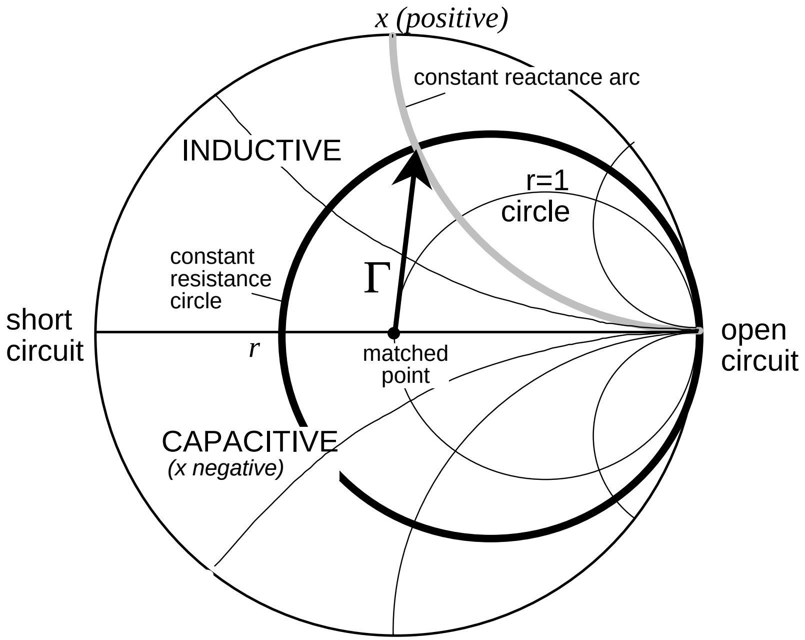
INTRODUCTION TO MATCHING TECHNIQUES & AMPLIFIER DESIGN

The Smith chart is a polar diagram on which the reflection coefficient of the load can be conveniently plotted. Since the reflection coefficient has a maximum value of ONE, the diagram is circular with a radius of one. It can be shown that circles of constant load resistance and reactance can be added to the graph : - this makes the chart immediately useful as a calculator for converting from Z to Γ and back again. To achieve this, the contours are all labelled with NORMALISED values of resistance, *r*, and reactance, *x*. Normalising means simply dividing by the system reference impedance Zo.

The key features with an example load are labelled below:-

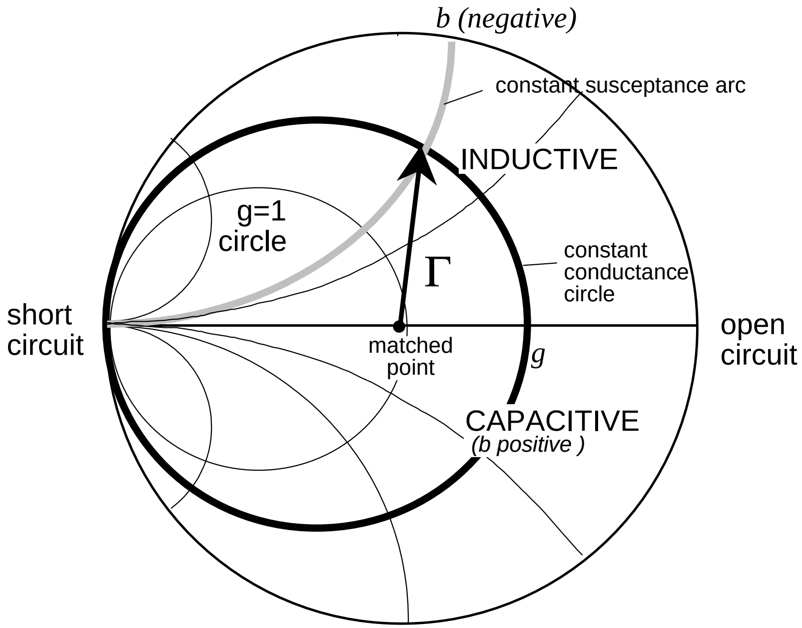


Impedance chart

The key feature that makes the Smith chart so valuable is that since it is based around Γ , it is very easy to cope with transmission-line problems: as you move along a transmission line the magnitude of Γ is fixed, so you are actually rotating around the centre of the chart with fixed radius. By convention, if you move away from the load you rotate clockwise, and vice versa.

The admittance chart

The Smith chart can also be used for admittance calculations. The admittance contours are shown below. Note that Γ has not changed, but the contours appear to have flipped over. The admittance chart is vital for shunt stub matching. **UNFORTUNATELY TEXT BOOKS DEAL WITH THE ADMITTANCE CHART IN DIFFERENT WAYS, CAUSING CONFUSION TO MANY STUDENTS.** It is possible to use the same chart (as per the previous page) for both Z and Y, and to convert from one to the other by mirror-imaging Γ through the centre (i.e. rotate 180 degrees). This has the same effect as flipping the contours, BUT the short/open points and all other features are also flipped, causing major confusion. Alternatively, use two charts, or a single chart with both admittance and impedance contours on (if you can handle it).

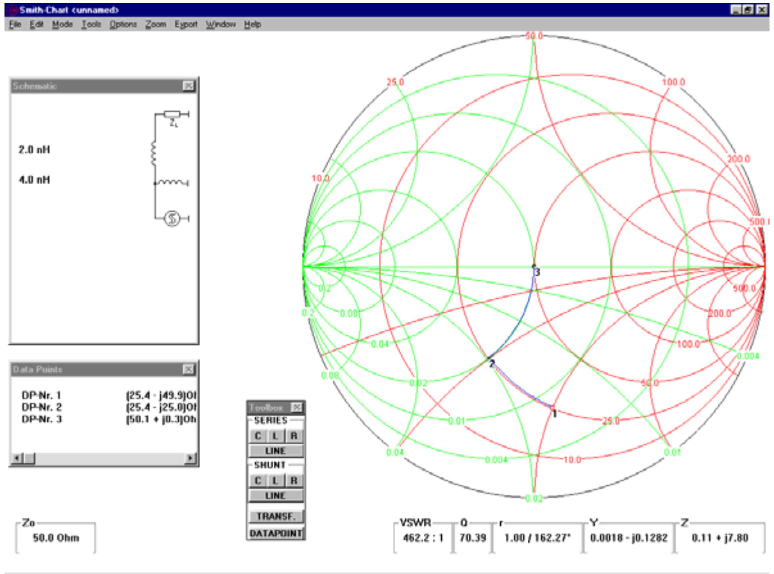


Admittance chart

“SMITH 182”   Windows Smith Chart Tool

This is the best free Smith chart tool that I have seen. It plots the impedance and admittance contours in read and green. They can individually be turned on or off. It is available from RF Globalnet.

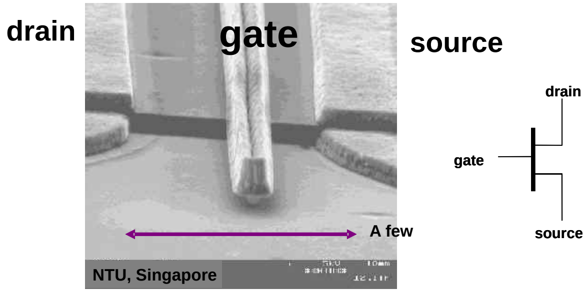
The example below shows a series/parallel inductor L-network. This will be studied as part of amplifier design.

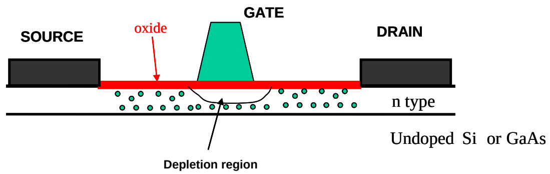


Screen dump of Smith182 programme showing Z and Y contours simultaneously

INTRODUCTION TO SMALL-SIGNAL AMPLIFIER DESIGN

Transistors: An RF FET:-

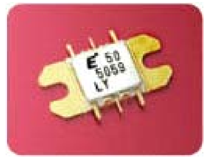




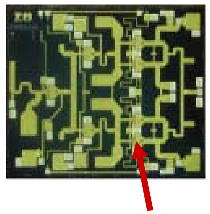
RF transistors come in many varieties and in different types of package:-

Bare chip with single transistor

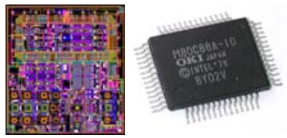
Single transistor in RF package



RF power transistors

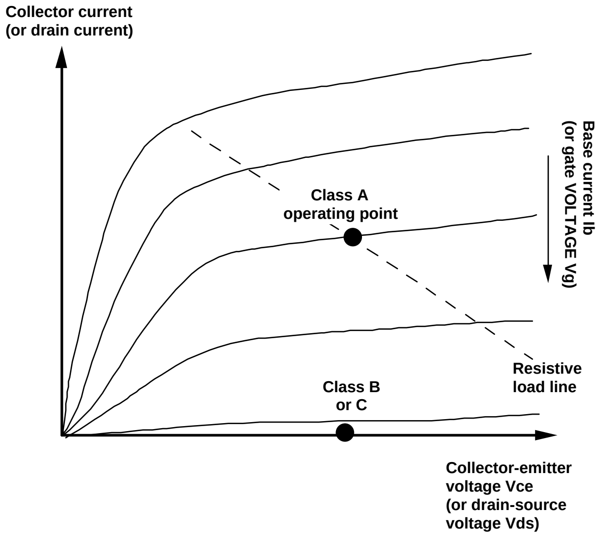


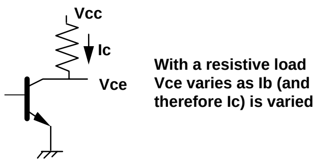
On a GaAs IC



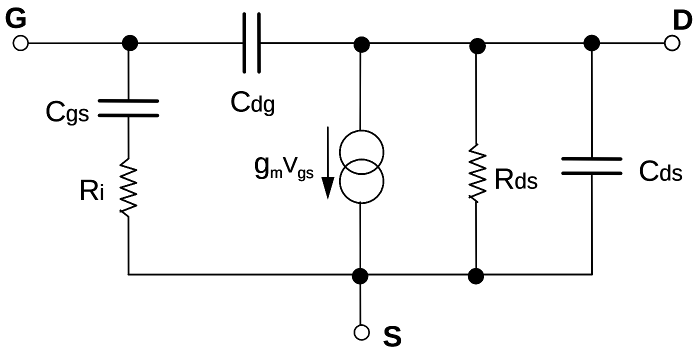
One of millions on a silicon chip

AMPLIFIER DESIGN: TRANSISTOR I-V CURVES AND CLASS **A**, **B**, and **C** OPERATION





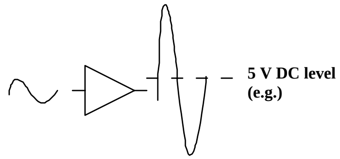
Small-signal behaviour can be summarised with an equivalent circuit model:-



Typical FET small signal equivalent circuit

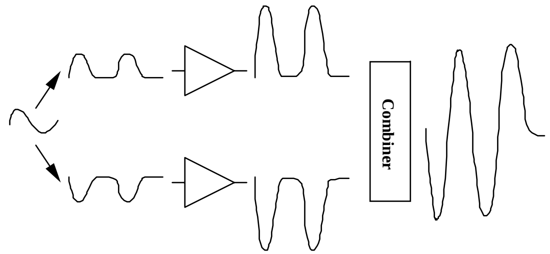
AMPLIFIER DESIGN: CLASS A, B and C OPERATION

In Class A operation the transistor is biased in the middle of the I-V curves so that the output voltage can swing up and down, around a non-zero DC level:-



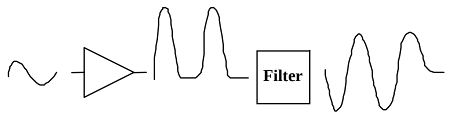
This is the simplest mode of operation to understand and is very linear. Unfortunately, the constant DC current drawn means that the DC-to-RF power conversion efficiency is poor – in fact it is limited theoretically to 50%. In battery-powered applications such as mobile phones this limits the talk time. Even in mains-powered applications such as base-stations (and indeed Hi-Fi amplifiers), the low efficiency is a major problem because it means the amplifier generates a lot of heat.

One solution is the Class-B push-pull amplifier, in which a pair of transistors is used. One amplifies the positive half-cycles and the other the negative half-cycles:-



This arrangement often uses a complementary pair of npn/pnp transistors (e.g. in Hi-Fi again), but at high frequencies it is hard to make pnp transistors match the performance of npn ones. An alternative is to use identical transistors but use baluns to create the anti-phase signals.

Class C amplifiers draw no constant DC current and are very efficient, but they are non-linear, so a filter is used to remove harmonics:-



Class A, AB, B, C (and other classes such as D, E & F) are studied in more detail in other modules and classified according to their conduction angles:-

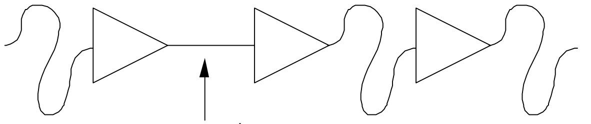
Class A; device conducts throughout the cycle; conduction angle = 360 degrees
Class B; device conducts for half the cycle; conduction angle = 180 degrees
Class C; device conducts for less than half a cycle; conduction angle <180 degrees

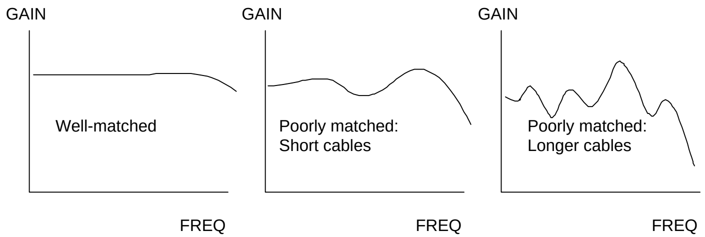
<u>The need for a consistent "system" impedance</u>

Consider an amplifier chain:

Unless the amplifiers are matched to the characteristic impedance of the cables and connectors, reflections are created between components. The reflected signals can interfere with the incoming signal. Furthermore, since the difference in phase between the incoming and reflected signals is different at different frequencies, the interference between them is also different: At some frequencies they may be in-phase and add constructively: At other frequencies they may become anti-phase and partially cancel each other out. The end result is GAIN RIPPLES: The gain changes with frequency, going up and down: Depending on the length of the cables, these ripples can be quite gentle (short cables) or very dramatic, with the gain going up and down rapidly as frequency varies (long cables---n.b. its all relative to the wavelength, which is why this is not generally considered in low frequency design)

(n.b. the signal generator and scope (e.g.) must also have 50 ohm impedance)



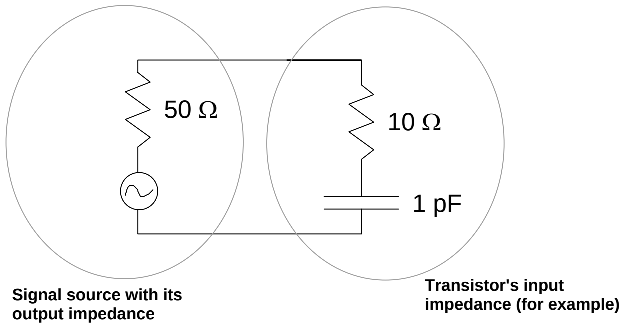


From inspection of the FET equivalent circuit earlier, the transistor has significant input and output capacitances and the input output resistances are very unlikely to be 50 Ohms anyway. Hence, a major part of RF amplifier design is “matching” the device; that is, adding L-C or transmission line networks which transform the transistor input/output impedances to 50 Ohms (most often; see history of 50 Ohms).

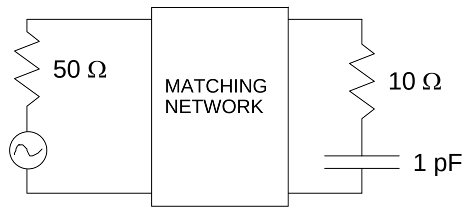
RF AMPLIFIER DESIGN

MATCHING TECHNIQUES

Without matching : maximum power is not transferred to the transistor resulting in reduced gain. In transmission-line circuits a reflected signal is created (more of this later)



With a matching network, the load impedance is transformed so that it appears to be 50 ohms as far as the source is concerned.
The matching network is REACTIVE so that it does not absorb power, but can only work over a limited frequency range. Design is performed with a SMITH CHART or with CAD tools.



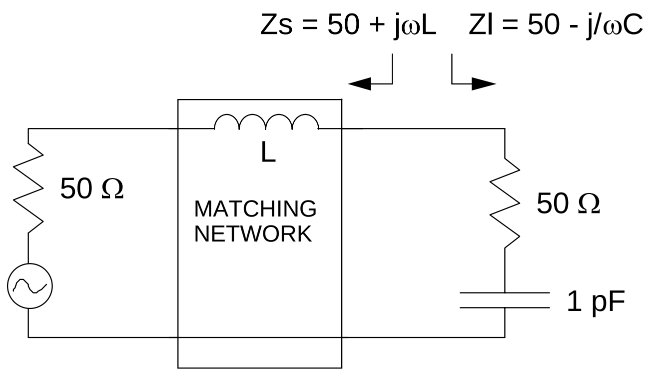
CONJUGATE MATCHING

RE[Zl] = RE[Zs]

IM[Zl] = - IM[Zs]

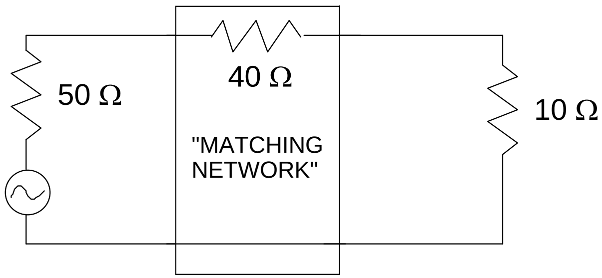
i.e. Zl = Zs*

MATCHING THE IMAGINARY PART
- an unrealistic simple example to illustrate what a matching network is



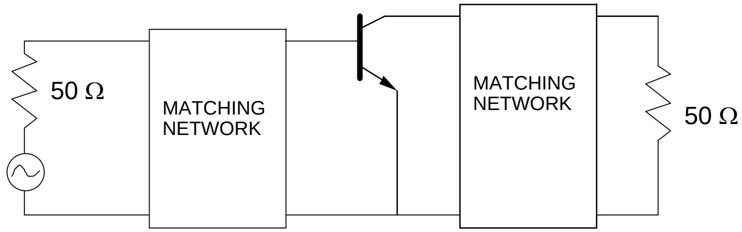
L is easily calculated from: $\omega L = 1/\omega C$

MATCHING THE REAL PART
- a common error



Whilst this IS matched, since Zin is 50 ohms, most of the signal is dissipated in the 40 ohm resistor, defeating the object of matching completely (i.e. maximum power transfer from source to load)

AMPLIFIER WITH INPUT OUTPUT MATCHING

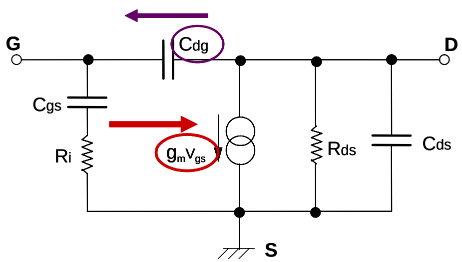


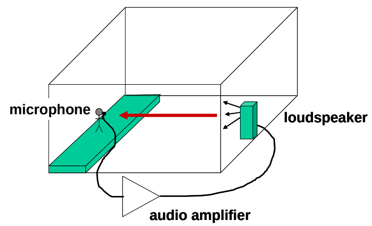
In reality, amplifier design is rarely this simple because of stability.

Real transistors allow some of the output signal to feedback to the input, creating the possibility of positive feedback and oscillation.

Also, the transistor usually has to be BIASED for proper amplification of the ac signal.

Stability is a major issue because of the feedback in the transistor:-





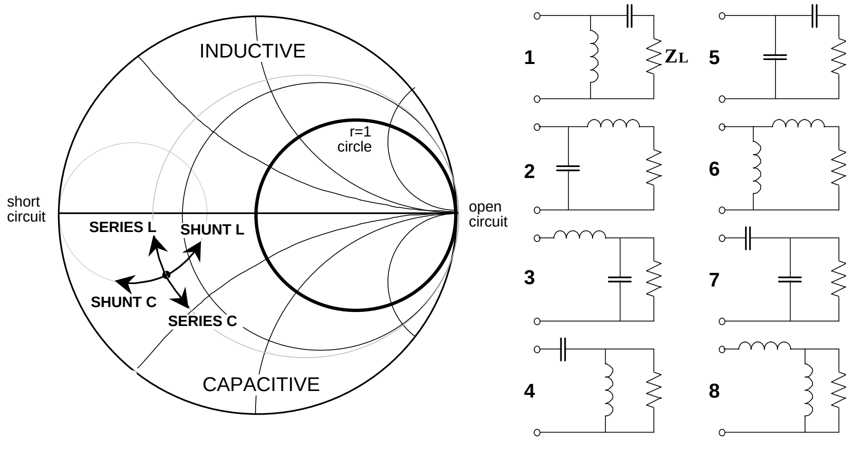
Analogy: Feedback in a PA system

Matching techniques

Lumped element matching with L-networks

In principle, any impedance can be matched to 50 Ω by using just two reactive lumped elements. On the Smith chart a series inductance will move the load clockwise along the constant resistance contour. A series capacitance will move the load anti-clockwise along the same constant resistance contour. A shunt inductance will move the load anti-clockwise along the constant conductance circle, and a shunt capacitance will move the load clockwise along the constant conductance circle. As the constant resistance and constant conductance contours are essentially orthogonal, a suitable choice of series and parallel components can move any load impedance into the centre of the chart.

In total there are eight series/shunt inductor/capacitor combinations, and at least one of these will be able to match any specific impedance to 50 Ω. The figure shows the eight combinations that can be used and summarises the required Smith chart operations for series and shunt inductors and capacitors. These networks are often called L-networks, and are the most basic type of lumped element matching network. Often, more than one of the L-networks can be used to match a given impedance and a choice must be made by considering factors such as the value of the components and the convenience of applying DC bias. For example, if it was suitable, network 4 would be convenient for a transistor matching network because the grounded end of the inductor can be used to apply DC bias and the series capacitor doubles up as the DC block.



Lumped element impedance matching with L-networks

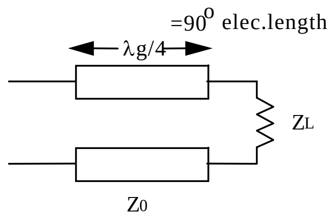
Distributed matching networks

As in filters, at high frequencies the stray capacitance, spurious resonances and distributed effects of spiral inductors mean that transmission-line matching elements are preferred to lumped elements. The only method that will be described here is the quarter-wave transformer. When using distributed matching elements, it should be remembered that these have a cyclical frequency response, and the amplifier's stability should be checked at the harmonics and sub-harmonics of the design frequency.

The quarter-wave transformer

This well known technique can match a load resistance, *R*, to Z_0 using a series transmission line one quarter-wavelength long (90°) with a characteristic impedance of $Z_T = \sqrt{Z_0 \times R}$. Since a transistor's impedance is rarely purely real, first a reactive element must be used to resonate out the imaginary part. This can be achieved with either a series or shunt inductor/capacitor or with a transmission-line stub.

Actually, it was already analysed before as the quarter-wave series line:-



$$Z_{in} = Z_0 \frac{Z_L + jZ_0 \tan\frac{\pi}{2}}{Z_0 + jZ_L \tan\frac{\pi}{2}} = \frac{Z_0^{\ 2}}{Z_L}$$ ***can be used as an impedance transformer***

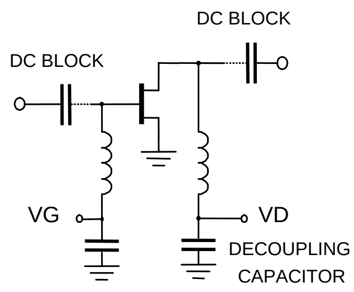
Zin is proportional to 1/ZL, so its often referred to as an IMPEDANCE INVERTER.

DC bias injection

The transistor must have DC bias applied to set the operating point on its I-V characteristic about which the AC microwave signal swings. It is important to appreaciate that a DC voltage source is a perfect short for AC signals, so you must isolate the DC power supplies from the RF signals in the amplifier.

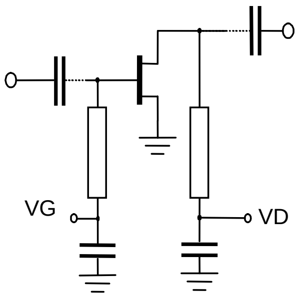
The two common methods are :-

(a) **Inductors as bias chokes.** An inductor has Z=jωL which means zero impedance at DC but a high impedance at RF. DC blocking capacitors are used at the input and output to isolate the bias from other circuits, and with decoupling capacitors to prevent the leakage of RF signals into the power supplies.



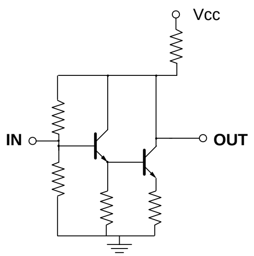
(b) **Short circuit microstrip stubs**. The short circuit end is grounded with a decoupling capacitor so that DC can be applied. If the bias stub is a quarter wavelength long, the RF short of the decoupling capacitor is transformed to an open circuit at the transistor end. However, remember that this is a very frequency-dependent element (it's used as a resonator in filters, remember!).

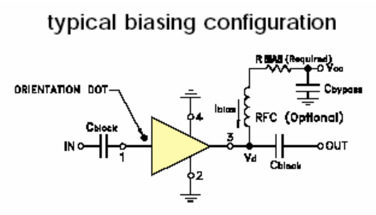
The decoupling at the end of these stubs is very critical, and care must be taken to avoid oscillations.



The "MODAMP"

The Darlington pair has become a very familiar amplifier configuration through its use in the ubiquitous silicon bipolar modular MMIC amplifier (e.g. the MODAMP™, which stands for *monolithic Darlington amplifier*). The Darlington pair has a current gain equal to the product of the individual transistor current gains and gives an increased input impedance and larger S_{21} bandwidth. The circuit diagram of a typical Darlington pair amplifier is shown.



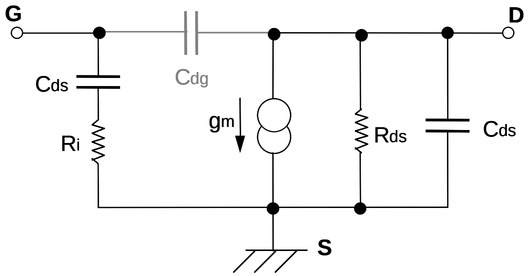


Darlington pair amplifier; ERA series DC-4 GHz amplifier from Minicircuits™

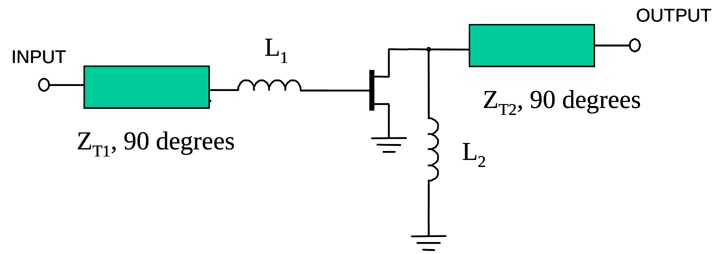
There are penalties for using this general purpose wideband amplifier; you don't just amplifier your desired signal but also a lot of noise, spurious signals, etc, and you may cause stability problems from having gain you don't need at other frequencies.

SIMPLE AMPLIFIER DESIGN EXAMPLE

Suppose our RF transistor has an input impedance consisting of a 20 Ohm resistance in series with a 1pF capacitance and an output impedance consisting of a 100 Ohm resistance in parallel with a 0.5pF capacitance. We want a 50 Ohm matched amplifier at 900 MHz. These values are made up, but of the sort of values you get for typical RF FET devices, which have a typical small-signal equivalent circuit shown below. Cdg (feedback capacitance) is ignored.



In this module only one matching technique is considered in detail; an inductor is used to resonate out the capacitance, then a quarter-wave transformer is used to match the real part:-



L1 is in series; when it resonates with the input capacitance, they go short circuit, leaving only Ri.

L2 is in parallel; when it resonates with the output capacitance, they go open circuit, leaving only Rds

In both cases, we want fr to be 900 MHz, so we use

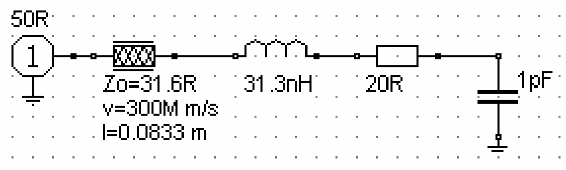
$$fr = \frac{1}{2\pi\sqrt{LC}}$$

To give $L = \dfrac{1}{4\pi^2 f^2 C}$: L1 = 31.3 nH L2 = 62.5 nH.

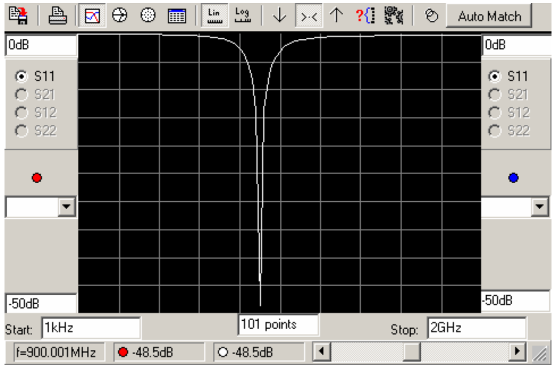
For the transformers, $Z_T = \sqrt{Z_0 \times R}$ gives ZT1 = 31.6 Ohm, ZT2 = 70.7 Ohm

In both cases, the electrical length is 90 degrees (one quarter wavelength).

The input matching can be checked on RFSIM99:-



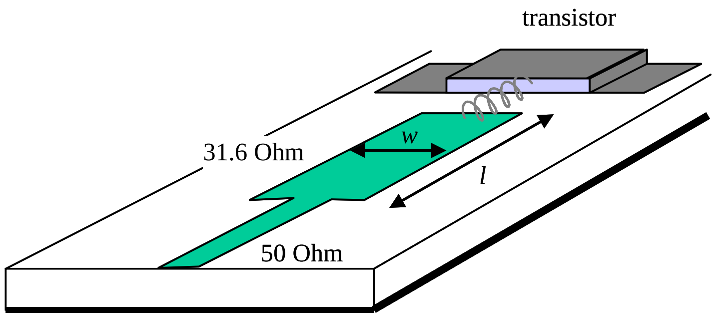
The schematic here has the TL length in metres, with velocity of propagation set to free space.



The S11 plot, from 1 kHz to 2 GHz confirms that the input is well matched to 50 Ohms at the design frequency.

This completes the basic RF design (ignoring bias networks) but the electrical parameters of the TLs must be converted to physical dimensions. The transformer ZT1, for example, in microstrip form, is shown below. It is a track of certain width (for a certain substrate the width can be calculated from the graphs shown earlier for microstrip). The **PHYSICAL LENGTH**, "*l*" is chosen, using the graph for Eeff, to give 90 degrees electrical length at 900 MHz (the design frequency).

At other frequencies, the electrical length is different.



3D view of input matching circuit in microstrip

Let's now investigate **electrical length** more…………consider what happens in the same 900MHz quarter wavelength line at 20 times and 1/20 th of the frequency:-

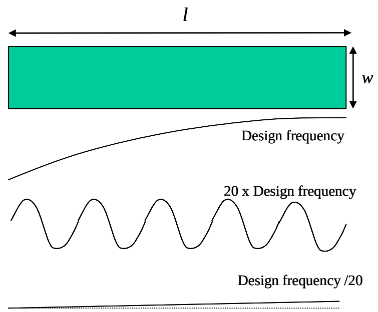
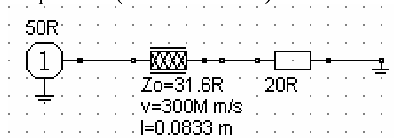


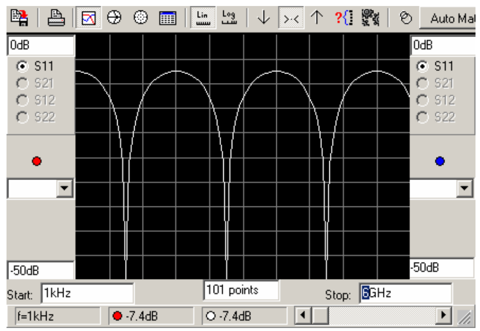
Illustration of change in electrical length with signal frequency

At 20X the frequency, the wavelength is shorter and 5 whole cycles fit in the same physical length of transmission line. The electrical length is "5 x 360 degrees" or "five wavelengths". At low frequency, the line length is very small compared with the wavelength.

Now, consider 20X the frequency again; on the Smith Chart, the extra cycles mean that the load impedance is rotated around and around the centre; in fact, in reflection calculations it is found that one HALF WAVELENGTH rotates right around the chart, back to where you started. For this reason, transmission lines have what is referred to as CYCLICAL behaviour – every half wavelength can be ignored when considering impedance, so because its ELECTRICAL LENGTH continually increases with frequency, the quarter wave transformer actually works at an infinite number of regularly-spaced frequencies (in the ideal case):-



Ideal quarter wave transformer matching 20 Ohm resistor



S11 plot showing cyclical behaviour; the matching works at 900MHz, 3 x 900 MHz, 5 x 900 MHz…. etc
This LOOKS like a VSWR plot, but it is not! The x-axis is FREQUENCY not distance.

AMPLIFIER NOISE FIGURE & DYNAMIC RANGE

There are a number of obvious figures of merit for amplifiers:-

Gain

Bandwidth

POWER HANDLING: Saturated output power

and some less obvious ones -
1 dB gain compression point
(the output power at which the gain has dropped by one dB).

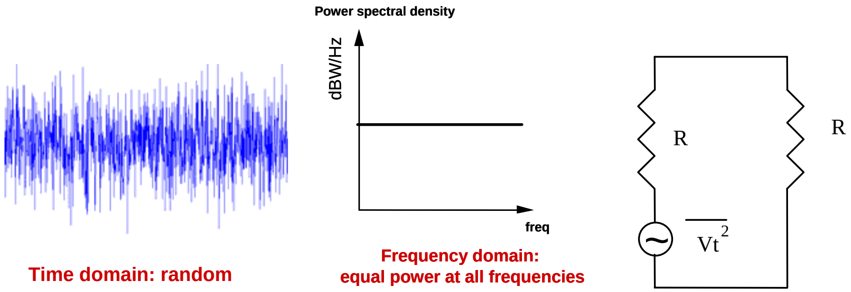
2nd and 3rd-order intercept point (measures of linearity)

these figures determine the largest signal that can be amplified.

The other major figure of merit is the NOISE FIGURE. This determines the weakest signal that can be received.

The noise figure and 1 dB gain compression point together determine the DYNAMIC RANGE of the amplifier (i.e. the signal power range over which it works).

Thermal Noise is generated as a result of the random thermal vibration of electrons.



Time domain: random

Frequency domain:
equal power at all frequencies

The mean square voltage of the noise generated by the resistor R is given by

$$\overline{Vt^2} = 4kTBR$$

where k is Boltzmann's Constant (1.38 x 10^{-23} J/K), T the temperature in Kelvin, and B the bandwidth of the measurement.

For a 50Ω system at room temperature you can remember that the thermal noise floor is at -204dBW/Hz. At this point it becomes necessary to be very familiar with the various forms of the deciBel unit (dB):

dB dB dB dB dB dB dB dB dB dB dB dB dB dB dB

This was discussed when S-parameters were described. It is so important that it is repeated here. Also, in noise calculations the dB units are very important and can become quite complicated.

dB is a unit based on the logarithmic ratio of powers

e.g.

$$\text{Gain in dB} = 10 \log \frac{\text{OUTPUT POWER}}{\text{INPUT POWER}}$$

Note that this is a ratio and has no absolute units.

+3dB = double the power. -3dB = half the power
+10dB = ten times the power -10dB = one tenth of the power

However, the input or output power must have some absolute units, e.g.:-

$$Power\ in\ dBW = 10\log(P\ in\ Watts)$$
$$Power\ in\ dBm = 10\log(P\ in\ mW)$$

Also useful for noise calculations is the bandwidth in dB units:-

Frequency in dBHz = 10 log (f in Hz)

<u>BUT</u> if using voltages etc ...
$P = V^2/R$

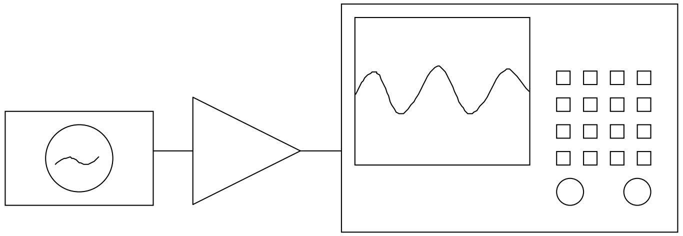
therefore if Voltage Gain = $\frac{VOUT}{VIN}$,

$$\text{Power Gain in dB} = 20 \log \frac{VOUT}{VIN}$$

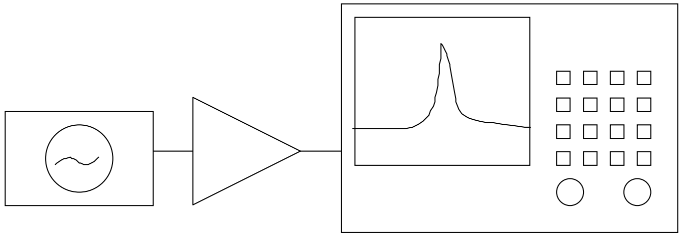
To aid the understanding of NOISE, now is a good time to introduce the SPECTRUM analyser.

There are 3 main instruments used in the RF lab:-

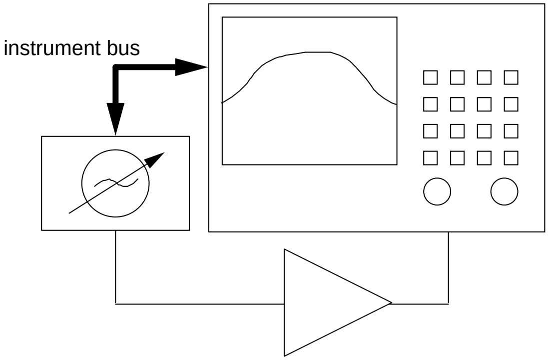
OSCILLOSCOPE Voltage vs. time



SPECTRUM ANALYSER: Power vs. frequency
n.b. POWER is averaged over a certain bandwidth



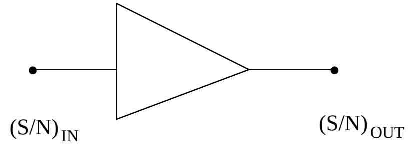
Network Analyser: GAIN vs. Frequency
(actually measures s-parameters)



Noise Factor and Noise Figure

The noise factor/figure is a measure of how much the signal-to-noise ratio is degraded by the amplifier. This degradation is inevitable since the amplifier will not only amplify the input signal, but it will amplify the input noise by the same amount, AND add some more noise of its own making.

The two terms are often used interchangeable, but they should not be confused.
It is recommended that Noise Factor is used to refer to the normal ratio of output and input signal-to-noise ratio, and Noise Figure used for the dB version.

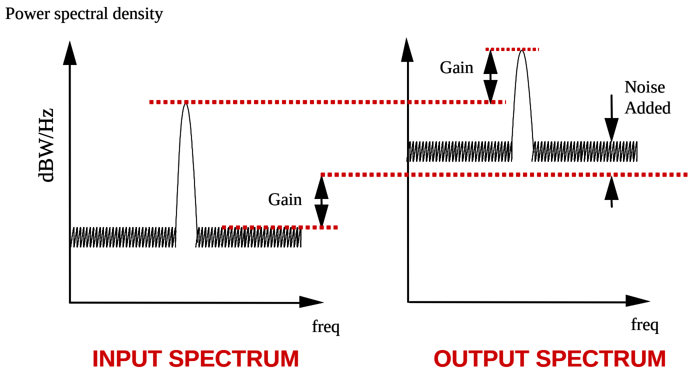


$$N_{OUT} = G \times N_{IN} + N_{ADDED}$$

$$NoiseFactor = \frac{\left(S/N\right)_{IN}}{\left(S/N\right)_{OUT}}$$

Noise figure is in dB and is given by $10\log_{10}$ (noise factor).

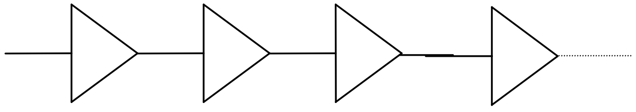
On the **spectrum analyser**, the degradation in signal-to-noise ratio can be easily seen, and the displays would look like this:-



White noise is random and equally distributed over the frequency band. The actual noise power you have is determined by how much BANDWIDTH there is in your system. So, the spectrum analyser Y-axis is labelled as Power Spectral Density and should have units dBm/Hz. i.e. there is a certain amount of noise power in each 1Hz worth of the band.

Cascaded Noise Figures

The noise factor of complete receivers is easily calculated from de Friis' Formula:



$$\text{noise factor } F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \ldots$$

Gain = $G_1\ G_2\ G_3 G_4$

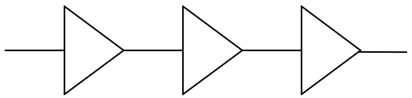
DO <u>NOT</u> USE dB's IN THESE FORMULAE

NOISE CALCULATION PROBLEMS

For the amplifier cascade shown below, calculate the following:

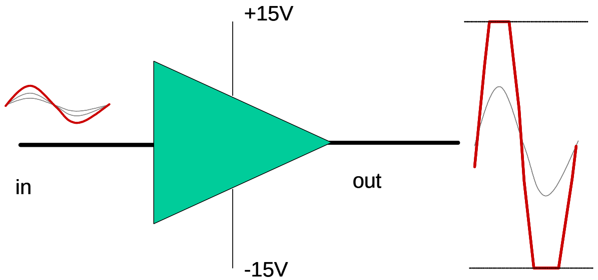
1. Noise power at the output if the amps are noiseless.
2. Gain and noise figure overall.
3. Noise power at the output including the added noise.
4. If they feed a mixer which has a 10dB noise figure, what then is the overall NF?
5. If only the first amp is used to feed the mixer what is the NF?

Bandwidth 1000 - 2000MHz

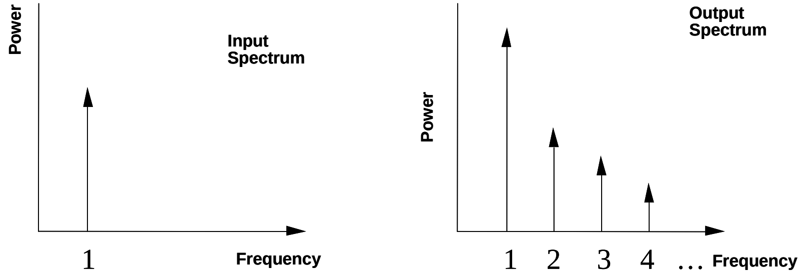


G=10dB 10dB 10dB
F=2dB 3dB 3dB

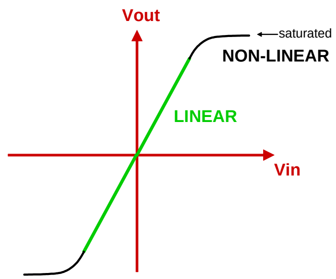
Distortion



Distortion viewed in the time domain



Distortion viewed in the frequency domain
In the limit, the output signal is a square wave and the harmonics follow a sinc function



$$V_{out} = a_0 + a_1 V_{in} + a_2 V_{in}^{\ 2} + a_3\, V_{in}^{\ 3} + \ldots.$$

Distortion viewed as a transfer function

Simple trigonometric relationships show how the non-linearity leads to harmonic frequency components:-

$$\underset{\uparrow\ \text{gain}}{}\quad V_{out} = a_0 + a_1 V_{in} + a_2 V_{in}^{\ 2} + a_3\ V_{in}^{\ 3} + \ldots.$$

$$\cos^3 \omega t = \tfrac{1}{4}(3\cos\omega t + \cos 3\omega t)$$

$$\cos^2 \omega t = \frac{1}{2}(1+\cos 2\omega t)$$

Taking this further, what happens if there are two input signals at different frequencies?
First consider the Squared term:-

$$V_{in} = \cos\omega_1 t + \cos\omega_2 t$$

$$V_{out} = a_0 + a_1 V_{in} + a_2 V_{in}^{\ 2} + a_3\ V_{in}^{\ 3} + \ldots.$$

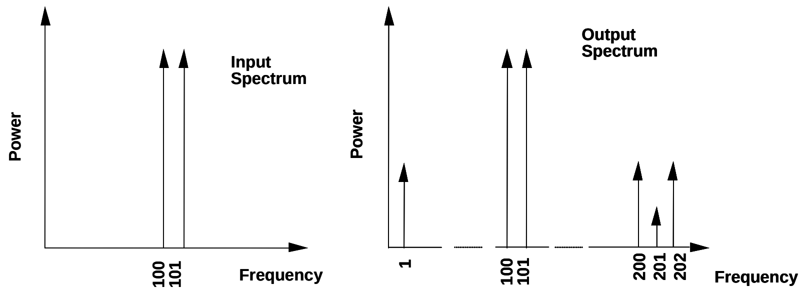
$$(\mathrm{co}\ A + \mathrm{co}\ B)^2 = \mathrm{co}^{\ 2} A + \mathrm{co}^{\ 2} B + 2\,\mathrm{co}\ A\,\mathrm{co}\ B$$

2ⁿ harmonics

$$2\,\mathrm{co}\ A\,\mathrm{co}\ B = \mathrm{co}\ (A+B) + \mathrm{co}\ (A-B)$$

$\boxed{2\omega_1,\ 2\omega_2,\ \ \omega_1+\omega_2 \text{ an } \ \omega_1\text{-}\omega_2}$

Second Order Two Tone Intermodulation Products



Good news: 2nd order products are easily removed with filters

(also, note that this is the same mechanism deliberately used in mixers to give sum and difference frequencies)

And now the Cubed term:-

$$V_{out} = a_0 + a_1 V_{in} + a_2 V_{in}^{\ 2} + a_3\ V_{in}^{\ 3} + \ldots.$$

$3\omega_1$, $3\omega_2$, $2\omega_1+\omega_2$, $2\omega_1-\omega_2$, $2\omega_2+\omega_1$ and $2\omega_2-\omega_1$

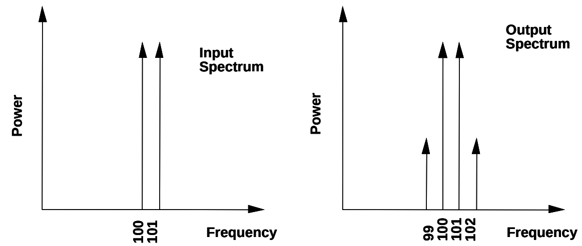
100 and 101 MHz input tones would give the following signals from the cubed term:-

300 MHz, 303 MHz

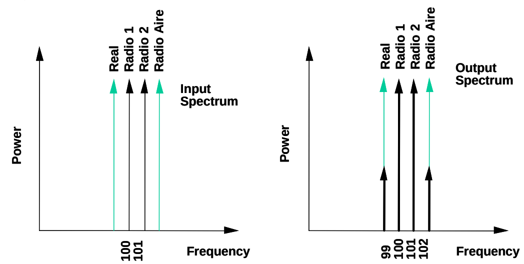
301 MHz, <u>99 MHz</u>, 302 MHz, <u>102 MHz</u>.

These are called the 3rd-order intermodulation products (IMPs)

99 and 102 are underlined because they are the nasty ones:-

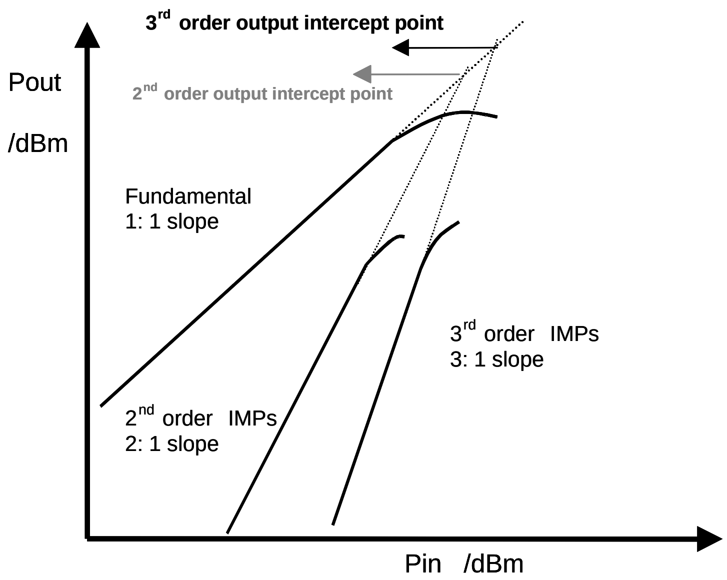


Now see what happens in (e.g.) an ordinary FM radio broadcasting transmitter (fictitious channels):-



3rd ORDER IMPs CANNOT BE REMOVED BY FILTERS

Approximately 50000 man/woman years have been spent around the world devising ways of reducing 3rd order IMPs (this may be an underestimate)



The Intercept Diagram