is recovered as a differential output current at the drains of M_3 and M_4.

[28] This analysis neglects body effect. Practical implementations do not work quite ideally as a result.

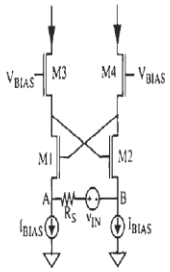


FIGURE 13.14. Cross-quad transconductor.

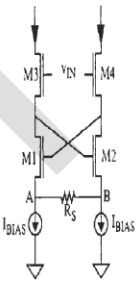


FIGURE 13.15. Alternate connection of cross-quad transconductor.

A variation of the cross-quad applies the input voltage across the gates of the top pair, as seen in Figure 13.15. The value of the transconductance is still equal to the conductance of R_s.

Feedforward is another linearization technique; it computes an estimate of the error at the same time the system processes the signal, thereby evading the bandwidth and stability problems of negative feedback. However, the error computation and cancellation then depend on matching, so the maximum practical distortion reduction tends to be substantially less than generally attainable with negative feedback. Feedforward is most attractive at high frequencies, where negative feedback becomes less effective owing to the insufficiency of loop transmission.

An example of feedforward correction applied to a transconductor is an adaptation of Pat Quinn's bipolar "cascomp" circuit[29] (Figure 13.16). As can be seen, this transconductor consists of a cascoded differential pair to which an additional

[29] "Feedforward Amplifier," U.S. Patent #4,146,844, issued 27 March 1979, reissued 1984.

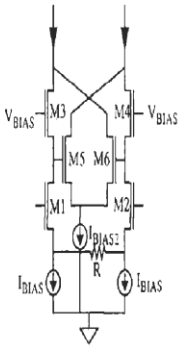


FIGURE 13.16. MOSFET cascomp.

differential pair has been added. Some linearization is provided by the source degeneration resistor R, but significant nonlinearity remains in the transconductance of inner differential pair M_1–M_2. To see this explicitly, consider that the voltage across the resistor is the input voltage minus the difference in gate-to-source voltages of M_1 and M_2:

$$V_R = v_{in} - (v_{gs1} - v_{gs2}) = v_{in} - \Delta v_{gs1}. \tag{23}$$

The goal is to have a differential output current precisely proportional to v_{in}, so any Δv_{gs} represents an error. The cascoding pair possesses the same Δv_{gs} as the input pair, which is measured by the inner differential pair. A current proportional to this error is subtracted from the main current to linearize the transconductance. The name "cascomp" derives from this combination of a cascode and error compensation. Although the inner pair is shown as an ordinary differential pair for simplicity, it is frequently advantageous to linearize it to increase the error correction range.

Another nonfeedback approach is piecewise approximation, which exploits the observation that virtually any system is linear over some sufficiently small range. It divides responsibility for linearity among several systems, each of which is active only over a small enough range so that the composite exhibits linearity over an extended range.

Gilbert's bipolar "multi-tanh"[30] arrangement is an example of piecewise approximation. In MOS form, it appears as shown in Figure 13.17. Each of the three differential pairs behaves as a reasonably linear transconductance over an input voltage range centered about V_B, 0, and $-V_B$, respectively. For input voltages near zero, the transconductance is provided by the middle pair and is roughly constant for small enough v_{IN}. As the input voltage deviates significantly from zero, the tail current

[30] The name derives from the fact that the transfer characteristic of a bipolar differential pair is a hyperbolic tangent.

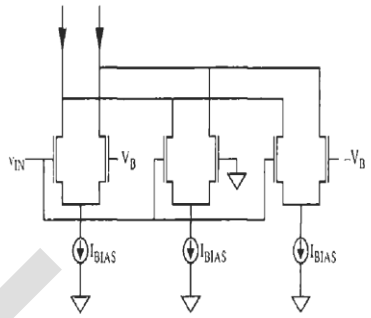


FIGURE 13.17. CMOS g_m cell.

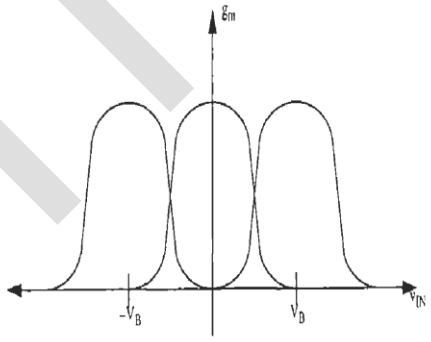


FIGURE 13.18. Illustration of linearization by piecewise approximation.

eventually steers almost completely to one side of the middle pair. However, with an appropriate selection of bias voltage V_B, one of the outer pairs takes over and continues to contribute an increase in output current; see Figure 13.18.

The overall transconductance is the sum of the individual offset transconductances and can be made roughly constant over an almost arbitrarily large range by using a sufficient number of additional differential pairs, each offset appropriately. The tradeoff is an increase in power dissipation and input capacitance.

13.4.3 POTENTIOMETRIC MIXERS

Gilbert-type mixers first convert an incoming RF voltage into a current through a transconductor, whose linearity and noise figure set a firm bound on the overall mixer linearity and noise figure. An alternative to using voltage-controlled current sources in *V–I* converters is to use voltage-controlled resistances. For example, consider varying the resistance of a triode-region MOSFET in a manner inversely proportional to the incoming RF signal. If the voltage between drain and source is maintained at a

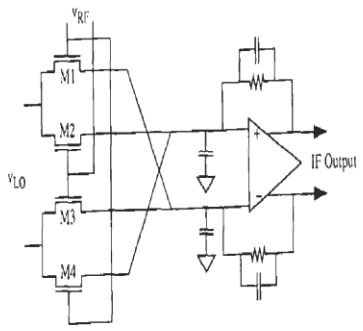


FIGURE 13.19. Potentiometric mixer.

fixed value, the current flowing through the device will be a faithful replica of the RF voltage, and if v_{ds} varies with the LO then the current will be proportional to the product of the LO and RF signals. One possible implementation of this idea is sketched in Figure 13.19.[31] The four MOSFETs perform the mixing, while the capacitors remove the sum frequency component as well as higher-order products.

The RF input drives the gates of the transistors, while the LO drives the sources. A simplified analysis assumes that the resistances of the transistors are inversely proportional to the RF signal. In that case, the current through the devices is

$$i_{in} = \frac{v_{LO}}{r_{ds}} \approx v_{LO} \cdot \mu C_{ox} \frac{W}{L}[(v_{RF} - V_T) - v_{LO}] \approx K \cdot v_{LO} \cdot v_{rf}. \qquad (24)$$

Because the current is then the result of a multiplication of the RF and LO signals, there are components at the sum and difference frequencies, as desired. This current flows through the feedback resistors so that the IF signal is available as an output voltage. The op-amp need only have enough bandwidth to handle the difference frequency component, since the sum component is filtered out by the four capacitances.

Note that, for good linearity, the gate overdrive must greatly exceed v_{LO}. Hence, v_{RF} must possess a sufficiently large DC component to satisfy this inequality for as large a value of v_{rf} that must be accommodated.

Practical mixers of this type may exhibit good linearity (e.g., 40 dBm IIP3) but high noise figures (e.g., 30 dB). The high noise figures are the result of the resistive thermal noise of the input FETs (which is worst when the signal levels are small) and the difficulty of providing a good noise match with the broadband op-amp. As a consequence, the overall dynamic range of this type of mixer is typically about the same as conventional Gilbert-type current-mode mixers.

[31] J. Crols and M. Steyaert, "A 1.5GHz Highly Linear CMOS Downconversion Mixer," *IEEE J. Solid-State Circuits*, v. 30, no. 7, July 1995, pp. 736–42.

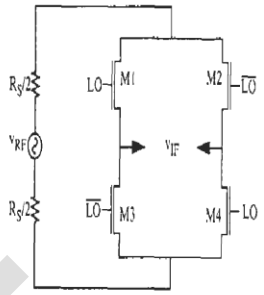


FIGURE 13.20. Simple double-balanced passive CMOS mixer.

13.4.4 PASSIVE DOUBLE-BALANCED MIXER

So far, we've examined active mixers only, with their attendant need for linear transconduction. However, passive mixers have some attractive properties, such as the potential for extremely low-power operation. Considering that CMOS technology offers excellent switches, high-performance multipliers based on switching are naturally realized in CMOS form.

In the active mixers considered so far, representations of the RF signal in the form of currents, rather than the RF voltages themselves, are effectively multiplied by a square-wave version of the local oscillator. An alternative that avoids the *V–I* conversion problem is to switch the RF signal directly in the voltage domain. This option is considerably easier to exercise in CMOS than bipolar form, which is why bipolar mixers are almost exclusively of the active, current-mode type.

The simplest passive commutating CMOS mixer consists of four switches in a bridge configuration (see Figure 13.20). The switches are driven by local oscillator signals in antiphase, so that only one diagonal pair of transistors is conducting at any given time. When M_1 and M_4 are on, v_{IF} equals v_{RF}, and when M_2 and M_3 are conducting, v_{IF} equals $-v_{RF}$. A fully equivalent description is that this mixer multiplies the incoming RF signal by a unit-amplitude square wave whose frequency is that of the local oscillator. Hence, the output contains many mixing products that result from the odd-harmonic Fourier components of the square wave.[32] Luckily, these are often readily filtered out, as discussed previously.

The voltage conversion gain of this basic cell is easy to compute from the foregoing description. Assuming multiplication by a unit-amplitude square wave, we may immediately write

$$G_C = 2/\pi. \qquad (25)$$

[32] This situation is the same as with the current-mode mixers, however. Also, even harmonics of the LO terms may be nonzero if the duty cycle of the square wave is not exactly 50%.

Here, the $2/\pi$ factor again results from splitting the IF energy evenly between the sum and difference components.[33]

In practice, the actual voltage conversion gain may differ somewhat from $2/\pi$ because real transistors do not switch in zero time. Hence, the incoming RF signal is not multiplied by a pure square-wave signal in general. Perhaps contrary to intuition, however, the effect of this departure from ideal assumptions is usually to *increase* the voltage conversion gain above $2/\pi$.

A more general expression for the voltage conversion gain is somewhat cumbersome to derive, so we will state only the relevant insights here.[34] The output of the mixer may be treated as the product of three time-varying components and a scaling factor:

$$v_{IF}(t) = v_{RF}(t) \cdot \left[\frac{g_T(t)}{g_{T\,max}} \cdot m(t)\right] \cdot \left[\frac{g_{T\,max}}{\overline{g_T}}\right]. \tag{26}$$

The function $g_T(t)$ is the time-varying Thévenin-equivalent conductance as viewed from the IF port, while $g_{T\,max}$ and $\overline{g_T}$ are the maximum and average values, respectively, of $g_T(t)$. The *mixing function*, $m(t)$, is defined by

$$m(t) = \frac{g(t) - g(t - T_{LO}/2)}{g(t) + g(t - T_{LO}/2)}, \tag{27}$$

where $g(t)$ is conductance of each switch and T_{LO} is the period of the LO drive. The mixing function has no DC component, is periodic in T_{LO}, and has only odd harmonic content because of its half-wave symmetry.

The Fourier transform of the first bracketed term in Eqn. 26 has a value of $2/\pi$ at the LO frequency for a square-wave drive (as asserted earlier) and a value of $1/2$ for a sinusoidal drive, so the effective mixing function indeed contributes a higher conversion gain for a square-wave drive. However, the second bracketed term is unity for a square-wave drive (because the peak and average conductances are equal) but $\pi/2$ for a sinusoidal drive. The overall conversion gain is greater with a sinusoidal drive because the second term more than compensates for the smaller contribution by the (effective) mixing function. The difference is not particularly large, however. With a sinusoidal drive, the conversion gain is $\pi/4$ (−2.1 dB), compared with the $2/\pi$ gain (−3.92 dB) obtained with the square-wave drive.

Because of the spectrum of the (effective) mixing function, undesirable products can appear at the IF port of this type of mixer. The subject of filtering therefore deserves careful consideration, especially in connection with the issue of input and

[33] If we assume equal source and load terminations, then this gain corresponds to a 3.92-dB voltage and power loss. Many practical implementations, such as the discrete passive mixers discussed in Section 13.6, typically exhibit a somewhat greater conversion loss than this theoretical limit because of additional sources of attenuation (e.g. nonzero switch drop, skin effect loss, etc.). Common conversion losses for mixers of this type are in the neighborhood of 5 dB to 6 dB.

[34] For a detailed derivation, see A. Shahani et. al., "A 12mW Wide Dynamic Range CMOS Front-End for a Portable GPS Receiver," *IEEE J. Solid-State Circuits*, December 1997.

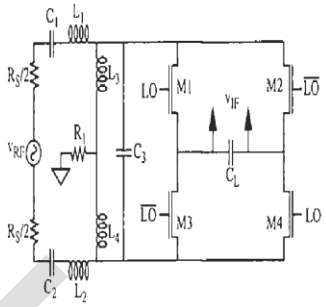


FIGURE 13.21. Low-noise, narrowband passive mixer.

output terminations. In discrete designs, the source and load impedances are usually real and well-defined (50 Ω, for example), but the sources and loads for IC mixers are usually on-chip and not at all standardized. Far from a liability, this lack of standardization is a degree of freedom that the IC engineer can exploit to improve performance. As a specific example, reactive source and load terminations might be preferable because they do not generate noise. Because it is difficult to obtain broadband operation with reactances, narrowband operation is implied for most practical mixers with reactive terminations. Fortunately, there are many applications for which this restriction is not a serious limitation.

In CMOS implementations, the load at the IF port of the mixer is frequently capacitive to an excellent approximation. In such cases, the loading is easily accommodated as forming a simple low-pass filter in conjunction with the resistance of the switches. A detailed analysis[35] reveals that the transfer function of this filter is simply

$$H(s) = \left[s\frac{C_L}{\overline{g_T}} + 1\right]^{-1}. \tag{28}$$

We see that the pole frequency is simply the ratio of the average conductance (again, as viewed from the IF port, back through the switches) to the load capacitance. This inherent filtering action may be exploited to provide a much desired attenuation of unwanted mixer products.

A somewhat more elaborate passive mixer that further exploits the freedom to select source and load terminations appears as Figure 13.21.[36] Note that this mixer assumes a capacitive load, represented as C_L in the schematic. This assumption reflects the typical situation in fully integrated CMOS circuits, and it stands in contrast with the resistive terminations common in discrete designs. A capacitive load

[35] Shahani et al., ibid.

[36] This example is adapted from A. Shahani et al., "A 12mW Wide Dynamic Range CMOS Front-End for a Portable GPS Receiver," *ISSCC Digest of Technical Papers*, February 1997, pp. 368–9.

generates no thermal noise of its own, and also helps filter out high-frequency noise and distortion.

The input network consists of an L-match in cascade with a parallel tank. The L-match, comprising L_1 and part of the tank capacitance, provides an impedance transformation that moderately boosts the RF signal voltage to help reduce the voltage conversion loss. The parallel tank, formed by L_3 and C_3+C_L, filters out-of-band noise and distortion components present at the input and generated by the mixer itself. Resistor R_1 sets the common-mode potential for the input circuit. Because any nonlinearity in the tank capacitance reduces IP3, C_3 is best implemented as a metal–metal capacitor. To reduce the area consumed, a good choice is to use a lateral flux or fractal capacitor.

Because of the small voltage boost provided by the L-match, the voltage conversion loss can be somewhat better than the 3.92 dB that a simple switch bridge would exhibit ideally. As an example, one implementation in a 0.35-μm technology exhibits a 3.6-dB voltage conversion loss with a 1.6-GHz RF and 1.4-GHz LO.[37]

Both noise figure and IP3 are strong functions of the LO drive, since the resistance of the switches in the "on" state must be kept low and constant to optimize both parameters. The IP3 is also a function of the amount of voltage boost provided by the L-match. This boost may be adjusted downward to trade conversion gain for improved IP3 and, in some cases, it may be appropriate to remove the L-match altogether. Typical SSB noise figures of 10 dB and input IP3 of 10 dBm are readily achievable with an LO drive amplitude of 300 mV.[38] As a crude estimate, the SSB noise figure of this type of mixer is approximately equal to the power conversion loss.

The first edition of this book contained an assertion that the absence of DC bias current in a passive mixer implied the absence of $1/f$ noise. That assertion is not quite correct, because a mixer is a periodically time-varying system. As such, noise centered at integer multiples of the local oscillator can fold down to DC, for example. Thus, $1/f$ noise may still appear at the output of the mixer without requiring any DC bias in the mixer itself.

In cases where it is important to minimize $1/f$ noise in the mixer output, it is generally helpful to (a) reduce the LO drive to the minimum value consistent with acceptable mixing action and (b) design the local oscillator carefully to minimize its close-in phase noise in particular (a topic we take up in detail in Chapter 18). These considerations are particularly important in the design of receivers that are sensitive to $1/f$ noise, which include the direct-conversion (also known as the homodyne or zero-IF) receiver and the low-IF receiver.

To reduce the power consumed by the LO drivers, the gate capacitance of the switches may be resonated with an inductor (for narrowband applications), resulting in a power reduction by a factor of Q^2. It is trivial to reduce the power to the order of a milliwatt or less, even at gigahertz frequencies.

[37] A. Shahani et al. (February 1997), ibid.
[38] This value applies to a sine-wave LO.

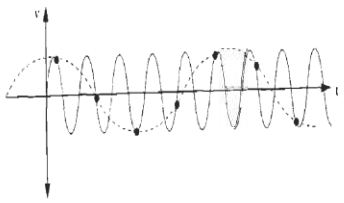


FIGURE 13.22. Illustration of subsampling.

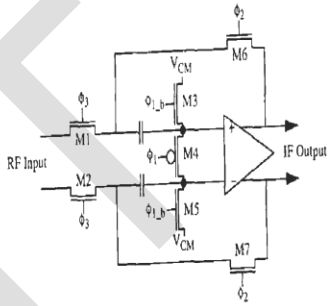


FIGURE 13.23. Track-and-hold subsampling mixer (simplified).

13.5 SUBSAMPLING MIXERS

The high quality of CMOS switches has also been exploited to realize what are sometimes called *subsampling* mixers. This type of mixer exploits the observation that the information bandwidth of modulation is necessarily lower than the carrier frequency. Hence, one may satisfy the Nyquist criterion with a sampling rate that is also lower than the carrier frequency, effecting downconversion in the process.

As can be seen in Figure 13.22, the higher-frequency signal is sampled at the instants indicated by the dots, while the downconverted signal is shown as the lower-frequency reconstruction. The theoretical advantage of this approach is that it may be easier to realize samplers that operate at a frequency well below that of the incoming RF signal. From the figure, it should be clear that a properly designed track-and-hold circuit (Figure 13.23) serves as a subsampling mixer.[39]

[39] This example is adapted from P. Chan et al., "A Highly Linear 1-GHz CMOS Downconversion Mixer," *IEEE J. Solid-State Circuits*, December 1993.

In the sample (track) mode, transistors M_1 through M_5 are turned on while transistors M_6 and M_7 are placed in the "off" state. Devices M_3, M_4, and M_5 put a voltage equal to the common-mode voltage level V_{CM} on the right-hand terminals of the sampling capacitors, while input switches M_1 and M_2 connect the capacitors to the RF input signal. Because M_6 and M_7 are open, the op-amp is irrelevant in this tracking mode, and the tracking bandwidth is simply set by the RC time constant formed by the total switch resistance and sampling (and parasitic) capacitance. Because the system operates open-loop in this mode, it is easy to obtain tracking bandwidths far in excess of what can be achieved with a feedback structure. For example, it is trivial to obtain tracking bandwidths greater than 1 GHz in a 1-μm technology.

In the hold mode, all switch states are reversed, so that the only conducting transistors are the two feedback devices M_6 and M_7. In this mode, the circuit degenerates to a pair of charged capacitors feeding back around the op-amp. The settling time of this system need only be fast relative to the (slow) sampling period, rather than to the RF signal period. Thus, the bandwidth penalty associated with feedback is not serious.

Although a subsampler is clocked at a relatively low frequency, the sampler must still possess good time resolution or else sampling errors result. Therefore, beyond an adequate tracking bandwidth, one must also have low aperture jitter (i.e., low uncertainty in the sampling instants), and this requirement places extraordinary demands on the phase noise of the sampling clock. Hence, even though the frequency of the sampling clock need only satisfy the Nyquist criterion applied to the *modulation* bandwidth, its absolute time jitter must be a tiny fraction of the *carrier* period.

Another problem is that the sampling operation converts more than just the signal. Noise at the input to the sampler undergoes folding into the IF band, resulting in an unfortunate noise boost roughly equal to the ratio of RF and IF bandwidths. Because the RF bandwidth typically exceeds the IF bandwidth by large amounts, subsampling mixers can exhibit large noise figures (e.g., 25-dB SSB NF). The large linearity implied by the high third-order intercepts often exhibited by these types of mixers is offset by their poor noise performance, so that the dynamic range of the mixer is frequently no better (or even worse) than what one may achieve with conventional architectures. In fact, the noise and IP3 performance of many subsampling mixers can be replicated by preceding a conventional mixer with a resistive divider. In principle, an LNA with sufficient gain may be used to overcome the mixer's noise, but it is difficult in practice to realize LNAs that provide simultaneously high gain and high linearity, so again overall (system) dynamic range may actually suffer. As a result of these problems, one must take great care in applying subsampling.

13.6 APPENDIX: DIODE-RING MIXERS

This appendix considers a number of passive mixers that are common in discrete implementations. The four-diode double-balanced mixer has particularly good characteristics and is nearly ubiquitous in high-performance discrete equipment.

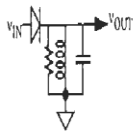


FIGURE 13.24. Simple diode mixer.

13.6.1 SINGLE-DIODE MIXER

The simplest and oldest passive mixer uses a single diode, as seen in Figure 13.24. In this circuit, the output RLC tank is tuned to the desired IF, and v_{IN} is the sum of RF, LO, and DC bias components. The nonlinear V–I characteristic of the diode provides diode currents at a number of harmonic and intermodulation frequencies, and the tank selects only those at the IF.

It is tempting to reject this circuit as hopelessly unsophisticated. It does not provide any isolation, and it doesn't provide any conversion gain, for example. However, at the highest frequencies, it may be difficult to exploit other types of nonlinearities, and such simple mixers may be suitable. In fact, all of the detectors[40] for radar sets developed in WWII were single-diode circuits.[41] Additionally, many early UHF television tuners also used mixers of this type. Much of the modern work in the millimeter-wave bands simply would not be possible without such mixers.

As another note on this circuit, it can be used as a crude demodulator for AM signals if the input signal is the AM signal (at either RF or IF). When used in this manner, the output inductor is removed entirely, no LO is used, and a simple RC network provides the output filtering. Millions of "crystal" radio sets used this type of detector (known in this context as an envelope detector), and even most AM superheterodyne radios built today use a single-diode demodulator.

13.6.2 TWO-DIODE MIXERS

There are several other ways to use diodes as mixers. As we'll see, it will appear that a diode bridge can be used as just about anything, depending on which terminals are defined as input and output and which way the diodes point.[42]

[40] We will use the terms "detector" and "demodulator" interchangeably.

[41] The birth of modern semiconductor technology can be traced directly to the development of microwave diodes for radar. By the end of WWII, point-contact microwave diodes capable of operation well into the gigahertz range became widely available.

[42] Diodes can even be used to provide *gain* by exploiting the nonlinear junction capacitance to make a thing known as a *parametric amplifier*. The nonlinearity can be used to transfer energy from a local oscillator (known as the *pump* in par-amp parlance) to the signal, instead of the more conventional transfer of power from a DC source to the signal frequency. Parametric amplifiers can be extremely low-noise devices, since only pure reactances are needed to make them work.

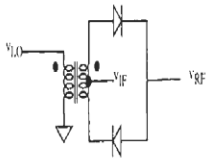


FIGURE 13.25. Single-balanced diode mixer.

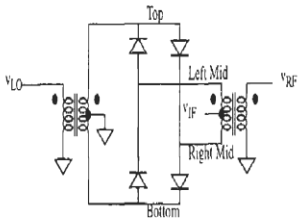


FIGURE 13.26. Double-balanced diode mixer.

With two diodes, it's possible to construct a single-balanced mixer. In this case, one may obtain isolation between LO and IF, but there is poor RF–IF isolation; see Figure 13.25. Assume that the LO drive is sufficient to make the diodes act as switches, regardless of the magnitude of the RF input. With a positive value for v_{LO}, both diodes will be on (note the reference dots on the transformer windings), effectively connecting v_{RF} to the IF output. When v_{LO} goes negative, the diodes open-circuit and disconnect v_{RF}. Hence, this mixer acts the same as the active commutating mixer studied previously.

The poor RF–IF isolation should be self-evident from the comment that the diodes connect the RF and IF ports together whenever the diodes are on. Similarly, it should be evident that symmetry guarantees excellent RF–LO isolation. Whenever the diodes are on, the RF voltage can only develop a common-mode voltage across the transformer windings, so no voltage can be induced at the LO port.

13.6.3 DOUBLE-BALANCED DIODE MIXER

By adding two more diodes and one more transformer, we can construct a double-balanced mixer to provide isolation among all ports (see Figure 13.26). Once again, assume that the LO drive is sufficient to cause the diodes to act as switches. In the circuit shown, the left pair of diodes is on whenever the LO drive is negative, whereas the right pair of diodes is on whenever the LO drive is positive.

With the LO drive positive, the voltage at "Right Mid" must be zero by symmetry, since the center tap of the input transformer is tied to ground. Thus, v_{IF} equals v_{RF} (again, note the polarity dots). With the LO drive negative, it is "Left Mid" that has a zero potential, and v_{IF} equals $-v_{RF}$. Hence, this mixer effectively multiplies v_{RF} by a unit-amplitude square wave whose frequency is that of the LO.

Isolation is guaranteed by the symmetry of the circuit. The LO drive forces a zero potential at either the top or bottom terminal of the output transformer, as noted previously. If the RF input is zero, there will be no IF output. Hence, this configuration provides LO–IF isolation. Similarly, we can show LO–RF isolation by considering a zero IF input. Since, again, there is a zero potential at either the top or bottom terminal of the output transformer, there will be no primary voltage and therefore no secondary voltage.

These passive mixers are available in discrete form, and perform exceptionally well. The upper limit on the dynamic range is typically constrained by diode breakdown, and isolation is a function of the matching levels achieved.

With a single quad of diodes, typical double-balanced mixers routinely achieve conversion losses in the neighborhood of 6 dB and isolation of at least 30 dB, and they can accommodate RF inputs of up to 1 dBm at the 1-dB compression point while requiring an LO drive of 7 dBm. Higher RF levels can be accommodated if series connections of diodes are used in place of each diode of Figure 13.26, the drawback being an increased LO drive requirement to guarantee switching operation of the diodes. Using a total of sixteen diodes, for example, extends the RF input range to around 9 dBm but also requires a whopping 13 dBm of LO drive.

13.6.4 FINAL NOTE ON DIODE MIXERS

When actually using such mixers, one should be aware that it is critically important to terminate all ports in the proper characteristic impedance – not only at the RF, IF, and desired LO frequencies, but at the image frequencies as well. If only narrowband terminations are used, it is possible for reflections of various intermodulation products to degrade performance seriously. Hence, it is generally insufficient merely to use a standard *RLC* tank as an output bandpass filter without an intermediate buffering stage to guarantee a broadband resistive termination. Failure to satisfy this condition can be the source of many perplexing phenomena.

PROBLEM SET FOR MIXERS

PROBLEM 1 Using the device models from Chapter 5, design a single-balanced mixer with inductive source degeneration to achieve an IIP3 of +6 dBm. What is the conversion transconductance?

PROBLEM 2 We've seen the utility of synthesizing a virtual short-circuit for linearizing transconductances. Suppose that someone were to propose the alternative circuit of Figure 13.27.

UNIT-4

RF POWER AMPLIFIERS

RF Power Amplifiers, Class A, AB, B, C amplifiers, Class D, E, F amplifiers, RF Power amplifier design examples, Voltage controlled oscillators, Resonators, Negative resistance oscillators, Phase locked loops, Linear zed PLL models, Phase detectors, charge pumps, Loop filters, and PLL design examples

Power Amplifiers:-

An audio **power amplifier** (or **power** amp) is an electronic **amplifier** that amplifies low-**power** electronic audio signals (signals composed primarily of frequencies between 20 - 20 000 Hz, the human range of hearing) to a level that is strong enough for driving loudspeakers and making the signal–whether it is recorded music.

Types:-

Class-A, B, C, AB, D, E, F Power amplifiers.

E Amplifiers:-

The class-E/F amplifier is a highly efficient switching power amplifier, typically used at such high frequencies that the switching time becomes comparable to the duty time. As said in the class-D amplifier, the transistor is connected via a serial LC circuit to the load, and connected via a large L (inductor) to the supply voltage. The supply voltage is connected to ground via a large capacitor to prevent any RF signals leaking into the supply. The class-E amplifier adds a C (capacitor) between the transistor and ground and uses a defined L_1 to connect to the supply voltage.

F Amplifiers:-For a **Class F** power **amplifier** where the active device is biased under **Class** B conditions (with conduction angle equal to p), the one-half sine-wave current waveform contains only even harmonics.

Voltage Controlled Oscillator (VCO):-

A **voltage-controlled oscillator** or **VCO** is an electronic oscillator whose oscillation frequency is controlled by a voltage input. The applied input voltage determines the instantaneous oscillation frequency.

Voltage Controlled Oscillator (VCO) A VCO is a voltage controlled oscillator, whose output frequency ωo is linearly proportional to the control voltage VC generated by the PD.

Resonators:-

1) Crystal Resonators 2) Ceramic Resonators 3) Dielectric Resonators

Quartz Crystal Resonator:-

Quartz crystals can be used as high-quality electromechanical resonators. Their piezoelectric properties allow them to be used as frequency-control elements in crystal oscillators. Quartz crystals offer high Q and superior frequency stability. In fact, their high Q is the main reason why crystal oscillators are often employed instead of LC oscillators.Piezoelectric materials have the capability to convert mechanical energy into electrical energy and vice versa. When a mechanical stress is applied, an electric charge is generated. This electric charge is proportional to the applied mechanical stress. The same material becomes strained when an electric field is applied.

Negative Resistance Oscillator:-

These are dynatron and tunnel diode **oscillators**. Dynatron operates in the **negative resistance** region of the characteristics of a tutored, which is coupled to an L-C tank circuit. Tunnel diode **oscillator** makes use of a tunnel diode for producing oscillations.

PLL (Phase-Locked Loop):-

It is basically a flip flop consisting of a phase detector, a low pass filter (LPF), and a Voltage Controlled Oscillator (VCO). The input signal V_i with an input frequency f_i is passed through a phase detector. The DC level is then passed on to a VCO.

Phase Detectors:-

A **phase detector** or **phase** comparator is a frequency mixer, analog multiplier or logic circuit that generates a voltage signal which represents the difference in **phase**between two signal inputs. It is an essential element of the**phase-locked loop** (PLL).

Loop Filters:-

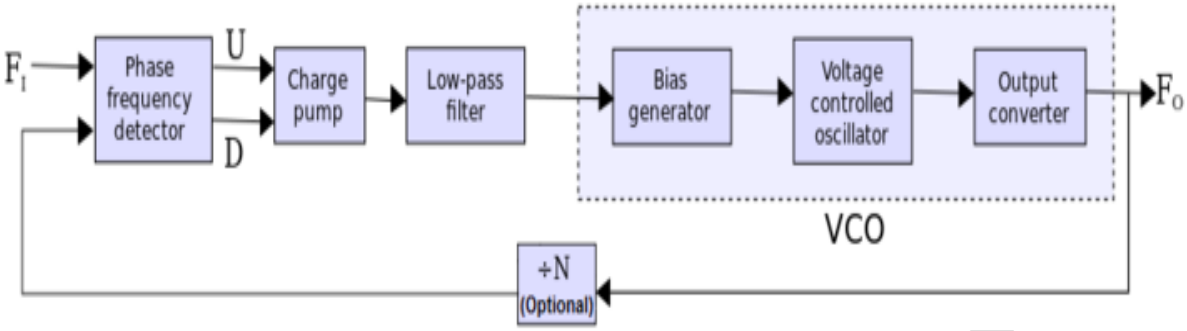
Loop Filter (LF) The filtering operation of the error voltage (coming out from the PD) is performed by the loop filter (LF). The output of PD consists of a dc component superimposed with an ac component.

The sum frequency is also produced and this will fall at a point equal to twice the frequency of the reference. If this signal is not attenuated it will reach the control voltage input to the VCO and give rise to spurious signals.

Charge Pumps:-

A **charge pump** is a kind of DC to DC converter that uses capacitors as energy-storage elements to create either a higher- or lower-voltage power source.

Block diagram of loop filter with PLL:-



The inclusion of a **loop filter** at the output of the **charge pump** serves two functions.

UNIT-5

FREQUENCY SYNTHESIS AND OSCILLATORS

Frequency synthesis and oscillators, Frequency division, integer-N synthesis, Fractional frequency, synthesis, Phase noise, General considerations, and Circuit examples, Radio architectures, GSM radio architectures, CDMA, UMTS radio architectures.

Frequency Divider:-

The IDT clock buffer (fan-out buffer), clock divider and clock multiplexer portfolio includes devices with up to 27 outputs. Differential outputs such as LVPECL, LVDS, HCSL, CML, HSTL, as well as selectable outputs, are supported for output frequencies up to 3.2 GHz and single-ended LVCMOS outputs for frequencies up to 350 MHz. Some buffers are available with mixed output signaling.

Applications of PLL's:-

The PLL is a very versatile building block and is suitable for a variety of applications including:

1.) Demodulation and modulation

2.) Signal Conditioning

3.) Frequency synthesis

4.) Clock and data recovery

5.) Frequency translation

Fractional-N PLL

An unavoidable occurrence in digital PLL synthesis is that frequency multiplication (by N), raises the signal's phase noise by 20Log(N) dB. The main source of this noise is the noise characteristics of the phase detector's active circuitry. Because the phase detectors are typically the

dominant source of close-in phase noise, N becomes a limiting factor when determining the lowest possible phase noise performance of the output signal. A multiplication factor of N = 30,000 will add about 90 dB to the phase detector noise floor. 30,000 is a typical N value used by an integer PLL synthesizer for a cellular transceiver with 30 kHz channel spacing.

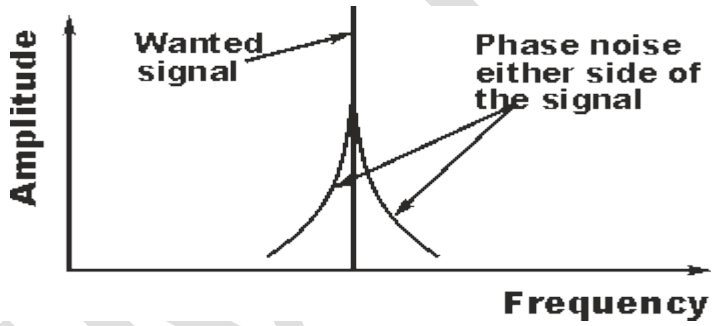
Integer-N PLL

Compared to the analog techniques used in the infancy of frequency synthesis, the modern PLL is now a mostly digital circuit. Figure 2 shows a typical block diagram of a PLL implemented with a TCXO reference.

This traditional digital PLL implementation will be termed "integer-N" to avoid confusion due to the addition of fractional-N technology.

Phase Noise:-

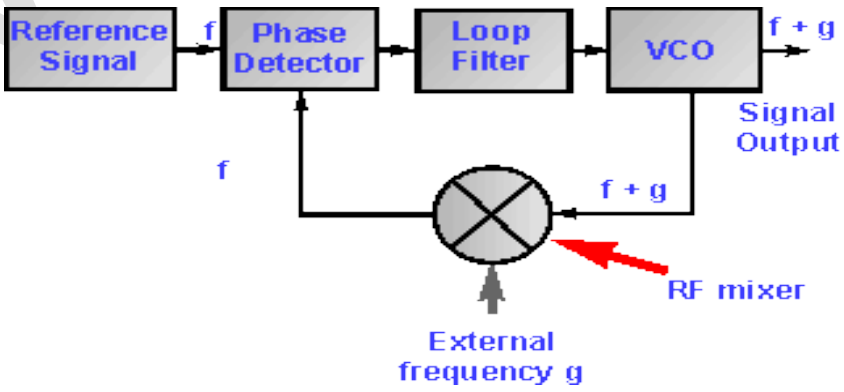
Phase noise is defined as the noise arising from the short term phase fluctuations that occur in a signal. The fluctuations manifest themselves as sidebands which appear as a noise spectrum spreading out either side of the signal.



Phase Jitter:-

Phase jitter: Phase jitter is the term used for looking at the phase fluctuations themselves, i.e. the deviations in the position of the phase against what would be expected from a pure signal at any given time. Accordingly phase jitter is measured in radians.

Basic block diagram of Frequency Synthesis:-



$$N_{eff} = \frac{A + B}{\frac{A}{N} + \frac{B}{N+1}}$$

Where:

Neff = overall division ratio

A = number of cycles divided by N

B = number of VCO cycles divided by N+1

Radio Architecture:-

The **Architecture** of **Radio** is data visualization, based on global open datasets of cell tower, Wi-Fi and satellite locations. The dataset includes almost 7 million cell towers, 19 million Wi-Fi routers and hundreds of satellites. The ***Architecture of Radio*** is data visualization, based on global open datasets of cell tower, Wi-Fi and satellite locations. Based on your GPS location the app shows a 360 degree visualization of signals around you.

GSM(Global System for Mobile):-

1. If you are in Europe or Asia and using a mobile phone, then most probably you are using GSM technology in your mobile phone.
2. GSM stands for **G**lobal **S**ystem for **M**obile Communication. It is a digital cellular technology used for transmitting mobile voice and data services.
3. The concept of GSM emerged from a cell-based mobile radio system at Bell Laboratories in the early 1970s.

Applications of GSM:-

Listed below are the features of GSM that account for its popularity and wide acceptance.

1) Improved spectrum efficiency

2) International roaming

3) Low-cost mobile sets and base stations (BSs)

4) High-quality speech.

CDMA (Code-Division Multiple Accesses):-

CDMA (Code-Division Multiple Access) refers to any of several protocols used in second-generation (2G) and third-generation (3G) wireless communications. As the term implies, CDMA is a form of multiplexing, which allows numerous signals to occupy a single transmission channel, optimizing the use of available bandwidth. The technology is used in ultra-high-frequency (UHF) cellular telephone systems in the 800-MHz and 1.9-GHz bands.

CDMA employs analog-to-digital conversion (ADC) in combination with spread spectrum technology.

The CDMA channel is nominally 1.23 MHz wide. CDMA networks use a scheme called soft handoff, which minimizes signal breakup as a handset passes from one cell to another. The combination of digital and spread-spectrum modes supports several times as many signals per unit bandwidth as analog modes. CDMA is compatible with other cellular technologies; this allows for nationwide roaming.

UMTS:-

The UMTS 3G architecture is required to provide a greater level of performance to that of the original GSM network. However as many networks had migrated through the use of GPRS and EDGE, they already had the ability to carry data. Accordingly many of the elements required for the WCDMA / UMTS network architecture were seen as a migration.

UMTS network architecture:-

The UMTS network architecture can be divided into three main elements:

1. ***User Equipment (UE):*** The User Equipment or UE is the name given to what was previous termed the mobile, or cellphone. The new name was chosen because the considerably greater functionality that the UE could have. It could also be anything between a mobile phone used for talking to a data terminal attached to a computer with no voice capability.
2. ***Radio Network Subsystem (RNS):*** The RNS also known as the UMTS Radio Access Network, UTRAN, is the equivalent of the previous Base Station Subsystem or BSS in GSM.
3. ***Core Network:*** The core network provides all the central processing and management for the system. It is the equivalent of the GSM Network Switching Subsystem or NSS.

The core network is then the overall entity that interfaces to external networks including the public phone network and other cellular telecommunications networks.