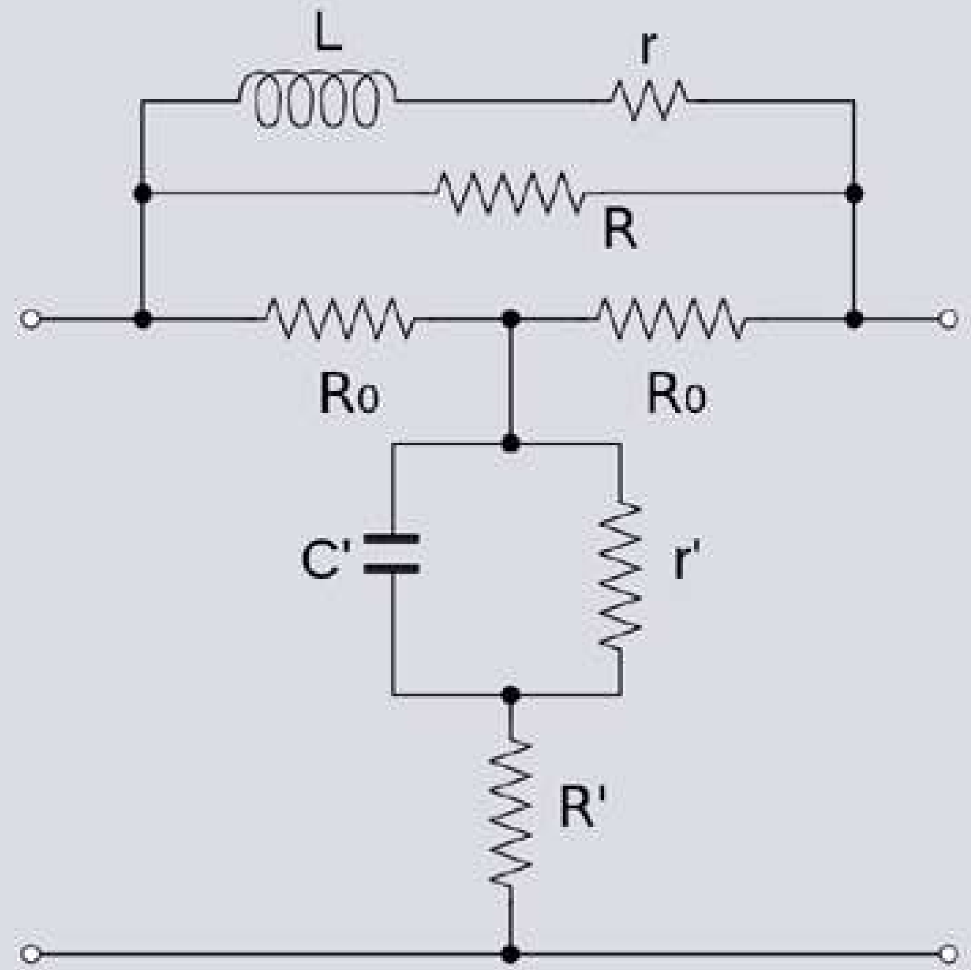
Network Synthesis, Simple and Image Impedance Filters

(Linear Analog Electronic Filters)



Collin Markham

January Drayton

First Edition, 2012

ISBN 978-81-323-1358-8

Published by:
College Publishing House
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Network Synthesis Filters

Network synthesis is a method of designing signal processing filters. It has produced several important classes of filter including the Butterworth filter, the Chebyshev filter and the Elliptic filter. It was originally intended to be applied to the design of passive linear analogue filters but its results can also be applied to implementations in active filters and digital filters. The essence of the method is to obtain the component values of the filter from a given mathematical polynomial ratio expression representing the desired transfer function.

Description of method

The method can be viewed as the inverse problem of network analysis. Network analysis starts with a network and by applying the various electric circuit theorems predicts the response of the network. Network synthesis on the other hand, starts with a desired response and its methods produce a network that outputs, or approximates to, that response.

Network synthesis was originally intended to produce filters of the kind formerly described as "wave filters" but now usually just called filters. That is, filters whose purpose is to pass waves of certain wavelengths while rejecting waves of other wavelengths. Network synthesis starts out with a specification for the transfer function of the filter, H(s), as a function of complex frequency, s. This is used to generate an expression for the input impedance of the filter (the driving point impedance) which then, by a process of continued fraction or partial fraction expansions results in the required values of the filter components. In a digital implementation of a filter, H(s) can be implemented directly.

The advantages of the method are best understood by comparing it to the filter design methodology that was used before it, the image method. The image method considers the characteristics of an individual filter section in an infinite chain (ladder topology) of identical sections. The filters produced by this method suffer from innaccuracies due to the theoretical termination impedance, the image impedance, not generally being equal to the actual termination impedance. This is not the case with network synthesis filters, the terminations are included in the design from the start. The image method also requires a certain amount of experience on the part of the designer. The designer must first decide how many sections and of what type should be used, and then after calculation, will obtain the transfer function of the filter. This may not be what is required and there can be a number of iterations. The network synthesis method, on the other hand, starts out with the required function and outputs the sections needed to build the corresponding filter.

In general, the sections of a network synthesis filter are identical topology (usually the simplest ladder type) but different component values are used in each section. By contrast, the structure of an image filter has identical values at each section - this is a consequence of the infinite chain approach - but may vary the topology from section to section to achieve various desirable characteristics. Both methods make use of low-pass prototype filters followed by frequency transformations and impedance scaling to arrive at the final desired filter.

Important filter classes

The class of a filter refers to the class of polynomials from which the filter is mathematically derived. The order of the filter is the number of filter elements present in the filter's ladder implementation. Generally speaking, the higher the order of the filter, the steeper the cut-off transition between passband and stopband. Filters are often named after the mathematician or mathematics on which they are based rather than the discoverer or inventor of the filter.

Butterworth filter

The Butterworth class of filter was first described in a 1930 paper by the British engineer Stephen Butterworth after whom it is named. The filter response is described by Butterworth polynomials, also due to Butterworth.

Chebyshev filter

A Chebyshev filter has a faster cut-off transition than a Butterworth, but at the expense of there being ripples in the frequency response of the passband. There is a compromise to be had between the maximum allowed attenuation in the passband and the steepness of the cut-off response. This is also sometimes called a type I Chebyshev, the type 2 being a filter with no ripple in the passband but ripples in the stopband. The filter is named after Pafnuty Chebyshev whose Chebyshev polynomials are used in the derivation of the transfer function.

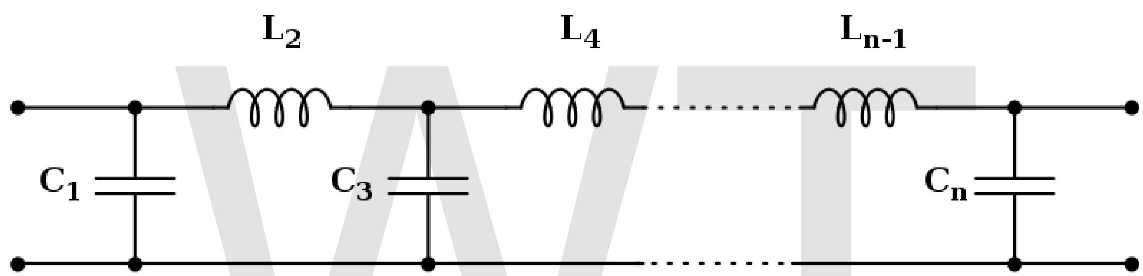
Cauer filter

Cauer filters have equal maximum ripple in the passband and the stopband. The Cauer filter has a faster transition from the passband to the stopband than any other class of network synthesis filter. The term Cauer filter can be used interchangeably with elliptical filter, but the general case of elliptical filters can have unequal ripples in the passband and stopband. An elliptical filter in the limit of zero ripple in the passband is identical to a Chebyshev Type 1 filter. An elliptical filter in the limit of zero ripple in the stopband is identical to a Chebyshev Type 2 filter. An elliptical filter in the limit of zero ripple in both passbands is identical to a Butterworth filter. The filter is named after Wilhelm Cauer and the transfer function is based on elliptic rational functions.

Bessel filter

- The Bessel filter has a maximally flat time-delay (group delay) over its passband. This gives the filter a linear phase response and results in it passing waveforms with minimal distortion. The Bessel filter has minimal distortion in the time domain due to the phase response with frequency as opposed to the Butterworth filter which has minimal distortion in the frequency domain due to the attenuation response with frequency. The Bessel filter is named after Friedrich Bessel and the transfer function is based on Bessel polynomials.

Driving point impedance



Low-pass filter implemented as a ladder (Cauer) topology

The driving point impedance is a mathematical representation of the input impedance of a filter in the frequency domain using one of a number of notations such as Laplace transform (s-domain) or Fourier transform (jω-domain). Treating it as a one-port network, the expression is expanded using continued fraction or partial fraction expansions. The resulting expansion is transformed into a network (usually a ladder network) of electrical elements. Taking an output from the end of this network, so realised, will transform it into a two-port network filter with the desired transfer function.

Not every possible mathematical function for driving point impedance can be realised using real electrical components. Wilhelm Cauer (following on from R. M. Foster) did much of the early work on what mathematical functions could be realised and in which filter topologies. The ubiquitous ladder topology of filter design is named after Cauer.

There are a number of canonical forms of driving point impedance that can be used to express all (except the simplest) realisable impedances. The most well known ones are;

- Cauer's first form of driving point impedance consists of a ladder of shunt capacitors and series inductors and is most useful for low-pass filters.
- Cauer's second form of driving point impedance consists of a ladder of series capacitors and shunt inductors and is most useful for high-pass filters.

- Foster's first form of driving point impedance consists of parallel connected LC resonators and is most useful for band-pass filters.
- Foster's second form of driving point impedance consists of series connected LC anti-resonators and is most useful for band-stop filters.

Prototype filters

Prototype filters are used to make the process of filter design less labour intensive. The prototype is usually designed to be a low-pass filter of unity nominal impedance and unity cut-off frequency, although other schemes are possible. The full design calculations from the relevant mathematical functions and polynomials are carried out only once. The actual filter required is obtained by a process of scaling and transforming the prototype.

Values of prototype elements are published in tables, one of the first being due to Sidney Darlington. Both modern computing power and the practice of directly implementing filter transfer functions in the digital domain have largely rendered this practice obsolete.

A different prototype is required for each order of filter in each class. For those classes in which there is attenuation ripple, a different prototype is required for each value of ripple. The same prototype may be used to produce filters which have a different bandform from the prototype. For instance low-pass, high-pass, band-pass and band-stop filters can all be produced from the same prototype.

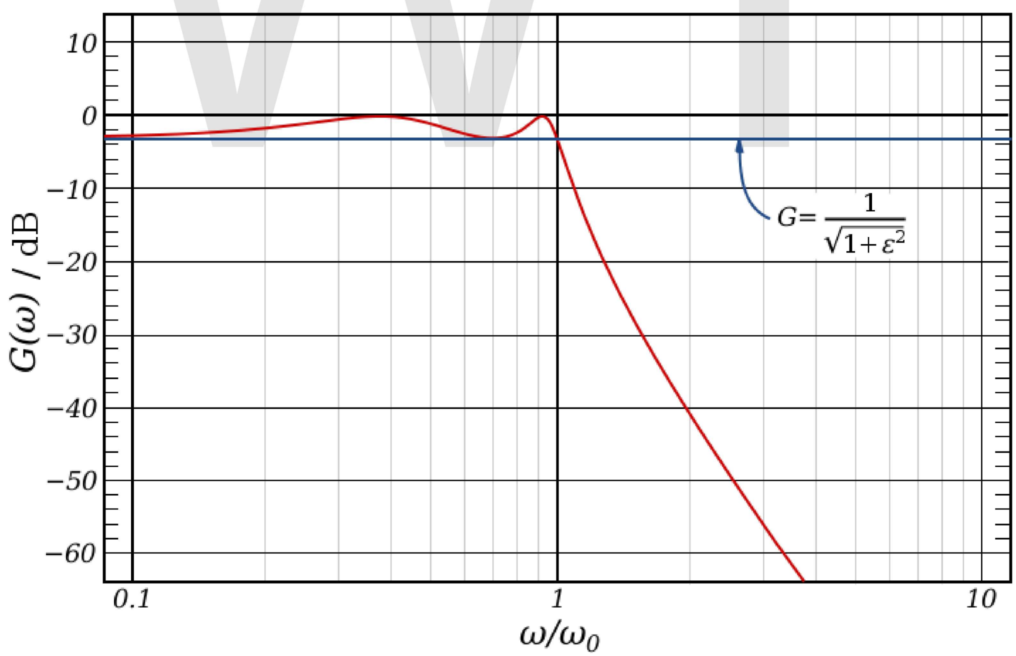
Chapter 2

Chebyshev Filter

Chebyshev filters are analog or digital filters having a steeper roll-off and more passband ripple (type I) or stopband ripple (type II) than Butterworth filters. Chebyshev filters have the property that they minimize the error between the idealized and the actual filter characteristic over the range of the filter, but with ripples in the passband. This type of filter is named in honor of Pafnuty Chebyshev because their mathematical characteristics are derived from Chebyshev polynomials.

Because of the passband ripple inherent in Chebyshev filters, filters which have a smoother response in the passband but a more irregular response in the stopband are preferred for some applications.

Type I Chebyshev filters



The frequency response of a fourth-order type I Chebyshev low-pass filter with $\varepsilon = 1$

These are the most common Chebyshev filters. The gain (or amplitude) response as a function of angular frequency ω of the *n*th order low pass filter is

$$G_n(\omega) = |H_n(j\omega)| = \frac{1}{\sqrt{1+\varepsilon^2 T_n^2\left(\frac{\omega}{\omega_0}\right)}}$$

where ε is the ripple factor, ω_0 is the cutoff frequency and $T_n()$ is a Chebyshev polynomial of the nth order.

The passband exhibits equiripple behavior, with the ripple determined by the ripple factor ε. In the passband, the Chebyshev polynomial alternates between 0 and 1 so the filter gain will alternate between maxima at $G = 1$ and minima at $G = 1/\sqrt{1+\varepsilon^2}$. At the cutoff frequency ω_0 the gain again has the value $1/\sqrt{1+\varepsilon^2}$ but continues to drop into the stop band as the frequency increases. This behavior is shown in the diagram on the right. (*note*: the common definition of the cutoff frequency to −3 dB does *not* hold for Chebyshev filters!)

The order of a Chebyshev filter is equal to the number of reactive components (for example, inductors) needed to realize the filter using analog electronics.

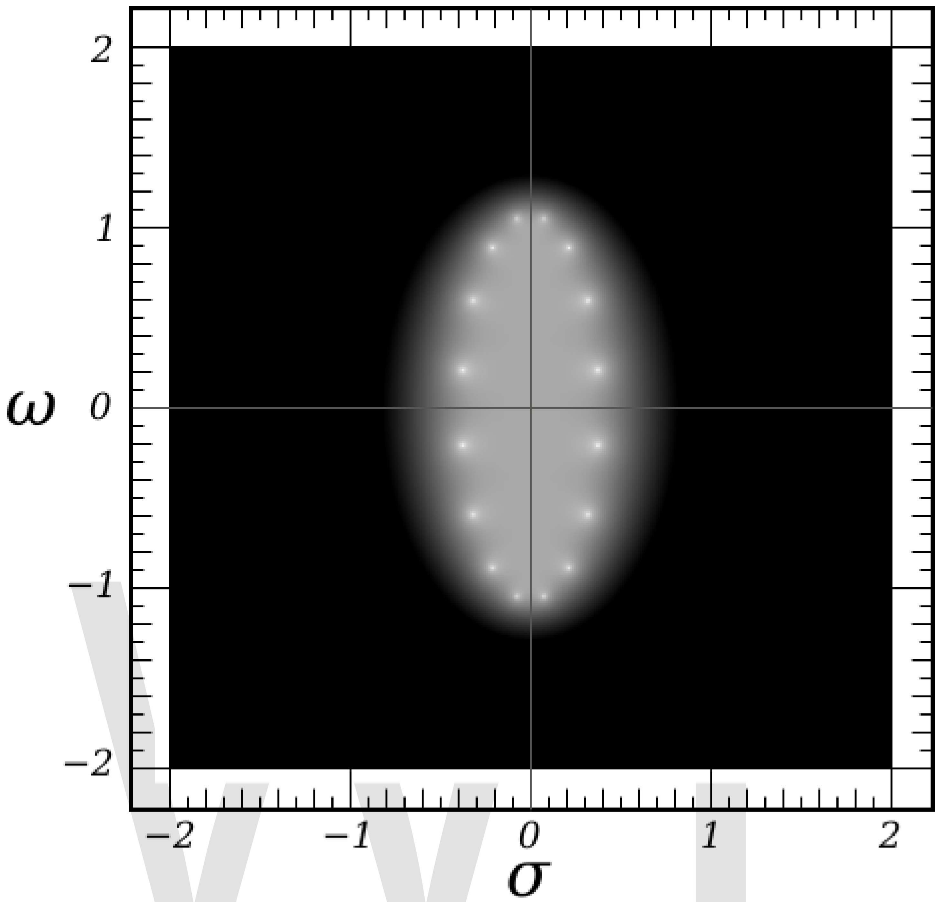
The ripple is often given in dB:

Ripple in dB = $20\log_{10}\frac{1}{\sqrt{1+\varepsilon^2}}$

so that a ripple amplitude of 3 dB results from $\varepsilon = 1.$

An even steeper roll-off can be obtained if we allow for ripple in the stop band, by allowing zeroes on the $j\omega$-axis in the complex plane. This will however result in less suppression in the stop band. The result is called an elliptic filter, also known as Cauer filters.

Poles and zeroes



Log of the absolute value of the gain of an 8th order Chebyshev type I filter in complex frequency space ($s = \sigma + j\omega$) with $\varepsilon = 0.1$ and $\omega_0 = 1$. The white spots are poles and are arranged on an ellipse with a semi-axis of 0.3836... in σ and 1.071... in ω. The transfer function poles are those poles in the left half plane. Black corresponds to a gain of 0.05 or less, white corresponds to a gain of 20 or more.

For simplicity, assume that the cutoff frequency is equal to unity. The poles (ω_{pm}) of the gain of the Chebyshev filter will be the zeroes of the denominator of the gain. Using the complex frequency s:

$$1 + \varepsilon^2 T_n^2(-js) = 0.$$

Defining $-js = \cos(\theta)$ and using the trigonometric definition of the Chebyshev polynomials yields:

$$1 + \varepsilon^2 T_n^2(\cos(\theta)) = 1 + \varepsilon^2 \cos^2(n\theta) = 0.$$

Solving for θ

$$\theta = \frac{1}{n}\arccos\left(\frac{\pm j}{\varepsilon}\right) + \frac{m\pi}{n}$$

where the multiple values of the arc cosine function are made explicit using the integer index *m*. The poles of the Chebyshev gain function are then:

$$\begin{aligned} s_{pm} &= j\cos(\theta) \\ &= j\cos\left(\frac{1}{n}\arccos\left(\frac{\pm j}{\varepsilon}\right) + \frac{m\pi}{n}\right) \end{aligned}$$

Using the properties of the trigonometric and hyperbolic functions, this may be written in explicitly complex form:

$$\begin{aligned} s_{pm}^{\pm} &= \pm\sinh\left(\frac{1}{n}\mathrm{arsinh}\left(\frac{1}{\varepsilon}\right)\right)\sin(\theta_m) \\ &+ j\cosh\left(\frac{1}{n}\mathrm{arsinh}\left(\frac{1}{\varepsilon}\right)\right)\cos(\theta_m) \end{aligned}$$

where $m = 1, 2, \ldots, n$ and

$$\theta_m = \frac{\pi}{2}\,\frac{2m-1}{n}$$

This may be viewed as an equation parametric in θ_n and it demonstrates that the poles lie on an ellipse in *s*-space centered at $s = 0$ with a real semi-axis of length $\sinh(\mathrm{arsinh}(1/\varepsilon)/n)$ and an imaginary semi-axis of length of $\cosh(\mathrm{arsinh}(1/\varepsilon)/n)$.

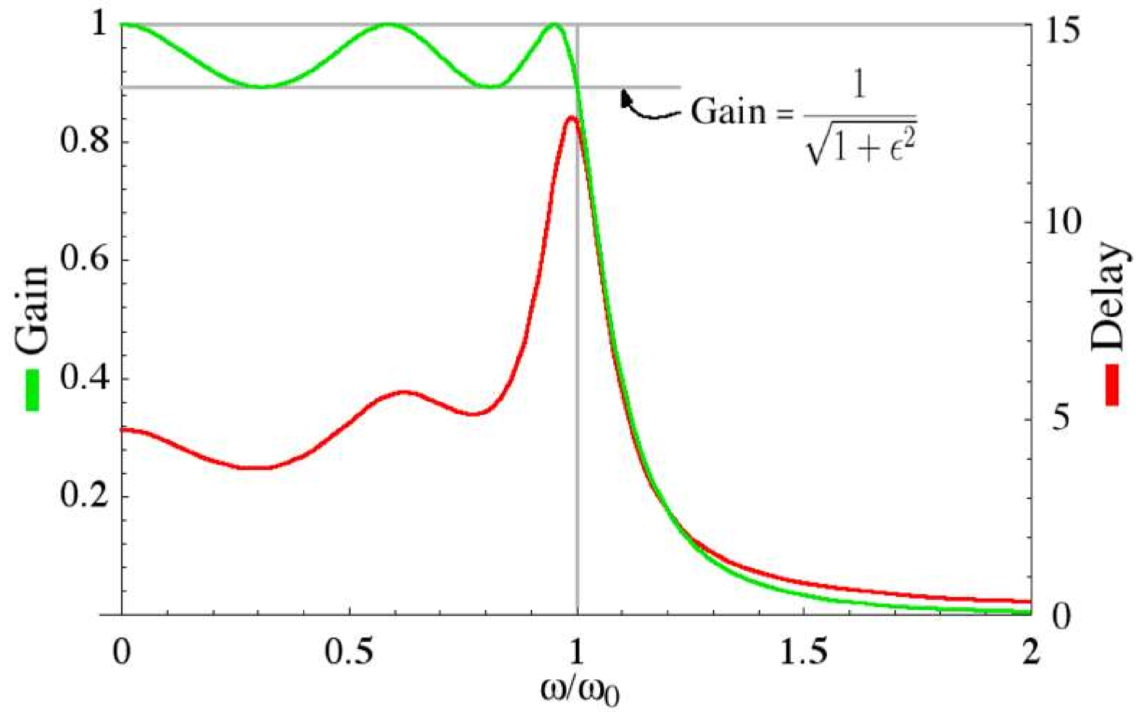
The transfer function

The above expression yields the poles of the gain *G*. For each complex pole, there is another which is the complex conjugate, and for each conjugate pair there are two more that are the negatives of the pair. The transfer function must be stable, so that its poles will be those of the gain that have negative real parts and therefore lie in the left half plane of complex frequency space. The transfer function is then given by

$$H(s) = \frac{1}{2^{n-1}\varepsilon}\prod_{m=1}^{n}\frac{1}{(s - s_{pm}^{-})}$$

where s_{pm}^{-} are only those poles with a negative sign in front of the real term in the above equation for the poles.

The group delay



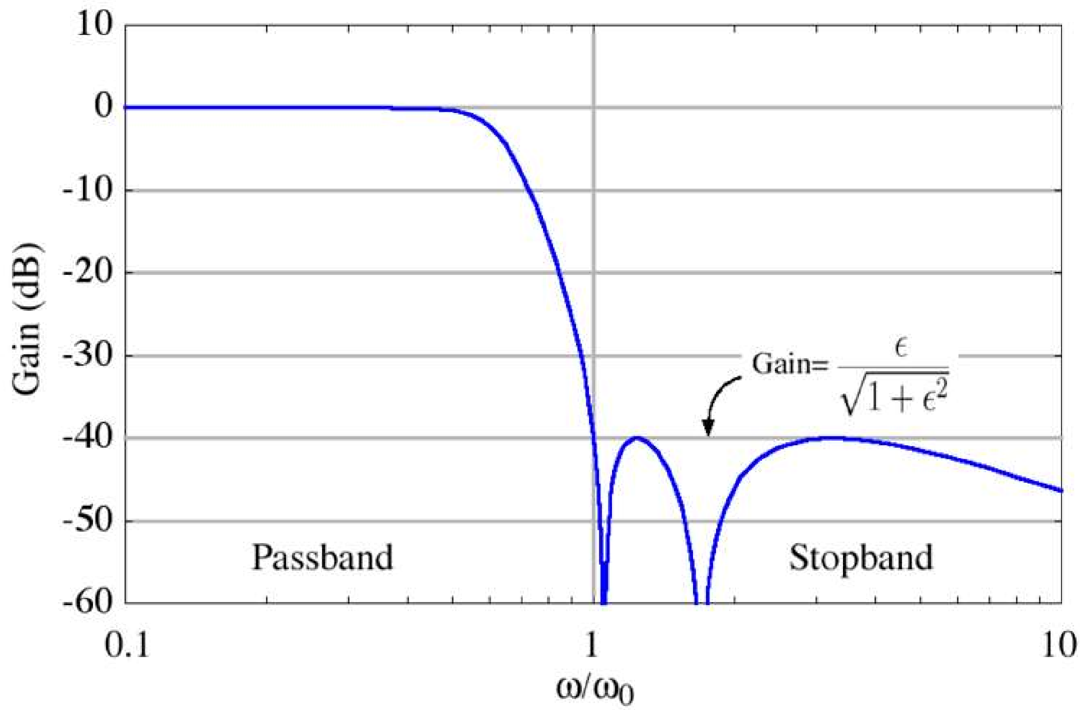
Gain and group delay of a fifth-order type I Chebyshev filter with ε = 0.5.

The group delay is defined as the derivative of the phase with respect to angular frequency and is a measure of the distortion in the signal introduced by phase differences for different frequencies.

$$\tau_g = -\frac{d}{d\omega}\arg(H(j\omega))$$

The gain and the group delay for a fifth order type I Chebyshev filter with ε=0.5 are plotted in the graph on the left. It can be seen that there are ripples in the gain and the group delay in the passband but not in the stop band.

Type II Chebyshev filters



The frequency response of a fifth-order type II Chebyshev low-pass filter with $\varepsilon = 0.01$

Also known as inverse Chebyshev, this type is less common because it does not roll off as fast as type I, and requires more components. It has no ripple in the passband, but does have equiripple in the stopband. The gain is:

$$G_n(\omega, \omega_0) = \frac{1}{\sqrt{1 + \frac{1}{\varepsilon^2 T_n^2(\omega_0/\omega)}}}.$$

In the stop band, the Chebyshev polynomial will oscillate between 0 and 1 so that the gain will oscillate between zero and

$$\frac{1}{\sqrt{1 + \frac{1}{\varepsilon^2}}}$$

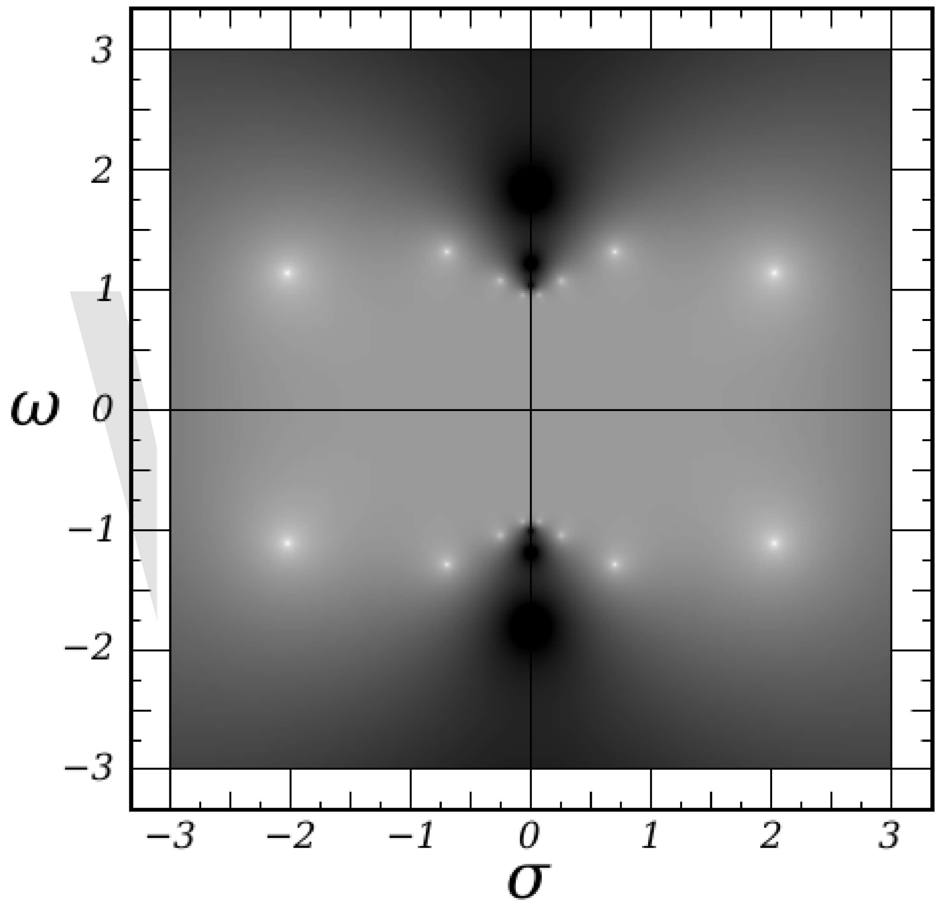
and the smallest frequency at which this maximum is attained will be the cutoff frequency ω_0. The parameter ε is thus related to the stopband attenuation γ in decibels by:

$$\varepsilon = \frac{1}{\sqrt{10^{0.1\gamma} - 1}}.$$

For a stopband attenuation of 5dB, ε = 0.6801; for an attenuation of 10dB, ε = 0.3333. The frequency $f_C = \omega_C/2\pi$ is the cutoff frequency. The 3dB frequency f_H is related to f_C by:

$$f_H = \frac{f_C}{\cosh\left(\frac{1}{n}\cosh^{-1}\frac{1}{\varepsilon}\right)}.$$

Poles and zeroes



Log of the absolute value of the gain of an 8th order Chebyshev type II filter in complex frequency space (s=σ+jω) with ε = 0.1 and $\omega_0 = 1$. The white spots are poles and the black spots are zeroes. All 16 poles are shown. Each zero has multiplicity of two, and 12 zeroes are shown and four are located outside the picture, two on the positive ω axis, and two on the negative. The poles of the transfer function will be poles on the left half plane and the zeroes of the transfer function will be the zeroes, but with multiplicity 1. Black corresponds to a gain of 0.05 or less, white corresponds to a gain of 20 or more.

Again, assuming that the cutoff frequency is equal to unity, the poles (ω_{pm}) of the gain of the Chebyshev filter will be the zeroes of the denominator of the gain:

$$1+\varepsilon^2 T_n^2(-1/js_{pm})=0$$

The poles of gain of the type II Chebyshev filter will be the inverse of the poles of the type I filter:

$$\frac{1}{s_{pm}^{\pm}}=\pm\sinh\left(\frac{1}{n}\mathrm{arsinh}\left(\frac{1}{\varepsilon}\right)\right)\sin(\theta_m)$$
$$+j\cosh\left(\frac{1}{n}\mathrm{arsinh}\left(\frac{1}{\varepsilon}\right)\right)\cos(\theta_m)$$

where $m = 1, 2, ..., n$. The zeroes (ω_{zm}) of the type II Chebyshev filter will be the zeroes of the numerator of the gain:

$$\varepsilon^2 T_n^2(-1/js_{zm})=0.$$

The zeroes of the type II Chebyshev filter will thus be the inverse of the zeroes of the Chebyshev polynomial.

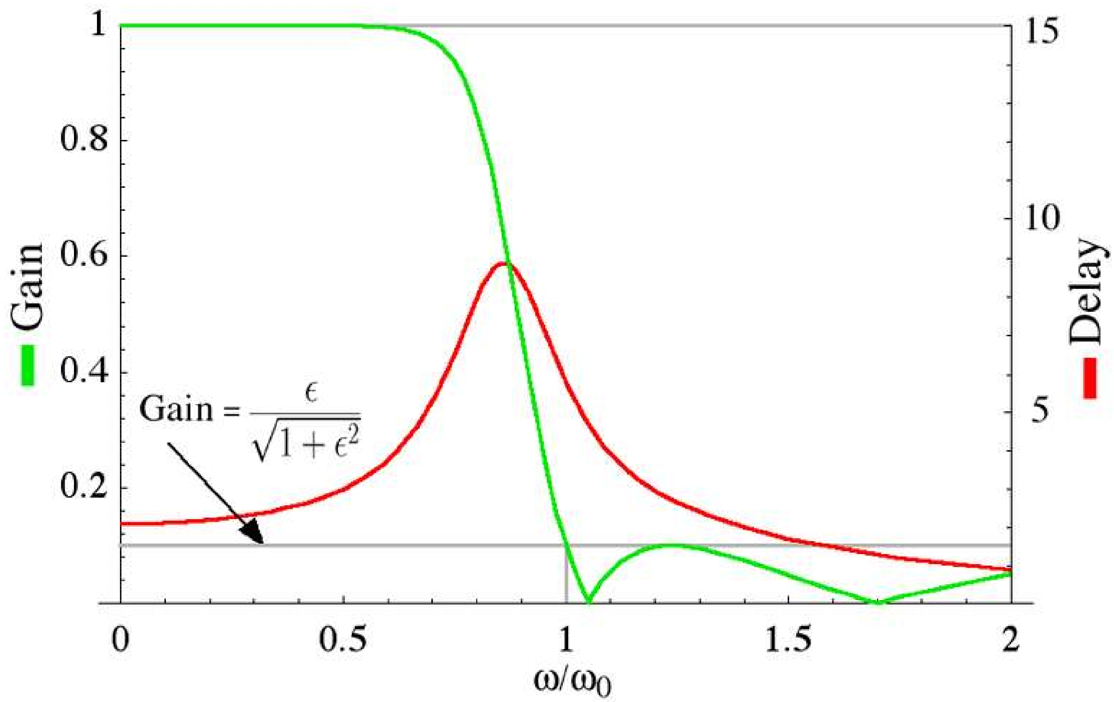
$$1/s_{zm}=-j\cos\left(\frac{\pi}{2}\,\frac{2m-1}{n}\right)$$

for $m = 1, 2, ..., n$.

The transfer function

The transfer function will be given by the poles in the left half plane of the gain function, and will have the same zeroes but these zeroes will be single rather than double zeroes.

The group delay



Gain and group delay of a fifth-order type II Chebyshev filter with ε = 0.1.

The gain and the group delay for a fifth order type II Chebyshev filter with ε=0.1 are plotted in the graph on the left. It can be seen that there are ripples in the gain in the stop band but not in the pass band.

Implementation

Cauer topology

A passive LC Chebyshev low-pass filter may be realized using a Cauer topology. Inductor or capacitor values of a nth-order Chebyshev filter may be calculated from the following equations:

$$G_1 = \frac{2A_1 \cosh(f_H)}{Y}$$

$$G_k = \frac{4A_{k-1}A_k \cosh^2(f_H)}{B_{k-1}G_{k-1}}, \qquad k = 1, 2, 3, \ldots, n$$

G_1, G_k are the capacitor or inductor element values.

f_H, the 3 dB frequency is calculated with: $f_H = f_C \cosh\left(\frac{1}{n}\cosh^{-1}\frac{1}{\varepsilon}\right)$

The coefficients A, Y, β, A_k, and B_k may be calculated from the following equations:

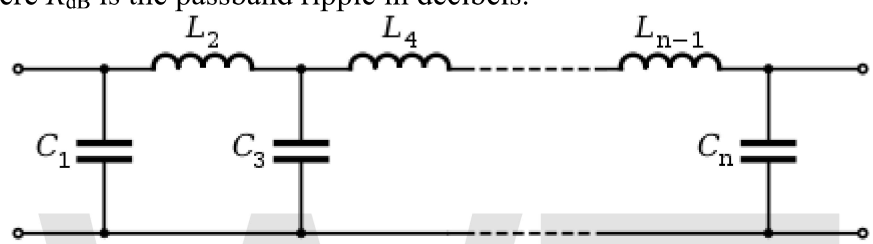
$$Y = \sinh\left(\frac{\beta}{2n}\right)$$

$$\beta = \ln\left(\coth(R_{db}/17.37)\right)$$

$$A_k = \sin\frac{(2k-1)\pi}{2n}, \qquad k = 1, 2, 3, \dots, n$$

$$B_k = Y^2 + \sin^2\left(\frac{k\pi}{n}\right), \qquad k = 1, 2, 3, \dots, n$$

where R_{dB} is the passband ripple in decibels.



The calculated G_k values may then be converted into shunt capacitors and top inductors as shown on the right, or they may be converted into top capacitors and shunt inductors.

- For example, $C_{1\ \text{shunt}} = G_1$, $L_{2\ \text{top}} = G_2$, ...
- or $L_{1\ \text{shunt}} = G_1$, $C_{1\ \text{top}} = G_2$, ...

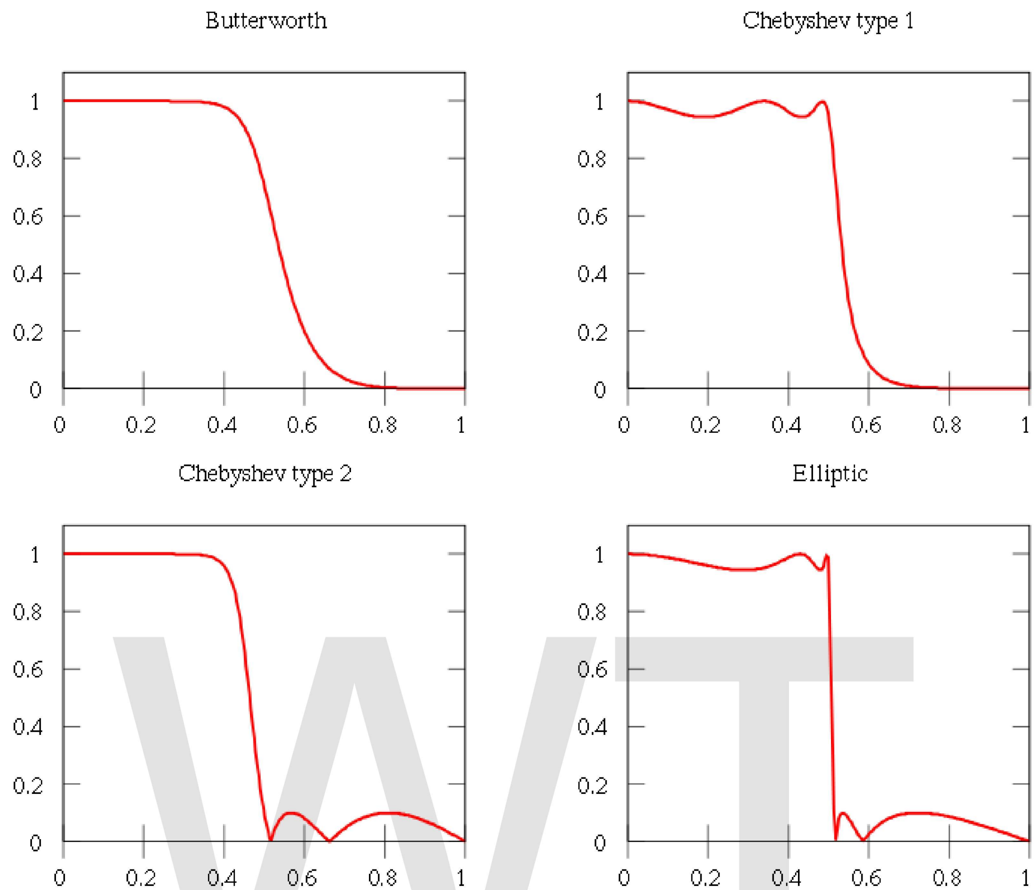
The resulting circuit is a normalized low-pass filter. Using frequency transformations and impedance scaling, the normalized low-pass filter may be transformed into high-pass, band-pass, and band-stop filters of any desired cutoff frequency or bandwidth.

Digital

As with most analog filters, the Chebyshev may be converted to a digital (discrete-time) recursive form via the bilinear transform. However, as digital filters have a finite bandwidth, the response shape of the transformed Chebyshev will be warped. Alternatively, the Matched Z-transform method may be used, which does not warp the response.

Comparison with other linear filters

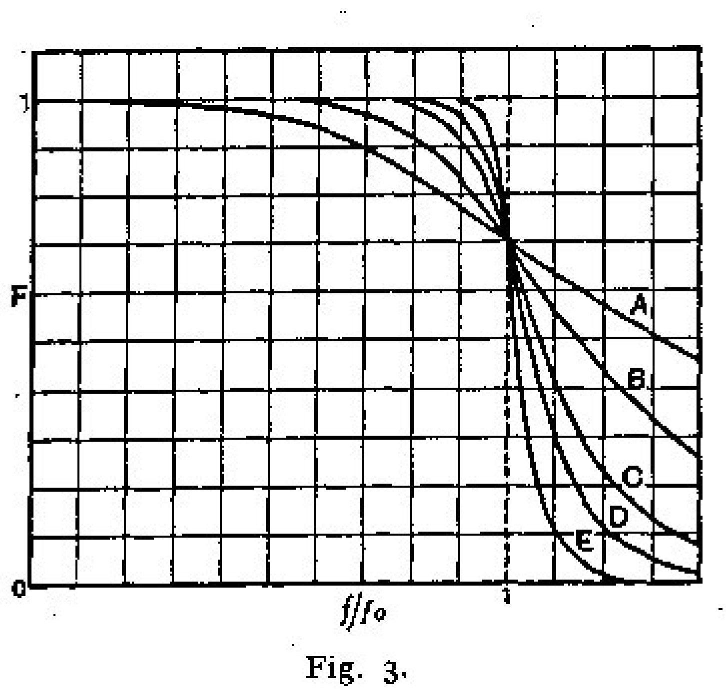
Here is an image showing the Chebyshev filters next to other common kind of filters obtained with the same number of coefficients (all filters are fifth order):



As is clear from the image, Chebyshev filters are sharper than the Butterworth filter; they are not as sharp as the elliptic one, but they show fewer ripples over the bandwidth.

Chapter 3

Butterworth Filter



The frequency response plot from Butterworth's 1930 paper

The **Butterworth filter** is a type of signal processing filter designed to have as flat a frequency response as possible in the passband so that it is also termed a **maximally flat magnitude filter**. It was first described by the British engineer Stephen Butterworth in his paper entitled "On the Theory of Filter Amplifiers".

Original paper

Butterworth had a reputation for solving "impossible" mathematical problems. At the time filter design was largely by trial and error because of their mathematical complexity. His paper was far ahead of its time: the filter was not in common use for over 30 years after its publication. Butterworth stated that;

"An ideal electrical filter should not only completely reject the unwanted frequencies but should also have uniform sensitivity for the wanted frequencies."

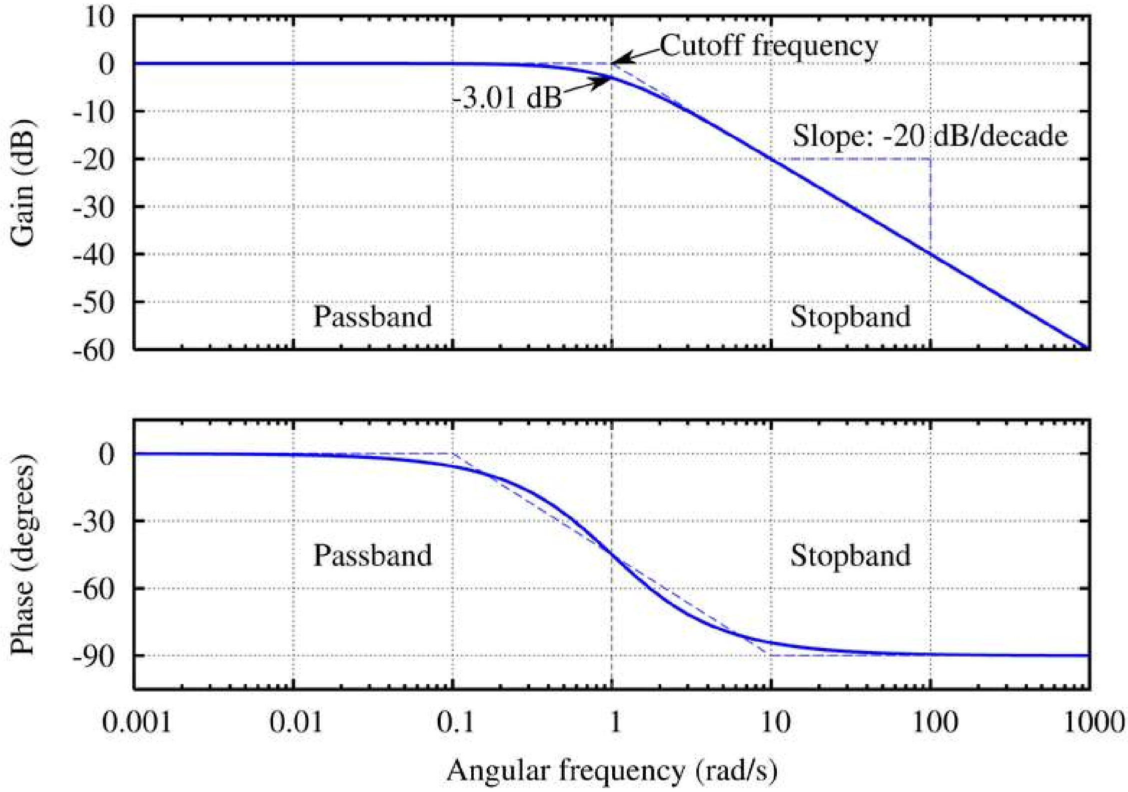
At the time filters generated substantial ripple in the passband and the choice of component values was highly interactive. Butterworth showed that low pass filters could be designed whose frequency response (gain) was;

$$G = \sqrt{\frac{1}{1+\omega^{2n}}}$$

where ω is the angular frequency in radians per second and n is the number of reactive elements (poles) in the filter. Butterworth only dealt with filters with an even number of poles in his paper: he may have been unaware that such filters could be designed with an odd number of poles. His plot of the frequency response of 2, 4, 6, 8, and 10 pole filters is shown as A, B, C, D, and E in his original graph.

Butterworth solved the equations for two- and four-pole filters, showing how the latter could be cascaded when separated by vacuum tube amplifiers and so enabling the construction of higher-order filters despite inductor losses. In 1930 low-loss core materials such as molypermalloy had not been discovered and air-cored audio inductors were rather lossy. Butterworth discovered that it was possible to adjust the component values of the filter to compensate for the winding resistance of the inductors. He also showed that his basic low-pass filter could be modified to give low-pass, high-pass, band-pass and band-stop functionality.

Overview

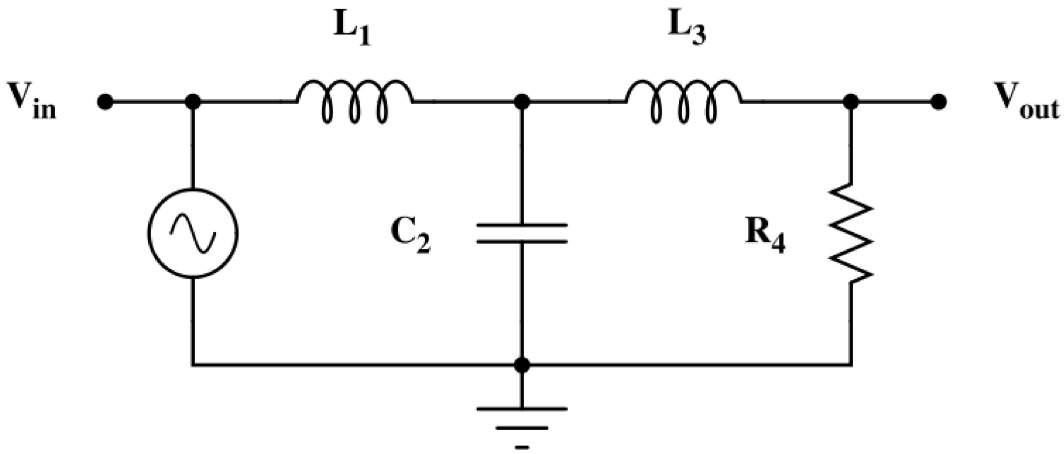


The Bode plot of a first-order Butterworth low-pass filter

The frequency response of the Butterworth filter is maximally flat (has no ripples) in the passband and rolls off towards zero in the stopband. When viewed on a logarithmic Bode plot the response slopes off linearly towards negative infinity. A first-order filter's response rolls off at −6 dB per octave (−20 dB per decade) (all first-order lowpass filters have the same normalized frequency response). A second-order filter decreases at −12 dB per octave, a third-order at −18 dB and so on. Butterworth filters have a monotonically changing magnitude function with ω, unlike other filter types that have non-monotonic ripple in the passband and/or the stopband.

Compared with a Chebyshev Type I/Type II filter or an elliptic filter, the Butterworth filter has a slower roll-off, and thus will require a higher order to implement a particular stopband specification, but Butterworth filters have a more linear phase response in the pass-band than Chebyshev Type I/Type II and elliptic filters can achieve.

A simple example



A third-order low-pass filter (Cauer topology). The filter becomes a Butterworth filter with cutoff frequency $\omega_c=1$ when (for example) $C_2=4/3$ farad, $R_4=1$ ohm, $L_1=3/2$ henry and $L_3=1/2$ henry.

A simple example of a Butterworth filter is the third-order low-pass design shown in the figure on the right, with $C_2 = 4 / 3$ farad, $R_4 = 1$ ohm, $L_1 = 3 / 2$ and $L_3 = 1 / 2$ henry. Taking the impedance of the capacitors C to be $1/Cs$ and the impedance of the inductors L to be Ls, where $s = \sigma + j\omega$ is the complex frequency, the circuit equations yield the transfer function for this device;

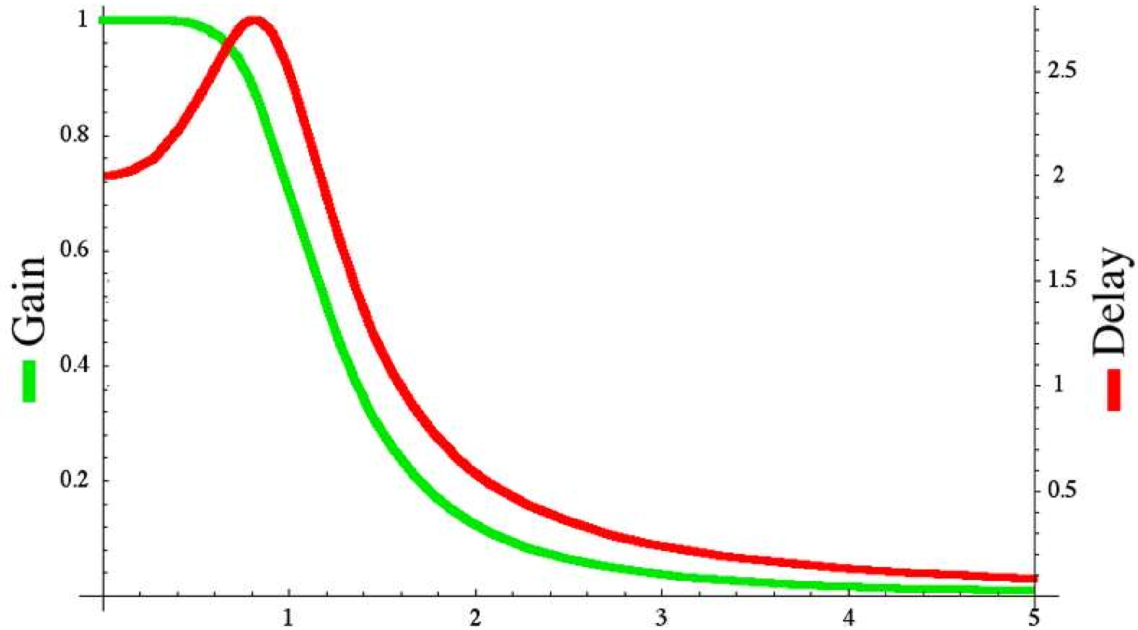
$$H(s) = \frac{V_o(s)}{V_i(s)} = \frac{1}{1 + 2s + 2s^2 + s^3}$$

The magnitude of the frequency response (gain) $G(\omega)$ is given by;

$$G^2(\omega) = |H(j\omega)|^2 = \frac{1}{1 + \omega^6}$$

and the phase is given by;

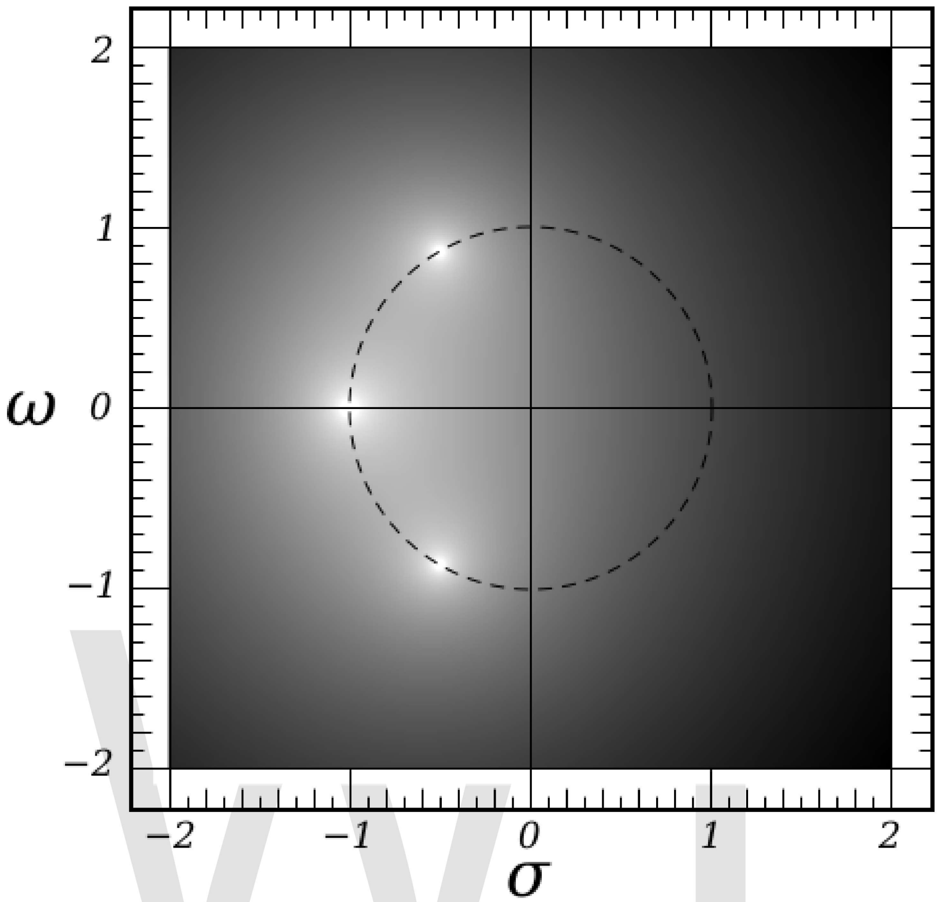
$$\Phi(\omega) = \arg(H(j\omega))$$



Gain and group delay of the third-order Butterworth filter with $\omega_c=1$

The group delay is defined as the derivative of the phase with respect to angular frequency and is a measure of the distortion in the signal introduced by phase differences for different frequencies. The gain and the delay for this filter are plotted in the graph on the left. It can be seen that there are no ripples in the gain curve in either the passband or the stop band.

The log of the absolute value of the transfer function *H(s)* is plotted in complex frequency space in the second graph on the right. The function is defined by the three poles in the left half of the complex frequency plane.



Log density plot of the transfer function H(s) in complex frequency space for the third-order Butterworth filter with ω_c=1. The three poles lie on a circle of unit radius in the left half-plane.

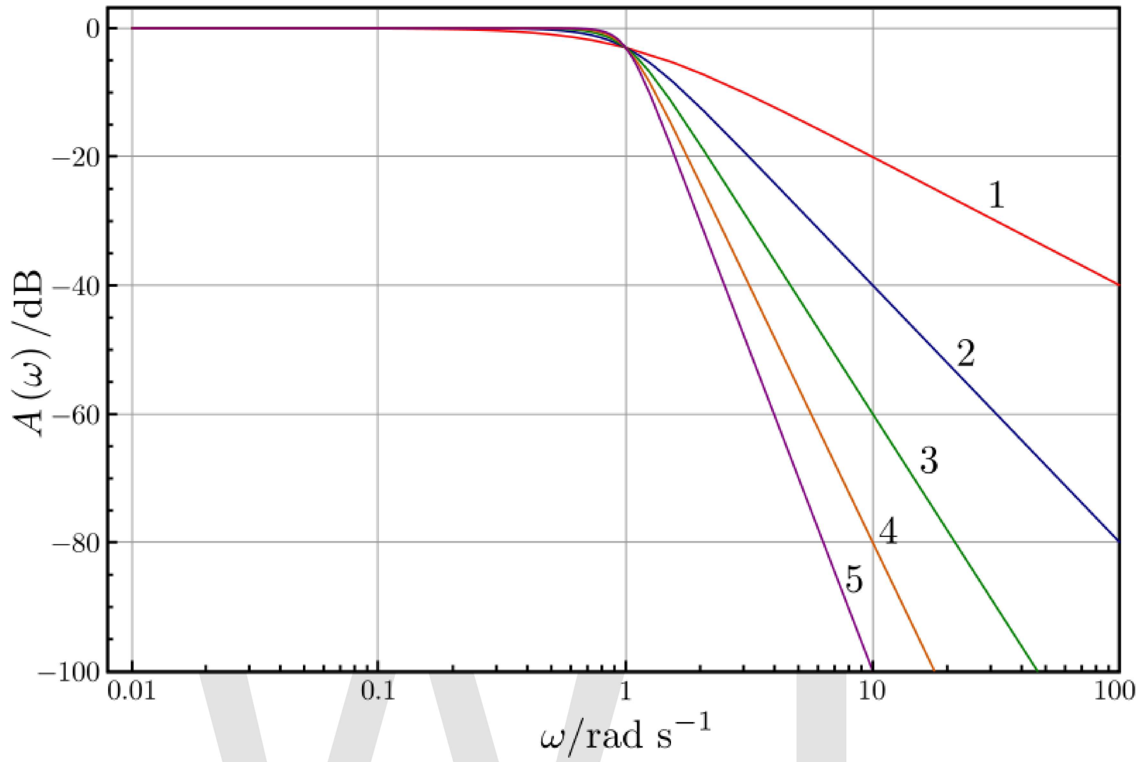
These are arranged on a circle of radius unity, symmetrical about the real *s* axis. The gain function will have three more poles on the right half plane to complete the circle.

By replacing each inductor with a capacitor and each capacitor with an inductor, a high-pass Butterworth filter is obtained.

A band-pass Butterworth filter is obtained by placing a capacitor in series with each inductor and an inductor in parallel with each capacitor to form resonant circuits. The value of each new component must be selected to resonate with the old component at the frequency of interest.

A band-stop Butterworth filter is obtained by placing a capacitor in parallel with each inductor and an inductor in series with each capacitor to form resonant circuits. The value of each new component must be selected to resonate with the old component at the frequency to be rejected.

Transfer function



Plot of the gain of Butterworth low-pass filters of orders 1 through 5, with cutoff frequency $\omega_0 = 1$. Note that the slope is 20*n* dB/decade where *n* is the filter order.

Like all filters, the typical prototype is the low-pass filter, which can be modified into a high-pass filter, or placed in series with others to form band-pass and band-stop filters, and higher order versions of these.

The gain $G(\omega)$ of an *n*-order Butterworth low pass filter is given in terms of the transfer function *H(s)* as;

$$G^2(\omega) = |H(j\omega)|^2 = \frac{{G_0}^2}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}$$

where

- n = order of filter
- ω_c = cutoff frequency (approximately the -3dB frequency)
- G_0 is the DC gain (gain at zero frequency)

It can be seen that as n approaches infinity, the gain becomes a rectangle function and frequencies below ω_c will be passed with gain G_0, while frequencies above ω_c will be suppressed. For smaller values of n, the cutoff will be less sharp.

We wish to determine the transfer function $H(s)$ where $s = \sigma + j\omega$ (from Laplace transform). Since $H(s)H(-s)$ evaluated at $s = j\omega$ is simply equal to $|H(j\omega)|^2$, it follows that;

$$H(s)H(-s) = \frac{{G_0}^2}{1 + \left(\frac{-s^2}{\omega_c^2}\right)^n}$$

The poles of this expression occur on a circle of radius ω_c at equally spaced points. The transfer function itself will be specified by just the poles in the negative real half-plane of s. The *k-th* pole is specified by;

$$-\frac{s_k^2}{\omega_c^2} = (-1)^{\frac{1}{n}} = e^{\frac{j(2k-1)\pi}{n}} \qquad \mathrm{k} = 1, 2, 3, \ldots, \mathrm{n}$$

and hence;

$$s_k = \omega_c e^{\frac{j(2k+n-1)\pi}{2n}} \qquad \mathrm{k} = 1, 2, 3, \ldots, \mathrm{n}$$

The transfer function may be written in terms of these poles as;

$$H(s) = \frac{G_0}{\prod_{k=1}^{n}(s - s_k)/\omega_c}$$

The denominator is a Butterworth polynomial in s.

Normalized Butterworth polynomials

The Butterworth polynomials may be written in complex form as above, but are usually written with real coefficients by multiplying pole pairs which are complex conjugates, such as s_1 and s_n. The polynomials are normalized by setting $\omega_c = 1$. The normalized Butterworth polynomials then have the general form;

$$B_n(s) = \prod_{k=1}^{\frac{n}{2}} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\,\pi\right) + 1\right]_{\text{for n even}}$$

$$B_n(s) = (s+1)\prod_{k=1}^{\frac{n-1}{2}} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\,\pi\right) + 1\right]_{\text{for n odd}}$$

To four decimal places, they are;

n	Factors of Polynomial $B_n(s)$
1	$(s+1)$
2	$s^2 + 1.4142s + 1$
3	$(s+1)(s^2+s+1)$
4	$(s^2 + 0.7654s + 1)(s^2 + 1.8478s + 1)$
5	$(s+1)(s^2 + 0.6180s + 1)(s^2 + 1.6180s + 1)$
6	$(s^2 + 0.5176s + 1)(s^2 + 1.4142s + 1)(s^2 + 1.9319s + 1)$
7	$(s+1)(s^2 + 0.4450s + 1)(s^2 + 1.2470s + 1)(s^2 + 1.8019s + 1)$
8	$(s^2 + 0.3902s + 1)(s^2 + 1.1111s + 1)(s^2 + 1.6629s + 1)(s^2 + 1.9616s + 1)$

The normalized Butterworth polynomials can be used to determine the transfer function for any low-pass filter cut-off frequency ω_c, as follows

$$H(s) = \frac{G_0}{B_n(a)}, \text{ where } a = \frac{s}{\omega_c}$$

Maximal flatness

Assuming $\omega_c = 1$ and $G_0 = 1$, the derivative of the gain with respect to frequency can be shown to be;

$$\frac{dG}{d\omega} = -nG^3\omega^{2n-1}$$

which is monotonically decreasing for all ω since the gain G is always positive. The gain function of the Butterworth filter therefore has no ripple. Furthermore, the series expansion of the gain is given by;

$$G(\omega) = 1 - \frac{1}{2}\omega^{2n} + \frac{3}{8}\omega^{4n} + \ldots$$

In other words all derivatives of the gain up to but not including the $2n$-th derivative are zero, resulting in "maximal flatness". If the requirement to be monotonic is limited to the passband only and ripples are allowed in the stopband, then it is possible to design a filter of the same order, such as the inverse Chebyshev filter, that is flatter in the passband than the "maximally flat" Butterworth.

High-frequency roll-off

Again assuming $\omega_c = 1$, the slope of the log of the gain for large ω is;

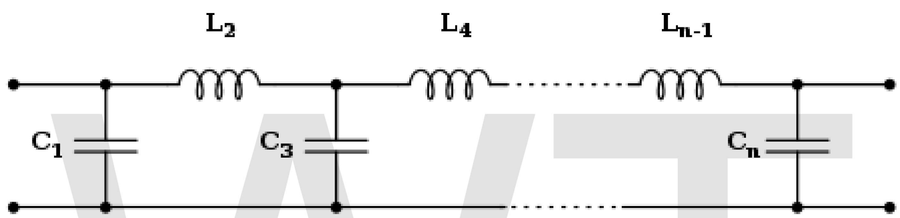
$$\lim_{\omega \to \infty} \frac{d\log(G)}{d\log(\omega)} = -n$$

In decibels, the high-frequency roll-off is therefore 20*n* dB/decade, or 6*n* dB/octave (The factor of 20 is used because the power is proportional to the square of the voltage gain.)

Filter design

There are a number of different filter topologies available to implement a linear analogue filter. The most often used topology for a passive realisation is Cauer topology and the most often used topology for an active realisation is Sallen-Key topology.

Cauer topology



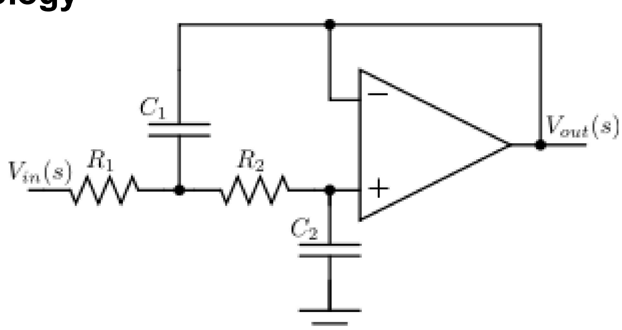
The Cauer topology uses passive components (shunt capacitors and series inductors) to implement a linear analog filter. The Butterworth filter having a given transfer function can be realised using a Cauer 1-form. The kth element is given by;

$$C_k = 2\sin\left[\frac{(2k-1)}{2n}\pi\right]; \text{ k = odd}$$

$$L_k = 2\sin\left[\frac{(2k-1)}{2n}\pi\right]; \text{ k = even}$$

The filter may start with a series inductor if desired, in which case the L_k are k odd and the C_k are k even.

Sallen–Key topology



The Sallen–Key topology uses active and passive components (noninverting buffers, usually op amps, resistors, and capacitors) to implement a linear analog filter. Each Sallen-Key stage implements a conjugate pair of poles; the overall filter is implemented by cascading all stages in series. If there is a real pole (in the case where *n* is odd), this must be implemented separately, usually as an RC circuit, and cascaded with the active stages.

For the second order Sallen-Key circuit shown to the right the transfer function is given by;

$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{1 + C_2(R_1 + R_2)s + C_1 C_2 R_1 R_2 s^2}$$

We wish the denominator to be one of the quadratic terms in a Butterworth polynomial. Assuming that $\omega_c = 1$, this will mean that;

$$C_1 C_2 R_1 R_2 = 1$$

and;

$$C_2(R_1 + R_2) = -2\cos\left(\frac{2k + n - 1}{2n}\pi\right)$$

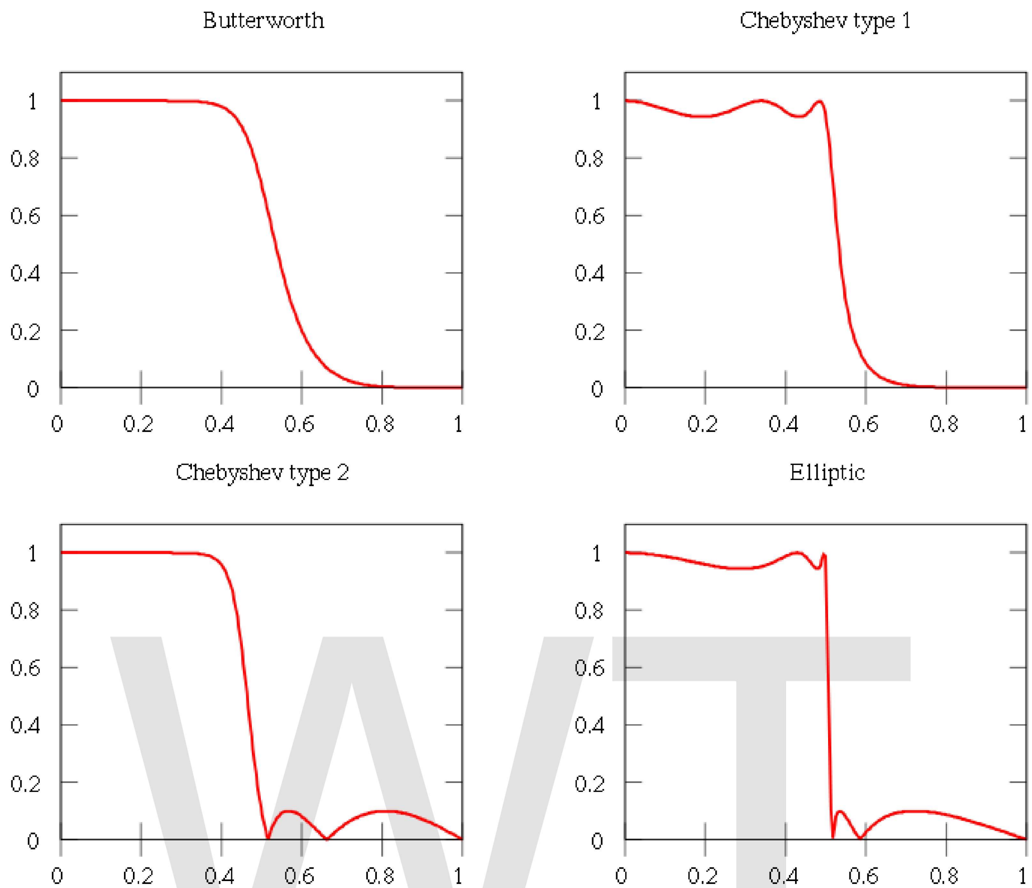
This leaves two undefined component values that may be chosen at will.

Digital implementation

Digital implementations of Butterworth and other filters are often based on the bilinear transform method or the matched Z-transform method, two different methods to discretize an analog filter design. In the case of all-pole filters such as the Butterworth, the matched Z-transform method is equivalent to the impulse invariance method. For higher orders, digital filters are sensitive to quantization errors, so they are often calculated as cascaded biquad sections, plus one first-order or third-order section for odd orders.

Comparison with other linear filters

Here is an image showing the gain of a discrete-time Butterworth filter next to other common filter types. All of these filters are fifth-order.



The Butterworth filter rolls off more slowly around the cutoff frequency than the Chebyshev filter or the Elliptic filter, but without ripple.

Chapter 4

Elliptic Filter and Optimum "L" Filter

Elliptic filter

An **elliptic filter** (also known as a **Cauer filter**, named after Wilhelm Cauer) is a signal processing filter with equalized ripple (equiripple) behavior in both the passband and the stopband. The amount of ripple in each band is independently adjustable, and no other filter of equal order can have a faster transition in gain between the passband and the stopband, for the given values of ripple (whether the ripple is equalized or not). Alternatively, one may give up the ability to independently adjust the passband and stopband ripple, and instead design a filter which is maximally insensitive to component variations.

As the ripple in the stopband approaches zero, the filter becomes a type I Chebyshev filter. As the ripple in the passband approaches zero, the filter becomes a type II Chebyshev filter and finally, as both ripple values approach zero, the filter becomes a Butterworth filter.

The gain of a lowpass elliptic filter as a function of angular frequency ω is given by:

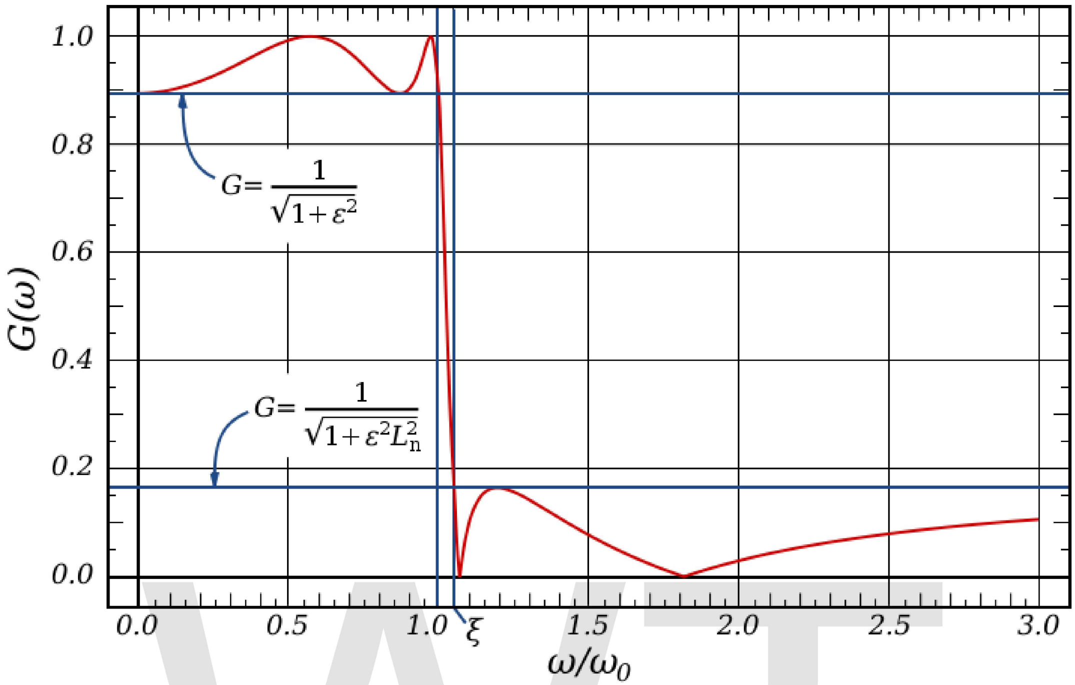
$$G_n(\omega) = \frac{1}{\sqrt{1+\epsilon^2 R_n^2(\xi, \omega/\omega_0)}}$$

where R_n is the *n*th-order elliptic rational function (sometimes known as a Chebyshev rational function) and

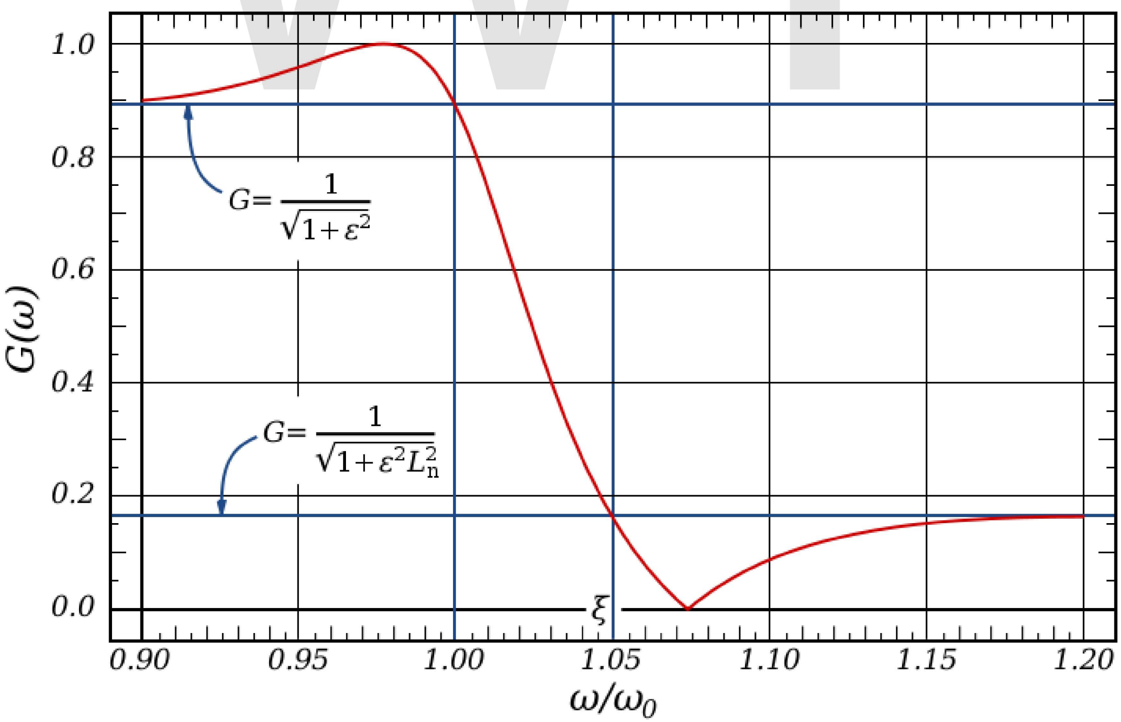
ω_0 is the cutoff frequency
ε is the ripple factor
ξ is the selectivity factor

The value of the ripple factor specifies the passband ripple, while the combination of the ripple factor and the selectivity factor specify the stopband ripple.

Properties



The frequency response of a fourth-order elliptic low-pass filter with ε=0.5 and ξ=1.05. Also shown are the minimum gain in the passband and the maximum gain in the stopband, and the transition region between normalized frequency 1 and ξ



A closeup of the transition region of the above plot

- In the passband, the elliptic rational function varies between zero and unity. The passband of the gain therefore will vary between 1 and $1/\sqrt{1+\epsilon^2}$.

- In the stopband, the elliptic rational function varies between infinity and the discrimination factor L_n which is defined as:

  $$L_n = R_n(\xi, \xi)$$

  The gain of the stopband therefore will vary between 0 and $1/\sqrt{1+\epsilon^2 L_n^2}$.

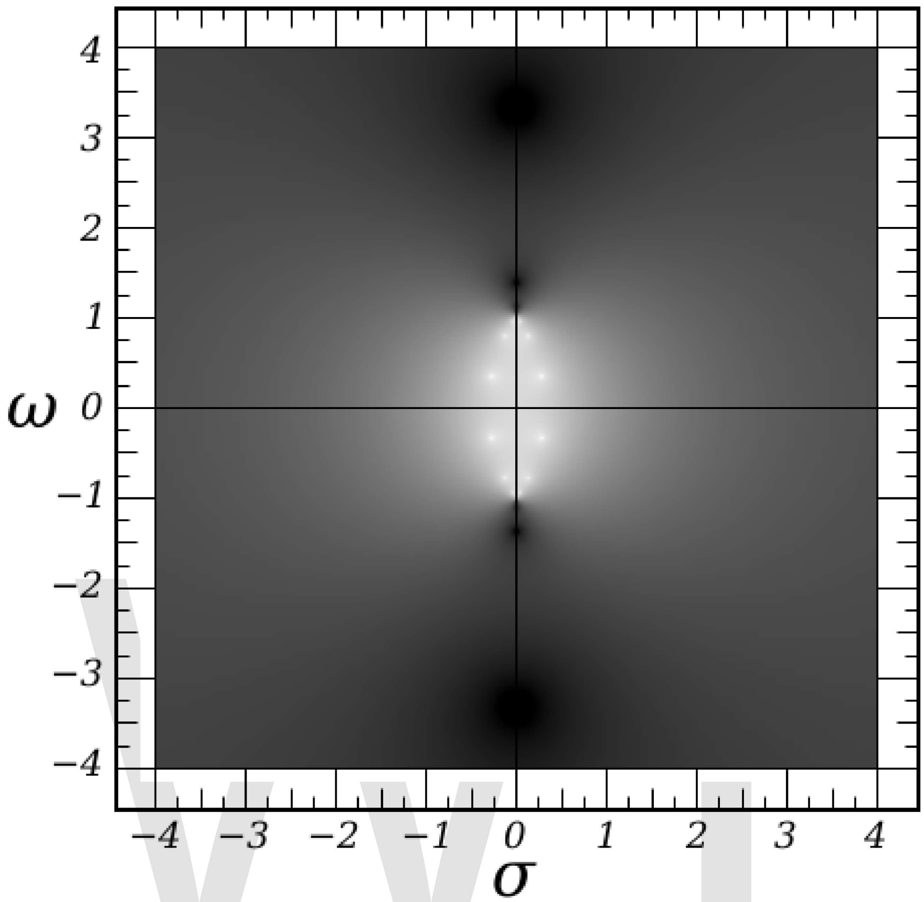
- In the limit of $\xi \rightarrow \infty$ the elliptic rational function becomes a Chebyshev polynomial, and therefore the filter becomes a Chebyshev type I filter, with ripple factor ε

- Since the Butterworth filter is a limiting form of the Chebyshev filter, it follows that in the limit of $\xi \rightarrow \infty$, $\omega_0 \rightarrow 0$ and $\epsilon \rightarrow 0$ such that $\epsilon\, R_n(\xi, 1/\omega_0) = 1$ the filter becomes a Butterworth filter

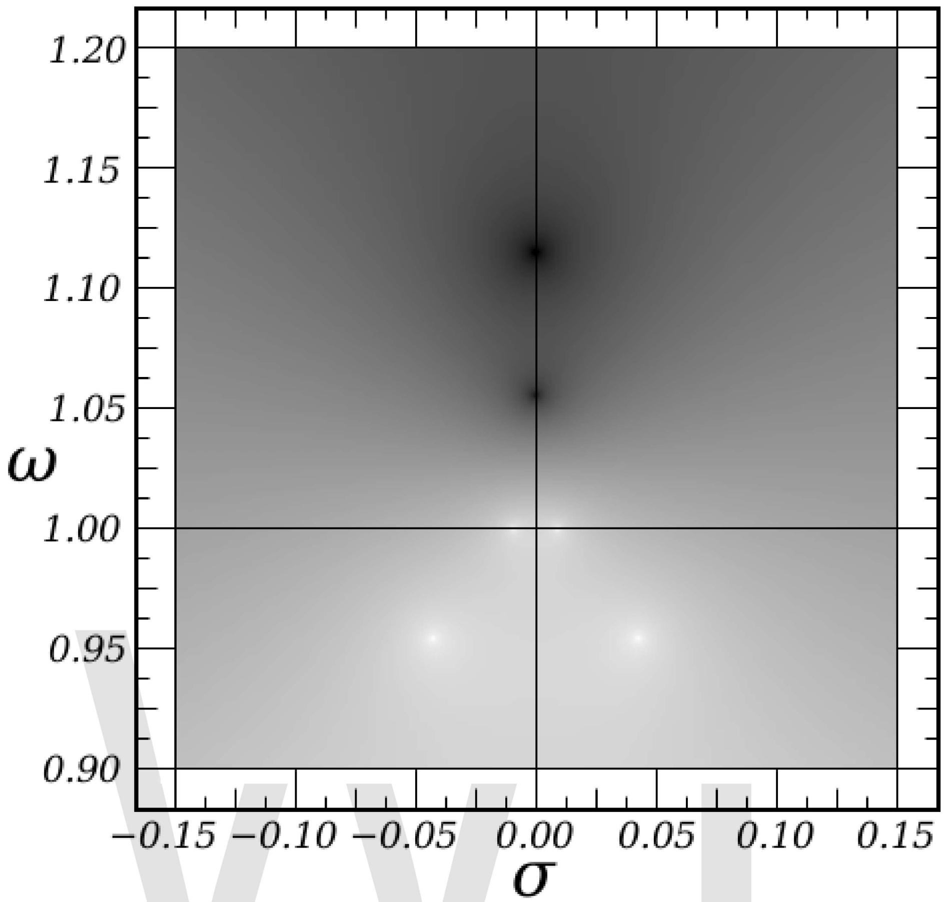
- In the limit of $\xi \rightarrow \infty$, $\epsilon \rightarrow 0$ and $\omega_0 \rightarrow 0$ such that $\xi\omega_0 = 1$ and $\epsilon L_n = \alpha$, the filter becomes a Chebyshev type II filter with gain

  $$G(\omega) = \frac{1}{\sqrt{1 + \frac{1}{\alpha^2 T_n^2(1/\omega)}}}$$

Poles and zeroes



Log of the absolute value of the gain of an 8th order elliptic filter in complex frequency space ($s=\sigma+j\omega$) with $\varepsilon=0.5$, $\xi=1.05$ and $\omega_0 = 1$. The white spots are poles and the black spots are zeroes. There are a total of 16 poles and 8 double zeroes. What appears to be a single pole and zero near the transition region is actually four poles and two double zeroes as shown in the expanded view below. In this image, black corresponds to a gain of 0.0001 or less and white corresponds to a gain of 10 or more.



An expanded view in the transition region of the above image, resolving the four poles and two double zeroes.

The zeroes of the gain of an elliptic filter will coincide with the poles of the elliptic rational function, which are derived in elliptic rational functions.

The poles of the gain of an elliptic filter may be derived in a manner very similar to the derivation of the poles of the gain of a type I Chebyshev filter. For simplicity, assume that the cutoff frequency is equal to unity. The poles (ω_{pm}) of the gain of the elliptical filter will be the zeroes of the denominator of the gain. Using the complex frequency $s = \sigma + j\omega$ this means that:

$$1+\epsilon^2 R_n^2(-js,\xi)=0$$

Defining $-js = \mathrm{cd}(w, 1/\xi)$ where cd() is the Jacobi elliptic cosine function and using the definition of the elliptic rational functions yields:

$$1+\epsilon^2\mathrm{cd}^2\left(\frac{nwK_n}{K},\frac{1}{L_n}\right)=0$$

where $K = K(1 / \xi)$ and $K_n = K(1 / L_n)$. Solving for w

$$w = \frac{K}{nK_n}\mathrm{cd}^{-1}\left(\frac{\pm j}{\epsilon}, \frac{1}{L_n}\right) + \frac{mK}{n}$$

where the multiple values of the inverse cd() function are made explicit using the integer index m.

The poles of the elliptic gain function are then:

$$s_{pm} = i\,\mathrm{cd}(w, 1/\xi)$$

As is the case for the Chebyshev polynomials, this may be expressed in explicitly complex form (Lutovac & et al. 2001, § 12.8)

$$\begin{aligned}
s_{pm} &= \frac{a + jb}{c} \\
a &= -\zeta_n\sqrt{1-\zeta_n^2}\sqrt{1-x_m^2}\sqrt{1-x_m^2/\xi^2} \\
b &= x_m\sqrt{1-\zeta_n^2(1-1/\xi^2)} \\
c &= 1-\zeta_n^2+x_i^2\zeta_n^2/\xi^2
\end{aligned}$$

where ζ_n is a function of n, ϵ and ξ and x_m are the zeroes of the elliptic rational function. ζ_n is expressible for all n in terms of Jacobi elliptic functions, or algebraically for some orders, especially orders 1,2, and 3. For orders 1 and 2 we have

$$\begin{aligned}
\zeta_1 &= \frac{1}{\sqrt{1+\epsilon^2}} \\
\zeta_2 &= \frac{2}{(1+t)\sqrt{1+\epsilon^2}+\sqrt{(1-t)^2+\epsilon^2(1+t)^2}}
\end{aligned}$$

where

$$t = \sqrt{1-1/\xi^2}$$

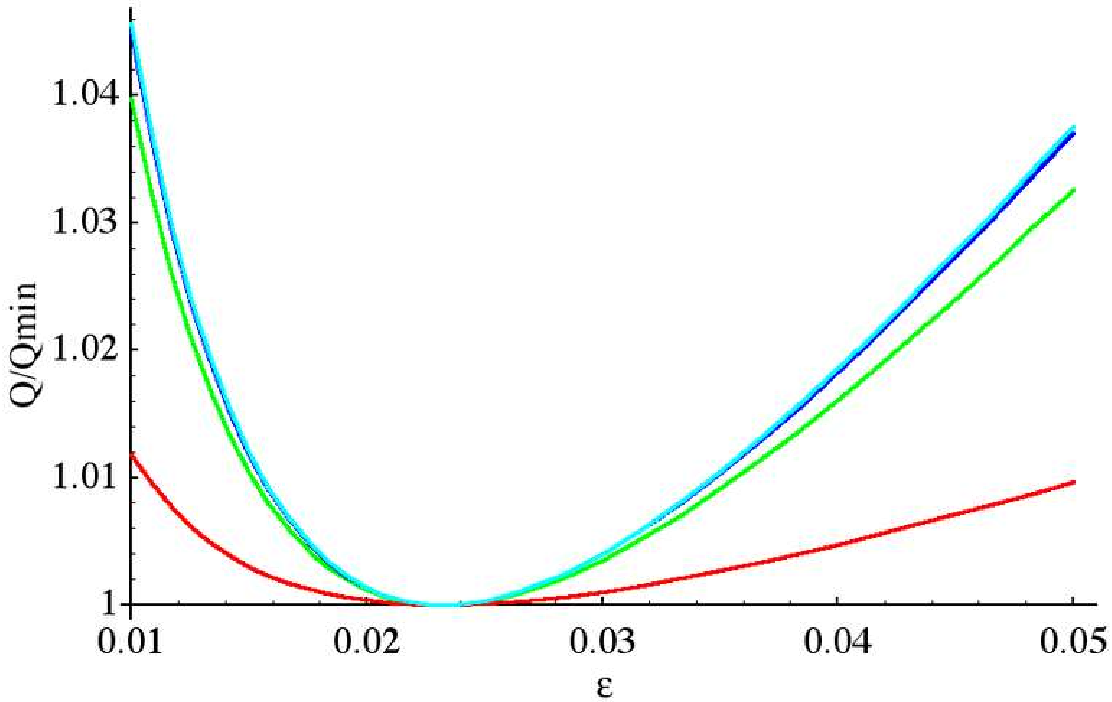
The algebraic expression for ζ_3 is rather involved.

The nesting property of the elliptic rational functions can be used to build up higher order expressions for ζ_n:

$$\zeta_{m\cdot n}(\xi, \epsilon) = \zeta_m\left(\xi, \sqrt{\frac{1}{\zeta_n^2(L_m, \epsilon)} - 1}\right)$$

where $L_m = R_m(\xi,\xi)$.

Minimum Q-factor elliptic filters



The normalized Q-factors of the poles of an 8-th order elliptic filter with ξ=1.1 as a function of ripple factor ε. Each curve represents four poles, since complex conjugate pole pairs and positive-negative pole pairs have the same Q-factor. (The blue and cyan curves nearly coincide). The Q-factor of all poles are simultaneously minimized at $\varepsilon_{Qmin}=1/\sqrt{L_n}=0.02323...$

Elliptic filters are generally specified by requiring a particular value for the passband ripple, stopband ripple and the sharpness of the cutoff. This will generally specify a minimum value of the filter order which must be used. Another design consideration is the sensitivity of the gain function to the values of the electronic components used to build the filter. This sensitivity is inversely proportional to the quality factor (Q-factor) of the poles of the transfer function of the filter. The Q-factor of a pole is defined as:

$$Q = -\frac{|s_{pm}|}{2\mathrm{Re}(s_{pm})} = -\frac{1}{2\cos(\arg(s_{pm}))}$$

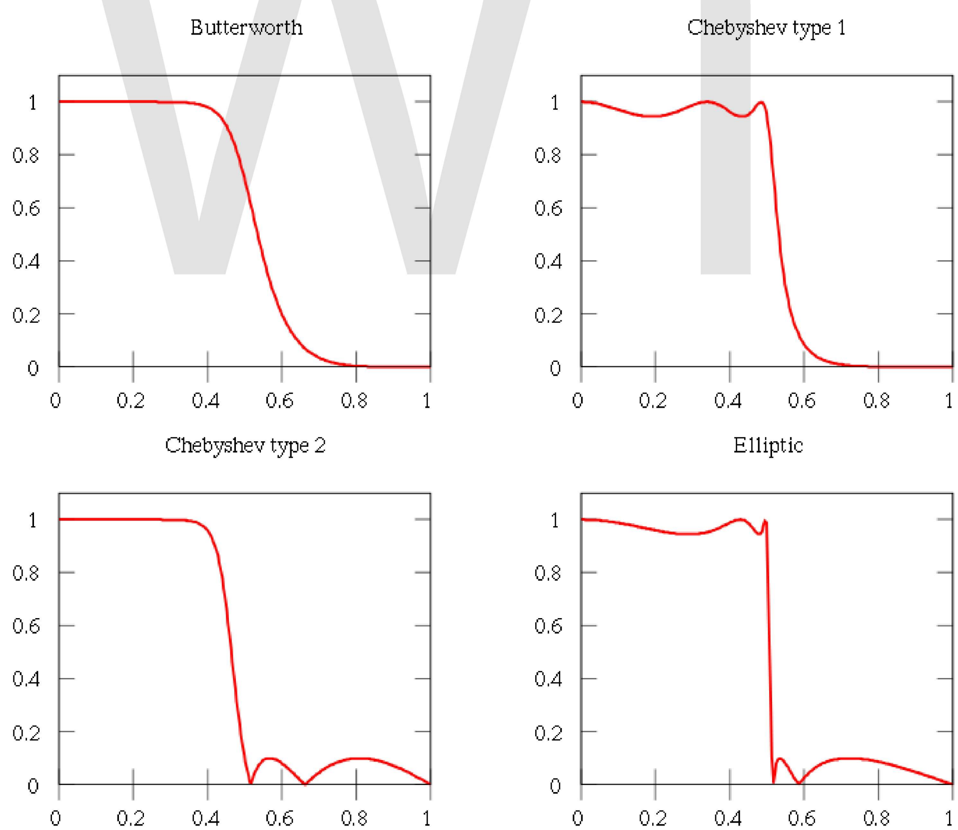
and is a measure of the influence of the pole on the gain function. For an elliptic filter, it happens that, for a given order, there exists a relationship between the ripple factor and selectivity factor which simultaneously minimizes the Q-factor of all poles in the transfer function:

$$\epsilon_{Qmin} = \frac{1}{\sqrt{L_n(\xi)}}$$

This results in a filter which is maximally insensitive to component variations, but the ability to independently specify the passband and stopband ripples will be lost. For such filters, as the order increases, the ripple in both bands will decrease and the rate of cutoff will increase. If one decides to use a minimum-Q elliptic filter in order to achieve a particular minimum ripple in the filter bands along with a particular rate of cutoff, the order needed will generally be greater than the order one would otherwise need without the minimum-Q restriction. An image of the absolute value of the gain will look very much like the image in the previous section, except that the poles are arranged in a circle rather than an ellipse. They will not be evenly spaced and there will be zeroes on the ω axis, unlike the Butterworth filter, whose poles are also arranged in a circle.

Comparison with other linear filters

Here is an image showing the elliptic filter next to other common kind of filters obtained with the same number of coefficients:



As is clear from the image, elliptic filters are sharper than all the others, but they show ripples on the whole bandwidth.

Optimum "L" filter

The **Optimum "L" filter** (also known as a **Legendre filter**) was proposed by Athanasios Papoulis in 1958. It has the maximum roll off rate for a given filter order while maintaining a monotonic frequency response. It provides a compromise between the Butterworth filter which is monotonic but has a slower roll off and the Chebyshev filter which has a faster roll off but has ripple in either the pass band or stop band. The filter design is based on Legendre polynomials which is the reason for its alternate name and the "L" in Optimum "L".

Chapter 5

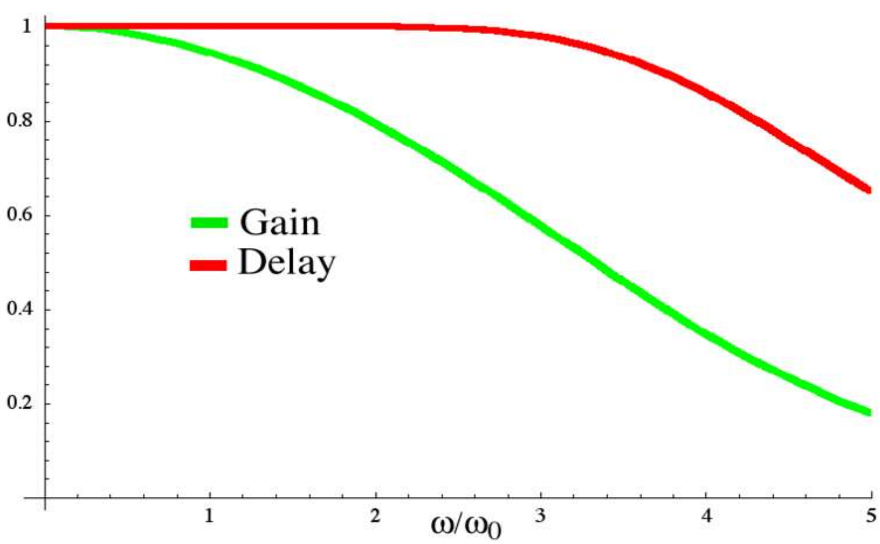
Bessel Filter, Gaussian Filter and Linkwitz–Riley Filter

Bessel filter

In electronics and signal processing, a **Bessel filter** is a type of linear filter with a maximally flat group delay (maximally linear phase response). Bessel filters are often used in audio crossover systems. Analog Bessel filters are characterized by almost constant group delay across the entire passband, thus preserving the wave shape of filtered signals in the passband.

The filter's name is a reference to Friedrich Bessel, a German mathematician (1784–1846), who developed the mathematical theory on which the filter is based. The filters are also called Bessel-Thomson filters in recognition of W. E. Thomson, who worked out how to apply Bessel functions to filter design.

The transfer function



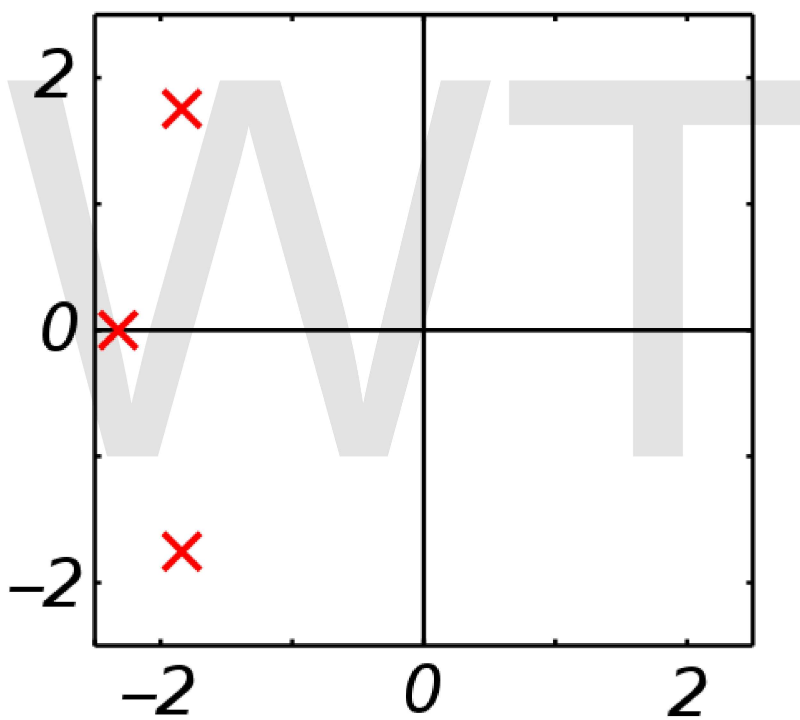
A plot of the gain and group delay for a fourth-order low pass Bessel filter. Note that the transition from the pass band to the stop band is much slower than for other filters, but the group delay is practically constant in the passband. The Bessel filter maximizes the flatness of the group delay curve at zero frequency.

A Bessel low-pass filter is characterized by its transfer function:

$$H(s) = \frac{\theta_n(0)}{\theta_n(s/\omega_0)}$$

where $\theta_n(s)$ is a reverse Bessel polynomial from which the filter gets its name and ω_0 is a frequency chosen to give the desired cut-off frequency. The filter has a low-frequency group delay of $1 / \omega_0$.

Bessel polynomials



The roots of the third-order Bessel polynomial are the poles of filter transfer function in the *s* plane, here plotted as crosses.

The transfer function of the Bessel filter is a rational function whose denominator is a reverse Bessel polynomial, such as the following:

$$\begin{aligned} n &= 1; \quad s+1 \\ n &= 2; \quad s^2+3s+3 \\ n &= 3; \quad s^3+6s^2+15s+15 \end{aligned}$$

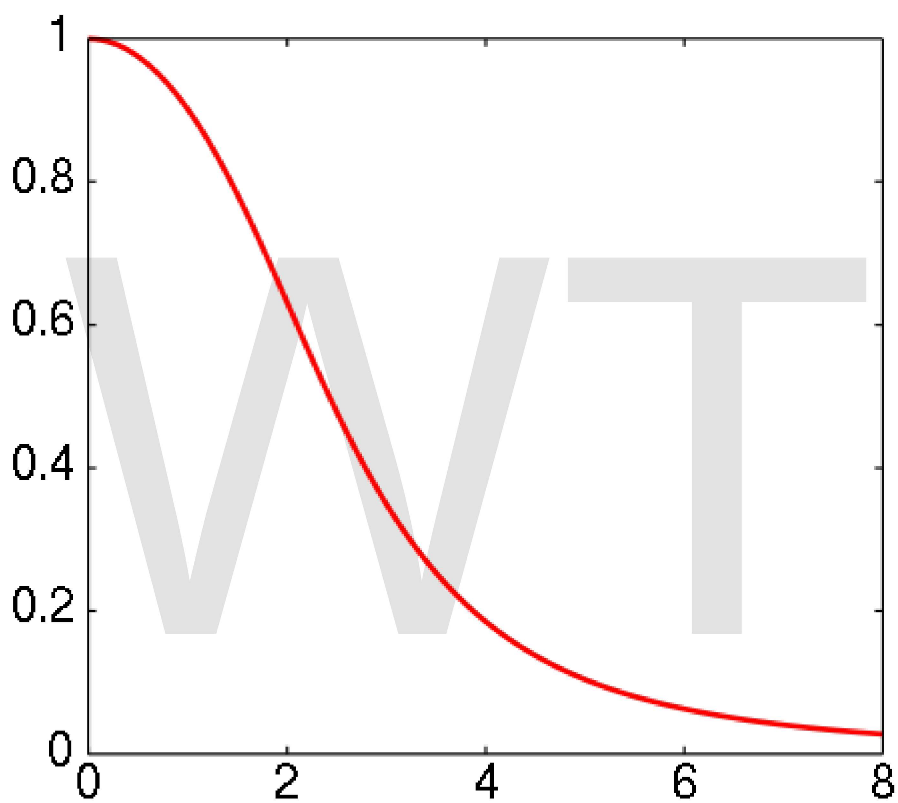
The reverse Bessel polynomials are given by:

$$\theta_n(s) = \sum_{k=0}^{n} a_k s^k$$

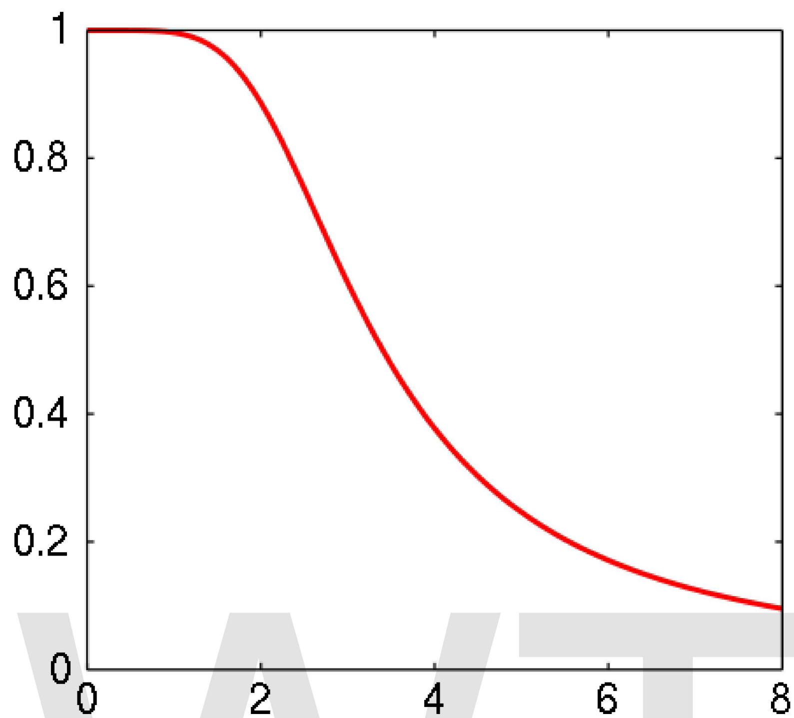
where

$$a_k = \frac{(2n-k)!}{2^{n-k} k!(n-k)!} \quad k = 0, 1, ..., n$$

Example



Gain plot of the third-order Bessel filter, versus normalized frequency



Group delay plot of the third-order Bessel filter, illustrating flat unit delay in the passband

The transfer function for a third-order (three-pole) Bessel low-pass filter, normalized to have unit group delay, is

$$H(s) = \frac{15}{s^3 + 6s^2 + 15s + 15}\,.$$

The roots of the denominator polynomial, the filter's poles, include a real pole at $s = -2.3222$, and a complex-conjugate pair of poles at $s = -1.8389 \pm j1.7544$, plotted above. The numerator 15 is chosen to give a gain of 1 at DC (at $s = 0$).

The gain is then

$$G(\omega) = |H(j\omega)| = \frac{15}{\sqrt{\omega^6 + 6\omega^4 + 45\omega^2 + 225}}\,.$$

The phase is

$$\phi(\omega) = -\arg(H(j\omega)) = -\arctan\left(\frac{15\omega - \omega^3}{15 - 6\omega^2}\right)\,.$$

The group delay is

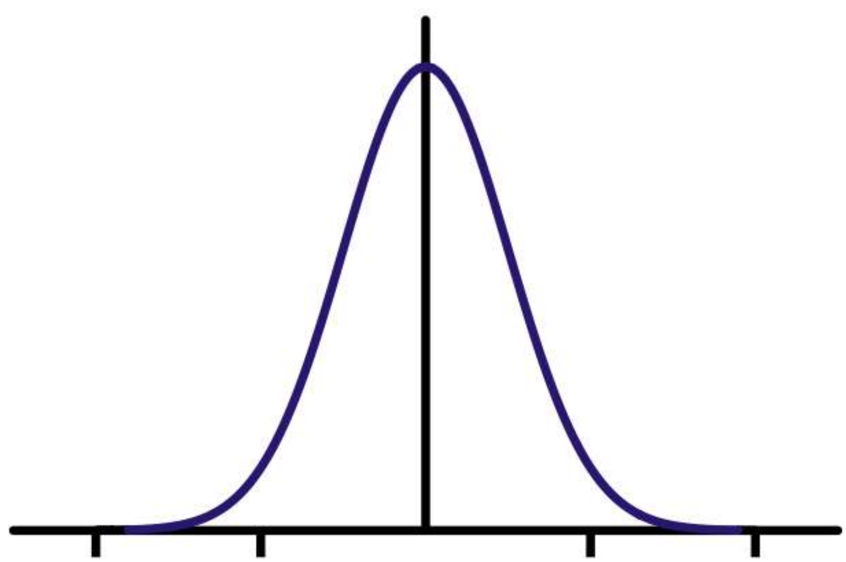
$$D(\omega) = -\frac{d\phi}{d\omega} = \frac{6\omega^4 + 45\omega^2 + 225}{\omega^6 + 6\omega^4 + 45\omega^2 + 225}\,.$$

The Taylor series expansion of the group delay is

$$D(\omega) = 1 - \frac{\omega^6}{225} + \frac{\omega^8}{1125} + \cdots\,.$$

Note that the two terms in ω^2 and ω^4 are zero, resulting in a very flat group delay at $\omega = 0$. This is the greatest number of terms that can be set to zero, since there are a total of four coefficients in the third order Bessel polynomial, requiring four equations in order to be defined. One equation specifies that the gain be unity at $\omega = 0$ and a second specifies that the gain be zero at $\omega = \infty$, leaving two equations to specify two terms in the series expansion to be zero. This is a general property of the group delay for a Bessel filter of order n: the first $n - 1$ terms in the series expansion of the group delay will be zero, thus maximizing the flatness of the group delay at $\omega = 0$.

Gaussian filter



Shape of a typical Gaussian filter

In electronics and signal processing, a **Gaussian filter** is a filter whose impulse response is a Gaussian function. Gaussian filters are designed to give no overshoot to a step function input while minimizing the rise and fall time. This behavior is closely connected to the fact that the Gaussian filter has the minimum possible group delay. Mathematically, a Gaussian filter modifies the input signal by convolution with a Gaussian function; this transformation is also known as the Weierstrass transform.

Definition

The one-dimensional Gaussian filter has an impulse response given by

$$g(x) = \sqrt{\frac{a}{\pi}} \cdot e^{-a \cdot x^2}$$

or with the standard deviation as parameter

$$g(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot e^{-\frac{x^2}{2\sigma^2}}$$

In two dimensions, it is the product of two such Gaussians, one per direction:

$$g(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \text{,}$$

where *x* is the distance from the origin in the horizontal axis, *y* is the distance from the origin in the vertical axis, and *σ* is the standard deviation of the Gaussian distribution.

Digital implementation

The Gaussian function is non-zero for $x \in [-\infty, \infty]$and would theoretically require an infinite window length. However, since it decays rapidly, it is often reasonable to truncate the filter window and implement the filter directly for narrow windows, in effect by using a simple rectangular window function. In other cases, the truncation may introduce significant errors.

Filtering involves convolution. The filter function is said to be the kernel of an integral transform. The Gaussian kernel is continuous. Most commonly, the discrete equivalent is the sampled Gaussian kernel that is produced by sampling points from the continuous Gaussian. An alternate method is to use the discrete Gaussian kernel which has superior characteristics for some purposes. Unlike the sample Gaussian kernel, the discrete Gaussian kernel is the solution to the discrete diffusion equation.

Since the Fourier transform of the Gaussian function yields a Gaussian function, the signal (preferably after being divided into overlapping windowed blocks) can be transformed with a Fast Fourier transform, multiplied with a Gaussian function and transformed back. This is the standard procedure of applying an arbitrary finite impulse response filter, with the only difference that the Fourier transform of the filter window is explicitly known.

Due to the central limit theorem, the Gaussian can be approximated by several runs of a very simple filter such as the moving average. The simple moving average corresponds to convolution with the constant B-spline, and, for example, four iterations of a moving average yields a cubic B-spline as filter window which approximates the Gaussian quite well.

Borrowing the terms from statistics, the standard deviation of a filter can be interpreted as a measure of its size. The cut-off frequency of the filter can be considered as the ratio between the sample rate F_s and the standard deviation σ

$$f_c = \frac{F_s}{\sigma}$$

A simple moving average corresponds to a uniform probability distribution and thus its filter width of size n has standard deviation $\sqrt{(\sigma^2 - 1)/12}$. Thus m moving averages with sizes $\sigma_1, \dots, \sigma_m$ yield a standard deviation of

$$\sigma = \sqrt{\frac{\sigma_1^2 + \cdots + \sigma_m^2 - m}{12}}.$$

(Note that standard deviations do not sum up, but variances do.)

When applied in two dimensions, this formula produces a Gaussian surface that has a maximum at the origin, whose contours are concentric circles with the origin as center. A two dimensional convolution matrix is precomputed from the formula and convolved with two dimensional data. Each element in the resultant matrix new value is set to a weighted average of that elements neighborhood. The focal element receives the heaviest weight (having the highest Gaussian value) and neighboring elements receive smaller weights as their distance to the focal element increases. In Image processing, each element in the matrix represents a pixel attribute such as brightness or a color intensity, and the overall effect is called Gaussian blur.

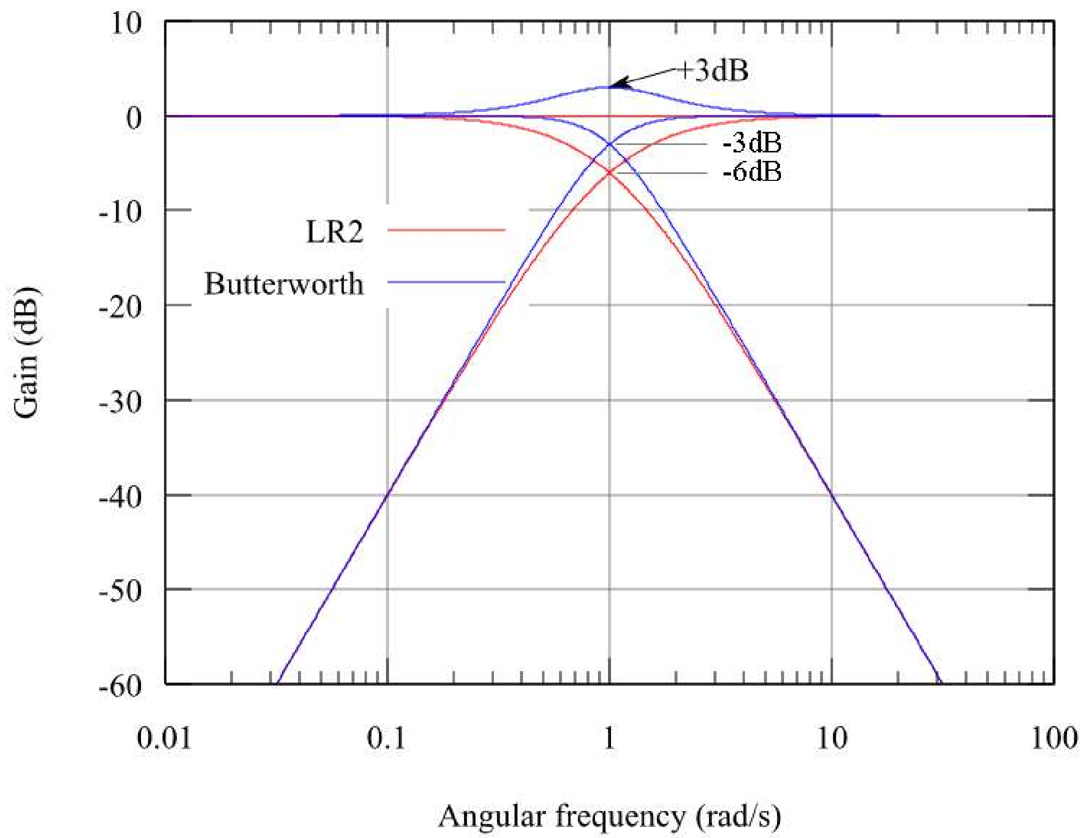
The Gaussian filter is non-causal which means the filter window is symmetric about the origin. This is usually of no consequence for most applications. In real-time systems, a delay is incurred because incoming samples need to fill the filter window before the filter can be applied to the signal.

Communications applications

- It is used in GSM since it applies GMSK modulation

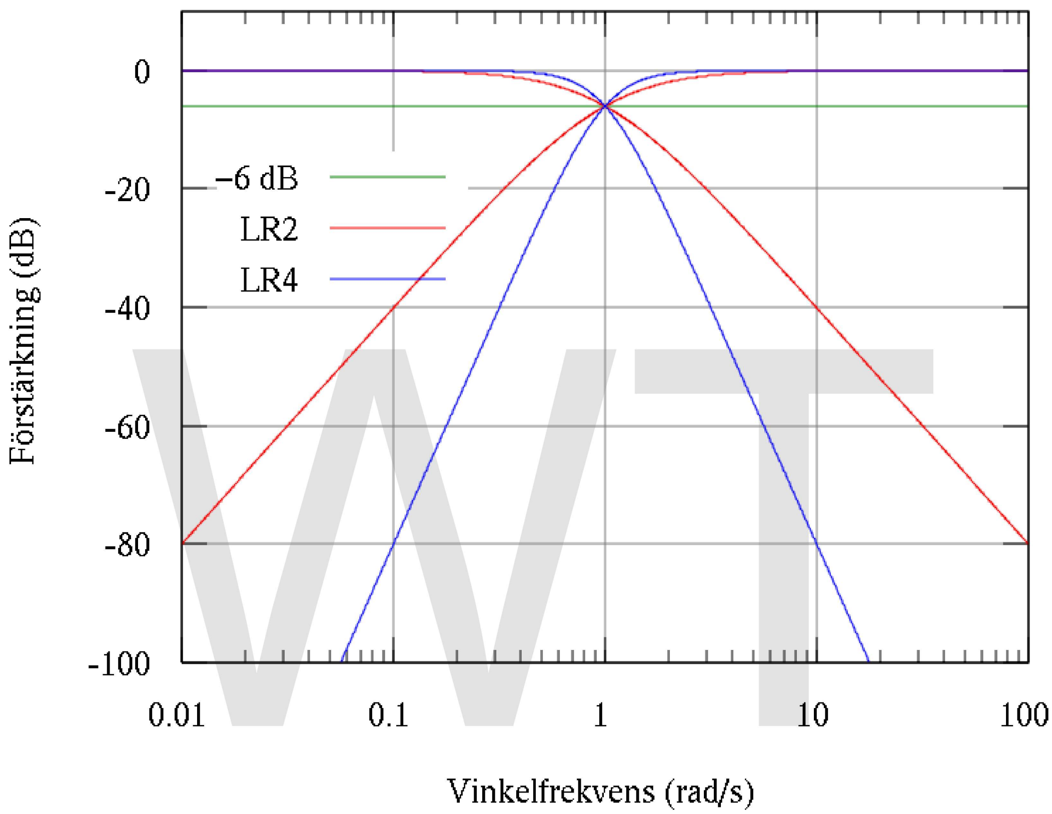
- the Gaussian filter is also used in GFSK.

Linkwitz–Riley filter

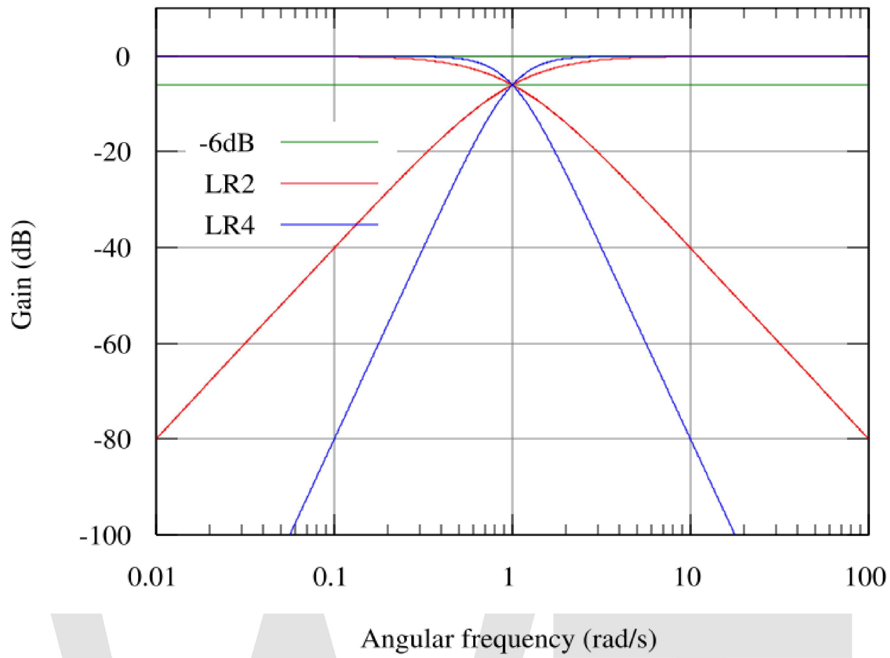


Comparison of the magitude response of the summed Butterworth and Linkwitz–Riley crossover filters. The Butterworth crossovers have a +3dB peak at the crossover frequency, whereas the L-R filters have a flat summed output.

A **Linkwitz–Riley (L-R) filter** is an infinite impulse response filter used in Linkwitz–Riley audio crossovers, named after its inventors Siegfried Linkwitz and Russ Riley, which was originally described in *Passive Crossover Networks for Noncoincident Drivers* in JAES Volume 26 Number 3 pp. 149-150; March 1978. It is also known as a *Butterworth squared* filter. An L-R crossover consists of a parallel combination of a low-pass and a high-pass L-R filter. The filters are usually designed by cascading two Butterworth filters, each of which has −3 dB gain at the cut-off frequency. The resulting Linkwitz–Riley filter has a −6 dB gain at the cutoff frequency. This means that summing the low-pass and high-pass outputs, the gain at the crossover frequency will be 0 dB, so the crossover behaves like an all-pass filter, having a flat amplitude response with a smoothly changing phase response. This is the biggest advantage of L-R crossovers compared to Butterworth crossovers, whose summed output has a +3 dB peak around the crossover frequency. Since cascading two nth order Butterworth filters will give a 2nth order Linkwitz–Riley filter, theoretically any 2nth order Linkwitz–Riley crossover can be designed. However, crossovers of higher order than 4th may have less usability due to their increasing peak in group delay around crossover frequency and complexity.



Linkwitz sv



Linkwitz

Common types

Second order Linkwitz–Riley crossover (LR2, LR-2)

Second order Linkwitz–Riley crossovers (LR2) have a 12 dB/octave (40 dB/decade) slope. They can be realized by cascading two one-pole filters, or using a Sallen Key filter topology with a Q_0 value of 0.5. There is a 180° phase difference between the lowpass and highpass output of the filter, which can be corrected by inverting one signal. In loudspeakers this is usually done by reversing the polarity of one driver if the crossover is passive. For active crossovers inversion is usually done using a unity gain inverting op-amp.

Fourth order Linkwitz–Riley crossover (LR4, LR-4)

Fourth order Linkwitz–Riley crossovers (LR4) are probably today's most commonly used type of audio crossover. They are constructed by cascading two second order Butterworth filters. Their steepness is 24 dB/octave (80 dB/decade). The phase difference amounts to 360°, i.e. the two drives appear in phase, albeit with a full period time delay for the low-pass section.

Eighth order Linkwitz–Riley crossover (LR8, LR-8)

Eighth order Linkwitz–Riley crossovers (LR8) have a very steep, 48 dB/octave (160 dB/decade) slope. They can be constructed by cascading two 4^{th} order Butterworth filters.

Chapter 6

RL Circuit

A **resistor-inductor circuit (RL circuit)**, or **RL filter** or **RL network**, is one of the simplest analogue infinite impulse response electronic filters. It consists of a resistor and an inductor, either in series or in parallel, driven by a voltage source.

Introduction

The fundamental passive linear circuit elements are the resistor (R), capacitor (C) and inductor (L). These circuit elements can be combined to form an electrical circuit in four distinct ways: the RC circuit, the RL circuit, the LC circuit and the RLC circuit with the abbreviations indicating which components are used. These circuits exhibit important types of behaviour that are fundamental to analogue electronics. In particular, they are able to act as passive filters.

In practice, however, capacitors (and RC circuits) are usually preferred to inductors since they can be more easily manufactured and are generally physically smaller, particularly for higher values of components.

Complex Impedance

The complex impedance Z_L (in ohms) of an inductor with inductance L (in henries) is

$$Z_L = Ls$$

The complex frequency s is a complex number,

$$s = \sigma + j\omega$$

where

- j represents the imaginary unit:

  $j^2 = -1$

- σ is the exponential decay constant (in radians per second), and

- ω is the angular frequency (in radians per second).

Eigenfunctions

The complex-valued **eigenfunctions** of ANY linear time-invariant (LTI) system are of the following forms:

$V(t) = \mathbf{A}e^{st} = \mathbf{A}e^{(\sigma+j\omega)t}$, or letting $\mathbf{A} = Ae^{j\phi}$ and rewriting;
$= Ae^{j\phi}e^{(\sigma+j\omega)t}$, and collecting terms is $= Ae^{\sigma t}e^{j(\omega t+\phi)}$

From Euler's formula, the **real-part** of these eigenfunctions are exponentially-decaying sinusoids:

$$v(t) = \mathrm{Re}\{V(t)\} = Ae^{\sigma t}\cos(\omega t + \phi)$$

Sinusoidal Steady State

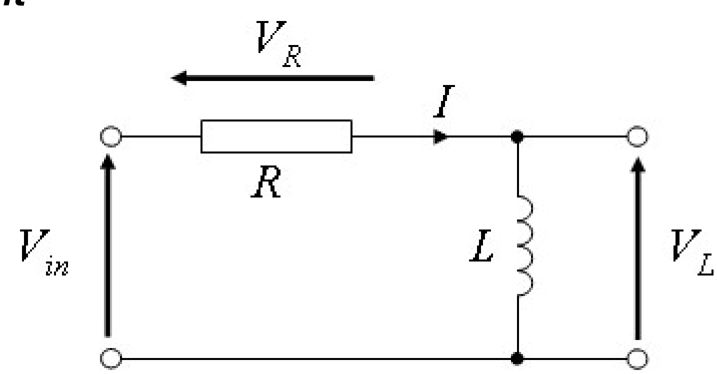
Sinusoidal steady state is a special case in which the input voltage consists of a pure sinusoid (with no exponential decay). As a result,

$$\sigma = 0$$

and the evaluation of *s* becomes

$$s = j\omega$$

Series circuit



Series RL circuit

By viewing the circuit as a voltage divider, we see that the voltage across the inductor is:

$$V_L(s) = \frac{Ls}{R + Ls}V_{in}(s)$$

and the voltage across the resistor is:

$$V_R(s) = \frac{R}{R + Ls} V_{in}(s).$$

Current

The current in the circuit is the same everywhere since the circuit is series:

$$I(s) = \frac{V_{in}(s)}{R + Ls}.$$

Transfer functions

The transfer function for the inductor is

$$H_L(s) = \frac{V_L(s)}{V_{in}(s)} = \frac{Ls}{R + Ls} = G_L e^{j\phi_L}$$

Similarly, the transfer function for the resistor is

$$H_R(s) = \frac{V_R(s)}{V_{in}(s)} = \frac{R}{R + Ls} = G_R e^{j\phi_R}$$

Poles and zeros

Both transfer functions have a single pole located at

$$s = -\frac{R}{L}$$

In addition, the transfer function for the inductor has a zero located at the origin.

Gain and phase angle

The gains across the two components are found by taking the magnitudes of the above expressions:

$$G_L = |H_L(s)| = \left|\frac{V_L(s)}{V_{in}(s)}\right| = \frac{\omega L}{\sqrt{R^2 + (\omega L)^2}}$$

and

$$G_R = |H_R(s)| = \left|\frac{V_R(s)}{V_{in}(s)}\right| = \frac{R}{\sqrt{R^2 + (\omega L)^2}},$$

and the phase angles are:

$$\phi_L = \angle H_L(s) = \tan^{-1}\left(\frac{R}{\omega L}\right)$$

and

$$\phi_R = \angle H_R(s) = \tan^{-1}\left(-\frac{\omega L}{R}\right).$$

Phasor notation

These expressions together may be substituted into the usual expression for the phasor representing the output:

$$V_L = G_L V_{in} e^{j\phi_L}$$
$$V_R = G_R V_{in} e^{j\phi_R}.$$

Impulse Response

The impulse response for each voltage is the inverse Laplace transform of the corresponding transfer function. It represents the response of the circuit to an input voltage consisting of an impulse or Dirac delta function.

The impulse response for the inductor voltage is

$$h_L(t) = \delta(t) - \frac{R}{L}e^{-tR/L}u(t) = \delta(t) - \frac{1}{\tau}e^{-t/\tau}u(t)$$

where $u(t)$ is the Heaviside step function and

$$\tau = \frac{L}{R}$$

is the time constant.

Similarly, the impulse response for the resistor voltage is

$$h_R(t) = \frac{R}{L}e^{-tR/L}u(t) = \frac{1}{\tau}e^{-t/\tau}u(t)$$

Zero input response (ZIR)

The **Zero input response**, also called the **natural response**, of an RL circuit describes the behavior of the circuit after it has reached constant voltages and currents and is disconnected from any power source. It is called the zero-input response because it requires no input.

The ZIR of an RL circuit is:

$$i(t) = i(0)e^{-(R/L)t} = i(0)e^{-t/\tau}.$$

Frequency domain considerations

These are frequency domain expressions. Analysis of them will show which frequencies the circuits (or filters) pass and reject. This analysis rests on a consideration of what happens to these gains as the frequency becomes very large and very small.

As $\omega \to \infty$:

$$G_L \to 1$$
$$G_R \to 0.$$

As $\omega \to 0$:

$$G_L \to 0$$
$$G_R \to 1.$$

This shows that, if the output is taken across the inductor, high frequencies are passed and low frequencies are attenuated (rejected). Thus, the circuit behaves as a *high-pass filter*. If, though, the output is taken across the resistor, high frequencies are rejected and low frequencies are passed. In this configuration, the circuit behaves as a *low-pass filter*. Compare this with the behaviour of the resistor output in an RC circuit, where the reverse is the case.

The range of frequencies that the filter passes is called its bandwidth. The point at which the filter attenuates the signal to half its unfiltered power is termed its cutoff frequency. This requires that the gain of the circuit be reduced to

$$G_L = G_R = \frac{1}{\sqrt{2}}.$$

Solving the above equation yields

$$\omega_c = \frac{R}{L} \text{ rad/s}$$

or

$$f_c = \frac{R}{2\pi L} \text{ Hz}$$

which is the frequency that the filter will attenuate to half its original power.

Clearly, the phases also depend on frequency, although this effect is less interesting generally than the gain variations.

As $\omega \to 0$:

$$\phi_L \to 90^\circ = \pi/2^c$$
$$\phi_R \to 0.$$

As $\omega \to \infty$:

$$\phi_L \to 0$$
$$\phi_R \to -90^\circ = -\pi/2^c$$

So at DC (0 Hz), the resistor voltage is in phase with the signal voltage while the inductor voltage leads it by 90°. As frequency increases, the resistor voltage comes to have a 90° lag relative to the signal and the inductor voltage comes to be in-phase with the signal.

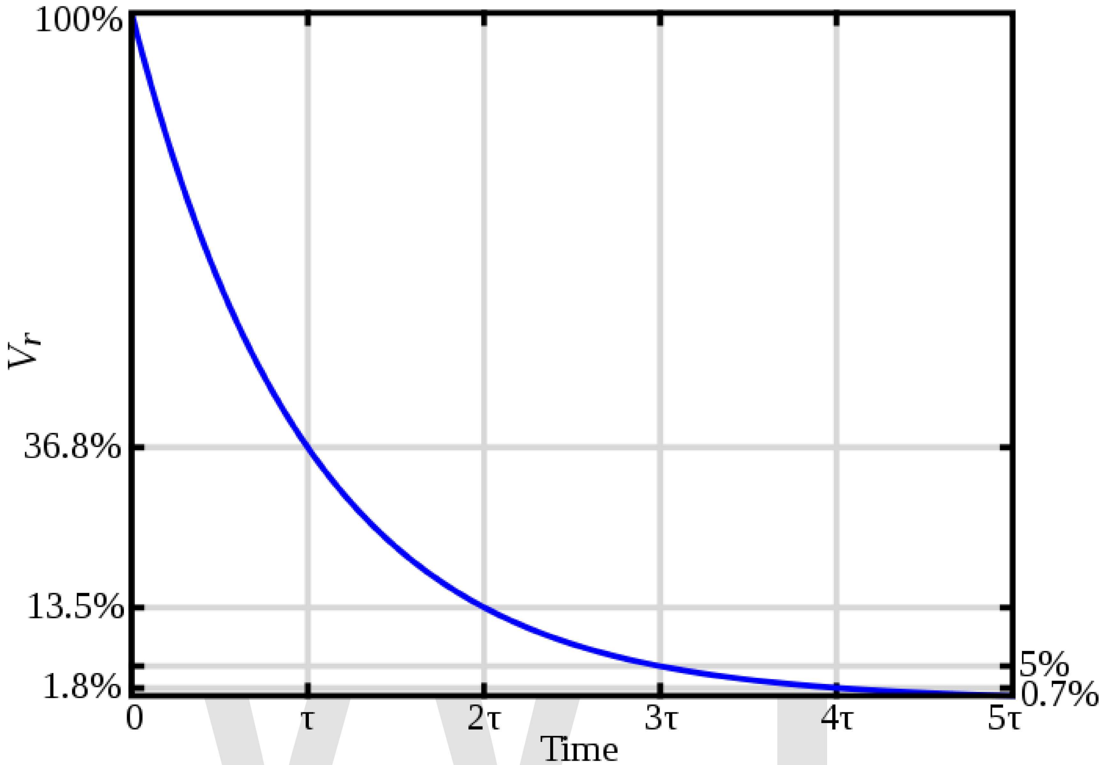
Time domain considerations

The most straightforward way to derive the time domain behaviour is to use the Laplace transforms of the expressions for V_L and V_R given above. This effectively transforms $j\omega \to s$. Assuming a step input (i.e. $V_{in} = 0$ before $t = 0$ and then $V_{in} = V$ afterwards):

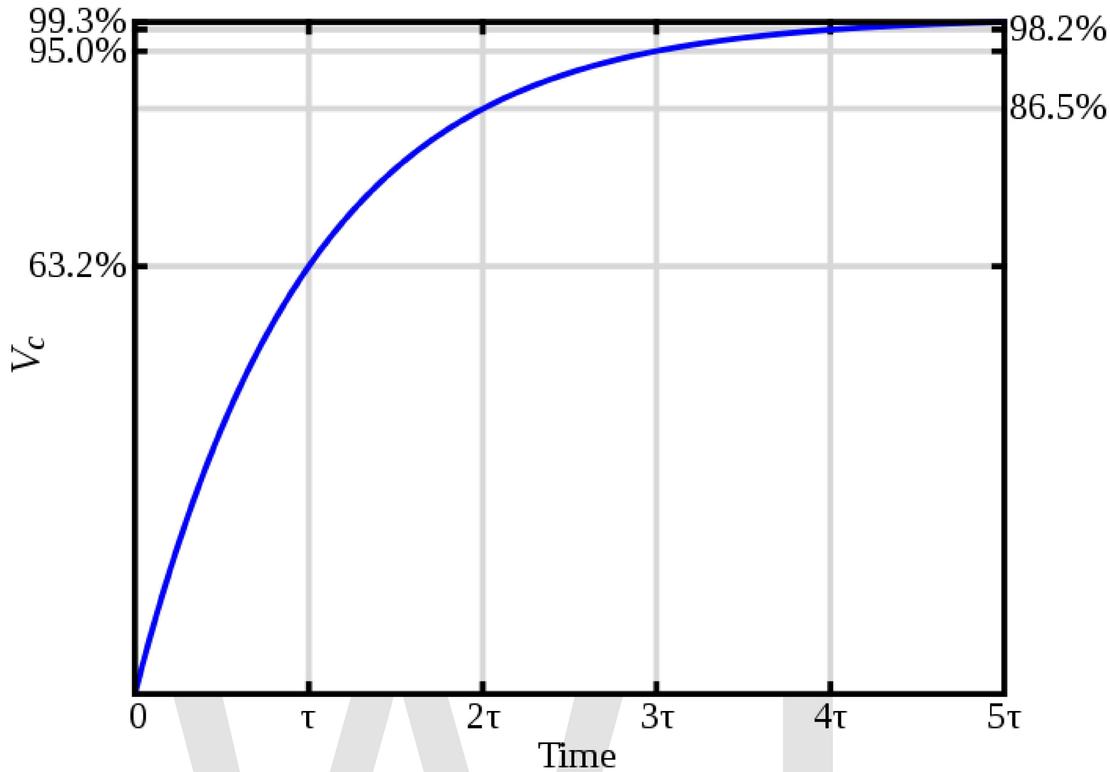
$$V_{in}(s) = V\frac{1}{s}$$
$$V_L(s) = V\frac{sL}{R + sL}\frac{1}{s}$$

and

$$V_R(s) = V\frac{R}{R+sL}\frac{1}{s}.$$



Inductor voltage step-response



Resistor voltage step-response

Partial fractions expansions and the inverse Laplace transform yield:

$$V_L(t) = Ve^{-tR/L}$$
$$V_R(t) = V\left(1 - e^{-tR/L}\right).$$

Thus, the voltage across the inductor tends towards 0 as time passes, while the voltage across the resistor tends towards V, as shown in the figures. This is in keeping with the intuitive point that the inductor will only have a voltage across as long as the current in the circuit is changing — as the circuit reaches its steady-state, there is no further current change and ultimately no inductor voltage.

These equations show that a series RL circuit has a time constant, usually denoted $\tau = L / R$ being the time it takes the voltage across the component to either fall (across L) or rise (across R) to within $1 / e$ of its final value. That is, τ is the time it takes V_L to reach $V(1 / e)$ and V_R to reach $V(1 - 1 / e)$.

The rate of change is a *fractional* $\left(1 - \frac{1}{e}\right)$ per τ. Thus, in going from $t = N\tau$ to $t = (N + 1)\tau$, the votage will have moved about 63% of the way from its level at $t = N\tau$ toward its final value. So the voltage across L will have dropped to about 37% after τ, and essentially to zero (0.7%) after about 5τ. Kirchhoff's voltage law implies that the voltage across the resistor will *rise* at the same rate. When the voltage source is then replaced with a short-circuit, the voltage across R drops exponentially with t from V towards 0. R will be discharged to about 37% after τ, and essentially fully discharged (0.7%) after about 5τ. Note that the current, I, in the circuit behaves as the voltage across R does, via Ohm's Law.

The delay in the rise/fall time of the circuit is in this case caused by the back-EMF from the inductor which, as the current flowing through it tries to change, prevents the current (and hence the voltage across the resistor) from rising or falling much faster than the time-constant of the circuit. Since all wires have some self-inductance and resistance, all circuits have a time constant. As a result, when the power supply is switched on, the current does not instantaneously reach its steady-state value, V / R. The rise instead takes several time-constants to complete. If this were not the case, and the current were to reach steady-state immediately, extremely strong inductive electric fields would be generated by the sharp change in the magnetic field — this would lead to breakdown of the air in the circuit and electric arcing, probably damaging components (and users).

These results may also be derived by solving the differential equation describing the circuit:

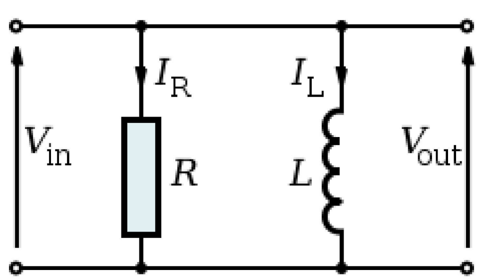
$$V_{in} = IR + L\frac{dI}{dt},$$

and

$$V_R = V_{in} - V_L.$$

The first equation is solved by using an integrating factor and yields the current which must be differentiated to give V_L; the second equation is straightforward. The solutions are exactly the same as those obtained via Laplace transforms.

Parallel circuit



Parallel RL circuit

The parallel RL circuit is generally of less interest than the series circuit unless fed by a current source. This is largely because the output voltage V_{out} is equal to the input voltage V_{in} — as a result, this circuit does not act as a filter for a voltage input signal.

With complex impedances:

$$I_R = \frac{V_{in}}{R}$$

and

$$I_L = \frac{V_{in}}{j\omega L} = -\frac{jV_{in}}{\omega L}.$$

This shows that the inductor lags the resistor (and source) current by 90°.

The parallel circuit is seen on the output of many amplifier circuits, and is used to isolate the amplifier from capacitive loading effects at high frequencies. Because of the phase shift introduced by capacitance, some amplifiers become unstable at very high frequencies, and tend to oscillate. This affects sound quality and component life (especially the transistors), and is to be avoided.

Chapter 7

RC Filter

A **resistor–capacitor circuit (RC circuit)**, or **RC filter** or **RC network**, is an electric circuit composed of resistors and capacitors driven by a voltage or current source. A first order RC circuit is composed of one resistor and one capacitor and is the simplest type of RC circuit.

RC circuits can be used to filter a signal by blocking certain frequencies and passing others. The four most common RC filters are the high-pass filter, low-pass filter, band-pass filter, and band-stop filter.

Introduction

There are three basic, linear passive lumped analog circuit components: the resistor (R), capacitor (C) and inductor (L). These may be combined in: the RC circuit, the RL circuit, the LC circuit and the RLC circuit with the abbreviations indicating which components are used. These circuits, between them, exhibit a large number of important types of behaviour that are fundamental to much of analog electronics. In particular, they are able to act as passive filters.

Natural response

The simplest RC circuit is a capacitor and a resistor in series. When a circuit consists of only a charged capacitor and a resistor, the capacitor will discharge its stored energy through the resistor. The voltage across the capacitor, which is time dependent, can be found by using Kirchhoff's current law, where the current through the capacitor must equal the current through the resistor. This results in the linear differential equation

$$C\frac{dV}{dt} + \frac{V}{R} = 0.$$

Solving this equation for V yields the formula for exponential decay:

$$V(t) = V_0 e^{-\frac{t}{RC}} ,$$

where V_0 is the capacitor voltage at time $t = 0$.

The time required for the voltage to fall to $\frac{V_0}{e}$ is called the RC time constant and is given by

$$\tau = RC \ .$$

Complex impedance

The complex impedance, Z_C (in ohms) of a capacitor with capacitance C (in farads) is

$$Z_C = \frac{1}{sC}$$

The complex frequency s is, in general, a complex number,

$$s \ = \ \sigma + j\omega$$

where

- j represents the imaginary unit:

  $j^2 = -1$

- σ is the exponential decay constant (in radians per second), and
- ω is the sinusoidal angular frequency (also in radians per second).

Sinusoidal steady state

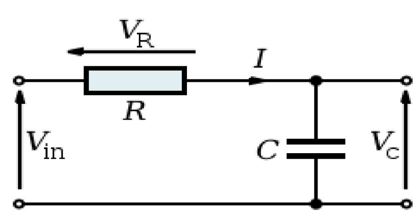
Sinusoidal steady state is a special case in which the input voltage consists of a pure sinusoid (with no exponential decay). As a result,

$$\sigma \ = \ 0$$

and the evaluation of s becomes

$$s \ = \ j\omega$$

Series circuit



Series RC circuit

By viewing the circuit as a voltage divider, the voltage across the capacitor is:

$$V_C(s) = \frac{1/Cs}{R + 1/Cs} V_{in}(s) = \frac{1}{1 + RCs} V_{in}(s)$$

and the voltage across the resistor is:

$$V_R(s) = \frac{R}{R + 1/Cs} V_{in}(s) = \frac{RCs}{1 + RCs} V_{in}(s).$$

Transfer functions

The transfer function for the capacitor is

$$H_C(s) = \frac{V_C(s)}{V_{in}(s)} = \frac{1}{1 + RCs}.$$

Similarly, the transfer function for the resistor is

$$H_R(s) = \frac{V_R(s)}{V_{in}(s)} = \frac{RCs}{1 + RCs}.$$

Poles and zeros

Both transfer functions have a single pole located at

$$s = -\frac{1}{RC}.$$

In addition, the transfer function for the resistor has a zero located at the origin.

Gain and phase angle

The magnitude of the gains across the two components are:

$$G_C = |H_C(j\omega)| = \left| \frac{V_C(j\omega)}{V_{in}(j\omega)} \right| = \frac{1}{\sqrt{1 + (\omega RC)^2}}$$

and

$$G_R = |H_R(j\omega)| = \left|\frac{V_R(j\omega)}{V_{in}(j\omega)}\right| = \frac{\omega RC}{\sqrt{1+(\omega RC)^2}},$$

and the phase angles are:

$$\phi_C = \angle H_C(j\omega) = \tan^{-1}(-\omega RC)$$

and

$$\phi_R = \angle H_R(j\omega) = \tan^{-1}\left(\frac{1}{\omega RC}\right).$$

These expressions together may be substituted into the usual expression for the phasor representing the output:

$$\begin{aligned} V_C &= G_C V_{in} e^{j\phi_C} \\ V_R &= G_R V_{in} e^{j\phi_R}. \end{aligned}$$

Current

The current in the circuit is the same everywhere since the circuit is in series:

$$I(s) = \frac{V_{in}(s)}{R+\frac{1}{Cs}} = \frac{Cs}{1+RCs} V_{in}(s)$$

Impulse response

The impulse response for each voltage is the inverse Laplace transform of the corresponding transfer function. It represents the response of the circuit to an input voltage consisting of an impulse or Dirac delta function.

The impulse response for the capacitor voltage is

$$h_C(t) = \frac{1}{RC} e^{-t/RC} u(t) = \frac{1}{\tau} e^{-t/\tau} u(t)$$

where $u(t)$ is the Heaviside step function and

$$\tau = RC$$

is the time constant.

Similarly, the impulse response for the resistor voltage is

$$h_R(t) = \delta(t) - \frac{1}{RC}e^{-t/RC}u(t) = \delta(t) - \frac{1}{\tau}e^{-t/\tau}u(t)$$

where $\delta(t)$ is the Dirac delta function

Frequency-domain considerations

These are frequency domain expressions. Analysis of them will show which frequencies the circuits (or filters) pass and reject. This analysis rests on a consideration of what happens to these gains as the frequency becomes very large and very small.

As $\omega \to \infty$:

$$G_C \to 0$$
$$G_R \to 1.$$

As $\omega \to 0$:

$$G_C \to 1$$
$$G_R \to 0.$$

This shows that, if the output is taken across the capacitor, high frequencies are attenuated (rejected) and low frequencies are passed. Thus, the circuit behaves as a *low-pass filter*. If, though, the output is taken across the resistor, high frequencies are passed and low frequencies are rejected. In this configuration, the circuit behaves as a *high-pass filter*.

The range of frequencies that the filter passes is called its bandwidth. The point at which the filter attenuates the signal to half its unfiltered power is termed its cutoff frequency. This requires that the gain of the circuit be reduced to

$$G_C = G_R = \frac{1}{\sqrt{2}}.$$

Solving the above equation yields

$$\omega_c = \frac{1}{RC} \text{ rad/s}$$

or

$$f_c = \frac{1}{2\pi RC} \text{ Hz}$$

which is the frequency that the filter will attenuate to half its original power.

Clearly, the phases also depend on frequency, although this effect is less interesting generally than the gain variations.

As $\omega \to 0$:

$$\phi_C \to 0$$
$$\phi_R \to 90^\circ = \pi/2^c.$$

As $\omega \to \infty$:

$$\phi_C \to -90^\circ = -\pi/2^c$$
$$\phi_R \to 0$$

So at DC (0 Hz), the capacitor voltage is in phase with the signal voltage while the resistor voltage leads it by 90°. As frequency increases, the capacitor voltage comes to have a 90° lag relative to the signal and the resistor voltage comes to be in-phase with the signal.

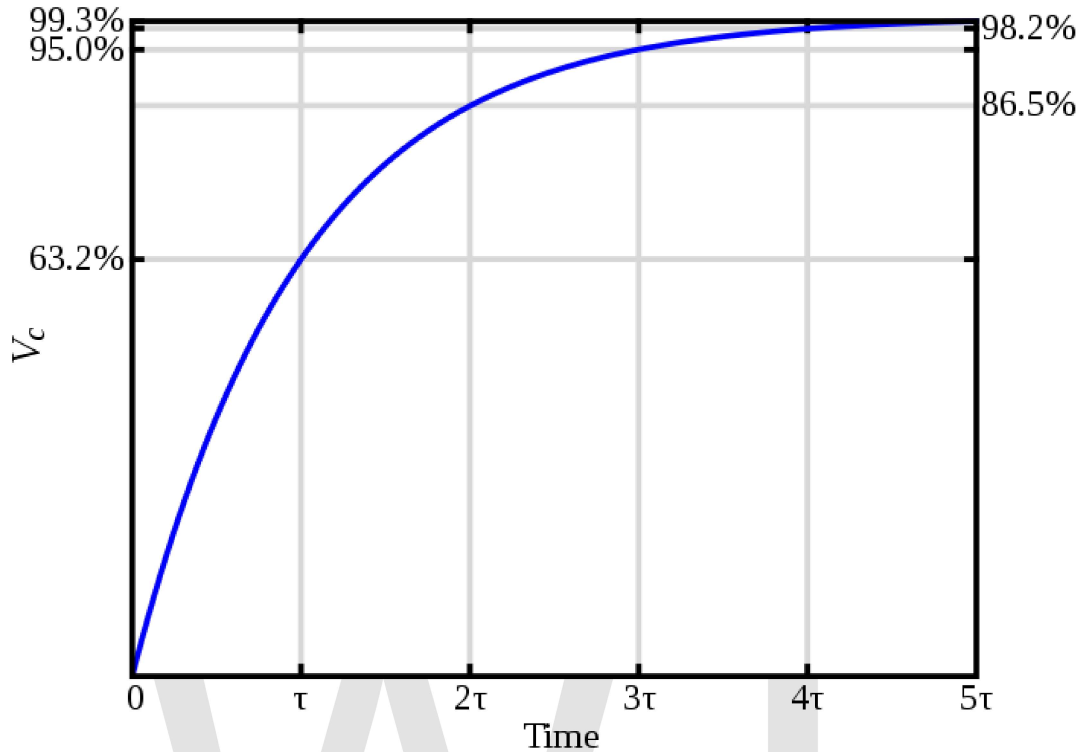
Time-domain considerations

The most straightforward way to derive the time domain behaviour is to use the Laplace transforms of the expressions for V_C and V_R given above. This effectively transforms $j\omega \to s$. Assuming a step input (i.e. $V_{in} = 0$ before $t = 0$ and then $V_{in} = V$ afterwards):

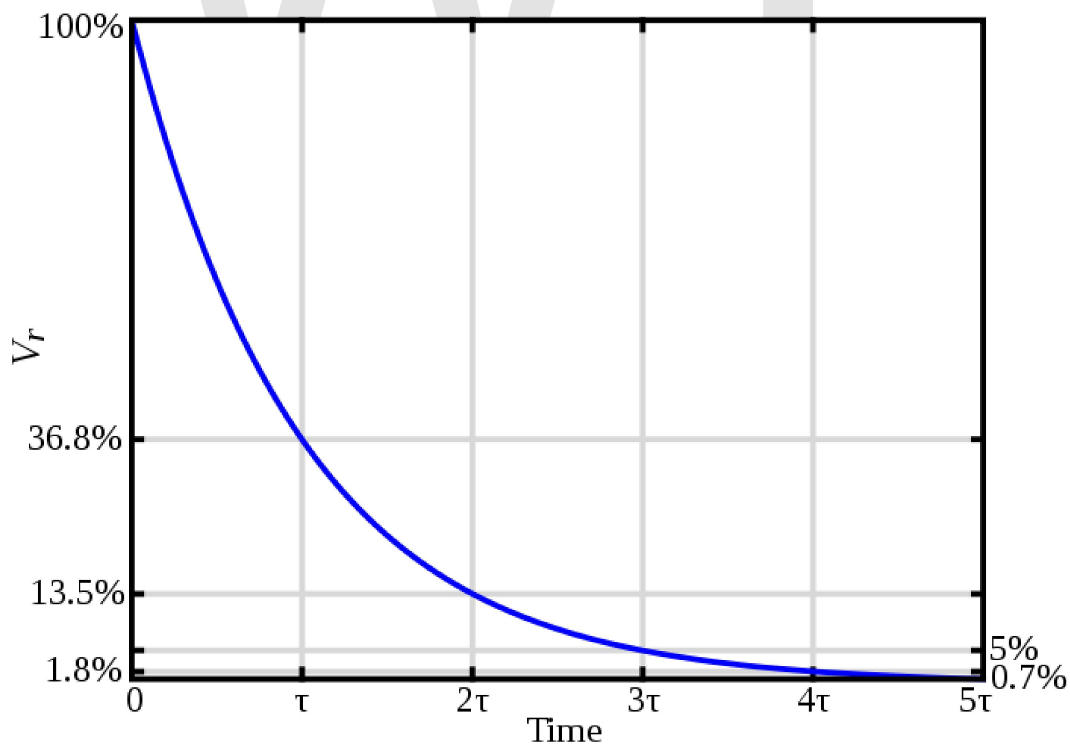
$$V_{in}(s) = V\frac{1}{s}$$
$$V_C(s) = V\frac{1}{1+sRC}\frac{1}{s}$$

and

$$V_R(s) = V\frac{sRC}{1+sRC}\frac{1}{s}.$$



Capacitor voltage step-response



Resistor voltage step-response

Partial fractions expansions and the inverse Laplace transform yield:

$$V_C(t) = V\left(1 - e^{-t/RC}\right)$$
$$V_R(t) = Ve^{-t/RC}.$$

These equations are for calculating the voltage across the capacitor and resistor respectively while the capacitor is charging; for discharging, the equations are vice-versa. These equations can be rewritten in terms of charge and current using the relationships C=Q/V and V=IR.

Thus, the voltage across the capacitor tends towards V as time passes, while the voltage across the resistor tends towards 0, as shown in the figures. This is in keeping with the intuitive point that the capacitor will be charging from the supply voltage as time passes, and will eventually be fully charged and form an open circuit.

These equations show that a series RC circuit has a time constant, usually denoted $\tau = RC$ being the time it takes the voltage across the component to either rise (across C) or fall (across R) to within $1 / e$ of its final value. That is, τ is the time it takes V_C to reach $V(1 - 1 / e)$ and V_R to reach $V(1 / e)$.

The rate of change is a *fractional* $\left(1 - \frac{1}{e}\right)$ per τ. Thus, in going from $t = N\tau$ to $t = (N + 1)\tau$, the voltage will have moved about 63.2 % of the way from its level at $t = N\tau$ toward its final value. So C will be charged to about 63.2 % after τ, and essentially fully charged (99.3 %) after about 5τ. When the voltage source is replaced with a short-circuit, with C fully charged, the voltage across C drops exponentially with t from V towards 0. C will be discharged to about 36.8 % after τ, and essentially fully discharged (0.7 %) after about 5τ. Note that the current, I, in the circuit behaves as the voltage across R does, via Ohm's Law.

These results may also be derived by solving the differential equations describing the circuit:

$$\frac{V_{in} - V_C}{R} = C\frac{dV_C}{dt}$$

and

$$V_R = V_{in} - V_C.$$

The first equation is solved by using an integrating factor and the second follows easily; the solutions are exactly the same as those obtained via Laplace transforms.

Integrator

Consider the output across the capacitor at *high* frequency i.e.

$$\omega \gg \frac{1}{RC}.$$

This means that the capacitor has insufficient time to charge up and so its voltage is very small. Thus the input voltage approximately equals the voltage across the resistor. To see this, consider the expression for *I* given above:

$$I = \frac{V_{in}}{R + 1/j\omega C}$$

but note that the frequency condition described means that

$$\omega C \gg \frac{1}{R}$$

so

$$I \approx \frac{V_{in}}{R}$$ which is just Ohm's Law.

Now,

$$V_C = \frac{1}{C} \int_0^t I dt$$

so

$$V_C \approx \frac{1}{RC} \int_0^t V_{in} dt,$$

which is an integrator *across the capacitor*.

Differentiator

Consider the output across the resistor at *low* frequency i.e.,

$$\omega \ll \frac{1}{RC}.$$

This means that the capacitor has time to charge up until its voltage is almost equal to the source's voltage. Considering the expression for *I* again, when

$$R \ll \frac{1}{\omega C},$$

so

$$I \approx \frac{V_{in}}{1/j\omega C}$$
$$V_{in} \approx \frac{I}{j\omega C} \approx V_C$$

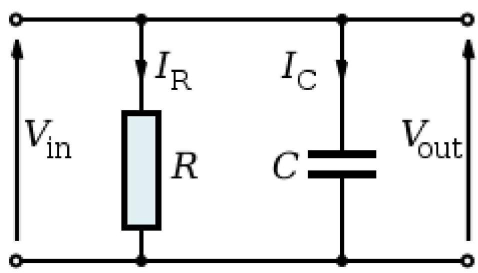
Now,

$$V_R = IR = C\frac{dV_C}{dt}R$$
$$V_R \approx RC\frac{dV_{in}}{dt}$$

which is a differentiator *across the resistor*.

More accurate integration and differentiation can be achieved by placing resistors and capacitors as appropriate on the input and feedback loop of operational amplifiers.

Parallel circuit



Parallel RC circuit

The parallel RC circuit is generally of less interest than the series circuit. This is largely because the output voltage V_{out} is equal to the input voltage V_{in} — as a result, this circuit does not act as a filter on the input signal unless fed by a current source.

With complex impedances:

$$I_R = \frac{V_{in}}{R}$$

and

$$I_C = j\omega C V_{in}.$$

This shows that the capacitor current is 90° out of phase with the resistor (and source) current. Alternatively, the governing differential equations may be used:

$$I_R = \frac{V_{in}}{R}$$

and

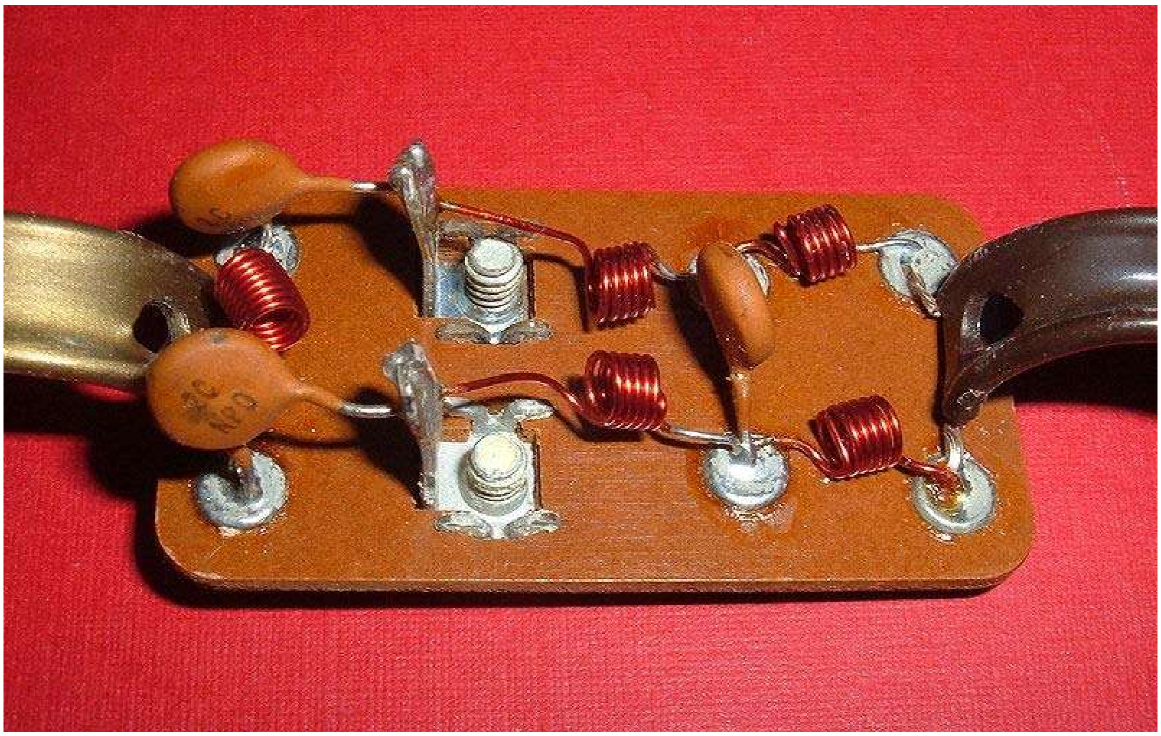
$$I_C = C\frac{dV_{in}}{dt}.$$

When fed by a current source, the transfer function of a parallel RC circuit is

$$\frac{V_{out}}{I_{in}} = \frac{R}{1 + sRC}.$$

Chapter 8

Electronic Filter



Television signal splitter consisting of a high-pass filter (left) and a low-pass filter (right). The antenna is connected to the screw terminals to the left of center.

Electronic filters are electronic circuits which perform signal processing functions, specifically to remove unwanted frequency components from the signal, to enhance wanted ones, or both. Electronic filters can be:

- passive or active
- analog or digital
- high-pass, low-pass, bandpass, band-reject (band reject; notch), or all-pass.
- discrete-time (sampled) or continuous-time
- linear or non-linear
- infinite impulse response (IIR type) or finite impulse response (FIR type)

The most common types of electronic filters are linear filters, regardless of other aspects of their design.

History

The oldest forms of electronic filters are passive analog linear filters, constructed using only resistors and capacitors or resistors and inductors. These are known as RC and RL single-pole filters respectively. More complex multipole LC filters have also existed for many years, and their operation is well understood.

Hybrid filters are also possible, typically involving a combination of analog amplifiers with mechanical resonators or delay lines. Other devices such as CCD delay lines have also been used as discrete-time filters. With the availability of digital signal processing, active digital filters have become common.

Classification by technology

Passive filters

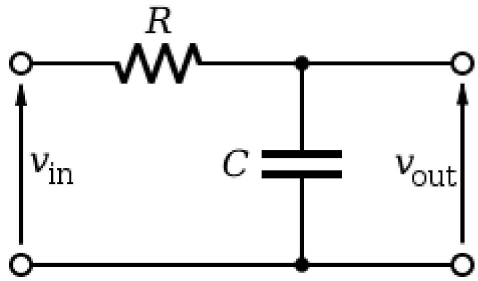
Passive implementations of linear filters are based on combinations of resistors (R), inductors (L) and capacitors (C). These types are collectively known as *passive filters*, because they do not depend upon an external power supply and/or they do not contain active components such as transistors.

Inductors block high-frequency signals and conduct low-frequency signals, while capacitors do the reverse. A filter in which the signal passes through an inductor, or in which a capacitor provides a path to ground, presents less attenuation to low-frequency signals than high-frequency signals and is a *low-pass filter*. If the signal passes through a capacitor, or has a path to ground through an inductor, then the filter presents less attenuation to high-frequency signals than low-frequency signals and is a *high-pass filter*. Resistors on their own have no frequency-selective properties, but are added to inductors and capacitors to determine the *time-constants* of the circuit, and therefore the frequencies to which it responds.

The inductors and capacitors are the reactive elements of the filter. The number of elements determines the order of the filter. In this context, an LC tuned circuit being used in a band-pass or band-stop filter is considered a single element even though it consists of two components.

At high frequencies (above about 100 megahertz), sometimes the inductors consist of single loops or strips of sheet metal, and the capacitors consist of adjacent strips of metal. These inductive or capacitive pieces of metal are called stubs.

Single element types



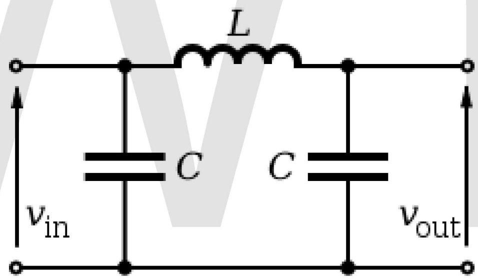
A low-pass electronic filter realised by an RC circuit

The simplest passive filters, RC and RL filters, include only one reactive element, except hybrid LC filter which is characterized by inductance and capacitance integrated in one element.

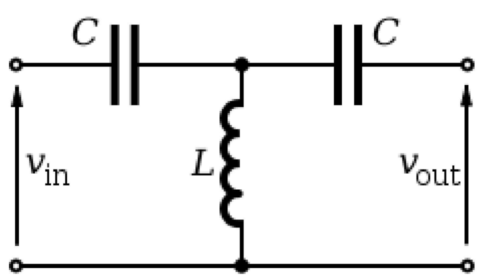
L filter

An L filter consists of two reactive elements, one in series and one in parallel.

T and π filters



Low-pass π filter



High-pass T filter

Three-element filters can have a 'T' or 'π' topology and in either geometries, a low-pass, high-pass, band-pass, or band-stop characteristic is possible. The components can be chosen symmetric or not, depending on the required frequency characteristics. The high-pass T filter in the illustration, has a very low impedance at high frequencies, and a very high impedance at low frequencies. That means that it can be inserted in a transmission line, resulting in the high frequencies being passed and low frequencies being reflected. Likewise, for the illustrated low-pass π filter, the circuit can be connected to a transmission line, transmitting low frequencies and reflecting high frequencies. Using m-derived filter sections with correct termination impedances, the input impedance can be reasonably constant in the pass band.

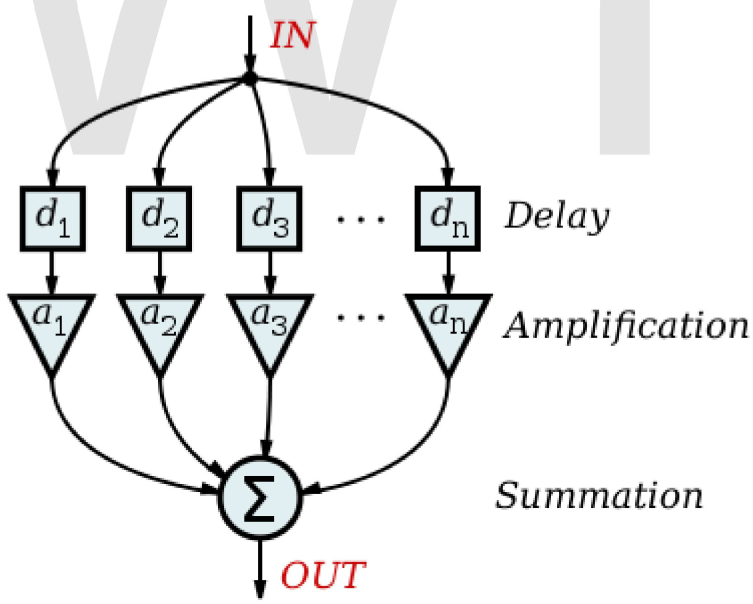
Multiple element types

Multiple element filters are usually constructed as a ladder network. These can be seen as a continuation of the L,T and π designs of filters. More elements are needed when it is desired to improve some parameter of the filter such as stop-band rejection or slope of transition from pass-band to stop-band.

Active filters

Active filters are implemented using a combination of passive and active (amplifying) components, and require an outside power source. Operational amplifiers are frequently used in active filter designs. These can have high Q factor, and can achieve resonance without the use of inductors. However, their upper frequency limit is limited by the bandwidth of the amplifiers used.

Digital filters



A general finite impulse response filter with *n* stages, each with an independent delay, d_i and amplification gain, a_i.

Digital signal processing allows the inexpensive construction of a wide variety of filters. The signal is sampled and an analog-to-digital converter turns the signal into a stream of numbers. A computer program running on a CPU or a specialized DSP (or less often running on a hardware implementation of the algorithm) calculates an output number stream. This output can be converted to a signal by passing it through a digital-to-analog converter. There are problems with noise introduced by the conversions, but these can be controlled and limited for many useful filters. Due to the sampling involved, the input signal must be of limited frequency content or aliasing will occur.

Other filter technologies

Quartz filters and piezoelectrics

In the late 1930s, engineers realized that small mechanical systems made of rigid materials such as quartz would acoustically resonate at radio frequencies, i.e. from audible frequencies (sound) up to several hundred megahertz. Some early resonators were made of steel, but quartz quickly became favored. The biggest advantage of quartz is that it is piezoelectric. This means that quartz resonators can directly convert their own mechanical motion into electrical signals. Quartz also has a very low coefficient of thermal expansion which means that quartz resonators can produce stable frequencies over a wide temperature range. Quartz crystal filters have much higher quality factors than LCR filters. When higher stabilities are required, the crystals and their driving circuits may be mounted in a "crystal oven" to control the temperature. For very narrow band filters, sometimes several crystals are operated in series.

Engineers realized that a large number of crystals could be collapsed into a single component, by mounting comb-shaped evaporations of metal on a quartz crystal. In this scheme, a "tapped delay line" reinforces the desired frequencies as the sound waves flow across the surface of the quartz crystal. The tapped delay line has become a general scheme of making high-*Q* filters in many different ways.

SAW filters

SAW (surface acoustic wave) filters are electromechanical devices commonly used in radio frequency applications. Electrical signals are converted to a mechanical wave in a device constructed of a piezoelectric crystal or ceramic; this wave is delayed as it propagates across the device, before being converted back to an electrical signal by further electrodes. The delayed outputs are recombined to produce a direct analog implementation of a finite impulse response filter. This hybrid filtering technique is also found in an analog sampled filter. SAW filters are limited to frequencies up to 3 GHz.

BAW filters

BAW (Bulk Acoustic Wave) filters are electromechanical devices. BAW filters can implement ladder or lattice filters. BAW filters typically operate at frequencies from around 2 to around 16 GHz, and may be smaller or thinner than equivalent SAW filters. Two main variants of BAW filters are making their way into devices, Thin film bulk acoustic resonator or FBAR and Solid Mounted Bulk Acoustic Resonators.

Garnet filters

Another method of filtering, at microwave frequencies from 800 MHz to about 5 GHz, is to use a synthetic single crystal yttrium iron garnet sphere made of a chemical combination of yttrium and iron (**YIGF**, or **yttrium iron garnet filter**). The garnet sits on a strip of metal driven by a transistor, and a small loop antenna touches the top of the sphere. An electromagnet changes the frequency that the garnet will pass. The advantage of this method is that the garnet can be tuned over a very wide frequency by varying the strength of the magnetic field.

Atomic filters

For even higher frequencies and greater precision, the vibrations of atoms must be used. Atomic clocks use caesium masers as ultra-high *Q* filters to stabilize their primary oscillators. Another method, used at high, fixed frequencies with very weak radio signals, is to use a ruby maser tapped delay line.

The transfer function

The transfer function $H(s)$ of a filter is the ratio of the output signal $Y(s)$ to that of the input signal $X(s)$ as a function of the complex frequency s:

$$H(s) = \frac{Y(s)}{X(s)}$$

with $s = \sigma + j\omega$.

The transfer function of all linear time-invariant filters, when constructed of discrete components, will be the ratio of two polynomials in s, i.e. a rational function of s. The order of the transfer function will be the highest power of s encountered in either the numerator or the denominator.

Classification by topology

Electronic filters can be classified by the technology used to implement them. Filters using passive filter and active filter technology can be further classified by the particular electronic filter topology used to implement them.

Any given filter transfer function may be implemented in any electronic filter topology.

Some common circuit topologies are:

- Cauer topology - Passive
- Sallen Key topology - Active

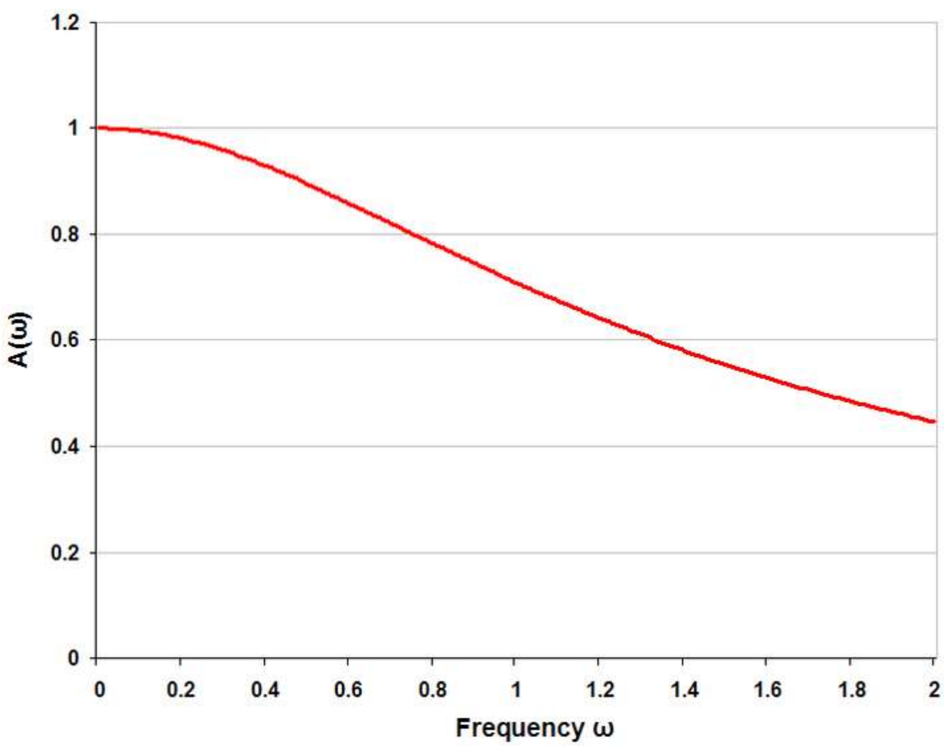
- Multiple Feedback topology - Active
- State Variable Topology - Active
- Biquadratic topology biquad filter - Active

Classification by design methodology

Historically, linear analog filter design has evolved through three major approaches. The oldest designs are simple circuits where the main design criterion was the Q factor of the circuit. This reflected the radio receiver application of filtering as Q was a measure of the frequency selectivity of a tuning circuit. From the 1920s filters began to be designed from the image point of view, mostly being driven by the requirements of telecommunications. After World War II the dominant methodology was network synthesis. The higher mathematics used originally required extensive tables of polynomial coefficient values to be published but modern computer resources have made that unnecessary.

Direct circuit analysis

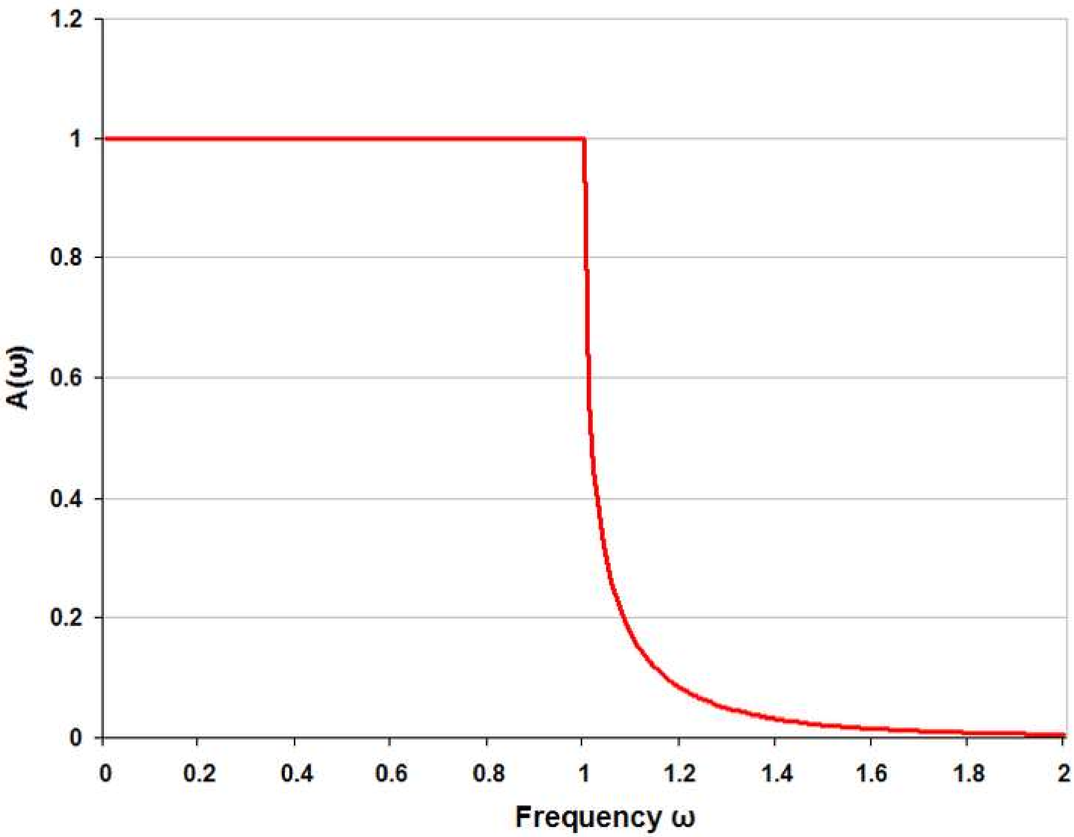
Low order filters can be designed by directly applying basic circuit laws such as Kirchoff's laws to obtain the transfer function. This kind of analysis is usually only carried out for simple filters of 1st or 2nd order.



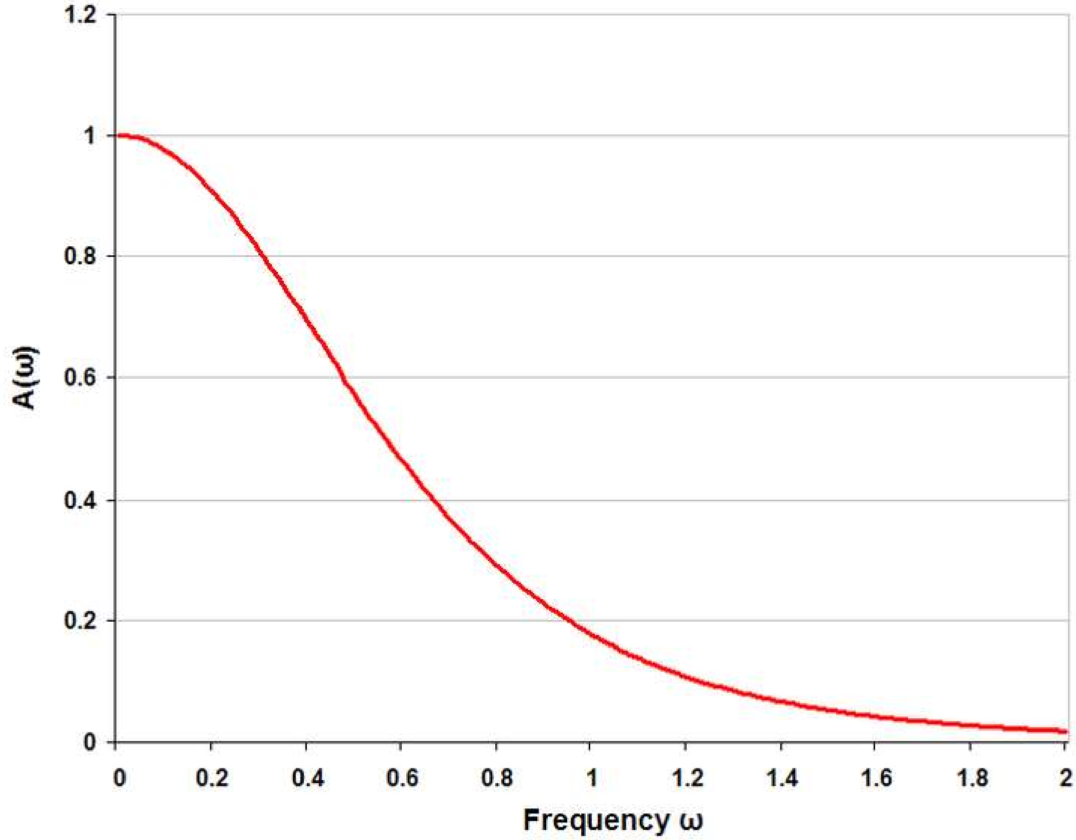
RL filter frequency response

Image impedance analysis

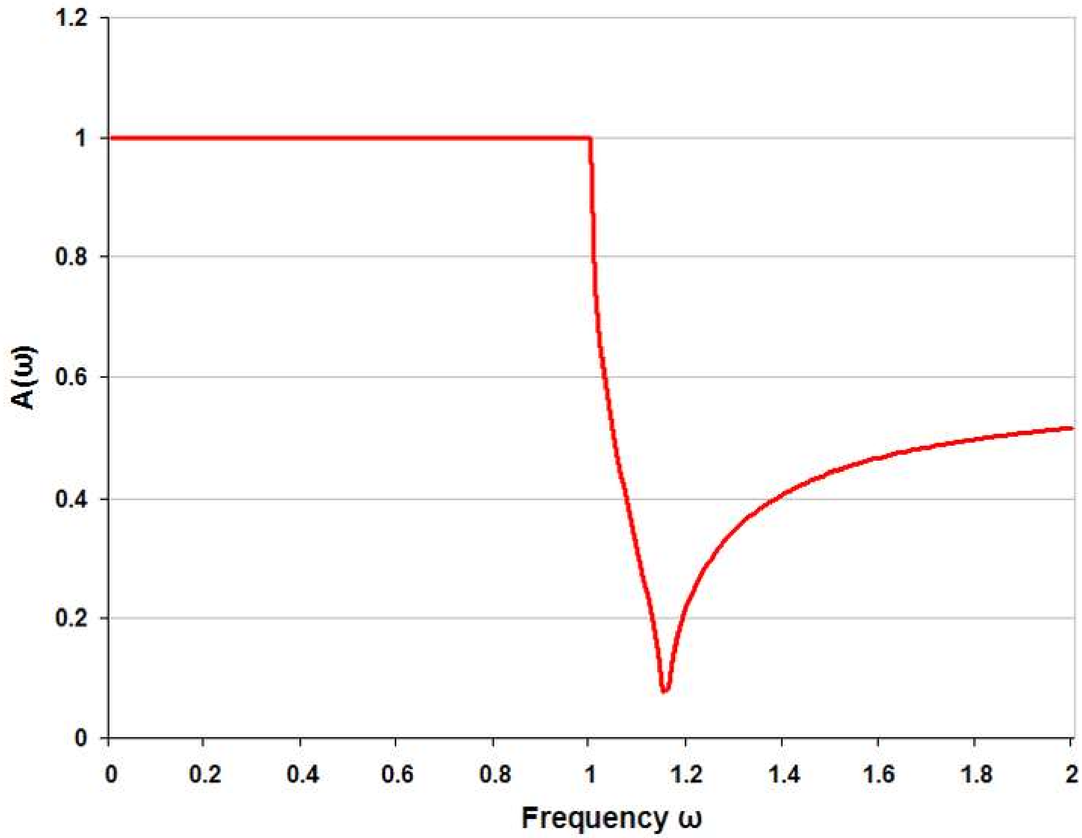
This approach analyses the filter sections from the point of view of the filter being in an infinite chain of identical sections. It has the advantages of simplicity of approach and the ability to easily extend to higher orders. It has the disadvantage that accuracy of predicted responses relies on filter terminations in the image impedance, which is usually not the case.



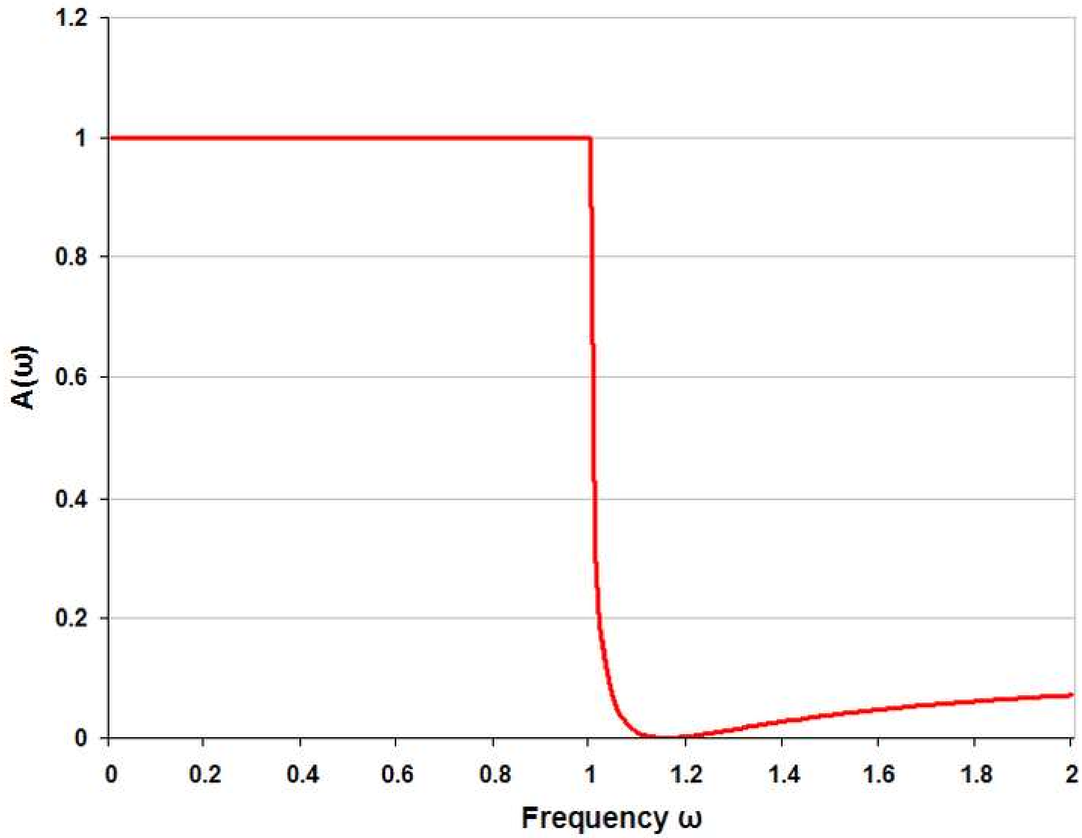
Constant k filter response with 5 elements



Zobel network (constant R) filter, 5 sections



m-derived filter response, m=0.5, 2 elements

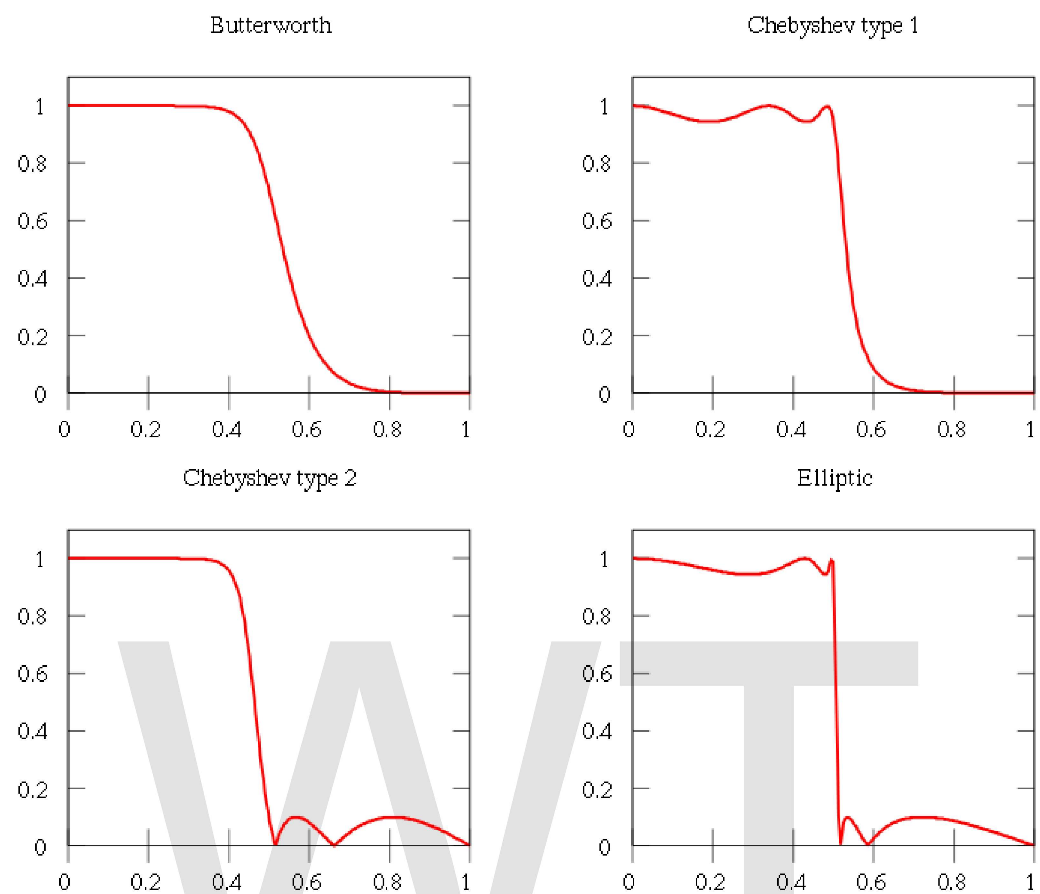


m-derived filter response, m=0.5, 5 elements

Network synthesis

The network synthesis approach starts with a required transfer function and then expresses that as a polynomial equation of the input impedance of the filter. The actual element values of the filter are obtained by continued-fraction or partial-fraction expansions of this polynomial. Unlike the image method, there is no need for impedance matching networks at the terminations as the effects of the terminating resistors are included in the analysis from the start.

Here is an image comparing Butterworth, Chebyshev, and elliptic filters. The filters in this illustration are all fifth-order low-pass filters. The particular implementation – analog or digital, passive or active – makes no difference; their output would be the same.



As is clear from the image, elliptic filters are sharper than all the others, but they show ripples on the whole bandwidth.

Chapter 9

Constant k Filter

Constant k filters, also **k-type filters**, are a type of electronic filter designed using the image method. They are the original and simplest filters produced by this methodology and consist of a ladder network of identical sections of passive components. Historically, they are the first filters that could approach the ideal filter frequency response to within any prescribed limit with the addition of a sufficient number of sections. However, they are rarely considered for a modern design, the principles behind them having been superseded by other methodologies which are more accurate in their prediction of filter response.

History

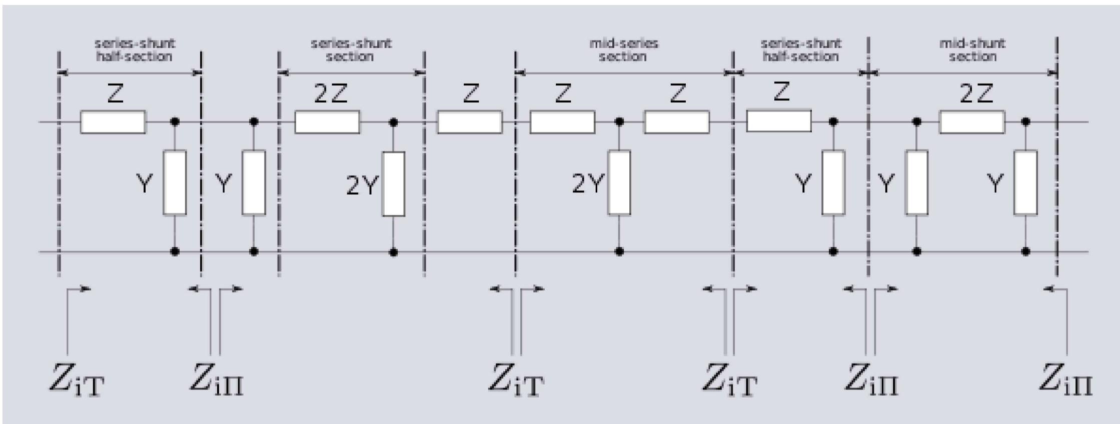
Constant k filters were invented by George Campbell. He published his work in 1922, but had clearly invented the filters some time before, as his colleague at AT&T Co, Otto Zobel, was already making improvements to the design at this time. Campbell's filters were far superior to the simpler single element circuits that had been used previously. Campbell called his filters electric wave filters, but this term later came to mean any filter that passes waves of some frequencies but not others. Many new forms of wave filter were subsequently invented; an early (and important) variation was the m-derived filter by Zobel who coined the term constant k for the Campbell filter in order to distinguish them.

The great advantage Campbell's filters had over the RL circuit and other simple filters of the time was that they could be designed for any desired degree of stop band rejection or steepness of transition between pass band and stop band. It was only necessary to add more filter sections until the desired response was obtained.

The filters were designed by Campbell for the purpose of separating multiplexed telephone channels on transmission lines, but their subsequent use has been much more widespread than that. The design techniques used by Campbell have largely been superseded. However, the ladder topology used by Campbell with the constant k is still in use today with implementations of modern filter designs such as the Tchebyscheff filter. Campbell gave constant k designs for low-pass, high-pass and band-pass filters. Band-stop and multiple band filters are also possible.

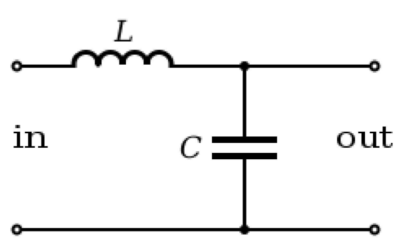
Terminology

Some of the impedance terms and section terms used here are pictured in the diagram below. Image theory defines quantities in terms of an infinite cascade of two-port sections, and in the case of the filters being discussed, an infinite ladder network of L-sections. Here "L" should not be confused with the inductance *L* – in electronic filter topology, "L" refers to the specific filter shape which resembles inverted letter "L".

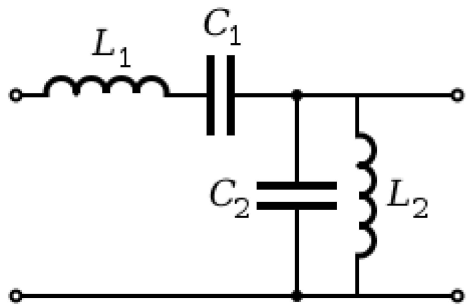


The sections of the hypothetical infinite filter are made of series elements having impedance 2*Z* and shunt elements with admittance 2*Y*. The factor of two is introduced for mathematical convenience, since it is usual to work in terms of half-sections where it disappears. The image impedance of the input and output port of a section will generally not be the same. However, for a *mid-series section* (that is, a section from halfway through a series element to halfway through the next series element) will have the same image impedance on both ports due to symmetry. This image impedance is designated $Z_{i\mathrm{T}}$ due to the "T" topology of a mid-series section. Likewise, the image impedance of a *mid-shunt section* is designated $Z_{i\Pi}$ due to the "Π" topology. Half of such a "T" or "Π" section is called a *half-section*, which is also an L-section but with half the element values of the full L-section. The image impedance of the half-section is dissimilar on the input and output ports: on the side presenting the series element it is equal to the mid-series $Z_{i\mathrm{T}}$, but on the side presenting the shunt element it is equal to the mid-shunt $Z_{i\Pi}$. There are thus two variant ways of using a half-section.

Derivation



Constant k low-pass filter half section. Here inductance *L* is equal Ck^2



Constant k band-pass filter half section.
$L_1 = C_2k^2$ and $L_2 = C_1k^2$

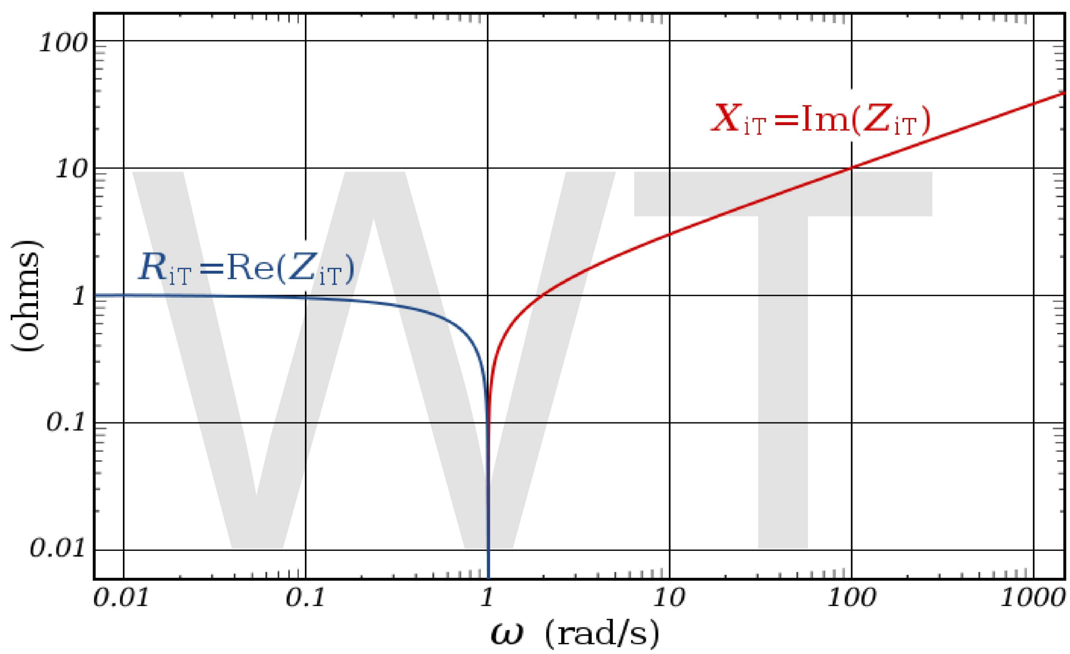


Image impedance Z_{iT} of a constant k prototype low-pass filter is plotted vs. frequency ω. The impedance is purely resistive (real) below ω_c, and purely reactive (imaginary) above ω_c.

The building block of constant k filters is the half-section "L" network, composed of a series impedance *Z*, and a shunt admittance *Y*. The "k" in "constant k" is the value given by,

$$k^2 = \frac{Z}{Y}$$

Thus, *k* will have units of impedance, that is, ohms. It is readily apparent that in order for *k* to be constant, *Y* must be the dual impedance of *Z*. A physical interpretation of k can be given by observing that *k* is the limiting value of Z_i as the size of the section (in terms of values of its components, such as inductances, capacitances, etc.) approaches zero, while keeping k at its initial value. Thus, k is the characteristic impedance, Z_0, of the transmission line that would be formed by these infinitesimally small sections. It is also the image impedance of the section at resonance, in the case of band-pass filters, or at $\omega = 0$ in the case of low-pass filters. For example, the pictured low-pass half-section has

$$k = \sqrt{\frac{i\omega L}{i\omega C}} = \sqrt{\frac{L}{C}}.$$

Elements L and C can be made arbitrarily small while retaining the same value of k. Z and Y however, are both approaching zero, and from the formulae (below) for image impedances,

$$\lim_{Z,Y \to 0} Z_\mathrm{i} = k.$$

Image impedance

The image impedances of the section are given by

$${Z_\mathrm{iT}}^2 = Z^2 + k^2$$

and

$$\frac{1}{{Z_\mathrm{i\Pi}}^2} = {Y_\mathrm{i\Pi}}^2 = Y^2 + \frac{1}{k^2}$$

Provided that the filter does not contain any resistive elements, the image impedance in the pass band of the filter is purely real and in the stop band it is purely imaginary. For example, for the pictured low-pass half-section,

$${Z_\mathrm{iT}}^2 = -(\omega L)^2 + \frac{L}{C}$$

The transition occurs at a cut-off frequency given by

$$\omega_c = \frac{1}{\sqrt{LC}}$$

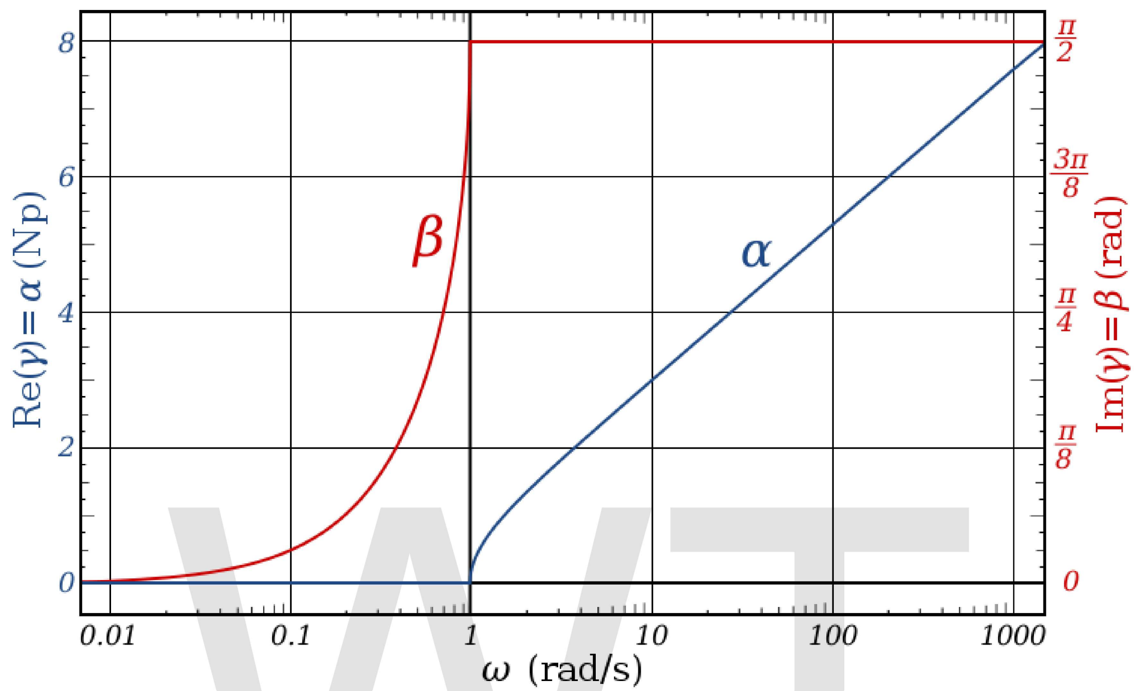
Below this frequency, the image impedance is real,

$$Z_\mathrm{iT} = L\sqrt{\omega_c^2 - \omega^2}$$

Above the cut-off frequency the image impedance is imaginary,

$$Z_{\mathrm{iT}} = iL\sqrt{\omega^2 - \omega_c^2}$$

Transmission parameters



The transfer function of a constant k prototype low-pass filter for a single half-section showing attenuation in nepers and phase change in radians.

The transmission parameters for a general constant k half-section are given by

$$\gamma = \sinh^{-1}\frac{Z}{k}$$

and for a chain of *n* half-sections

$$\gamma_n = n\gamma$$

For the low-pass L-shape section, below the cut-off frequency, the transmission parameters are given by

$$\gamma = \alpha + i\beta = 0 + i\sin^{-1}\frac{\omega}{\omega_c}$$

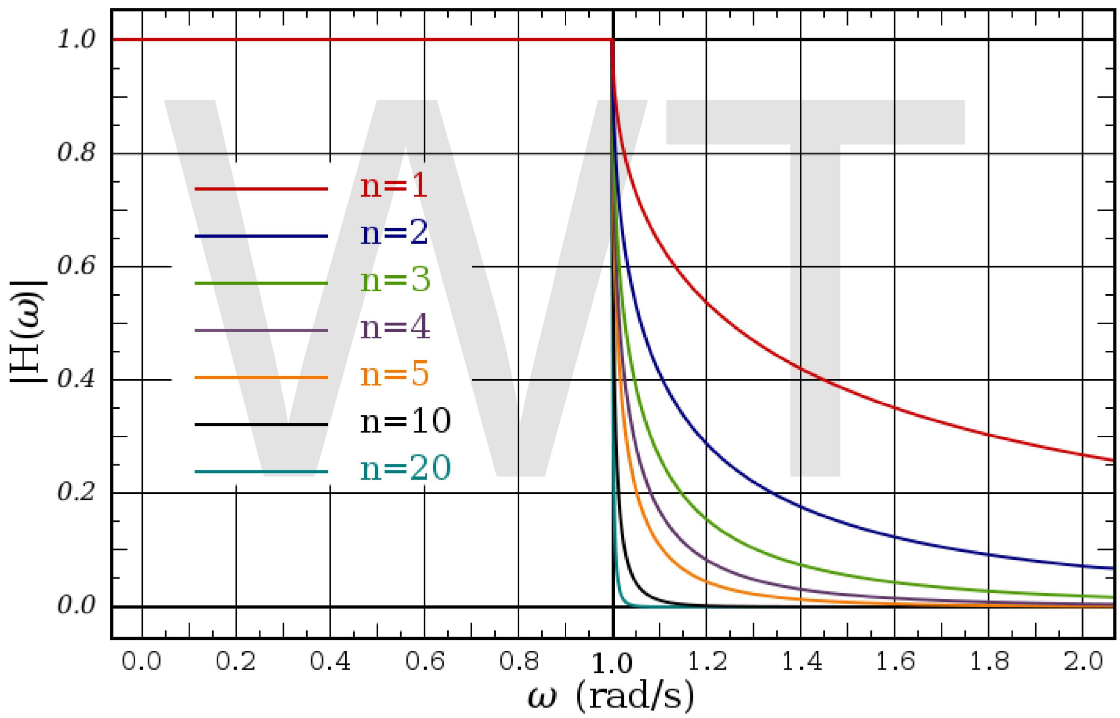
That is, the transmission is lossless in the pass-band with only the phase of the signal changing. Above the cut-off frequency, the transmission parameters are:

$$\gamma = \alpha + i\beta = \cosh^{-1}\frac{\omega}{\omega_c} + i\frac{\pi}{2}$$

Prototype transformations

The presented plots of image impedance, attenuation and phase change correspond to a low-pass prototype filter section. The prototype has a cut-off frequency of $\omega_c = 1$ rad/s and a nominal impedance $k = 1\ \Omega$. This is produced by a filter half-section with inductance $L = 1$ henry and capacitance $C = 1$ farad. This prototype can be impedance scaled and frequency scaled to the desired values. The low-pass prototype can also be transformed into high-pass, band-pass or band-stop types by application of suitable frequency transformations.

Cascading sections



Gain response, H(ω) for a chain of *n* low-pass constant-k filter half-sections

Several L-shape half-sections may be cascaded to form a composite filter. Like impedance must always face like in these combinations. There are therefore two circuits that can be formed with two identical L-shaped half-sections. Where a port of image impedance Z_{iT} faces another Z_{iT}, the section is called a Π section. Where $Z_{i\Pi}$ faces $Z_{i\Pi}$ the section so formed is a T section. Further additions of half-sections to either of these section forms a ladder network which may start and end with series or shunt elements.

It should be borne in mind that the characteristics of the filter predicted by the image method are only accurate if the section is terminated with its image impedance. This is usually not true of the sections at either end, which are usually terminated with a fixed resistance. The further the section is from the end of the filter, the more accurate the prediction will become, since the effects of the terminating impedances are masked by the intervening sections.

Chapter 10

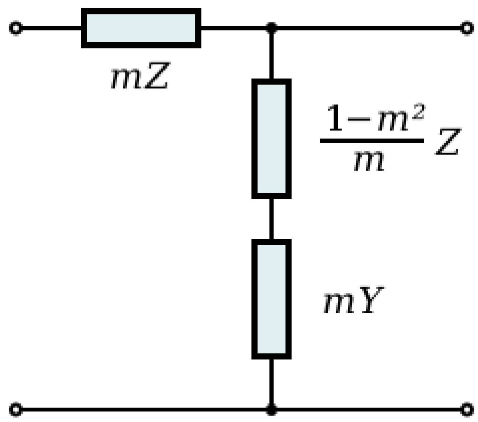
m-derived Filter

m-derived filters or **m-type filters** are a type of electronic filter designed using the image method. They were invented by Otto Zobel in the early 1920s. This filter type was originally intended for use with telephone multiplexing and was an improvement on the existing constant k type filter. The main problem being addressed was the need to achieve a better match of the filter into the terminating impedances. In general, all filters designed by the image method fail to give an exact match, but the m-type filter is a big improvement with suitable choice of the parameter m. The m-type filter section has a further advantage in that there is a rapid transition from the cut-off frequency of the pass band to a pole of attenuation just inside the stop band. Despite these advantages, there is a drawback with m-type filters; at frequencies past the pole of attenuation, the response starts to rise again, and m-types have poor stop band rejection. For this reason, filters designed using m-type sections are often designed as composite filters with a mixture of k-type and m-type sections and different values of m at different points to get the optimum performance from both types.

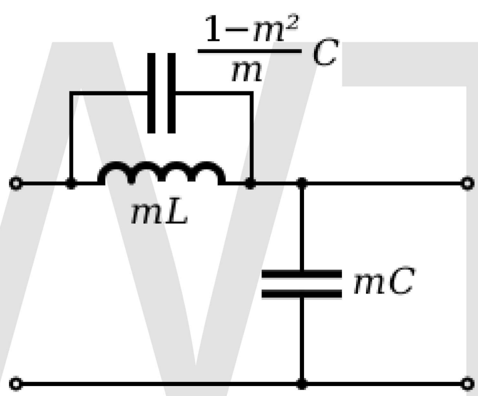
Background

Zobel patented an impedance matching network in 1920 which, in essence, used the topology of what are now called m-type filters, but Zobel did not name them as such or analyse them by the image method. This pre-dated George Campbell's publication of his constant k-type design in 1922 on which the m-type filter is based. Zobel published the image analysis theory of m-type filters in 1923. Once popular, M-type filters and image parameter designed filters in general are now rarely designed, having been superseded by more advanced network synthesis methods.

Derivation



m-derived series general filter half section



m-derived shunt low-pass filter half section

$$C = \frac{L}{R_0^2}$$

The building block of m-derived filters, as with all image impedance filters, is the "L" network, called a half-section and composed of a series impedance *Z*, and a shunt admittance *Y*. The m-derived filter is a derivative of the constant k filter. The starting point of the design is the values of *Z* and *Y* derived from the constant k prototype and are given by

$$k^2 = \frac{Z}{Y}$$

where *k* is the nominal impedance of the filter, or R_0. The designer now multiplies *Z* and *Y* by an arbitrary constant *m* ($0 < m < 1$). There are two different kinds of m-derived section; series and shunt. To obtain the m-derived series half section, the designer determines the impedance that must be added to 1/mY to make the image impedance Z_{iT} the same as the image impedance of the original constant k section. From the general formula for image impedance, the additional impedance required can be shown to be

$$\frac{1-m^2}{m}Z.$$

To obtain the m-derived shunt half section, an admittance is added to 1/mZ to make the image impedance $Z_{i\Pi}$ the same as the image impedance of the original half section. The additional admittance required can be shown to be

$$\frac{1-m^2}{m}Y.$$

The general arrangements of these circuits are shown in the diagrams to the right along with a specific example of a low pass section.

A consequence of this design is that the m-derived half section will match a k-type section on one side only. Also, an m-type section of one value of m will not match another m-type section of another value of m except on the sides which offer the Z_i of the k-type.

Operating frequency

For the low-pass half section shown, the cut-off frequency of the m-type is the same as the k-type and is given by

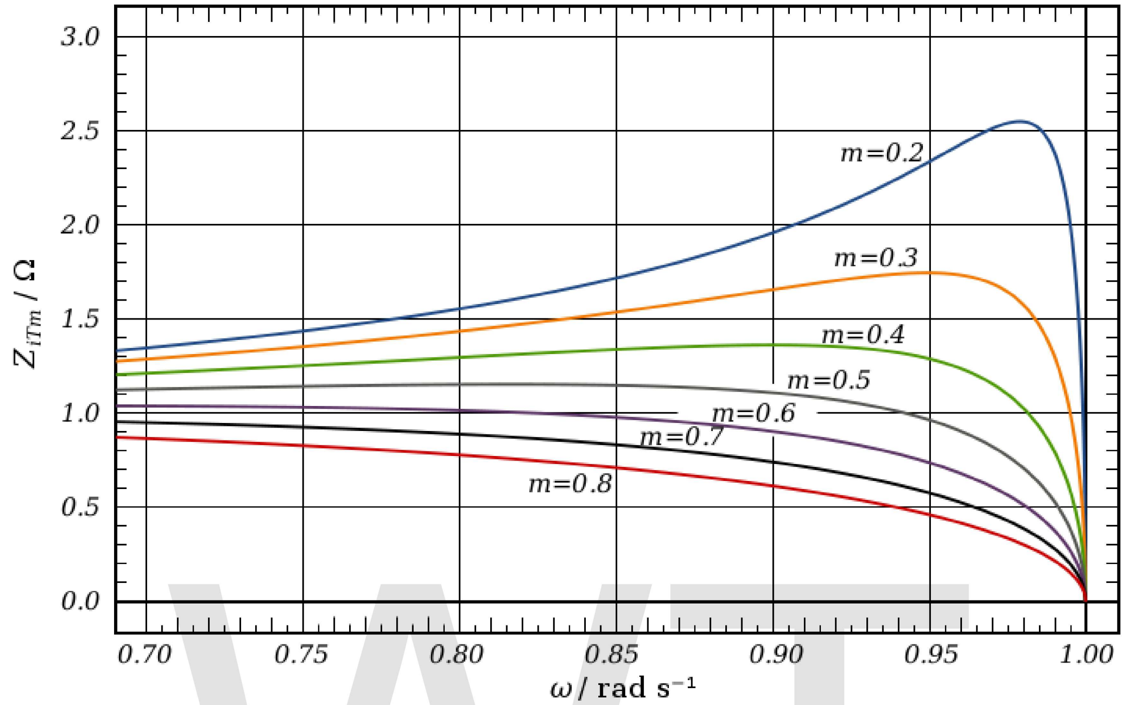
$$\omega_c = \frac{1}{\sqrt{LC}}.$$

The pole of attenuation occurs at;

$$\omega_\infty = \frac{\omega_c}{\sqrt{1-m^2}}.$$

From this it is clear that smaller values of m will produce ω_∞closer to the cut-off frequency $\omega_c$and hence will have a sharper cut-off. Despite this cut-off, it also brings the unwanted stop band response of the m-type closer to the cut-off frequency, making it more difficult for this to be filtered with subsequent sections. The value of m chosen is usually a compromise between these conflicting requirements. There is also a practical limit to how small m can be made due to the inherent resistance of the inductors. This has the effect of causing the pole of attenuation to be less deep (that is, it is no longer a genuinely infinite pole) and the slope of cut-off to be less steep. This effect becomes more marked as $\omega_\infty$is brought closer to ω_c, and there ceases to be any improvement in response with an m of about 0.2 or less.

Image impedance



m-derived prototype shunt low-pass filter Z_{iTm} image impedance for various values of *m*. Values below cut-off frequency only shown for clarity.

The following expressions for image impedances are all referenced to the low-pass prototype section. They are scaled to the nominal impedance $R_0 = 1$, and the frequencies in those expressions are all scaled to the cut-off frequency $\omega_c = 1$.

Series sections

The image impedances of the series section are given by

$$Z_{iT} = \sqrt{1-\omega^2}$$

and is the same as that of the constant k section

$$Z_{i\Pi m} = \frac{1-(\omega/\omega_\infty)^2}{\sqrt{1-\omega^2}}.$$

Shunt sections

The image impedances of the shunt section are given by

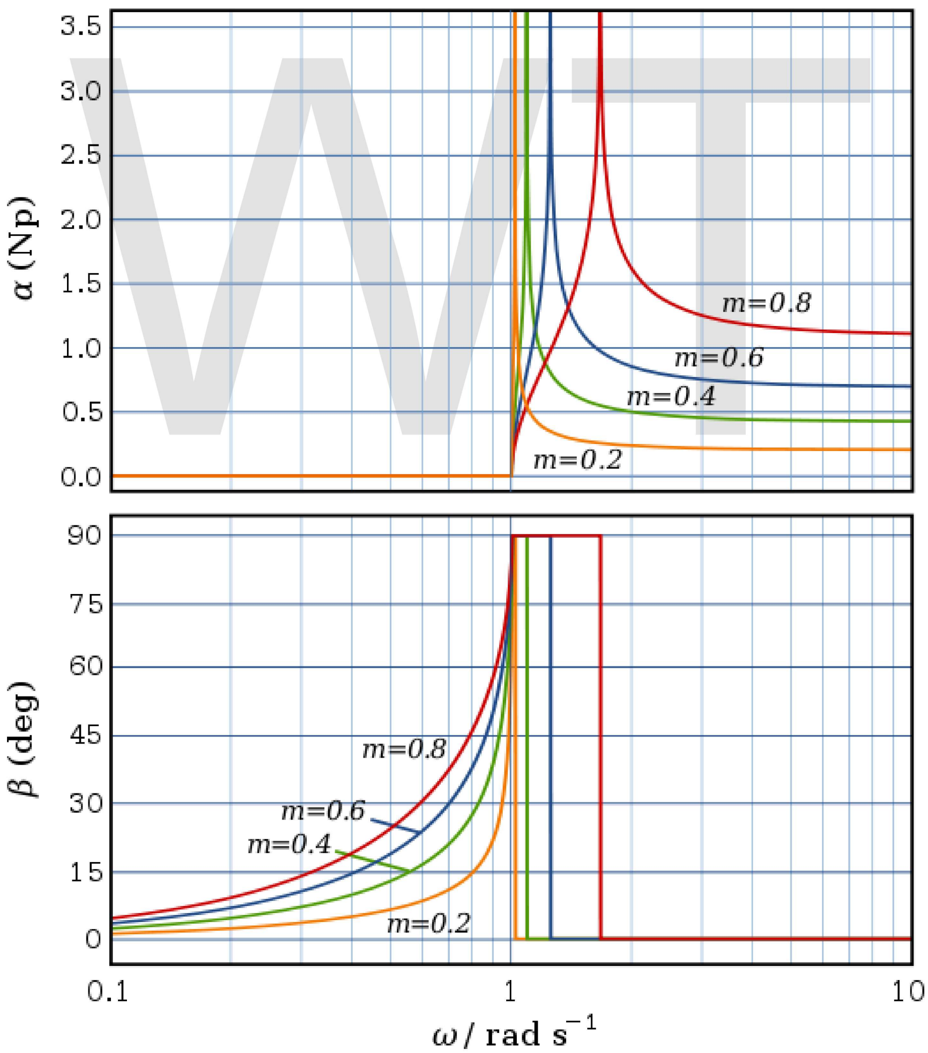
$$Z_{i\Pi} = \frac{1}{\sqrt{1-\omega^2}}$$

and is the same as that of the constant k section

$$Z_{iTm} = \frac{\sqrt{1-\omega^2}}{1-(\omega/\omega_\infty)^2}$$

As with the k-type section, the image impedance of the *m*-type low-pass section is purely real below the cut-off frequency and purely imaginary above it. From the chart it can be seen that in the passband the closest impedance match to a constant pure resistance termination occurs at approximately *m* = 0.6.

Transmission parameters



m-Derived low-pass filter transfer function for a single half-section

For an m-derived section in general the transmission parameters for a half-section are given by

$$\gamma = \sinh^{-1} \frac{mZ}{\sqrt{k^2 + (1-m^2)Z^2}}$$

and for n half-sections

$$\gamma_n = n\gamma$$

For the particular example of the low-pass L section, the transmission parameters solve differently in three frequency bands.

For $0 < \omega < \omega_c$ the transmission is lossless:

$$\gamma = \alpha + i\beta = 0 + i\frac{1}{2}\cos^{-1}\left(1 - \frac{2m^2}{\left(\frac{\omega_c}{\omega}\right)^2 - \left(\frac{\omega_c}{\omega_\infty}\right)^2}\right)$$

For $\omega_c < \omega < \omega_\infty$ the transmission parameters are

$$\gamma = \alpha + i\beta = \frac{1}{2}\cosh^{-1}\left(\frac{2m^2}{\left(\frac{\omega_c}{\omega}\right)^2 - \left(\frac{\omega_c}{\omega_\infty}\right)^2} - 1\right) + i\frac{\pi}{2}$$

For $\omega_\infty < \omega < \infty$ the transmission parameters are

$$\gamma = \alpha + i\beta = \frac{1}{2}\cosh^{-1}\left(1 - \frac{2m^2}{\left(\frac{\omega_c}{\omega}\right)^2 - \left(\frac{\omega_c}{\omega_\infty}\right)^2}\right) + i0$$

Prototype transformations

The plots shown of image impedance, attenuation and phase change are the plots of a low-pass prototype filter section. The prototype has a cut-off frequency of $\omega_c = 1$ rad/s and a nominal impedance $R_0 = 1\ \Omega$. This is produced by a filter half-section where L = 1 henry and C = 1 farad. This prototype can be impedance scaled and frequency scaled to the desired values. The low-pass prototype can also be transformed into high-pass, band-pass or band-stop types by application of suitable frequency transformations.

Cascading sections

Several L half-sections may be cascaded to form a composite filter. Like impedance must always face like in these combinations. There are therefore two circuits that can be formed with two identical L half-sections. Where Z_{iT} faces Z_{iT}, the section is called a Π section. Where $Z_{i\Pi}$ faces $Z_{i\Pi}$ the section formed is a T section. Further additions of half-sections to either of these forms a ladder network which may start and end with series or shunt elements.

It should be born in mind that the characteristics of the filter predicted by the image method are only accurate if the section is terminated with its image impedance. This is usually not true of the sections at either end which are usually terminated with a fixed resistance. The further the section is from the end of the filter, the more accurate the prediction will become since the effects of the terminating impedances are masked by the intervening sections. It is usual to provide half half-sections at the ends of the filter with m = 0.6 as this value gives the flattest Z_i in the passband and hence the best match in to a resistive termination.

Chapter 11

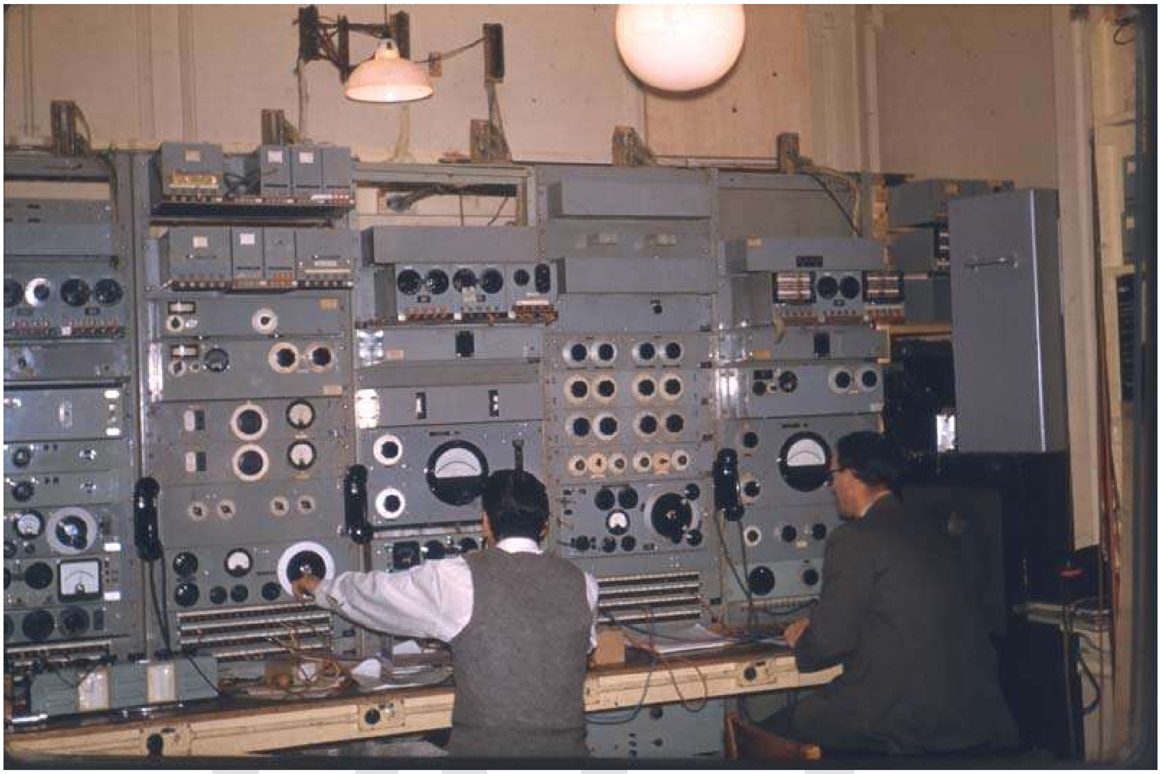
Zobel Network

Zobel networks are a type of filter section based on the image impedance design principle. They are named after Otto Zobel of Bell Labs who published a much referenced paper on image filters in 1923. The distinguishing feature of Zobel networks is that the input impedance is fixed in the design independently of the transfer function. This characteristic is achieved at the expense of a much higher component count compared to other types of filter sections. The impedance would normally be specified to be constant and purely resistive. For this reason they are also known as constant resistance networks. However, any impedance achievable with discrete components is possible.

Zobel networks were formerly widely used in telecommunications to flatten and widen the frequency response of copper land lines, producing a higher quality line from one originally intended for ordinary telephone use. However, as analogue technology has given way to digital they are now little used.

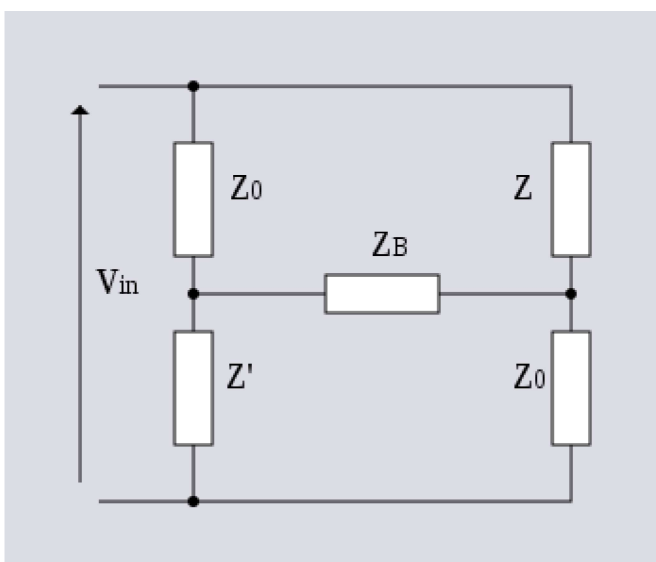
When used to cancel out the reactive portion of loudspeaker impedance, the design is sometimes called a Boucherot cell. In this case, only half the network is implemented as fixed components, the other half being the real and imaginary components of the loudspeaker impedance. This network is more akin to the power factor correction circuits used in electrical power distribution, hence the association with Boucherot's name.

A common circuit form of Zobel networks is in the form of a bridged T. This term is often used to mean a Zobel network, sometimes incorrectly when the circuit implementation is, in fact, something other than a bridged T.



BBC engineers equalising audio landlines circa 1959. The boxes with two large black dials towards the top of the equipment racks are adjustable Zobel equalisers. They are used both for temporary outside broadcast lines and for checking the engineer's calculations prior to building permanent units

Derivation



The basis of a Zobel network is a balanced bridge circuit as shown in the circuit to the right. The condition for balance is that;

$$\frac{Z}{Z_0} = \frac{Z_0}{Z'}.$$

If this is expressed in terms of a normalised $Z_0 = 1$ as is conventionally done in filter tables, then the balance condition is simply;

$$Z = \frac{1}{Z'}.$$

In other words, Z' is simply the inverse, or dual impedance of Z.

The bridging impedance Z_B is across the balance points and hence has no potential across it. Consequently, it will draw no current and its value makes no difference to the function of the circuit. However, its value is often chosen to be Z_0 for reasons which will become clear in the discussion of bridged T circuits.

Input impedance

The input impedance is given by

$$\frac{1}{Z_{\text{in}}} = \frac{1}{Z_0 + Z'} + \frac{1}{Z + Z_0}.$$

Substituting the balance condition,

$$Z' = \frac{Z_0^2}{Z},$$

yields

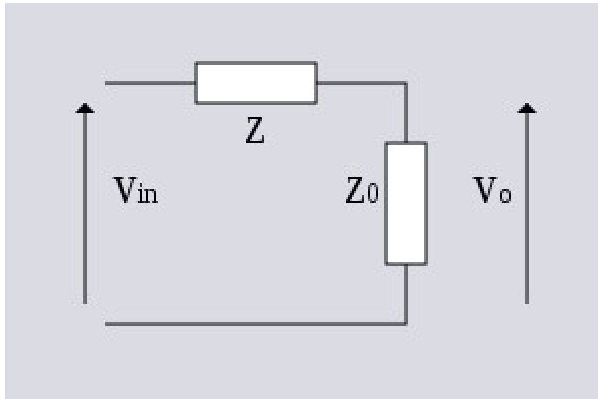
$$Z_{\text{in}} = Z_0$$

The input impedance can be designed to be purely resistive by setting

$$Z_0 = R_0.$$

The input impedance will then be real and independent of ω in band and out of band no matter what complexity of filter section is chosen.

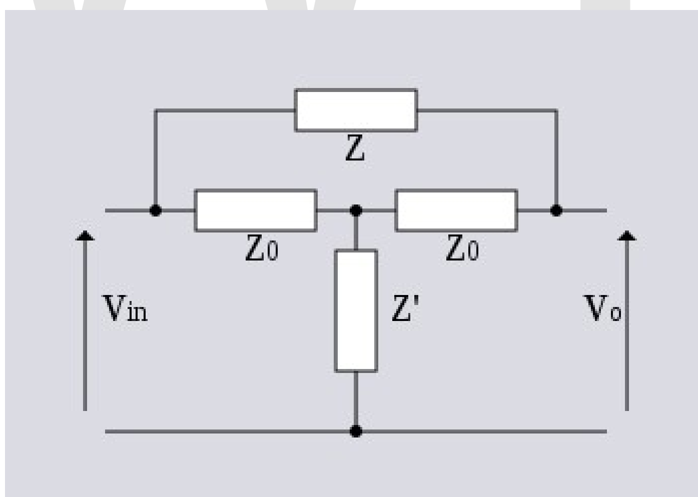
Transfer function



If the Z_0 in the bottom right of the bridge is taken to be the output load then a transfer function of V_{in}/V_o can be calculated for the section. Only the rhs branch needs to be considered in this calculation. The reason for this can be seen by considering that there is no current flow through R_B. None of the current flowing through the lhs branch is going to flow into the load. The lhs branch therefore, cannot possibly affect the output. It certainly affects the input impedance (and hence the input terminal voltage) but not the transfer function. The transfer function can now easily be seen to be;

$$A(\omega) = \frac{Z_0}{Z + Z_0}.$$

Bridged T implementation



The load impedance is actually the impedance of the following stage or of a transmission line and can sensibly be omitted from the circuit diagram. If we also set;

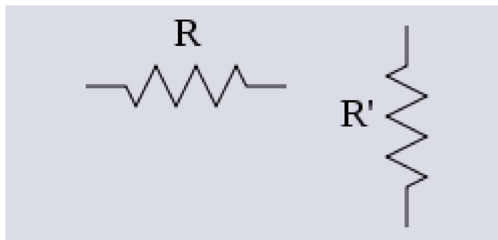
$$Z_B = Z_0$$

then the circuit to the right results. This is referred to as a bridged T circuit because the impedance Z is seen to "bridge" across the T section. The purpose of setting $Z_B = Z_0$ is to make the filter section symmetrical. This has the advantage that it will then present the same impedance, Z_0, at both the input and the output port.

Types of section

A Zobel filter section can be implemented for low-pass, high-pass, band-pass or band-stop. It is also possible to implement a flat frequency response attenuator. This last is of some importance for the practical filter sections described later.

Attenuator



Z and *Z'* for a Zobel attenuator

For an attenuator section, Z is simply

$$Z = R$$

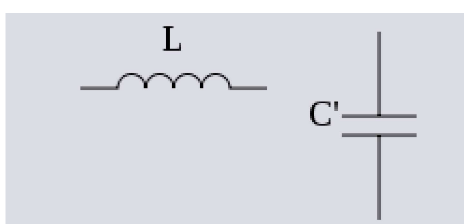
and,

$$Z' = R' = \frac{R_0^2}{R}$$

The attenuation of the section is given by;

$$L = 20\log\left(\frac{R}{R_0} + 1\right)\ \mathrm{dB}.$$

Low pass



Z and *Z'* for a Zobel low-pass filter section

For a low-pass filter section, *Z* is an inductor and *Z'* is a capacitor;

$$Z = i\omega L,$$

and

$$Z' = \frac{1}{i\omega C'}$$

where

$$C' = \frac{L}{R_0^2}.$$

The transfer function of the section is given by

$$A(\omega) = \frac{R_0}{i\omega L + R_0}.$$

The 3dB point occurs when $\omega L = R_0$ *so the 3dB cut-off frequency is given by*

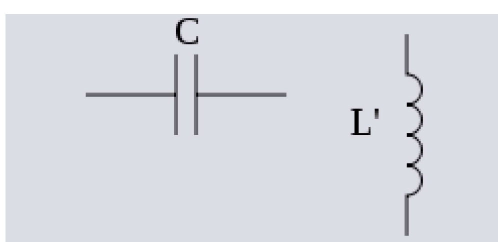
$$\omega_c = \frac{R_0}{L},$$

where ω is in the stop band well above ω_c,

$$A(\omega) \approx \frac{R_0}{i\omega L},$$

it can be seen from this that $A(\omega)$ is falling away in the stop band at the classic 6dB/8ve (or 20dB/decade).

High pass



Z and *Z'* for a Zobel high-pass filter section

For a high-pass filter section, *Z* is a capacitor and *Z'* is an inductor:

$$Z = \frac{1}{i\omega C},$$

and

$$Z' = i\omega L'$$

where

$$L' = CR_0^2.$$

The transfer function of the section is given by

$$A(\omega) = \frac{i\omega CR_0}{1 + i\omega CR_0}.$$

The 3dB point occurs when $\omega C = 1/R_0$ so the 3dB cut-off frequency is given by

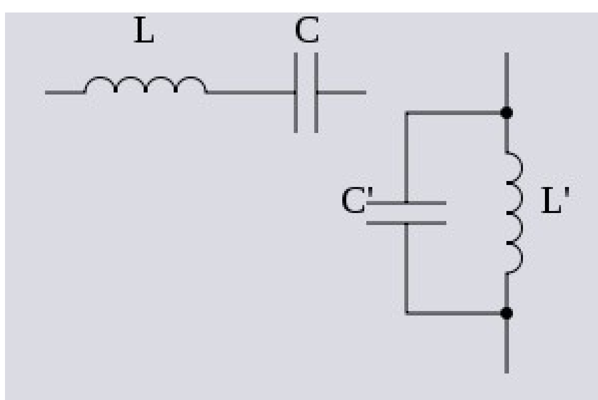
$$\omega_c = \frac{1}{CR_0}.$$

In the stop band,

$$A(\omega) \approx i\omega CR_0$$

falling at 6dB/8ve with decreasing frequency.

Band pass



Z and *Z'* for a Zobel band-pass filter section

For a band-pass filter section, *Z* is a series resonant circuit and *Z'* is a shunt resonant circuit;

$$Z = i\omega L + \frac{1}{i\omega C},$$

and

$$Y' = \frac{1}{Z'} = i\omega C' + \frac{1}{i\omega L'}.$$

The transfer function of the section is given by

$$A(\omega) = \frac{i\omega C R_0}{1 + i\omega C R_0 - \omega^2 LC}.$$

The 3dB point occurs when $|1 - \omega^2 LC| = \omega C R_0$ so the 3dB cut-off frequencies are given by

$$\omega_c = \frac{\pm R_0 C + \sqrt{R_0^2 C^2 + 4LC}}{2LC},$$

from which the centre frequency, ω_m, and bandwidth, $\Delta\omega$, can be determined:

$$\Delta\omega = \frac{R_0}{L}$$
$$\omega_m = \sqrt{\frac{R_0^2}{4L^2} + \frac{1}{LC}}.$$

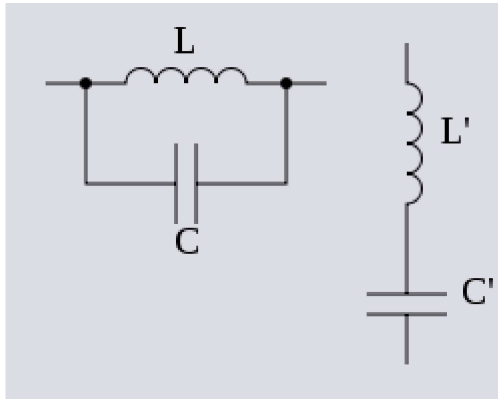
Note that this is different from the resonant frequency

$$\omega_0 = \sqrt{\frac{1}{LC}};$$

the relationship between them being given by

$$\omega_m^2 = \left(\frac{\Delta\omega}{2}\right)^2 + \omega_0^2.$$

Band stop



Z and *Z'* for a Zobel band-stop filter section

For a band-stop filter section, *Z* is a shunt resonant circuit and *Z'* is a series resonant circuit:

$$Y = \frac{1}{Z} = i\omega C + \frac{1}{i\omega L},$$

and

$$Z' = i\omega L' + \frac{1}{i\omega C'}.$$

The transfer function and bandwidth can be found by analogy with the band-pass section.

$$\Delta\omega = \frac{1}{CR_0}.$$

And,

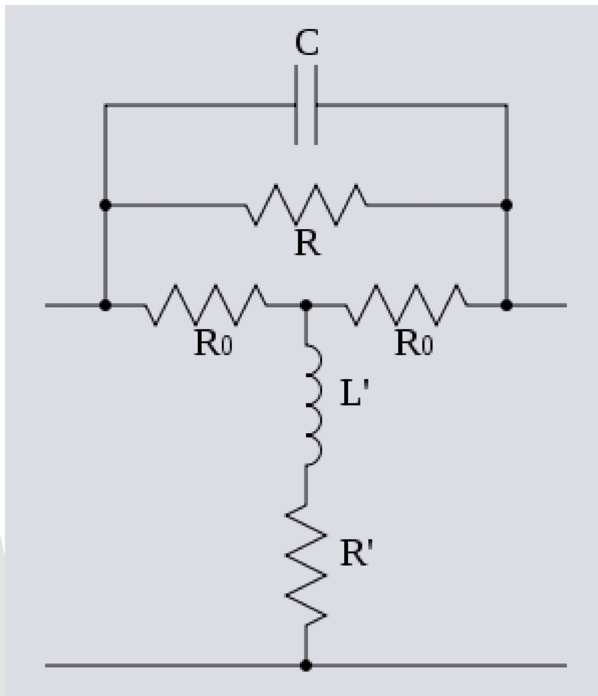
$$\omega_m = \sqrt{\left(\frac{1}{2R_0C}\right)^2 + \frac{1}{LC}}.$$

Practical sections

Zobel networks are rarely used for traditional frequency filtering. Other filter types are significantly more efficient for this purpose. Where Zobels come into their own is in frequency equalisation applications, particularly on transmission lines. The difficulty with transmission lines is that the impedance of the line varies in a complex way across the band and is tedious to measure. For most filter types, this variation in impedance will cause a significant difference in response to the theoretical, and is mathematically difficult to compensate for, even assuming that the impedance is known precisely. If

Zobel networks are used however, it is only necessary to measure the line response into a fixed resistive load and then design an equaliser to compensate it. It is entirely unnecessary to know anything at all about the line impedance as the Zobel network will present exactly the same impedance to line as the measuring instruments. Its response will therefore be precisely as theoretically predicted. This is a tremendous advantage where high quality lines with flat frequency responses are desired.

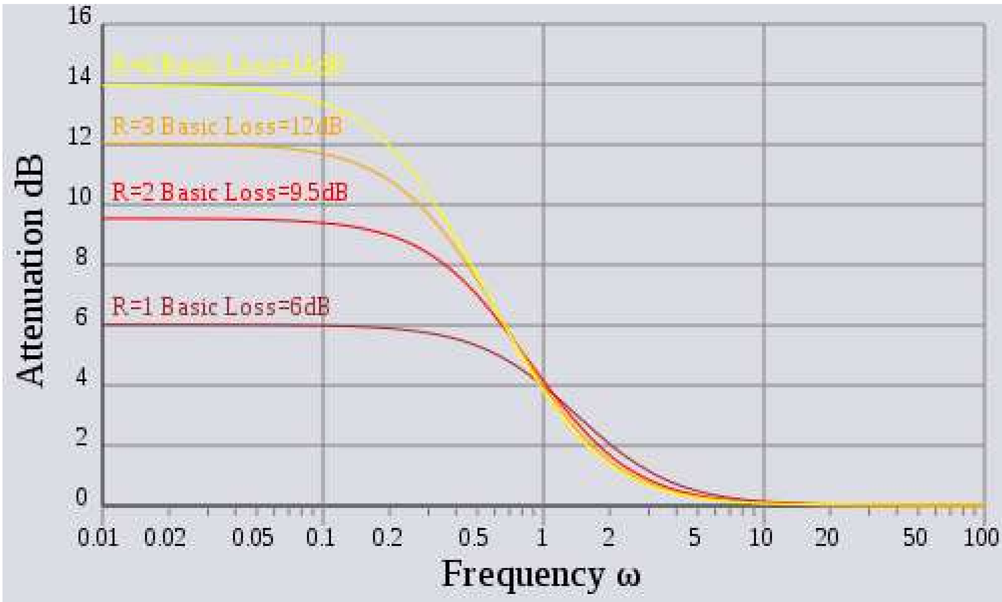
Basic loss



A practical high pass section incorporating basic loss used to correct high end roll-off

For audio lines, it is invariably necessary to combine L/C filter components with resistive attenuator components in the same filter section. The reason for this is that the usual design strategy is to require the section to attenuate all frequencies down to the level of the frequency in the passband with the lowest level. Without the resistor components, the filter, at least in theory, would increase attenuation without limit. The attenuation in the stop band of the filter (that is, the limiting maximum attenuation) is referred to as the "basic loss" of the section. In other words, the flat part of the band is attenuated by the basic loss down to the level of the falling part of the band which it is desired to equalise. The following discussion of practical sections relates in particular to audio transmission lines.

6dB/octave roll-off



High-pass Zobel network response for various basic losses. Normalised to $R_0 = 1$ and $\omega_c = 1$

The most significant effect that needs to be compensated for is that at some cut-off frequency the line response starts to roll-off like a simple low-pass filter. The effective bandwidth of the line can be increased with a section that is a high-pass filter matching this roll-off, combined with an attenuator. In the flat part of the pass-band only the attenuator part of the filter section is significant. This is set at an attenuation equal to the level of the highest frequency of interest. All frequencies up to this point will then be equalised flat to an attenuated level. Above this point, the output of the filter will again start to roll-off.

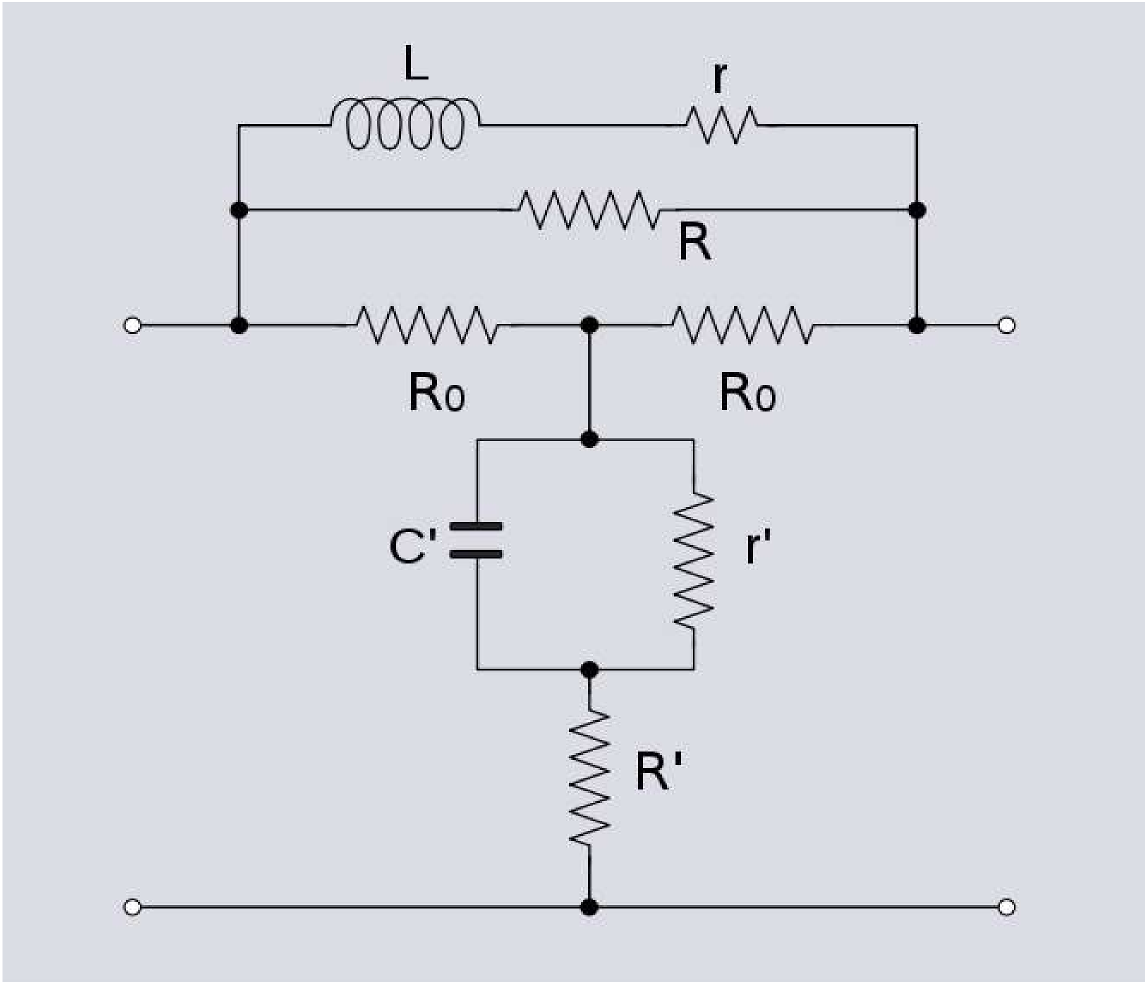
Mismatched lines

Quite commonly in telecomms networks, a circuit is made up of two sections of line which do not have the same characteristic impedance. For instance 150Ω and 300Ω. One effect of this is that the roll-off can start at 6dB/octave at an initial cut-off frequency f_{c1}, but then at f_{c2} can become suddenly steeper. This situation then requires (at least) two high-pass sections to compensate each operating at a different f_c.

Bumps and dips

Bumps and dips in the passband can be compensated for with band-stop and band-pass sections respectively. Again, an attenuator element is also required, but usually rather smaller than that required for the roll-off. These anomalies in the pass-band can be caused by mismatched line segments as described above. Dips can also be caused by ground temperature variations.

Transformer roll-off



Low frequency equaliser section with compensation for inductor resistance. The resistance **r** represents the stray resistance of the non-ideal inductor. The resistance **r'** is a real resistor calculated to compensate for **r**.

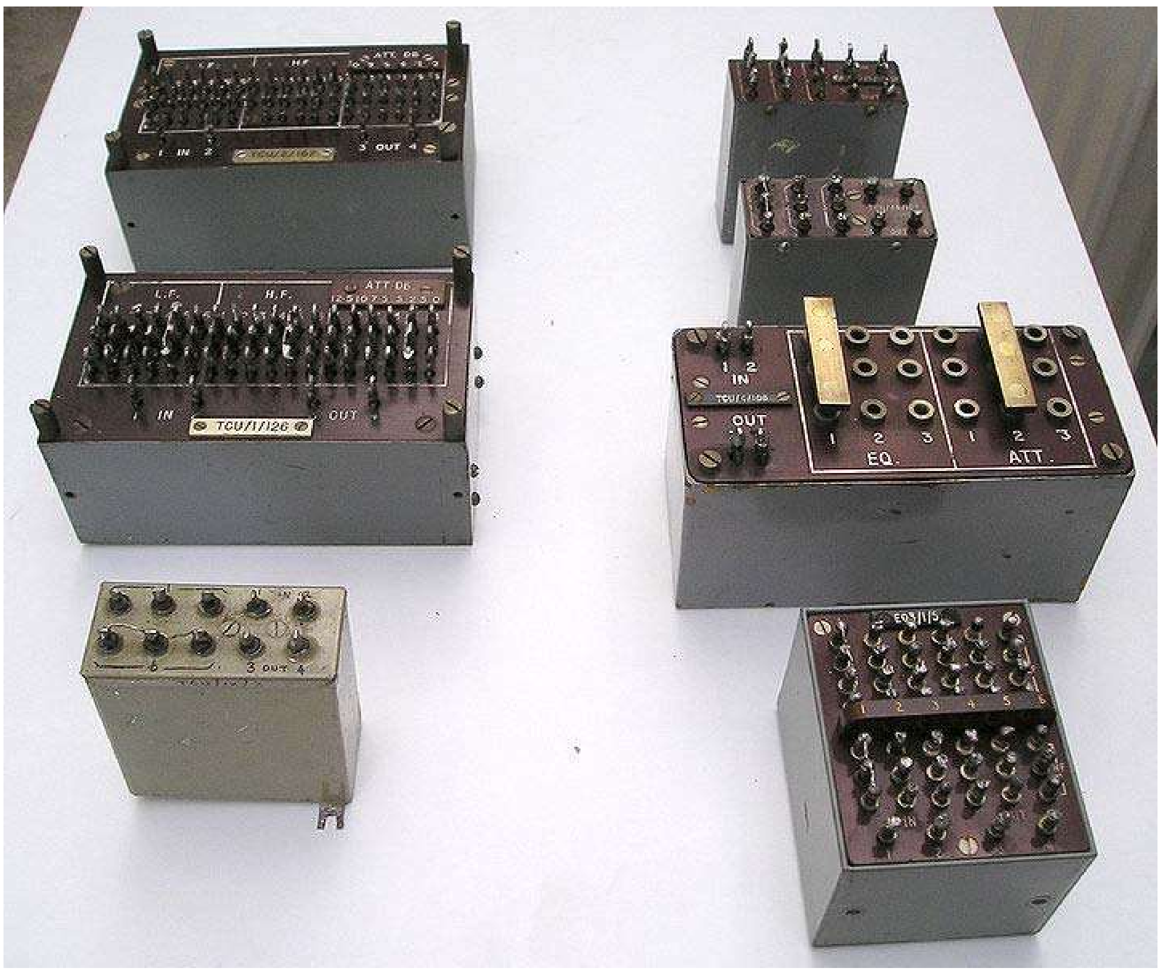
Occasionally, a low-pass section is included to compensate for excessive line transformer roll-off at the low frequency end. However, this effect is usually very small compared to the other effects noted above.

Low frequency sections will usually have inductors of high values. Such inductors have many turns and consequently tend to have significant resistance. In order to keep the section constant resistance at the input, the dual branch of the bridge T must contain a dual of the stray resistance, that is, a resistor in parallel with the capacitor. Even with the compensation, the stray resistance still has the effect of inserting attenuation at low frequencies. This in turn has the effect of slightly reducing the amount of LF lift the section would otherwise have produced. The basic loss of the section can be increased by the same amount as the stray resistance is inserting and this will return the LF lift achieved to that designed for.

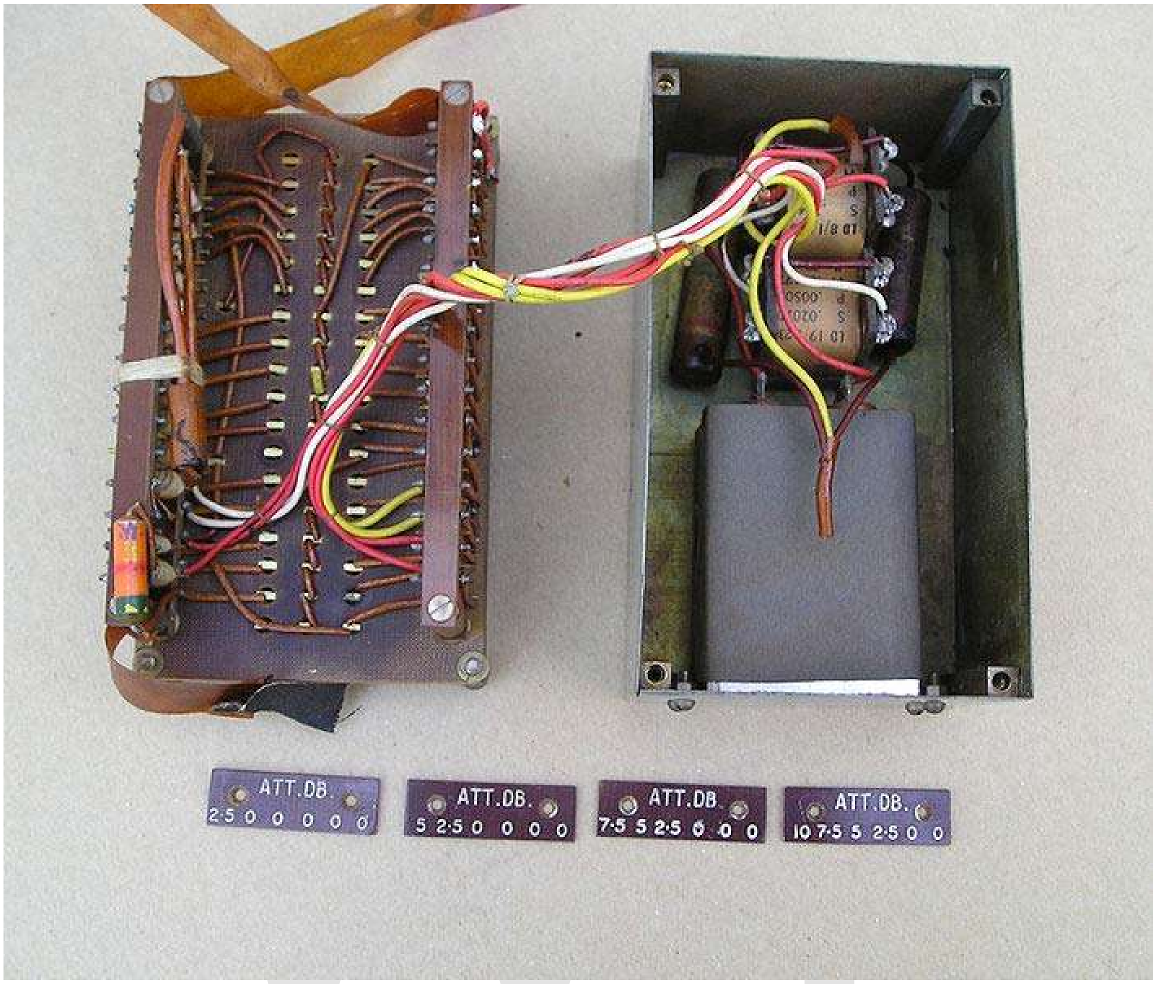
Compensation of inductor resistance is not such an issue at high frequencies were the inductors will tend to be smaller. In any case, for a high-pass section the inductor is in series with the basic loss resistor and the stray resistance can merely be subtracted from that resistor. On the other hand, the compensation technique may be required for resonant sections, especially a high Q resonator being used to lift a very narrow band. For these sections the value of inductors can also be large.

Temperature compensation

An adjustable attenuation high-pass section can be used to compensate for changes in ground temperature. Ground temperature is very slow varying in comparison to surface temperature. Adjustments are usually only required 2-4 times per year for audio applications.

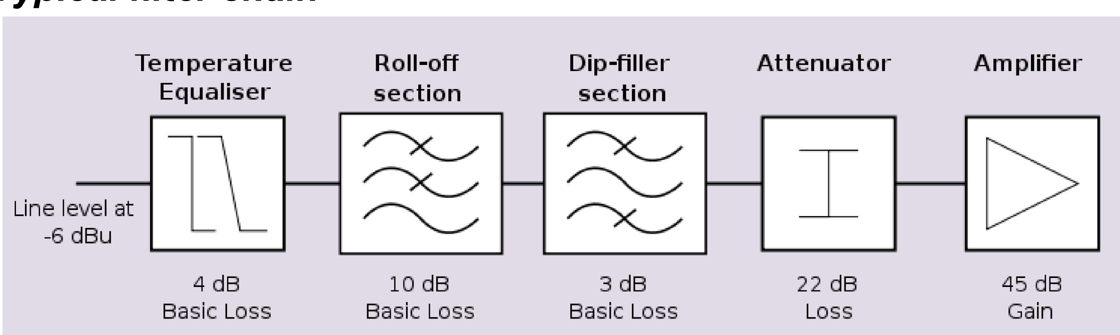


A collection of various designs of temperature compensation equaliser. Some can be adjusted with plugable links, others require soldering. Adjustment is not very frequent.



The internal components of a temperature equaliser. The inductor and capacitor on the right set the frequency at which the equaliser starts to operate, the banks of selectable resistors on the left set the basic loss and hence amount of equalisation.

Typical filter chain



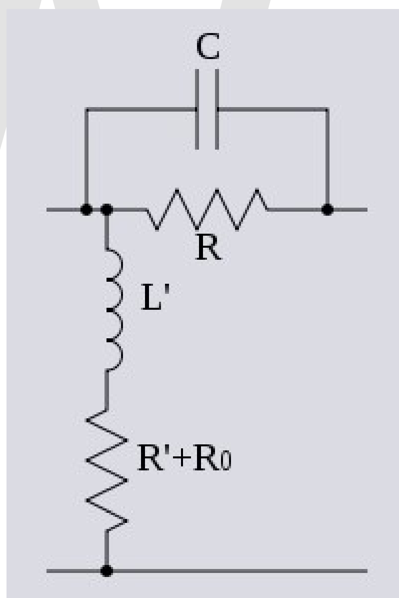
An example of a typical chain of Zobel networks being used for line equalisation

A typical complete filter will consist of a number of Zobel sections for roll-off, frequency dips and temperature followed by a flat attenuator section to bring the level down to a standard attenuation. This is followed by a fixed gain amplifier to bring the signal back up to a usable level, typically 0dBu. The gain of the amplifier is usually no more than 45dB maximum. Any more and the amplification of line noise will tend to cancel out the quality benefits of improved bandwidth. This limit on amplification essentially limits how much the bandwidth can be increased by these techniques. It should also be noted that no one part of the incoming signal band will be amplified by the full 45dB. The 45dB is made up of the line loss in the flat part of its spectrum plus the basic loss of each section. In general, each section will be minimum loss at a different frequency band, hence the amplification in that band will be limited to the basic loss of just that one filter section, assuming insignificant overlap. A typical choice for R_0 is 600 Ω. A good quality transformer (usually essential, but not shown on the diagram), known as a repeating coil, is at the beginning of the chain where the line terminates.

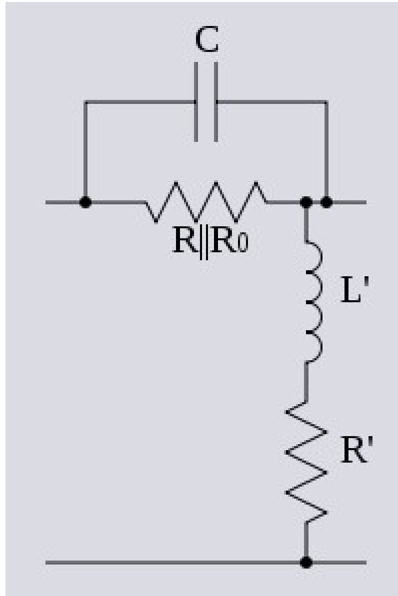
Other section implementations

Besides the Bridged T, there are a number of other possible section forms that can be used.

L-sections



Open circuit derived Zobel L-section for a high-pass section with basic loss

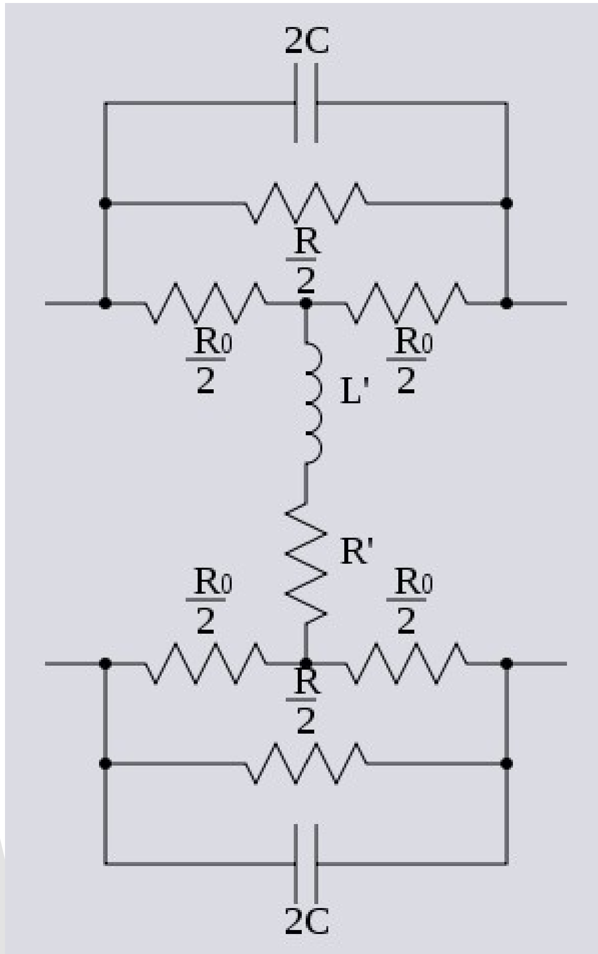


Short circuit derived Zobel L-section for a high-pass section with basic loss

As mentioned above, Z_B can be set to any desired impedance without affecting the input impedance. In particular, setting it as either an open circuit or a short circuit results in a simplified section circuit, called L–sections. These are shown above for the case of a high pass section with basic loss.

The input port still presents an impedance of R_0 (provided that the output is terminated in R_0) but the output port no longer presents a constant impedance. Both the open-circuit and the short-circuit L–sections are capable of being reversed so that R_0 is then presented at the output and the variable impedance is presented at the input.

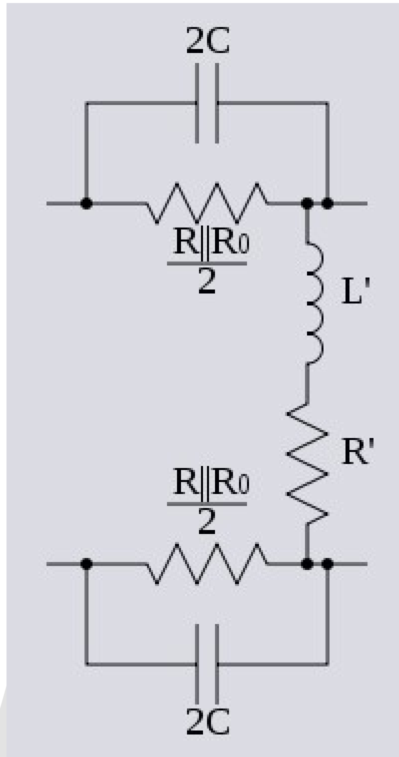
To retain the benefit of Zobel networks constant impedance, the variable impedance port must not face the line impedance. Nor should it face the variable impedance port of another half section. Facing the amplifier is acceptable since the input impedance of the amplifier is normally arranged to be R_0 within acceptable tolerances. In other words, variable impedance must not face variable impedance.



A balanced bridged T high-pass full section with basic loss

Balanced bridged T

The Zobel networks described here can be used to equalise land lines composed of twisted pair or star quad cables. The balanced circuit nature of these lines delivers a good common mode rejection ratio (CMRR). To maintain the CMRR, circuits connected to the line should maintain the balance. For this reason, balanced versions of Zobel networks are sometimes required. This is achieved by halving the impedance of the series components and then putting identical components in the "earthy" leg of the circuit.



A balanced Zobel high-pass short-circuit derived C-section with basic loss

Balanced C-sections

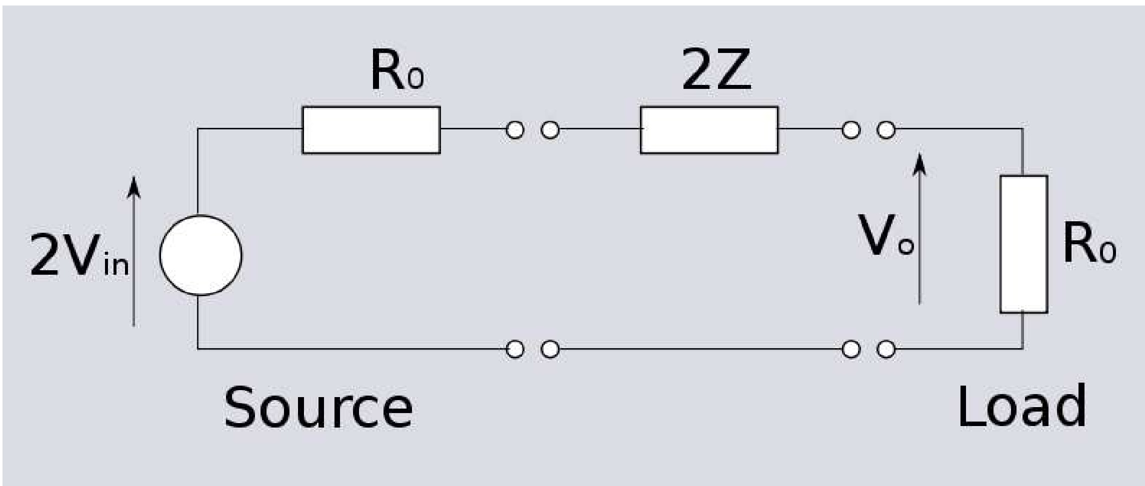
A C–section is a balanced version of an L–section. The balance is achieved in the same way as a balanced full bridged T section by placing half of the series impedance in, what was, the common conductor. C–sections, like the L–section from which they are derived, can come in both open-circuit and short circuit varieties. The same restrictions apply to C–sections regarding impedance terminations as to L–sections.

X-section

It is possible to transform a bridged–T section into a Lattice, or X–section. The X–section is a kind of bridge circuit, but usually drawn as a lattice, hence the name. Its topology makes it intrinsically balanced but it is never used to implement the constant resistance filters of the kind described here because of the increased component count. The component count increase arises out of the transformation process rather than the balance. There is however, one common application for this topology, the lattice phase equaliser, which is also constant resistance and also invented by Zobel. This circuit differs from those described here in that the bridge circuit is not generally in the balanced condition.

Half sections

In respect of constant resistance filters, the term half section has a somewhat different meaning to other kinds of image filter. Generally, a half section is formed by cutting through the mid-point of the series impedance and shunt admittance of a full section of a ladder network. It is literally half a section. Here, however, there is a somewhat different definition. A half section is either the series impedance (series half-section) or shunt admittance (shunt half-section) that, when connected between source and load impedances of R_0, will result in the same transfer function as some arbitrary constant resistance circuit. The purpose of using half sections is that the same functionality is achieved with a drastically reduced component count.



A general Zobel series half section showing the equality of transfer function to an equivalent constant resistance section

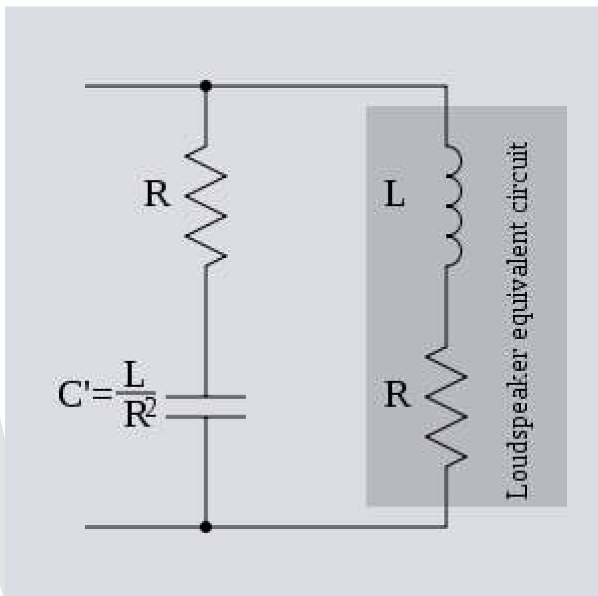
If a constant resistance circuit has an input V_{in}, then a generator with an impedance R_0 must have an open-circuit voltage of $E=2V_{in}$ in order to produce V_{in} at the input of the constant resistance circuit. If now the constant resistance circuit is replaced by an impedance of 2Z, as in the diagram above, it can be seen by simple symmetry that the voltage V_{in} will appear half way along the impedance 2Z. The output of this circuit can now be calculated as,

$$\frac{V_O}{V_{in}} = \frac{R_0}{Z + R_0}$$

which is precisely the same as a bridged T section with series element Z. The series half-section is thus a series impedance of 2Z. By corresponding reasoning, the shunt half-section is a shunt impedance of ½Z' (or twice the admittance).

It must be emphasised that these half sections are far from being constant resistance. They have the same transfer function as a constant resistance network, but only when correctly terminated. An equaliser will not give good results if a half-section is positioned facing the line since the line will have a variable (and probably unknown) impedance. Likewise, two half-sections cannot be connected directly to each other as these both will have variable impedances. However, if a sufficiently large attenuator is placed between the two variable impedances, this will have the effect of masking the effect. A high value attenuator will have an input impedance $\approx R_0$ no matter what the terminating impedance on the other side. In the example practical chain shown above there is a 22dB attenuator required in the chain. This does not need to be at the end of the chain, it can be placed anywhere desired and used to mask two mismatched impedances. It can also be split into two or more parts and used for masking more than one mismatch.

Zobel networks and loudspeaker drivers



Zobel network correcting loudspeaker impedance

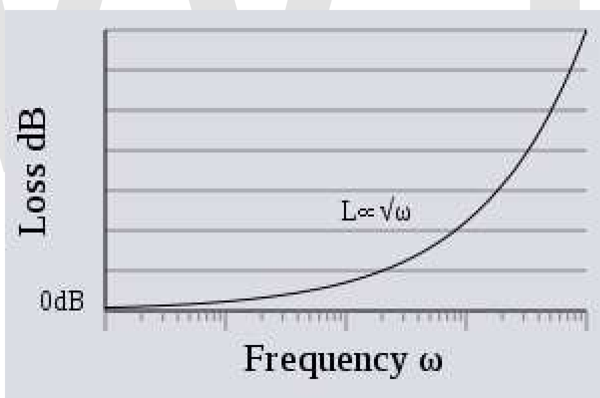
Zobel networks can be used to make the impedance a loudspeaker presents to its amplifier output appear as a steady resistance. This is beneficial to the amplifier performance. The impedance of a loudspeaker is partly resistive. The resistance representing the energy transferred from the amplifier to the sound output plus some heating losses in the loudspeaker. However, the speaker also possesses inductance due to the windings of its coil. The impedance of the loudspeaker is thus typically modelled as a series resistor and inductor. A parallel circuit of a series resistor and capacitor of the correct values will form a Zobel bridge. It is obligatory to choose $R_B = \infty$ because the centre point between the inductor and resistor is inaccessible (and, in fact, fictitious - the resistor and inductor are distributed quantities as in a transmission line). The loudspeaker may be modelled more accurately by a more complex equivalent circuit. The compensating Zobel network will also become more complex to the same degree.

Note that the circuit will work just as well if the capacitor and resistor are interchanged. In this case the circuit is no longer a Zobel balanced bridge but clearly the impedance has not changed. The same circuit could have been arrived at by designing from Boucherot's minimising reactive power point of view. From this design approach there is no difference in the order of the capacitor and the resistor and Boucherot cell might be considered a more accurate description.

Video equalisers

Zobel networks can be used for the equalisation of video lines as well as audio lines. There is, however, a noticeably different approach taken with the two types of signal. The difference in the cable characteristics can be summarised a s follows;

- Video commonly uses co-axial cable which requires an unbalanced topology for the filters whereas audio commonly uses twisted pair which requires a balanced topology.
- Video requires a wider bandwidth and tighter differential phase specification which in turn results in a tighter dimensional specification for the cable.
- The tighter specifications for video cable tends to produce a substantially constant characteristic impedance over a wide band (usually nominally 75 Ω). On the other hand, audio cable may be nominally 600 Ω (300 Ω and 150 Ω are also standard values), but it will only actually measure this value at 800 Hz. At a lower frequencies it will be much higher and at higher frequencies will be lower and more reactive.

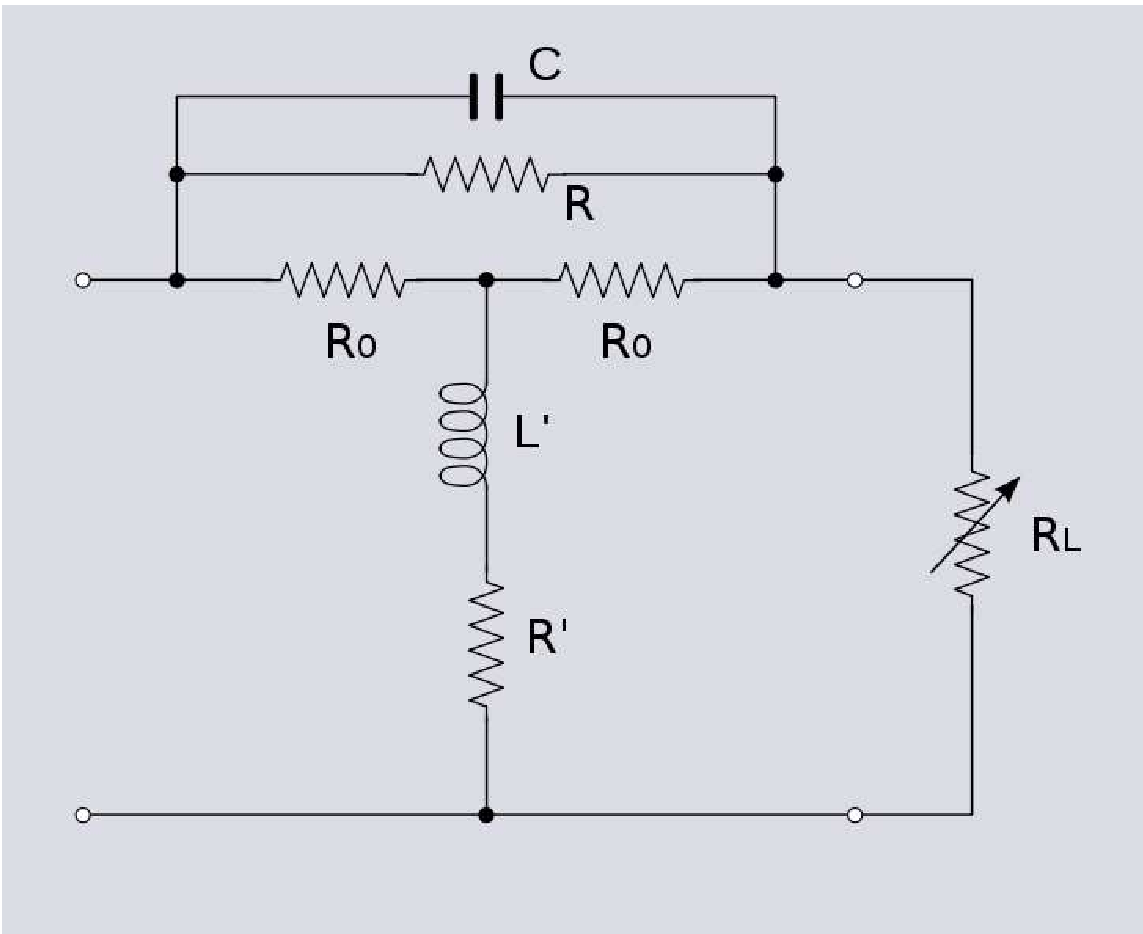


Response plot of a video line showing the typical $\sqrt{f}$ response

- These characteristics result in a smoother, more well behaved response for video lines with none of the nasty discontinuities typically found with audio lines. These discontinuities in the frequency response are often caused by the habit of the telecom companies of forming a connection by joining two shorter lines of differing characteristic impedance. Video lines on the other hand tend to roll off smoothly with frequency in a predictable way.

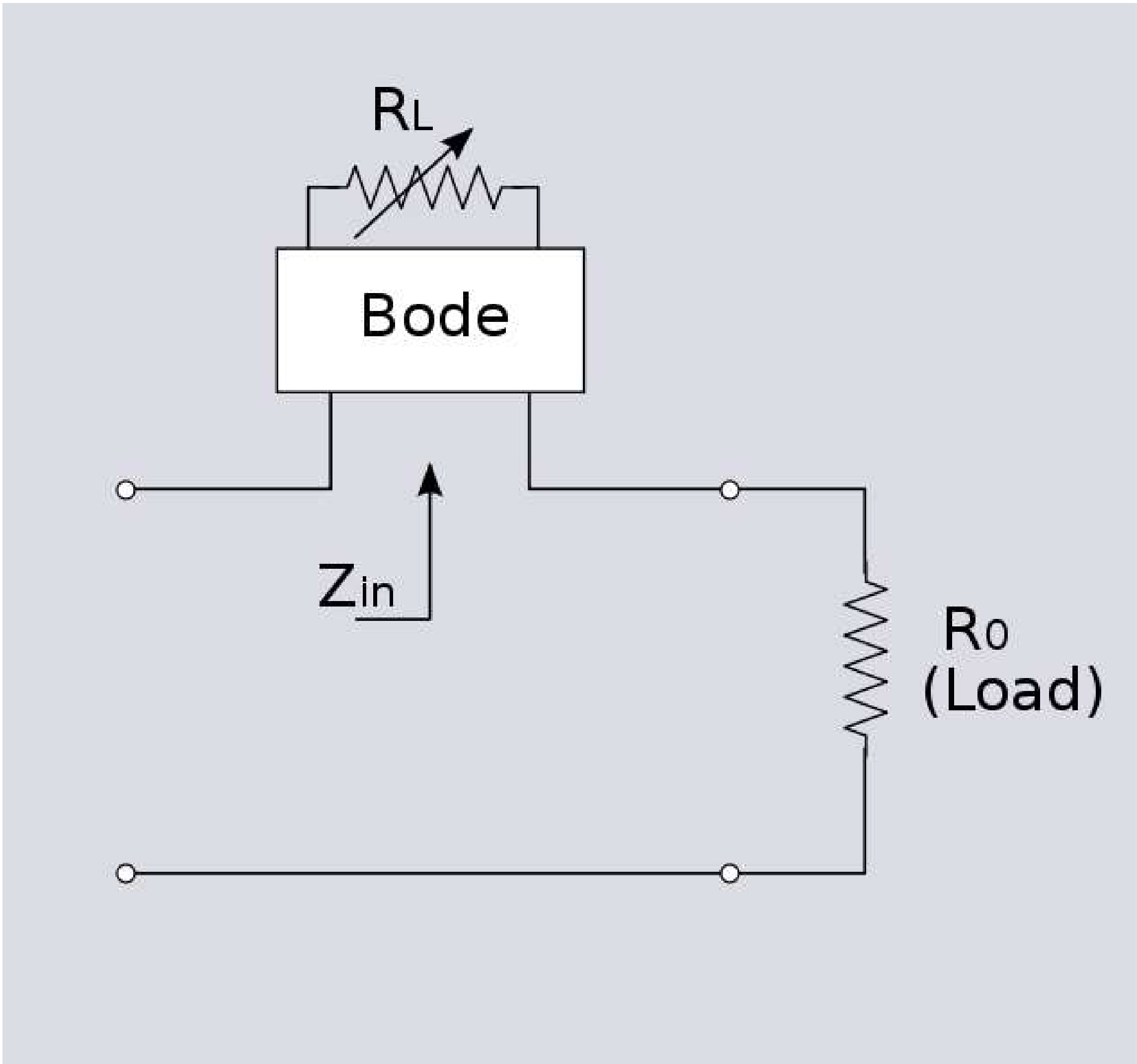
This more predictable response of video allows a different design approach. The video equaliser is built as a single bridged T section but with a rather more complex network for Z. For short lines, or for a trimming equaliser, a Bode filter topology might be used. For longer lines a network with Cauer filter topology might be used. Another driver for this approach is the fact that a video signal occupies a large number of octaves, around 20 or so. If equalised with simple basic sections, a large number of filter sections would be required. Simple sections are designed, typically, to equalise a range of one or two octaves.

Bode equaliser

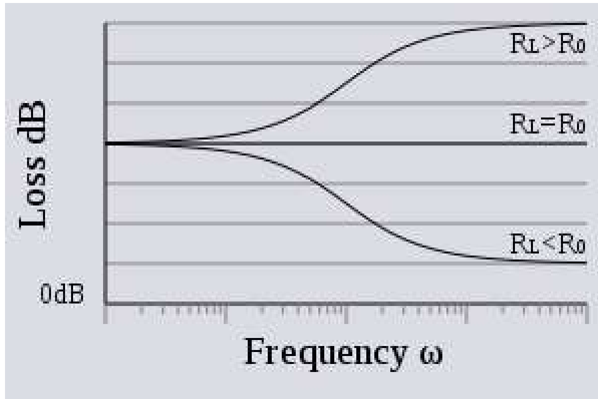


Bode network for high frequency equalisation

A Bode network, as with a Zobel network, is a symmetrical bridge T network which meets the constant k condition. It does not however meet the constant resistance condition, that is, the bridge is not in balance. Any impedance network, Z, can be used in a Bode network, just as with a Zobel network, but the high pass section shown for correcting high-end frequencies is the most common. A Bode network terminated in a variable resistor can be used to produce a variable impedance at the input terminals of the network. A useful property of this network is that the input impedance can be made to vary from a capacitive impedance through a purely resistive impedance to an inductive impedance all by adjusting the single load potentiometer, R_L. The bridging resistor, R_0, is chosen to equal the nominal impedance so that in the special case when R_L is set to R_0 the network behaves as a Zobel network and Z_{in} is also equal to R_0.



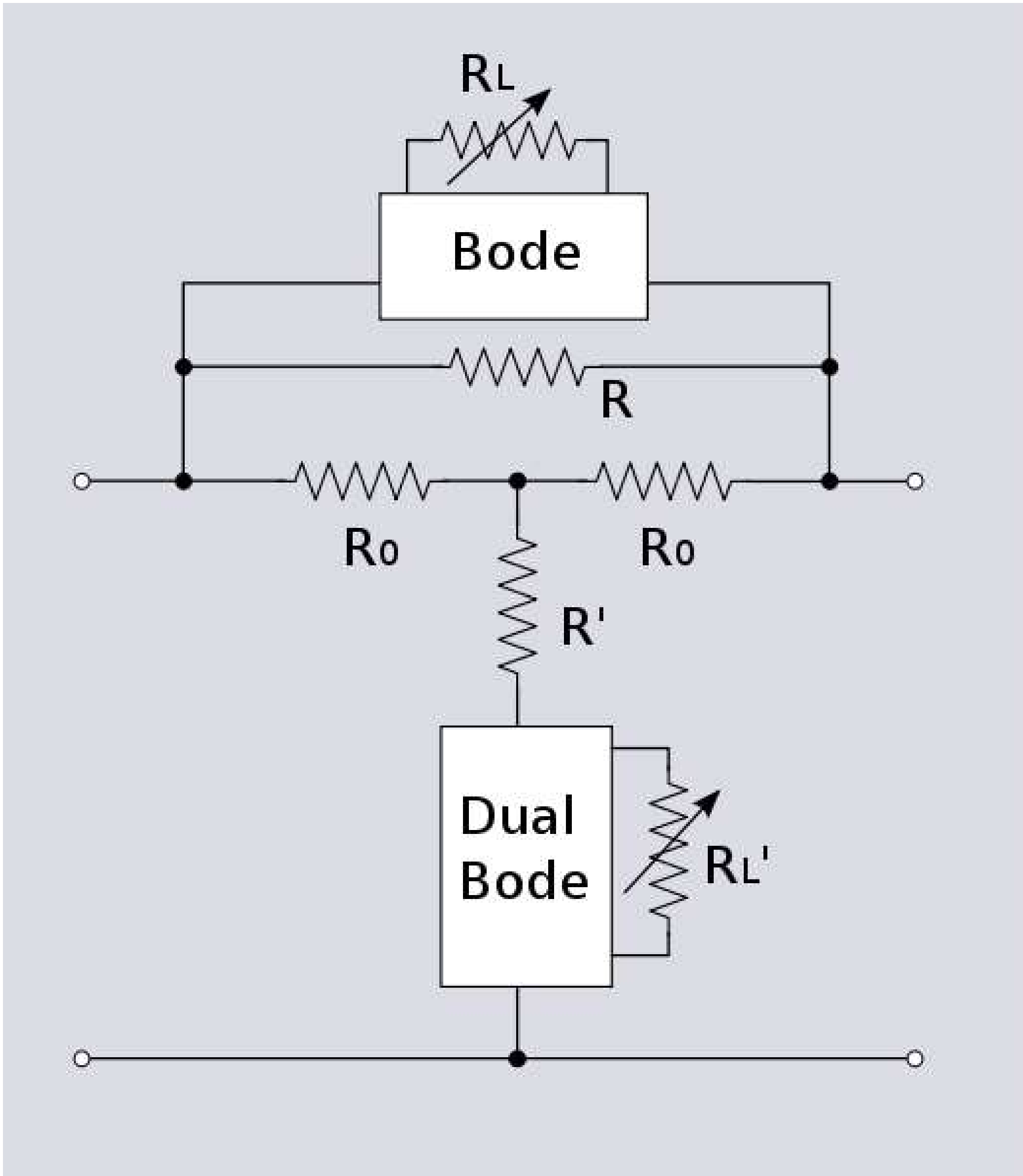
Bode network used in an equaliser circuit



Bode trimming equaliser response plot

The Bode network is used in an equaliser by connecting the whole network such that the input impedance of the Bode network, Z_{in}, is in series with the load. Since the impedance of the Bode network can be either capacitive or inductive depending on the position of the adjustment potentiometer, the response may be a boost or a cut to the band of frequencies it is acting on. The transfer function of this arrangement is:

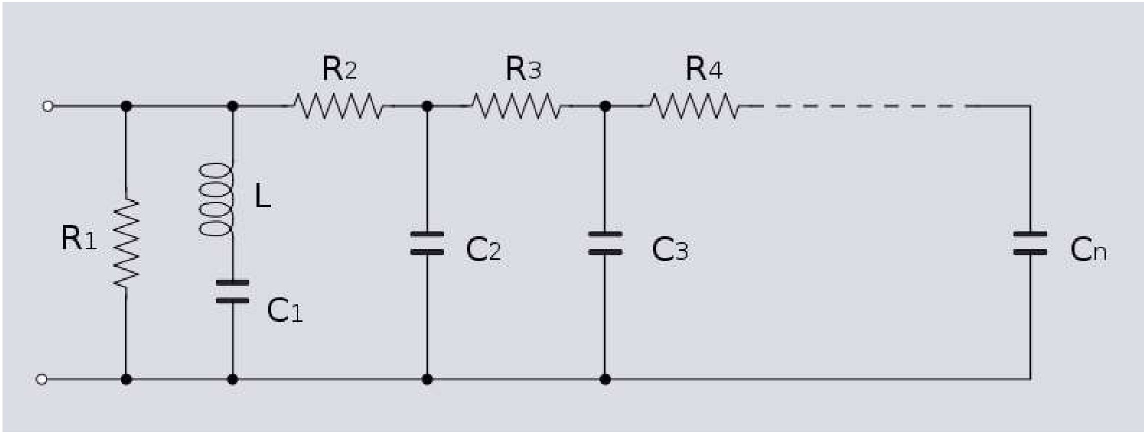
$$A(\omega) = \frac{R_0}{Z_{in} + R_0}$$



Bode equaliser implemented as a Zobel constant resistance equaliser

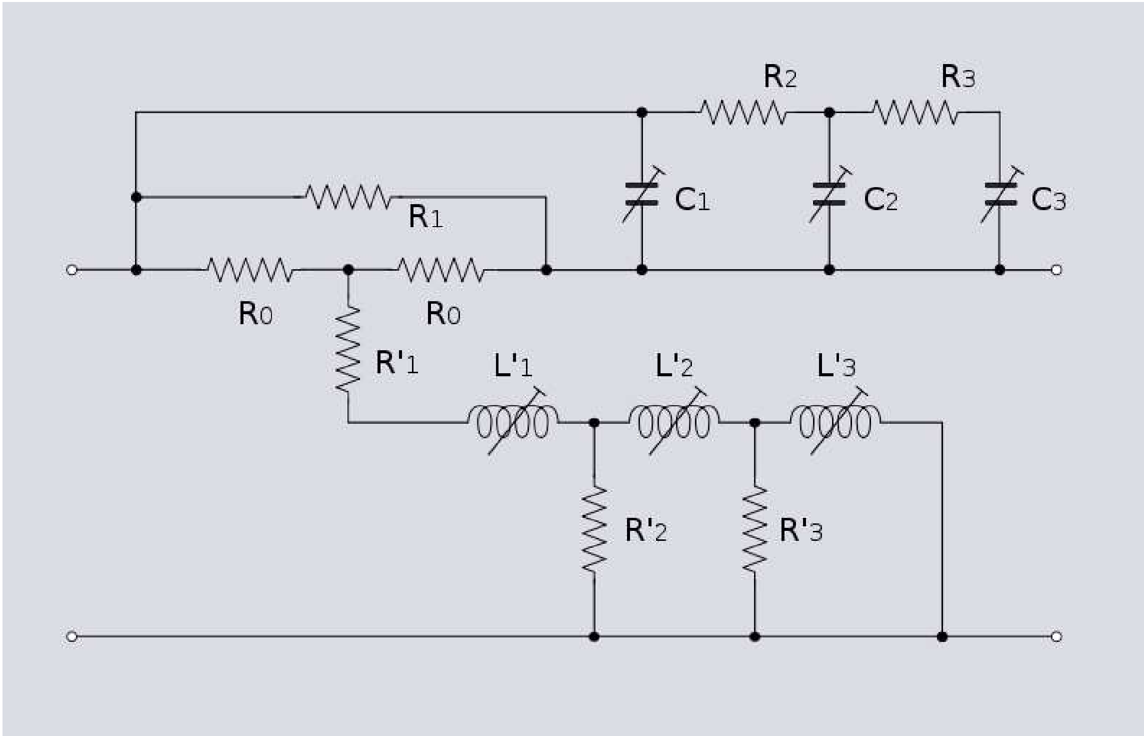
The Bode equaliser can be converted into a constant resistance filter by using the entire Bode network as the Z branch of a Zobel network, resulting in a rather complex network of bridge T networks embedded in a larger bridge T. It can be seen that this results in the same transfer function by noting that the transfer function of the Bode equaliser is identical to the transfer function of the general form of Zobel equaliser. Note that the dual of a constant resistance bridge T network is the identical network. The dual of a Bode network is therefore the same network except for the load resistance R_L, which must be the inverse, R_L', in the dual circuit. To adjust the equaliser R_L and R_L' must be ganged, or otherwise kept in step such that as R_L increases R_L' will decrease and vice versa.

Cauer equaliser



Cauer topology network to be used as the Z impedance of a Zobel network equaliser

To equalise long video lines, a network with Cauer topology is used as the Z impedance of a Zobel constant resistance network. Just as the input impedance of a Bode network is used as the Z impedance of a Zobel network to form a Zobel Bode equaliser, so the input impedance of a Cauer network is used to make a Zobel Cauer equaliser. The equaliser is required to correct an attenuation increasing with frequency and for this a Cauer ladder network consisting of series resistors and shunt capacitors is required. Optionally, there may be an inductor included in series with the first capacitor which increases the equalisation at the high end due to the steeper slope produced as resonance is approached. This may be required on longer lines. The shunt resistor R_1 provides the basic loss of the Zobel network in the usual way.



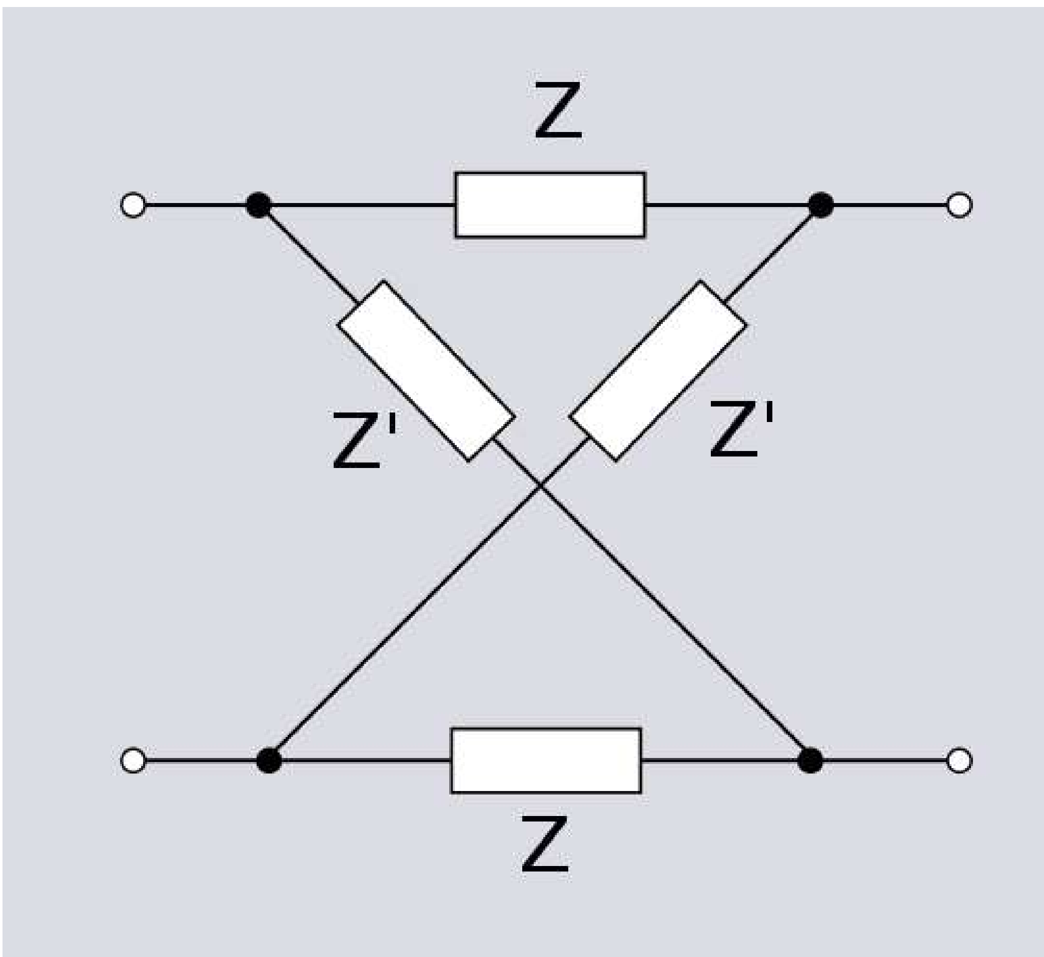
Cauer Zobel bridge T implemented video equaliser. The impedance Z of this example consists of a three section ladder and is suitable for the equalisation of short lines (between nearby buildings for instance)

The dual of a RC Cauer network is a LR Cauer network which is required for the Z' impedance as shown in the example. Adjustment is a bit problematic with this equaliser. In order to maintain the constant resistance, the pairs of components C_1/L_1', C_2/L_2' etc., must remain dual impedances as the component is adjusted, so both parts of the pair must be adjusted together. With the Zobel Bode equaliser, this is a simple matter of ganging two pots together - a component configuration available off-the-shelf. Ganging together a variable capacitor and inductor is not, however, a very practical solution. These equalisers tend to be "hand built", one solution being to select the capacitors on test and fit fixed values according to the measurements and then adjust the inductors until the required match is achieved. The furthest element of the ladder from the driving point is equalising the lowest frequency of interest. This is adjusted first as it will also have an effect on higher frequencies and from there progressively higher frequencies are adjusted working along the ladder towards the driving point.

Chapter 12

Lattice Phase Equalizer and General mn-type Image Filter

Lattice phase equaliser



Lattice filter topology

A **lattice phase equaliser** or **lattice filter** is an example of an all-pass filter. That is, the attenuation of the filter is constant at all frequencies but the relative phase between input and output varies with frequency. The lattice filter topology has the particular property of being a constant-resistance network and for this reason is often used in combination with other constant resistance filters such as bridge-T equalisers. The topology of a lattice filter, also called an **X-section** is identical to bridge topology. The lattice phase equaliser was invented by Otto Zobel. using a filter topology proposed by George Campbell.

The characteristic impedance of this structure is given by;

$$Z_o^2 = ZZ'$$

and the transfer function is given by;

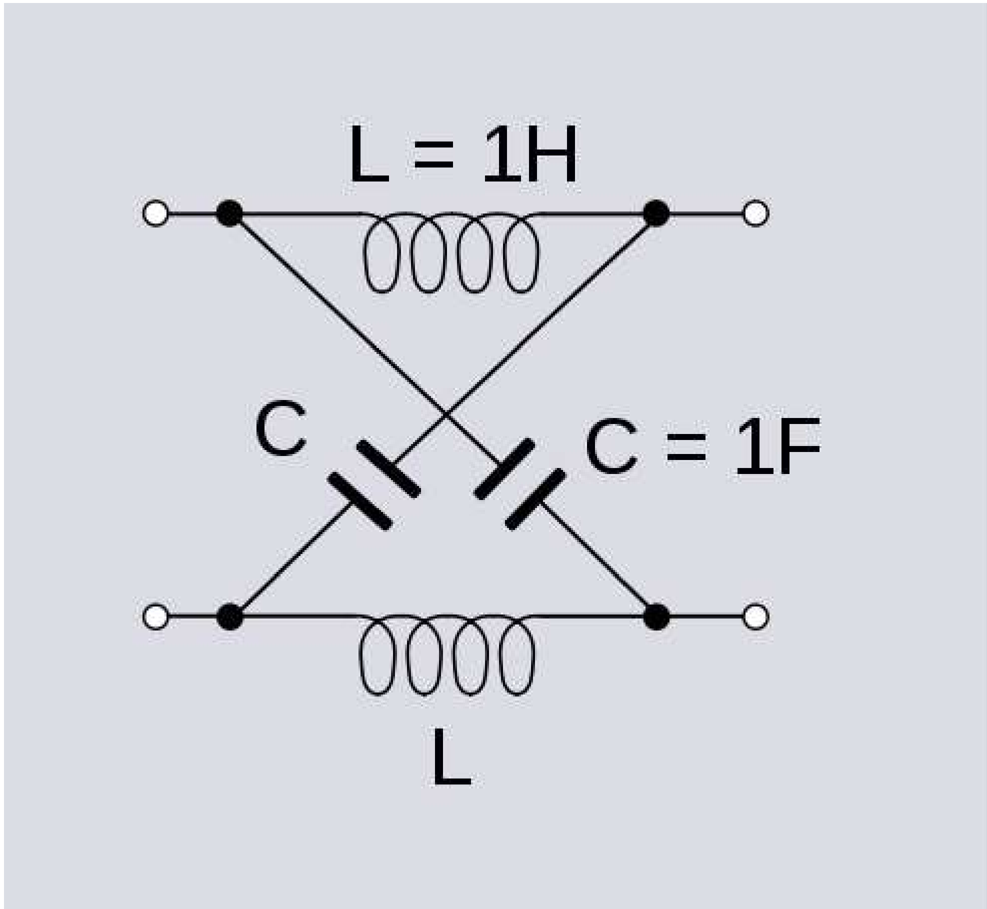
$$H(\omega) = \frac{Z_o - Z}{Z_o + Z}$$

Applications

The lattice filter has an important application on lines used by broadcasters for stereo audio feeds. Phase distortion on a monophonic line does not have a serious effect on the quality of the sound unless it is very large. The same is true of the absolute phase distortion on each leg (left and right channels) of a stereo pair of lines. However, the differential phase between legs has a very dramatic effect on the stereo image. This is because the formation of the stereo image in the brain relies on the phase difference information from the two ears. A phase difference translates to a delay, which in turn can be interpreted as a direction the sound came from. Consequently, landlines used by broadcasters for stereo transmissions are equalised to very tight differential phase specifications.

Another property of the lattice filter is that it is an intrinsically balanced topology. This is useful when used with landlines which invariably use a balanced format. Many other types of filter section are intrinsically unbalanced and have to be transformed into a balanced implementation in these applications which increases the component count. This is not required in the case of lattice filters.

Design

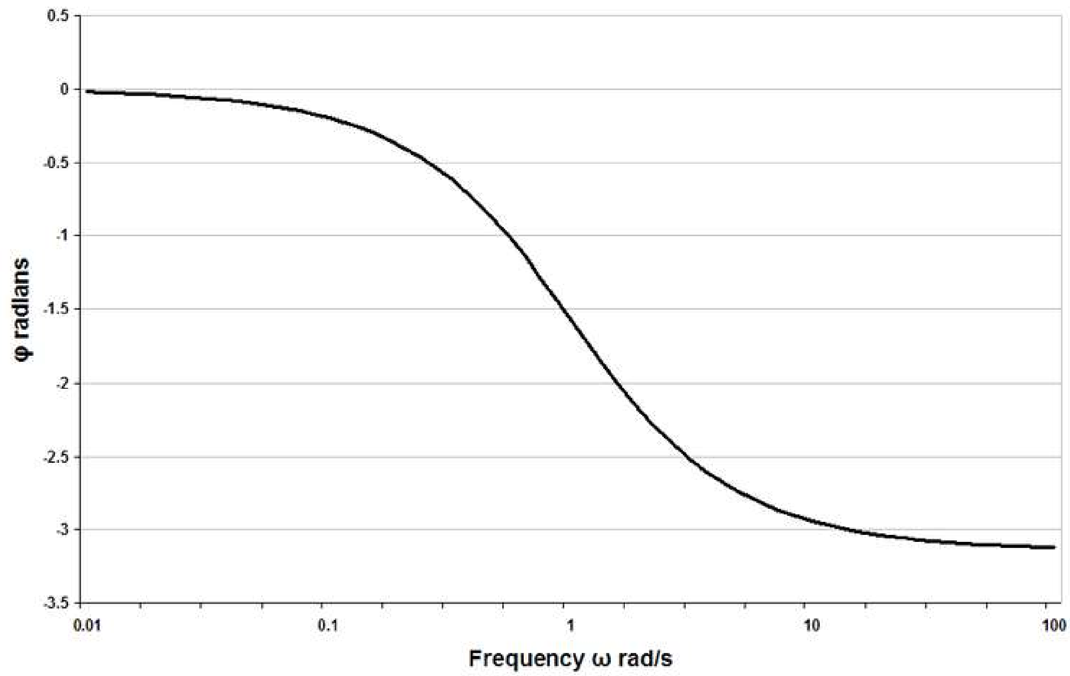


A prototype lattice filter which passes low frequencies without phase shifting

The essential requirement for a lattice filter is that for it to be constant resistance, the lattice element of the filter must be the dual of the series element with respect to the characteristic impedance. That is,

$$\frac{Z}{R_0} = \frac{R_0}{Z'}$$

Such a network, when terminated in R_0, will have an input resistance of R_0 at all frequencies. If the impedance Z is purely reactive such that Z = iX then the phase shift, φ, inserted by the filter is given by,

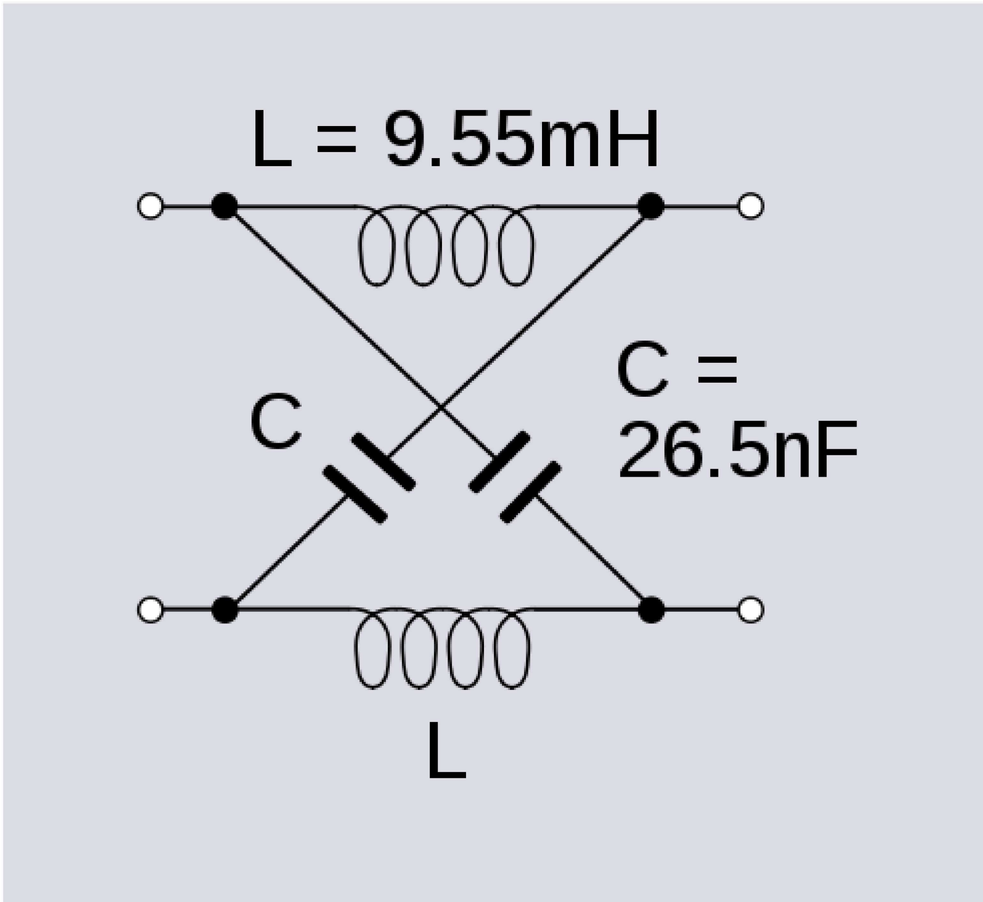


Prototype lattice filter response ranging from 0 radians at low frequencies to -π radians at high frequencies

$$\tan\frac{\varphi}{2} = -\frac{X}{R_0}$$

The prototype lattice filter shown here passes low frequencies without modification but phase shifts high frequencies. That is, it is phase correction for the high end of the band. At low frequencies the phase shift is 0° but as the frequency increases the phase shift approaches 180°. It can be seen qualitatively that this is so by replacing the inductors with open circuits and the capacitors with short circuits, which is what they become at high frequency. At high frequency the lattice filter is a cross-over network and will produce 180° phase shift. A 180° phase shift is the same as an inversion in the frequency domain, but is a delay in the time domain. At an angular frequency of ω = 1 rad/s the phase shift is exactly 90° and this is the mid-point of the filter's transfer function.

Low-in-phase section



Lattice filter transformed from the prototype to operate at 10 kHz mid-point and 600Ω terminations

The prototype section can be scaled and transformed to the desired frequency, impedance and bandform by applying the usual prototype filter transforms. A filter which is in-phase at low frequencies (that is, one that is correcting phase at high frequencies) can be obtained from the prototype with simple scaling factors.

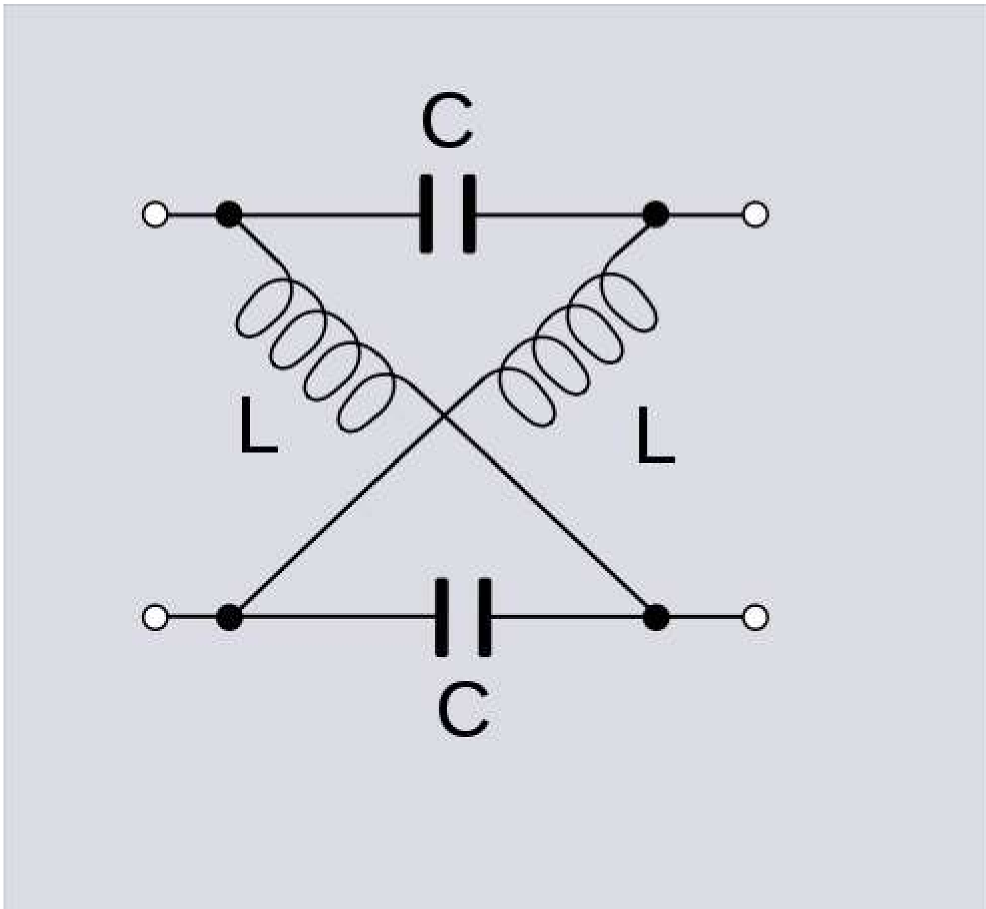
The phase response of a scaled filter is given by,

$$\tan\frac{\varphi}{2} = -\frac{\omega}{\omega_m}$$

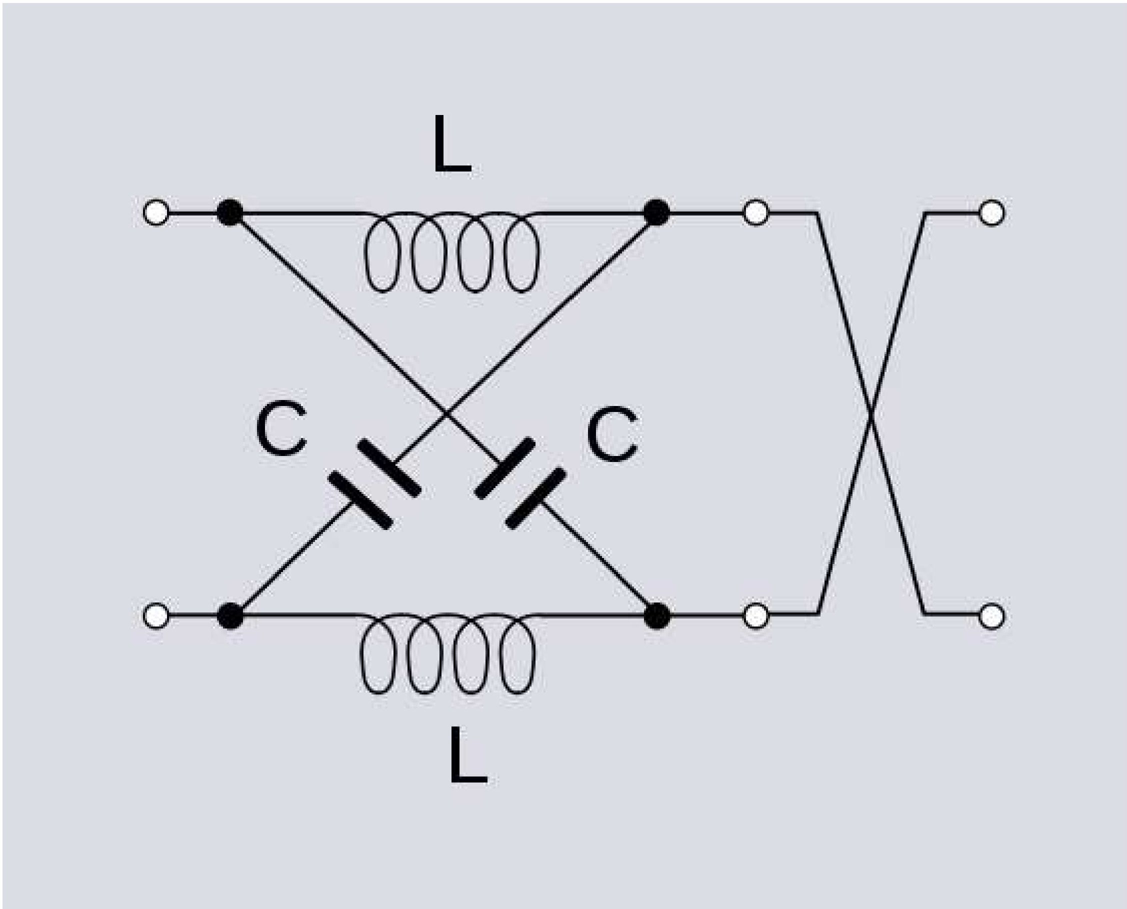
where ω_m is the mid-point frequency and is given by,

$$\omega_m = \frac{1}{\sqrt{LC}}$$

High-in-phase section



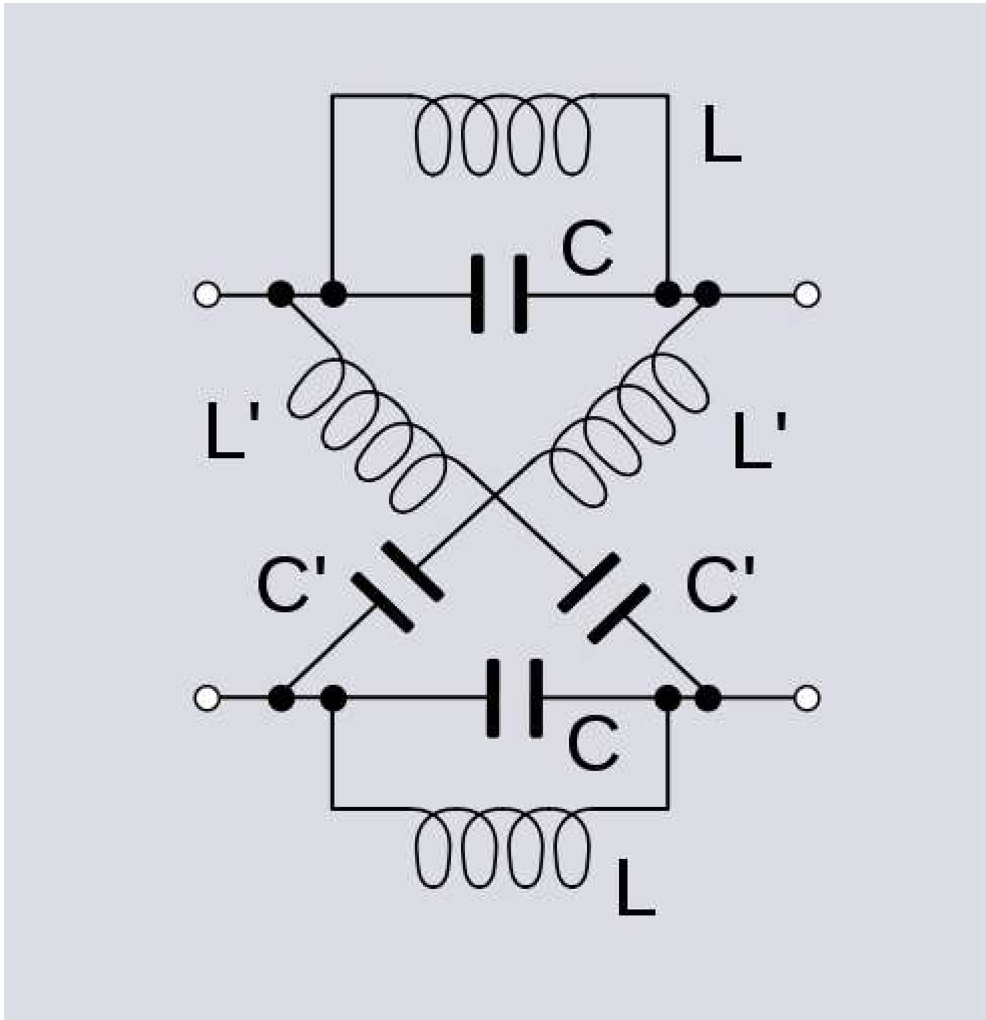
Lattice filter for low end phase correction



Demonstration that a low-in-phase section in cascade with a crossover is equivalent to a high-in-phase section

A filter that is in-phase at high frequencies (that is, a filter to correct low-end phase) can be obtained by applying the high-pass transformation to the prototype filter. However, it can be seen that due to the lattice topology this is also equivalent to a crossover on the output of the corresponding low-in-phase section. This second method may not only make calcualtion easier but it is also a useful property where lines are being equalised on a temporary basis, for instance for outside broadcasts. It is desirable to keep the number of different types of adjustable sections to a minimum for temporary work and being able to use the same section for both high end and low end correction is a distinct advantage.

Band equalise section

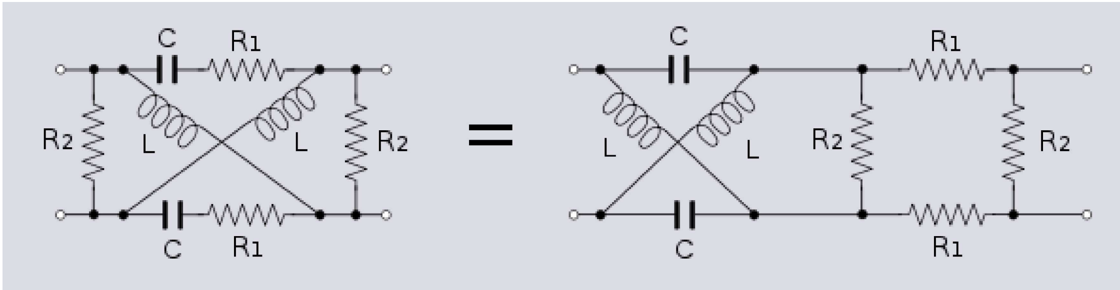


Lattice filter for phase correction of a limited band

A filter that corrects a limited band of frequencies (that is, a filter that is in phase everywhere except in the band being corrected) can be obtained by applying the band-stop transformation to the prototype filter. This results in resonant elements appearing in the filter's network.

An alternative, and possibly more accurate, view of this filter's response is to describe it as a phase change that varies from 0° to 360° with increasing frequency. At 360° phase shift, of course, the input and output are now back in phase with each other.

Resistance compensation

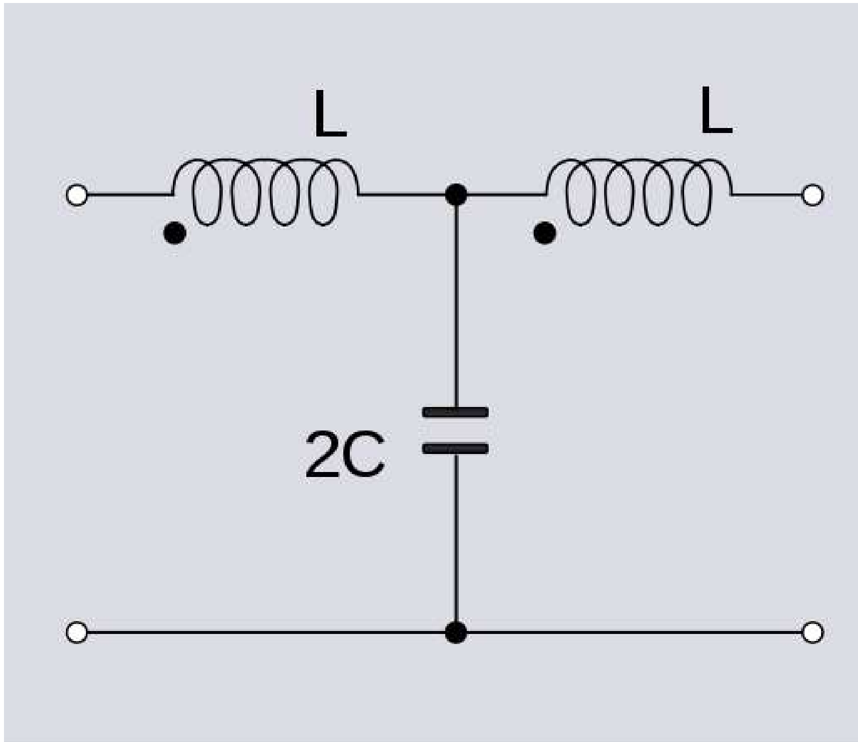


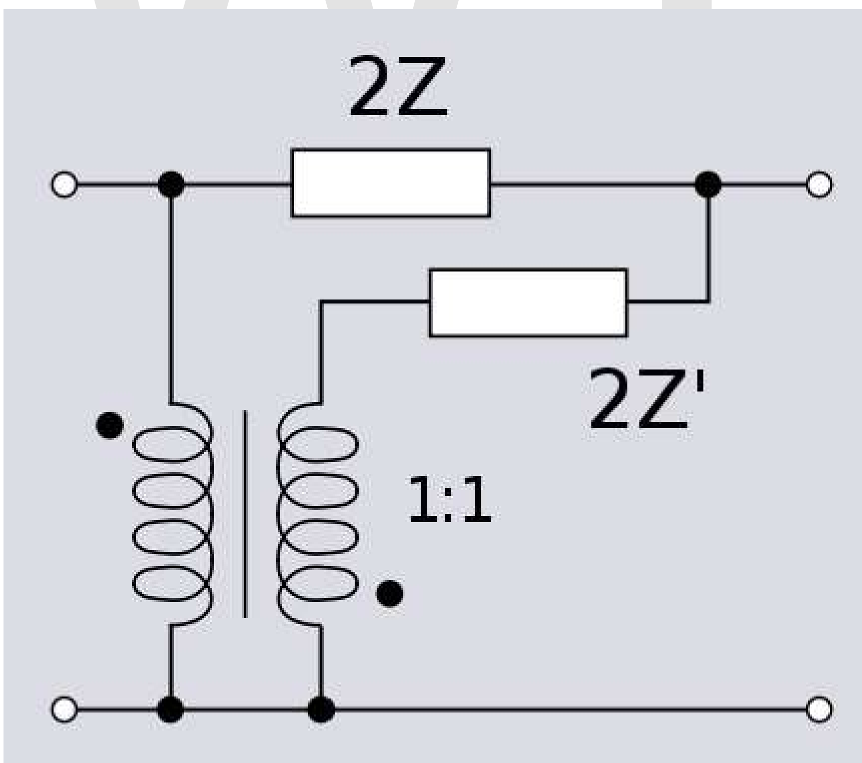
A lattice filter with compensation for the resistance of its inductors and its equivalent circuit

With ideal components there is no need to use resistors in the design of lattice filters. However, practical considerations of properties of real components leads to resistors being incorporated. Sections designed to equalise low audio frequencies will have larger inductors with a high number of turns. This results in significant resistance being in the inductive branches of the filter, which in turn causes attenuation at low frequencies.

In the example diagram, the resistors placed in series with the capacitors, R_1, are made equal to the unwanted stray resistance present in the inductors. This ensures that the attenuation at high frequency is the same as the attenuation at low frequency and brings the filter back to a flat response. The purpose of the shunt resistors, R_2, is to bring the image impedance of the filter back to the original design R_0. The resulting filter is the equivalent of a box attenuator formed from the R_1's and R_2's connected in cascade with an ideal lattice filter as shown in the diagram.

Unbalanced topology



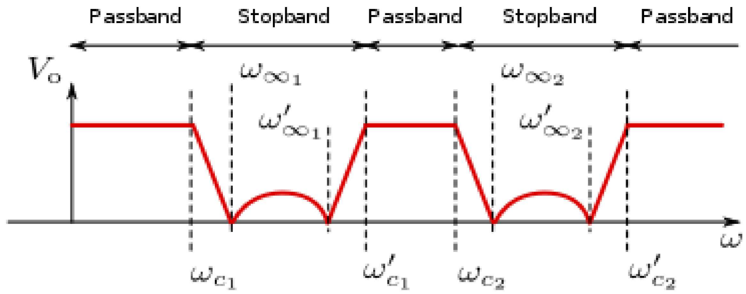


The lattice phase equaliser cannot be directly transformed into T-section topology without introducing active components. However, a T-section is possible if ideal transformers are introduced. Transformer action can be conveniently achieved in the low-in-phase T-section by winding both inductors on a common core. The response of this section is identical to the original lattice, however, the input is no longer constant resistance. This circuit was first used by George Washington Pierce who needed a delay line as part of the improved sonar he developed between the world wars. Pierce used a cascade of these sections to provide the required delay. The circuit can be considered a low-pass m-derived filter with $m>1$ which puts the transmission zero on the $j\omega$ axis of the complex frequency plane. Other unbalanced transformations utilising ideal transformers are possible, one such is shown on the right.

General mn-type image filter

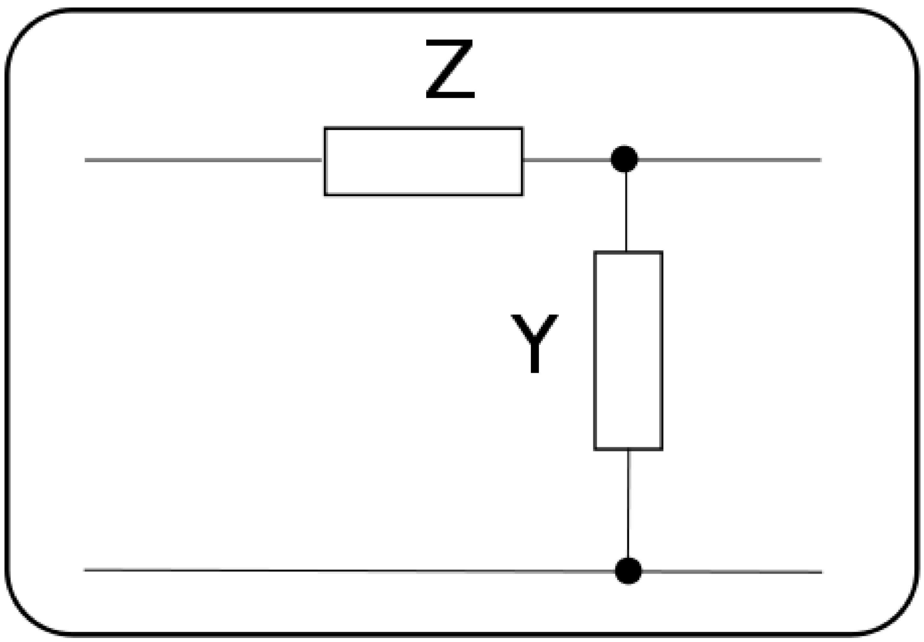
These filters are electrical wave filters designed using the image method. They are an invention of Otto Zobel at AT&T Corp.. They are a generalisation of the m-type filter in that a transform is applied that modifies the transfer function while keeping the image impedance unchanged. For filters that have only one stopband there is no distinction with the m-type filter. However, for a filter that has multiple stopbands, there is the possibility that the form of the transfer function in each stopband can be different. For instance, it may be required to filter one band with the sharpest possible cut-off, but in another to minimise phase distortion while still achieving some attenuation. If the form is identical at each transition from passband to stopband the filter will be the same as an m-type filter (k-type filter in the limiting case of $m=1$). If they are different, then the general case described here pertains.

The k-type filter acts as a prototype for producing the general m_n designs. For any given desired bandform there are two classes of m_n transformation that can be applied, namely, the mid-series and mid-shunt derived sections; this terminology being more fully explained in the m-derived filter article. Another feature of m-type filters that also applies in the general case is that a half section will have the original k-type image impedance on one side only. The other port will present a new image impedance. The two transformations have equivalent transfer functions but different image impedances and circuit topology.



Bandform diagram showing frequency response of a general image filter. The ω_c are the critical frequencies (the frequency where cut-off begins) and the ω_∞ are the poles of attenuation in the stop bands.

Mid-series multiple stopband



If Z and Y are the series impedance and shunt admittance of a constant k half section and;

$$Z = Z_1 + Z_2 + Z_3 + \ldots + Z_N$$

where Z_1, Z_2 etc are a cascade of antiresonators,

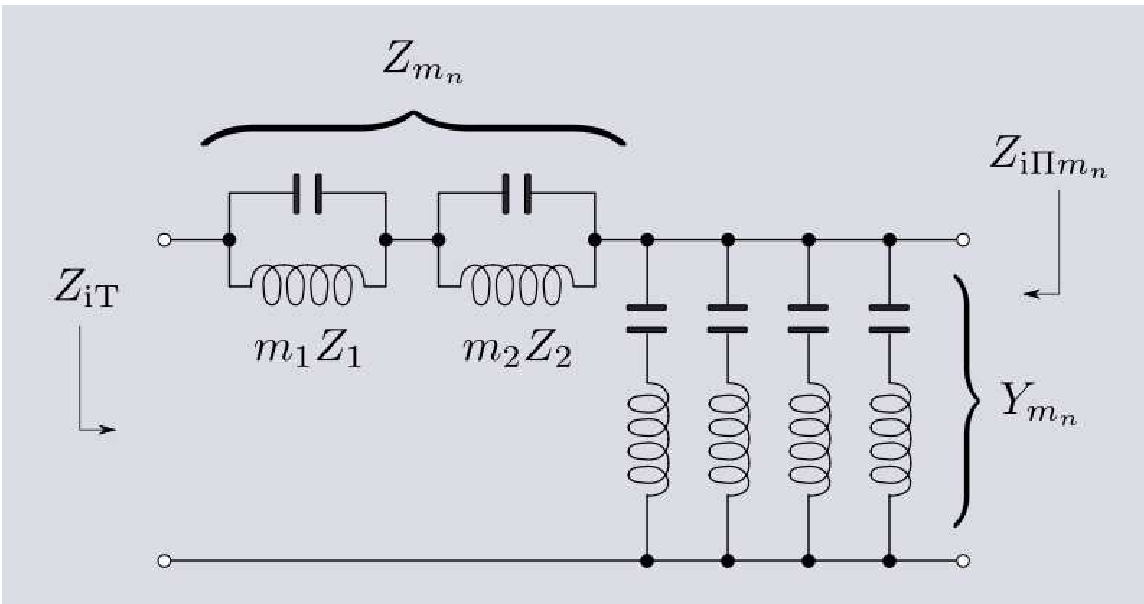
the transformed series impedance for a mid-series derived filter becomes;

$$Z_{mn} = m_1 Z_1 + m_2 Z_2 + m_3 Z_3 + \ldots + m_N Z_N$$

Where the m_n are arbitrary positive coefficients. For an invariant image impedance Z_{iT} and invariant bandform (that is, invariant cut-off frequencies ω_c) the transformed shunt admittance, expressed in terms of Z_{mn}, is given by;

$$Y_{m_n} = \frac{Z_{m_n}}{k^2 + Z^2 - Z_{m_n}^2}$$

where $k = \sqrt{\frac{Z}{Y}}$ and is a constant by definition. When the m_n are all equal this reduces to the expression for an m-type filter and where they are all equal to one it reduces further to the *k*-type filter.



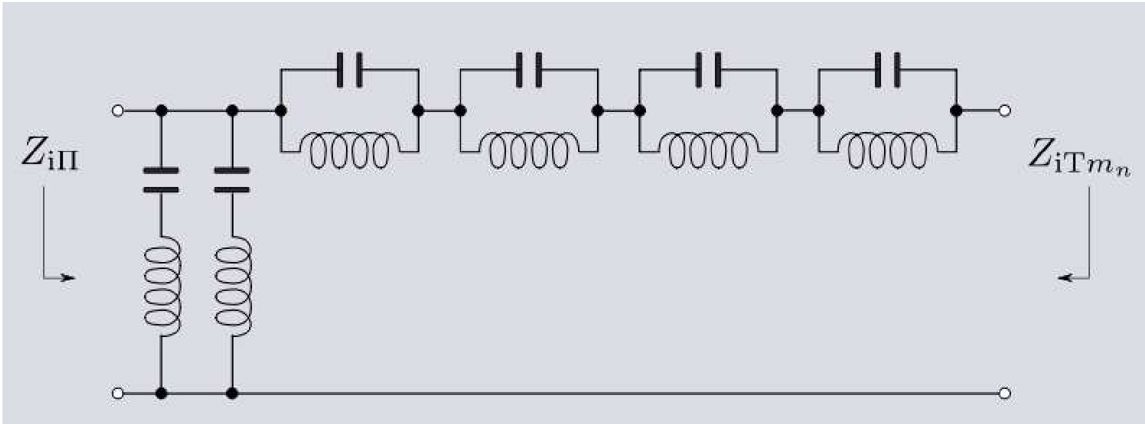
2-bandstop mid-series derived m_1, m_2 filter half-section

A result of this relationship is that the *N* antiresonators in Z_{mn} will transform into $2N$ resonators in Y_{mn}. The coefficients m_n can be adjusted by the designer to set the frequency of one of the two poles of attenuation, ω_∞, in each stopband. The second pole of attenuation is dependant and cannot be set separately.

Special cases

In the case of a filter with a stopband extending to zero frequency, one of the antiresonators in *Z* will reduce to a single inductor. In this case the resonators in Y_{mn} are reduced by one to $2N$-1. Similarly, for a filter with a stopband extending to infinity, one antiresonator will reduce to a single capacitor and the resonators will again be reduced by one. In a filter where both conditions pertain, the number of resonators will be $2N$-2. For these end stopbands, there is only one pole of attenuation in each, as would be expected from the reduced number of resonators. These forms are the maximum allowable complexity while maintaining invariance of bandform and one image impedance.

Mid-shunt multiple stopband



2-bandstop mid-shunt derived m1, m2 filter half-section. Frequency response is equivalent to the corresponding mid-series derived filter

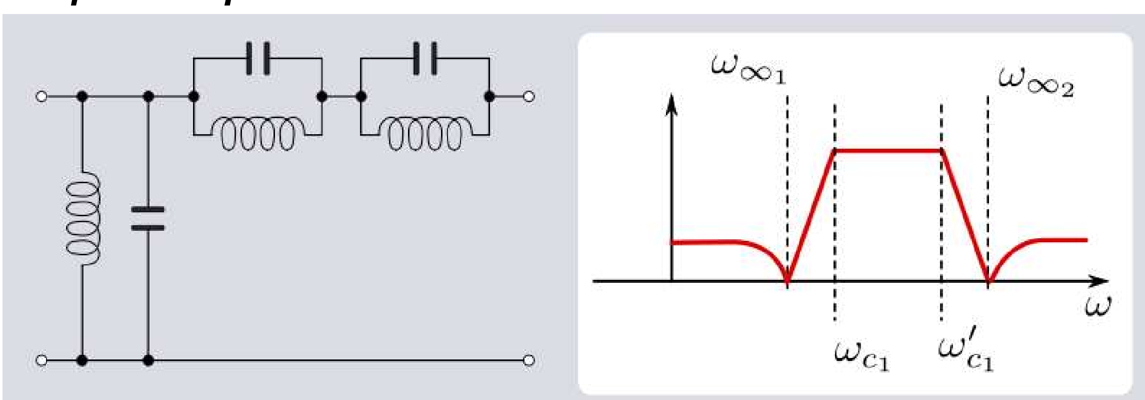
By dual analogy, the shunt derived filter starts from;

$$Y_{m_n} = m_1 Y_1 + m_2 Y_2 + m_3 Y_3 + \ldots + m_N Y_N$$

For an invariant image admittance $Y_{\mathrm{i}\Pi}$ and invariant bandform the transformed series impedance is given by;

$$Z_{m_n} = \frac{Y_{m_n}}{k^2 + Y^2 - Y_{m_n}^2}$$

Simple bandpass section



General image bandpass filter, mid-shunt derived

The bandpass filter can be characterised as a 2-bandstop filter with $\omega_c = 0$ for the lower critical frequency of the lower band and $\omega_c = \infty$ for the upper critical frequency of the upper band. The two resonators reduce to an inductor and a capacitor respectively. The number of antiresonators reduces to two.

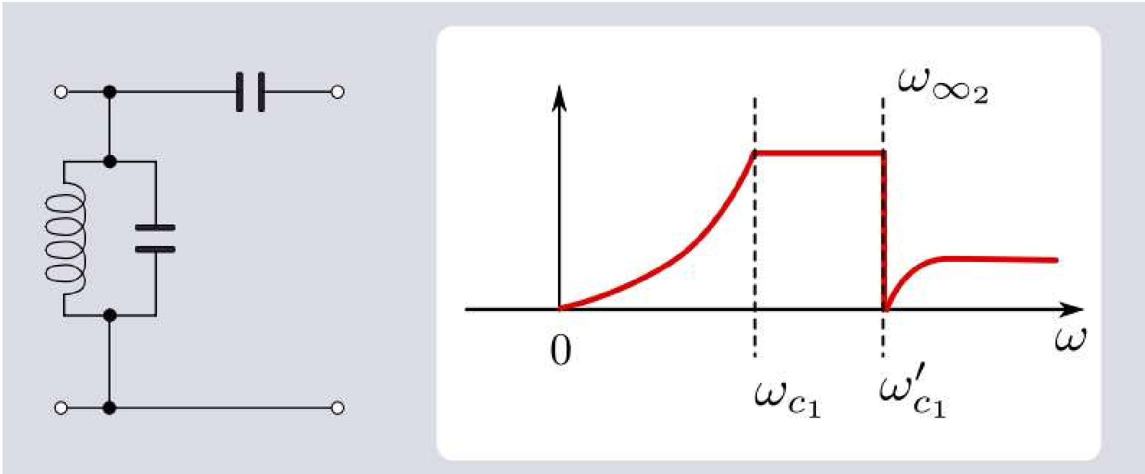


Image bandpass filter with $\omega_{\infty 1}$ set to zero and $\omega_{\infty 2}$ set to correspond to ω'_{c1}.

If, however, $\omega_{\infty 1}$ is set to zero (that is, there is no pole of attenuation in the lower stopband) and $\omega_{\infty 2}$ is set to correspond to the upper critical frequency ω'_{c1}, then a particularly simple form of the bandpass filter is obtained consisting of just antiresonators coupled by capacitors. This was a popular topology for multi-section band-pass filters due its low component count, particularly of inductors. Many other such reduced forms are possible by setting one of the poles of attenuation to correspond to one of the critical frequencies for various classes of basic filter.