

Cathode decoupling for an inductively loaded triode

Marcel van de Gevel, version 1, 30 December 2024

1. Introduction

By following tonescout's diyAudio thread "Cathode bypass calculator help", <https://www.diyaudio.com/community/threads/cathode-bypass-calculator-help.421677/post-7882156>, I learned that the way I always calculated the required cathode bypass capacitance for a valve was a bit too simplistic. I usually neglected the effect of the internal resistance of the valve, but that is not necessarily correct, especially not for triodes. When the triode is resistively loaded, this only leads to overestimating the required cathode decoupling capacitance, but in post #12, Miles Prower warned that incorrect capacitive cathode decoupling of a triode driving an inductive load can in fact lead to nasty resonances.

I later did a more extensive calculation and found that Miles Prower was absolutely right. That is, you can capacitively decouple the cathode without getting sharp resonances, but then you need to account properly for the internal resistance of the valve and choose the right capacitance. The calculation is described in this document.

As not everyone on diyAudio is used to calculate and manipulate transfer functions, section 2 contains some general information about electrical network theory and transfer functions; please skip this part if you are familiar with such theory. The derivation of the transfer function of an inductively loaded triode with cathode decoupling is in section 3. Section 4 deals with how to get a reasonably flat response. Numerical examples are found in section 5.

I assume that the reader has a good working knowledge of elementary algebra, differential calculus and complex numbers, and is familiar with the concept of linearizing a circuit around a bias point.

2. Electrical network theory

2.1. Modelling the circuit

Calculating the small-signal transfer of an electric or electronic circuit usually goes in two steps:

1. Draw a network model, that is a mathematical abstraction of the real-life circuit
2. Calculate the transfer of this model

Regarding the first step: the simplest model for a capacitor is an ideal linear capacitor (which my former network theory professor Fred Neerhoff always called a lineaire capaciteit, a linear capacity), the simplest model for a resistor is an ideal linear resistor, the simplest model for an inductor is an ideal linear inductor (linear inductivity). If you are worried about second-order effects, you need to use more elaborate models.

To give an example that is relevant to audio: an inductor in a passive crossover could be represented by the series connection of a linear inductivity and an ideal linear resistor if you

are worried about the effect of its DC resistance. In this document, however, all resistors, capacitors and inductors will be treated as ideal.

A triode is usually linearized around its bias point and then modelled with a voltage-controlled current source, and if needed, with an internal differential resistance from anode to cathode and with capacitances between its terminals. As we are concerned with the low-frequency behaviour in this document, the capacitances between the terminals will be neglected. The internal resistance will not, because its effect is much greater than I originally thought.

2.2. Transfer functions

The transfer from an input to an output of a linear(ized) network can be described by a transfer function consisting of the ratio between two polynomials in s , which in older literature is denoted as p . (This actually only applies when the network is linear, time invariant, continuous time and lumped, but analogue circuits used for audio are usually close enough to being all of that to model them with a linear, time-invariant, continuous-time lumped network). Personally I think the old notation p is much clearer than s , because s is too similar to s , the SI symbol for second. Still, since s is the more usual notation, I will stick to it.

Depending on the type of calculation one wants to do, s can be regarded as the Laplace variable (outside the scope of this article), as Oliver Heaviside's differentiation to time operator, or as $j\omega$, where ω is the radian frequency ($\omega=2\pi f$) and j is the imaginary unit number ($j^2=-1$).

As an example, suppose the transfer function of a circuit is:

$$H(s) = \frac{V_{\text{out}}}{V_{\text{in}}} = K \frac{a_2 s^2 + a_1 s + 1}{b_2 s^2 + b_1 s + 1}$$

Interpreting s as a differentiation-to-time operator, this means that the relation between the input and output voltage is given by this differential equation:

$$b_2 \frac{d^2 v_{\text{out}}}{dt^2} + b_1 \frac{dv_{\text{out}}}{dt} + v_{\text{out}} = K \left(a_2 \frac{d^2 v_{\text{in}}}{dt^2} + a_1 \frac{dv_{\text{in}}}{dt} + v_{\text{in}} \right)$$

If the circuit has more inputs, this applies when the other input voltages and currents are set to zero.

Using complex numbers, calculating the output signal becomes relatively simple when the circuit does not oscillate and when the input signal is a stationary sine or cosine wave (that is, a sine or cosine that has been there long enough for initial transients to damp out). A cosine equals the sum of two complex exponential functions:

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

As the network is linear, we may use the superposition principle. That is, we may calculate the response to each complex exponential signal independently and simply add the results.

The time derivative of a complex exponential signal is:

$$\frac{de^{j\omega t}}{dt} = j\omega e^{j\omega t}$$

Hence, when

$$v_{\text{in}} = \frac{1}{2} e^{j\omega t}$$

we get

$$b_2 \frac{d^2 v_{\text{out}}}{dt^2} + b_1 \frac{dv_{\text{out}}}{dt} + v_{\text{out}} = K(a_2(j\omega)^2 + a_1 j\omega + 1) \frac{1}{2} e^{j\omega t}$$

This equation is satisfied when the output signal is also a complex exponential signal of the same frequency, but multiplied with some complex multiplication factor. That is, assume that

$$v_{\text{out}} = X e^{j\omega t}$$

This results in

$$(b_2(j\omega)^2 + b_1 j\omega + 1) X e^{j\omega t} = K(a_2(j\omega)^2 + a_1 j\omega + 1) \frac{1}{2} e^{j\omega t}$$

$$X = \frac{1}{2} \cdot \frac{K(a_2(j\omega)^2 + a_1 j\omega + 1)}{b_2(j\omega)^2 + b_1 j\omega + 1} = \frac{1}{2} H(j\omega)$$

and

$$\frac{v_{\text{out}}}{v_{\text{in}}} = \frac{X}{\frac{1}{2}} = H(j\omega)$$

So when you substitute $s=j\omega$, the transfer function turns into a complex-valued gain factor for complex exponential input signals. Note that $H(j\omega)$ can also be written in a polar form:

$$H(j\omega) = |H(j\omega)| e^{j\phi}$$

with

$$\phi = \arctan(\text{Im}(H(j\omega)) / \text{Re}(H(j\omega))) + k\pi$$

when the real part of $H(j\omega)$ is not zero and where k is an integer. The factor $|H(j\omega)|$ represents the actual gain, while the factor $e^{j\phi}$ just gives a phase shift of ϕ .

In the end, we are interested in the response to the real-valued cosine wave rather than to the complex exponential waveform. Hence, when

$$v_{\text{in}} = \cos(\omega t) = \frac{1}{2} e^{j\omega t} + \frac{1}{2} e^{-j\omega t}$$

then

$$v_{\text{out}} = \frac{1}{2} e^{j\omega t} H(j\omega) + \frac{1}{2} e^{-j\omega t} H(-j\omega)$$

It can be shown that $H(j\omega)$ and $H(-j\omega)$ must have equal real parts and opposite imaginary parts for any filter that produces a real-valued output signal for each real-valued input signal. This results in equal magnitudes, but opposite phases for $H(j\omega)$ and $H(-j\omega)$. Hence,

$$v_{\text{out}} = \frac{1}{2} e^{j\omega t} |H(j\omega)| e^{j\phi} + \frac{1}{2} e^{-j\omega t} |H(j\omega)| e^{-j\phi} = |H(j\omega)| \frac{1}{2} (e^{j(\omega t + \phi)} + e^{-j(\omega t + \phi)}) = |H(j\omega)| \cos(\omega t + \phi)$$

That is, also for a normal cosine wave $|H(j\omega)|$ represents the gain, while the phase shift is ϕ .

2.3. Admittance and impedance of ideal capacitors and inductors

As mentioned, one of the interpretations of s is a differentiation-to-time operator. The current through an ideal linear capacitor is C times the derivative to time of the voltage across it, so in terms of s , the admittance of an ideal capacitor is $Y = sC$ and its impedance is $Z = 1/(sC)$.

Similarly, the voltage across an ideal linear inductor is L times the derivative to time of the current through it, so in terms of s , the impedance of an ideal inductor is $Z = sL$ and its admittance is $Y = 1/(sL)$.

2.4. Poles and zeros

The values of s for which the denominator of the transfer function is zero are called the poles of the transfer function. The values of s for which the numerator of the transfer function is zero are called the zeros. The number of poles of a lumped system can never exceed the number of energy-storing parts (such as ideal capacitors and inductors in an electrical network, or masses and springs in a mechanical system).

An interesting property is that all transfers of a system with multiple in- or outputs have the same poles, although in some of these transfers some poles may be unobservable because they are covered by zeros.

2.5. General method for calculating the transfer function of a given network

The transfer of any network consisting of impedances, independent sources, linear controlled sources and various other types of components can be calculated with a method called modified nodal analysis - although we normally only use it when we can't think of a simpler method. The procedure is as follows:

A. For a network with n nodes, number the nodes from 0 up to and including $n - 1$. Node 0 will be the reference node (also known as the datum), all voltages are with respect to node 0. The ground node is usually taken as node 0, although this is not necessary (in fact the term ground has no meaning in network theory).

B. For each of the $n - 1$ nodes that have a number different from 0, write down the nodal equation. This is an equation expressing how much current flows into and out of the node as a function of the node voltages. Of course all current that flows into a node also has to flow out of the node again (Kirchhoff's current law). For example, when node 3 is connected to node 2 via an ideal resistor with resistance R , to node 0 via an ideal capacitor with capacitance C and to an ideal current source that injects a current I into node 3, the nodal equation for node 3 will be

$$-\frac{1}{R}V_2 + \left(\frac{1}{R} + sC\right)V_3 = I$$

where sC is the admittance (reciprocal of the impedance) in the Laplace domain of an ideal capacitor with capacitance C . There is no $-sCV_0$ term anywhere because the voltage on node 0 is zero by definition.

C. For each voltage source, you have to introduce additional equations. For example, when there is an independent voltage source connected between nodes 5 and 0, the extra equation states that the voltage on node 5 is simply the voltage of the independent voltage source.

D. The last step is to solve the unknown node voltages from the resulting set of equations. With linear equations there is a simple trick for this, which is known as Gaussian elimination. The trick is best illustrated by an example. Suppose you have the following set of equations:

$$\begin{aligned} 3a + 2b + 3c &= 0 \\ 2a + 5b + 8c &= 0 \\ a + 6b + 4c &= 12 \end{aligned}$$

By subtracting two thirds of the first equation from the second equation and one third of the first equation from the third equation, you can eliminate a from the second and third equations:

$$\begin{aligned} 3a + 2b + 3c &= 0 \\ \frac{11}{3}b + 6c &= 0 \\ \frac{16}{3}b + 3c &= 12 \end{aligned}$$

By subtracting 16/11 times the second equation from the third equation, b is eliminated from the third equation and we get:

$$\begin{aligned} 3a + 2b + 3c &= 0 \\ \frac{11}{3}b + 6c &= 0 \\ -\frac{63}{11}c &= 12 \end{aligned}$$

The third equation now has only one unknown left (c) and is quite easily solved. Once the third equation is solved, you can substitute the result in the second equation, which then also

becomes a simple equation with only one unknown (b). Substituting b and c in the first equation then results in a simple equation with only the unknown a .

$$\begin{aligned} -\frac{63}{11}c &= 12 \Leftrightarrow c = -\frac{11 \cdot 12}{63} = -\frac{132}{63} = -\frac{44}{21} \\ \frac{11}{3}b + 6 \cdot \left(-\frac{44}{21}\right) &= 0 \Leftrightarrow b = \frac{3}{11} \cdot 6 \cdot \frac{44}{21} = \frac{3 \cdot 6 \cdot 4}{21} = \frac{24}{7} \\ 3a + 2 \cdot \frac{24}{7} + 3 \cdot \left(-\frac{44}{21}\right) &= 0 \Leftrightarrow a = \frac{1}{3} \left(-2 \cdot \frac{24}{7} + 3 \cdot \frac{44}{21}\right) = \frac{1}{3} \left(-\frac{4}{7}\right) = -\frac{4}{21} \end{aligned}$$

It is clear that the amount of work increases rapidly with increasing number of equations (though not as rapidly as with other algorithms for solving systems of linear equations). As the number of equations in modified nodal analysis depends on the number of nodes, it is advisable to eliminate any node that can easily be eliminated before doing the analysis. In fact, it is advisable to avoid modified nodal analysis altogether when methods that involve less work can easily be found.

3. Transfer of the inductively loaded triode with cathode decoupling

Trying to avoid modified nodal analysis, the zeros will first be determined by inspection (section 3.1), then the impedance looking into the cathode of the triode will be calculated (section 3.2), then this impedance is combined with the cathode decoupling to find the poles, or rather their natural frequency and quality factor (section 3.3).

3.1. Zeros

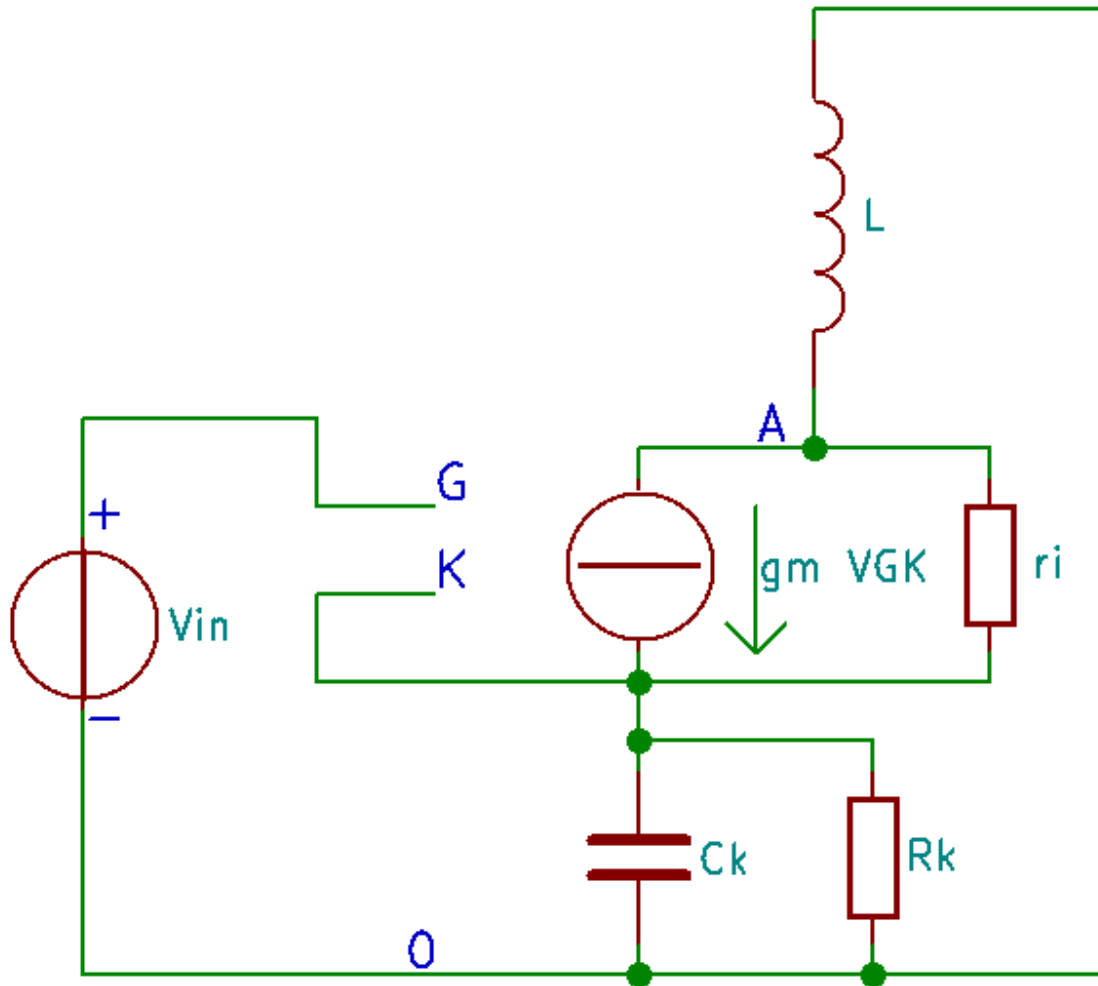


Figure 1: Small-signal network model for an inductively loaded triode

Figure 1 shows a linear network model for a triode with capacitively decoupled cathode resistor and purely inductive load. Instead of numbering the nodes, I have named the node connected to the anode A, the node connected to the cathode K and the node connected to the grid G. The transfer of interest is the transfer from voltage source V_{in} to the voltage at node A (V_A). There are two values of s where this transfer is zero:

At $s = 0$, the inductor is a short-circuit, so the transfer must be 0.

At $s = -\frac{1}{R_k C_k}$, the admittance sC_k of C_k is the opposite of the admittance $1/R_k$ of R_k . The admittance of their parallel connection is the sum of their admittances and is therefore zero (impedance infinite). There is therefore no path from the cathode to ground (node 0) for this value of s , leading to a zero transfer.

3.2. Impedance "looking" into the cathode

The left part of Figure 2 shows a model for an inductively loaded triode that gets driven at its cathode by an independent voltage source. This model will be used to determine the admittance looking into the cathode, the impedance looking into the cathode is then by definition its reciprocal.

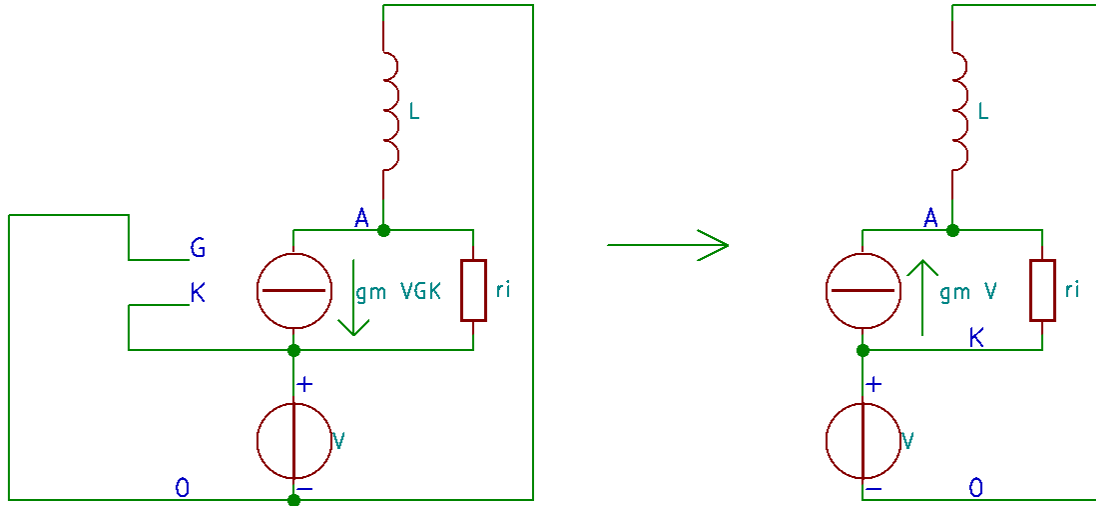


Figure 2: Network model for calculating the admittance "looking" into the cathode, and a simplified version. Apologies for the double node name in the left circuit.

It is clear that the voltage-controlled current source will always conduct a current of $g_m V$ in the direction opposite to the arrow, as the independent voltage source with voltage V is connected straight to its controlling terminals, but with the polarity swapped. Using the substitution theorem, it can be replaced with an independent current source that conducts the exact same current. This will not change any of the voltages and currents in the network. See the right side of Figure 2.

As the circuit is linear, we can now use superposition: calculate the response to each of the two independent sources separately (with the other temporarily set to zero) and add them.

With the voltage source set to zero and the current source to $g_m V$, the equation for current division with impedances will show that there is a current $g_m V r_i / (sL + r_i)$ flowing through L . This same current has to flow into the negative terminal and out of the positive terminal of the voltage source, as it has no other way to complete the current loop.

With the current source switched off, the circuit simplifies into a voltage source driving the series connection of r_i and L . Hence, the current out of the voltage source will now be $V / (sL + r_i)$.

Adding these terms, the total current out of voltage source V is

$$I = V \frac{g_m r_i + 1}{sL + r_i}$$

so the admittance looking into the cathode is

$$Y = \frac{g_m r_i + 1}{sL + r_i}$$

and the impedance looking into the cathode is

$$Z = \frac{sL + r_i}{g_m r_i + 1} = s \frac{L}{g_m r_i + 1} + \frac{r_i}{g_m r_i + 1}$$

This is the impedance of the series connection of an inductance $\frac{L}{g_m r_i + 1}$ and a resistance

$$\frac{r_i}{g_m r_i + 1} \approx \frac{1}{g_m}$$

where the approximation holds for $g_m r_i \gg 1$. The product $g_m r_i$ is also known as the voltage gain μ of the triode.

It should be noted that if, instead of an inductance L , we would have used some arbitrary impedance Z_A as load for the anode, the impedance looking into the cathode would have been

$$\frac{Z_A}{g_m r_i + 1} \text{ in series with } \frac{r_i}{g_m r_i + 1} \approx \frac{1}{g_m}.$$

3.3. Poles

As mentioned, all transfers of a circuit with multiple inputs and outputs have the same poles, although some may be unobservable in some transfers. Calculating the poles can therefore be simplified by finding a combination of input and output for which the poles are easy to calculate and in which all poles are observable.

For example, one can add a current source between nodes 0 and K, and observe the voltage between these nodes when all independent sources except this current source are set to 0. The impedance between K and 0 is then another transfer function of the circuit. The normal transfer is not affected, because when the new current source is set to 0, it does not affect the behaviour. See the left side of Figure 3.

The circuit can have two poles, as there are two energy-storing parts, C_k and L . If two poles are found when calculating the impedance between nodes K and 0, those are also the poles of the normal signal transfer.

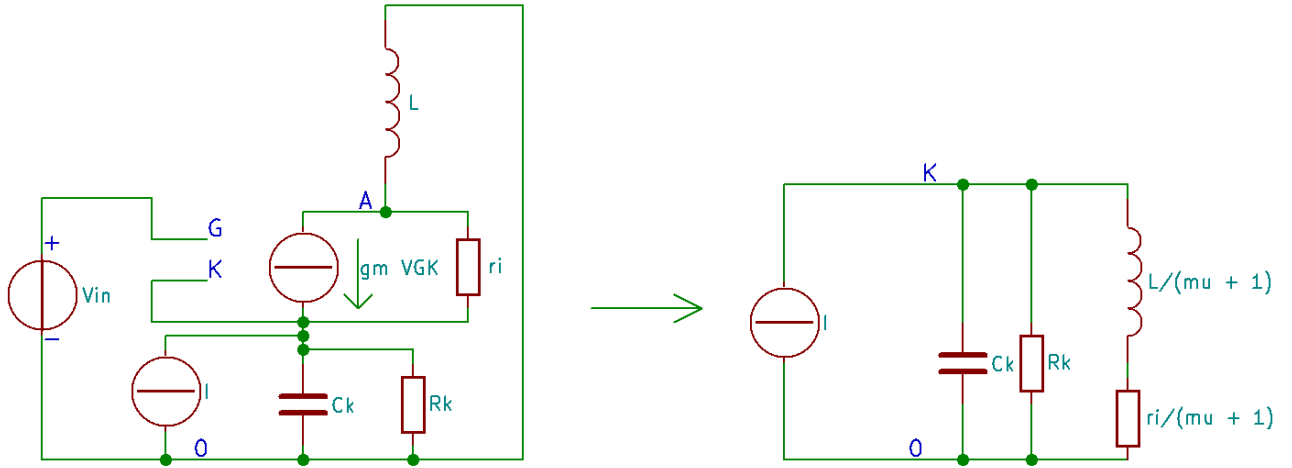


Figure 3: Finding the poles by calculating the impedance at the cathode including the cathode resistor and cathode decoupling

The impedance looking into the cathode was calculated in section 3.2. This in parallel with R_k and C_k gives the total impedance between nodes K and 0. This is shown schematically in the right part of Figure 3.

It would be easy if there were only one damping resistor, all textbook texts about RLC parallel and series circuits would then apply directly. Unfortunately, there are two of them.

The admittance of the circuit on the right part of Figure 3 is

$$Y = \frac{1}{R_k} + sC_k + \frac{\frac{\mu+1}{r_i}}{s\frac{L}{r_i} + 1}$$

which can be rewritten as

$$Y = \frac{\frac{1}{R_k} s\frac{L}{r_i} + \frac{1}{R_k} + s^2 C_k \frac{L}{r_i} + sC_k + \frac{\mu+1}{r_i}}{s\frac{L}{r_i} + 1}$$

so the impedance is

$$Z = \frac{s\frac{L}{r_i} + 1}{\frac{1}{R_k} s\frac{L}{r_i} + \frac{1}{R_k} + s^2 C_k \frac{L}{r_i} + sC_k + \frac{\mu+1}{r_i}}$$

One can now calculate the poles by equating the denominator to zero and solving for s , but it is actually more useful to bring the denominator into the standard form that is always used for second-order transfers:

$$\frac{1}{\omega_n^2} s^2 + \frac{1}{Q \omega_n} s + 1$$

because the quality factor Q is then a measure for the shape of the response, while the so-called natural frequency $f_n = \frac{\omega_n}{2\pi}$ is then a measure for the cut-off frequency.

Dividing all factors by $\frac{1}{R_k} + \frac{\mu+1}{r_i}$, which is exactly the same as multiplying all factors by the parallel value of R_k and $\frac{r_i}{\mu+1}$, brings the denominator polynomial in the desired form without changing the pole positions.

Defining that R_p is the parallel value of R_k and $\frac{r_i}{\mu+1}$, so

$$R_p = \frac{1}{\frac{1}{R_k} + \frac{\mu+1}{r_i}}$$

the result is

$$s^2 C_k \frac{L}{r_i} R_p + s \left(C_k + \frac{L}{R_k r_i} \right) R_p + 1$$

Hence,

$$\omega_n = \sqrt{\frac{r_i}{C_k L R_p}}$$

and

$$Q = \frac{1}{\left(C_k + \frac{L}{R_k r_i} \right) R_p \sqrt{\frac{r_i}{C_k L R_p}}} = \frac{1}{\left(C_k + \frac{L}{R_k r_i} \right) \sqrt{\frac{r_i R_p}{C_k L}}}$$

4. Obtaining a reasonably or even very flat transfer

If both zeros were at 0, the transfer function would be a standard second-order high-pass transfer. Bringing the two poles into Butterworth locations would then suffice to get a maximally-flat magnitude response. The poles are in second-order Butterworth locations when $Q = \frac{1}{2} \sqrt{2} \approx 0.7071068$. With the zeros at 0 and the poles in second-order Butterworth locations, the half-power point (-3.01 dB point) would be the natural frequency f_n .

However, one zero is not in the origin, but at $s = -\frac{1}{R_k C_k}$. This results in a peak in the response even when the poles are in second-order Butterworth locations. The peak is not large when the corner frequency of the zero lies well below f_n , though. Examples:

When $\frac{1}{R_k C_k} = 0.55 \omega_n$, the peak is +0.0961 dB at $2.6 f_n$.

When $\frac{1}{R_k C_k} = \frac{2}{3} \omega_n$, the peak is +0.20012 dB at $2.16 f_n$.

The whole response changes somewhat, with a -3.01 dB point that is lower than f_n :

When $\frac{1}{R_k C_k} = 0.55 \omega_n$:

Response peak: +0.0961 dB at $2.6 f_n$,

Gain drops to -0.1 dB at $1.64 f_n$,

-0.2 dB at $1.53 f_n$,

-0.5 dB at $1.33 f_n$,

-1 dB at $1.16 f_n$,

-1.86 dB at f_n ,

-3.01 dB at $0.86084 f_n$,

-10 dB at $0.453 f_n$,

-20 dB at $0.173 f_n$.

When $\frac{1}{R_k C_k} = \frac{2}{3} \omega_n$:

Response peak: +0.20012 dB at $2.16 f_n$,

-0.2 dB at $1.34 f_n$,

-0.5 dB at $1.21 f_n$,

-1 dB at $1.07 f_n$,

-1.41 dB at f_n ,

-3.01 dB at $0.80588 f_n$,

-10 dB at $0.41 f_n$,

-20 dB at $0.146 f_n$.

With both zeros at 0 (ideal second-order Butterworth):

No peak at all,

-0.1 dB at $2.55 f_n$,

-0.2 dB at $2.14 f_n$,

-0.5 dB at $1.69 f_n$,

-1 dB at $1.4 f_n$,

-3.01 dB at f_n ,

-10 dB at $0.577 f_n$,

-20 dB at $0.316 f_n$.

If there is a first-order high-pass somewhere else in the signal path, for example an AC coupling capacitor somewhere, one can make the combined response a perfect second-order Butterworth high-pass response by putting the pole of this extra first-order high-pass on top of

the zero. In the case of an AC coupling, the RC time of the AC coupling then has to be made equal to $R_k C_k$.

5. Numerical examples

The numerical examples are based on a 6C45 biased at 12.9 mA with a 124 Ω cathode resistor driving a 100 H inductive load (which is actually an interstage transformer - its datasheet says 130 H, but the DC current is higher than it should be).

To get reasonably damped poles (say, Q not much above $\sqrt{2}/2$), you either have to make the decoupling capacitor so large that the LC circuit is damped adequately by the series resistance $r_i/(\mu + 1)$ of the inductance you see at the cathode, or so small that it is damped adequately by the parallel resistor R_k . For the first option, 1000 μF is already on the low side, for the second 47 μF or 68 μF works well.

That is, I calculated the characteristic polynomial assuming that the primary inductance of the interstage transformer is 100 H due to the somewhat too large current (wild guess), that the voltage gain of the valve is 52 (datasheet value) and that the internal resistance is either 1684.91525 Ω (extrapolated from datasheet value using Child's law) or 2500 Ω (value from post #1 of tonescout's thread). For 47 μF , I find either $f_n \approx 18.94389481$ Hz and $Q \approx 0.6316674341$, or $f_n \approx 19.85691976$ Hz and $Q \approx 0.6346589069$.

With 47 μF and 124 Ω , the second zero lies at -171.5854496 rad/s, corner frequency of the zero 27.30867246 Hz. This is higher than f_n , so it will cause a bump when it is not compensated. If, for example, the output transformer causes a first-order roll-off below a corner frequency close to this value, you actually get a flatter response with 47 μF than with a large capacitance.

With 68 μF , the values are either $f_n \approx 15.74939031$ Hz and $Q \approx 0.7306017495$, or $f_n \approx 16.33365341$ Hz and $Q \approx 0.7300749122$. The corner frequency of the zero becomes 18.87511185 Hz, again higher than f_n .

With 1000 μF , I find either $f_n \approx 4.106938529$ Hz and $Q \approx 1.035762987$, or $f_n \approx 4.30487762$ Hz and $Q \approx 0.8180474716$. The second zero lies at -8.064516129 rad/s, corner frequency of the zero 1.283507606 Hz. In this case, the effect of the second zero not being at 0 will be negligible, but you will get a small bump anyway because the Q is somewhat too high. It occurs at deep subsonic frequencies, though.