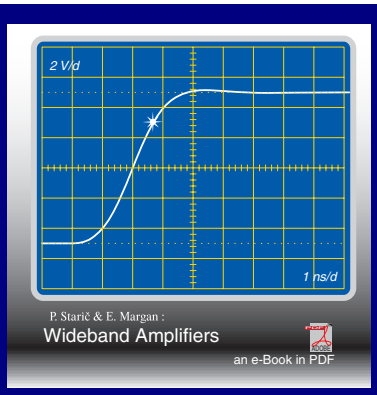


Peter Starič and Erik Margan

Wideband Amplifiers

Springer
http://www.springeronline.com/0-387-28340-4

Demo Excerpts

Foreword to Wideband Amplifiers

With the exception of the tragedy on September 11, the year 2001 was relatively normal and uneventful — remember, this should have been the year of the *Clarke's* and *Kubrick's* Space Odyssey, mission to Jupiter; it should have been the year of the infamous HAL-9000 computer.

Today, the Personal Computer is as ubiquitous and omnipresent as was HAL on the Discovery spaceship. And the rate of technology development and market growth in electronics industry still follows the famous 'Moore Law', almost four decades after it has been first formulated: in 1965, *Gordon Moore* of Intel Corporation predicted the doubling of the number of transistors on a chip every 2 years, corrected to 18 months in 1967; at that time, the landing on the Moon was in full preparation.

Curiously enough, today no one cares to go to the Moon again, let alone Jupiter. And, in spite of all the effort in digital engineering, we still do not have anything close to 0.1% of HAL's capacity (fortunately?!). Whilst there are many research labs striving to put artificial intelligence into a computer, there are also rumors that this has already happened (with Windows-95, of course!).

In the early 1990s it was felt that digital electronics will eventually render analog systems obsolete. This never happened. Not only is the analog sector vital as ever, the job market demands are expanding in all fields, from high-speed measurement instrumentation and data acquisition, telecommunications and radio frequency engineering, high-quality audio and video, to grounding and shielding, electromagnetic interference suppression and low-noise printed circuit board design, to name a few. And it looks like this demand will be going on for decades to come.

But while the proliferation of digital systems attracted a relatively high number of hardware and software engineers, analog engineers are still rare birds. So, for creative young people, who want to push the envelope, there are lots of opportunities in the analog field.

However, analog electronics did not earn its "Black-Magic Art" attribute in vain. If you have ever experienced the problems and frustrations from circuits found in too many 'cook-books' and 'sure-working schemes' in electronics magazines, and if you became tired of performing exorcism on every circuit you build, then it is probably the time to try it in a different way: in our own experience, the HARD way of doing the correct math first often turns out to be the EASY way!

Here is the book **"Wideband Amplifiers"**. The book was intended to serve both as a design manual to more experienced engineers and as a good learning guide to beginners. It should help you to improve your analog design, making better and faster amplifier circuits, especially if time-domain performance is of major concern. We have strieved to provide the complete math for every design stage. And, to make learning a joyful experience, we explain the derivation of important math relations from a design engineer point of view, in an intuitive and self-evident manner (rigorous mathematicians might not like our approach). We have included many practical applications, schematics, performance plots, and a number of computer routines.

However, as with any interesting subject, the greatest problem was never what to include, but rather what to leave out!

In the foreword of his popular book "A Brief History of Time", *Steven Hawking* wrote that his publisher warned him not to include any math, since the number of readers would be halved by each formula. So he included the $E = m\,c^2$ and bravely cut one half of the world population out.

We went further: there are some 220 formulae in Part 1 only. By estimating the current world population to some 6×10^9, of which 0.01% could be electronics engineers and assuming an average lifetime interest in the subject of, say, 30 years, if the publisher's rule holds, there ought to be one reader of our book once every:

$$2^{220}/(6\times10^{9}\times10^{-4}\times30\times356\times24\times3600)\approx3\times10^{51}\ \text{seconds}$$

or something like $6.6\times10^{33}\times$ the total age of the Universe!

Now, whatever you might think of it, this book is **not** about math! It is about getting your design to run right first time! Be warned, though, that it will be not enough to just read the book. To have any value, a theory must be put into practice. Although there is no theoretical substitute for hands-on experience, this book should help you to significantly shorten the trial-and-error phase.

We hope that by studying this book thoroughly you will find yourself at the **beginning of a wonderful journey**!

Peter Starič and Erik Margan,

Ljubljana, June 2003

Important Note:

We would like to reassure the Concerned Environmentalists that during the writing of this book, no animal or plant had suffered any harm whatsoever, either in direct or indirect form (excluding the authors, one computer 'mouse' and countless computation 'bugs'!).

Release Notes (Aug. 2005)

The manuscript of **Wideband Amplifiers** appeared first in spring of 1988.

Since then the text has been extended and revised several times, with some minor errors corrected and figures redrawn, in particular in Part 2 where inductive peaking networks are analyzed. Some of the topics have been updated to reflect the latest developments in the field, mainly in Part 5, dealing with modern high speed circuits. The Part 6, where a number of computer algorithms are developed, and Part 7, containing several algorithm application examples, have also been brought up to date.

The book also comes in the Adobe Portable Document Format (PDF), readable by the Adobe Acrobat™ Reader program (the latest version can be downloaded free of charge from http://www.adobe.com/products/Acrobat/readstep2/).

One of the advantages offered by the PDF format and the Acrobat Reader program are numerous links (blue underlined text) , which enable easy access to related topics by pointing the 'mouse' cursor on the link and clicking the left 'mouse' button. Returning to the original reading position is possible by clicking the right 'mouse' button and selecting 'Go Back' from the pop up menu (see the AR Help menu). There are also numerous highlights (green underlined text) relating to the content within the same page.

The cross-file links (red underlined text) relate to the contents of other parts, contained in different PDF files, which open by clicking the link in the same way. The Internet and World Wide Web links are in violet (dark magenta) and are accessed by opening the default browser installed on your computer. Likewise, the e-mail links open your default e-mail program.

The writing of the text and the math formating was done by using a program which, in our opinion, is the best for this job: 𝔼𝕏ℙ ™ the Scientific Word Processor, version 5.0 (made by Simon L. Smith, see http://www.expswp.com/).

The computer algorithms, the development of which we present in Part 6 and 7, are intended as tools for the process of amplifier design and analysis. Written for Matlab™ the Language of Technical Computing (made by The MathWorks, Inc., http://www.mathworks.com/), the algorithms have all been revised to conform with the newer version of Matlab (version 5.3 For Students), but still retaining downward compatibility (to version 1) as much as possible. We have used Matlab to check all the calculations and draw most of the figures. Before importing them into 𝔼𝕏ℙ the figures were finalized and converted to vector graphics format by using the Adobe Illustrator™ version 8 (see http://www.adobe.com/products/Illustrator/).

All circuit designs were checked using Micro-CAP ™, the Microcomputer Circuit Analysis Program, v. 5 (Spectrum Software, http://www.spectrum-soft.com/). Many circuits have also been checked in practice; unfortunately, in spite of all our effort, we could not include hyperlinks pointing to that ...

Peter Starič and Erik Margan

Part 1: Laplace Transform

Part 2: Inductive Peaking Circuits

Part 3: Wideband Amplifier Stages with Semiconductor Devices

Part 4: Cascading of Amplifier Stages, Selection of Poles

Part 5: System Synthesis And Integration

Part 6: Computer Algorithms For System Analysis And Synthesys

Part 7: Algorithm Application Examples

1.5 Examples of Direct Laplace Transform

Now let us put our new tools to use and calculate the $\mathcal{L}$ transform of several simple functions. The results may also be used for the $\mathcal{L}^{-1}$ transform and the reader is encouraged to learn the most basic of them by heart, because they are used extensively in the other parts of the book and, of course, in the analysis of the most common electronics circuits.

1.5.1 Example 1

Most of our calculations will deal with the step response of a network. To do so our excitation function will be a simple unit step $h(t)$, or the *Heaviside* function (after *Oliver Heaviside*, 1850–1925) as is shown in Fig. 1.5.1. This function is defined as:

$$f(t) = h(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ 1 & \text{for} \quad t > 0 \end{cases}$$

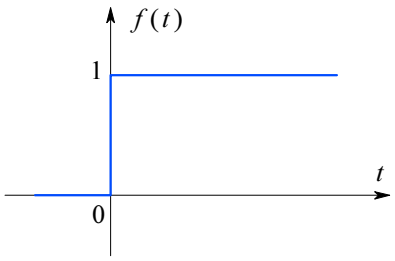


Fig. 1.5.1: Unit step function.

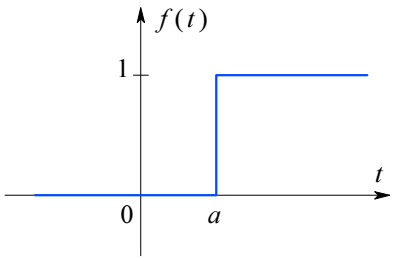


Fig. 1.5.2: Unit step function starting at $t = a$.

As we agreed in the previous section, $f(t) = 0$ for $t < 0$ for all the following functions, and we will not repeat this statement in further examples. At the same time let us mention that for our calculations of $\mathcal{L}$ transform it is not important what is the actual value of $f(0)$, providing it is finite [Ref. 1.3].

The $\mathcal{L}$ transform for the unit step function $f(t) = h(t)$ is:

$$F(s) = \mathcal{L}\{f(t)\} = \int_0^\infty [1]\, \mathrm{e}^{-st}\, dt = \frac{1}{-s}\, \mathrm{e}^{-st} \Big|_{t=0}^{t=\infty} = \frac{1}{s} \tag{1.5.1}$$

1.5.2 Example 2

The function is the same as in Example 1, except that the step does not start at $t = 0$ but at $t = a > 0$ (Fig. 1.5.2):

$$f(t) = \begin{cases} 0 & \text{for} \quad t < a \\ 1 & \text{for} \quad t > a \end{cases}$$

Solution:

$$F(s) = \int_a^\infty [1]\, \mathrm{e}^{-st}\, dt = \frac{1}{-s}\, \mathrm{e}^{-st} \Big|_{t=a}^{t=\infty} = \frac{1}{s}\, \mathrm{e}^{-as} \tag{1.5.2}$$

1.5.3 Example 3

The exponential decay function is shown in Fig. 1.5.3; its mathematical expression:

$$f(t) = \mathrm{e}^{-\sigma_1 t}$$

is defined for $t > 0$, as agreed, and σ_1 is a constant.
Solution:

$$F(s) = \int_0^\infty \mathrm{e}^{-\sigma_1 t}\,\mathrm{e}^{-st}\,dt = \int_0^\infty \mathrm{e}^{-(\sigma_1+s)t}\,dt$$

$$= \frac{-1}{\sigma_1 + s}\,\mathrm{e}^{-(\sigma_1+s)t}\Big|_{t=0}^{t=\infty} = \frac{1}{\sigma_1 + s} \tag{1.5.3}$$

Later, we shall meet this and the following function and also their product very often.

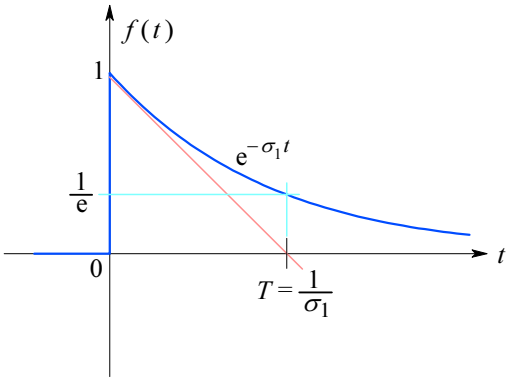


Fig. 1.5.3: Exponential function.

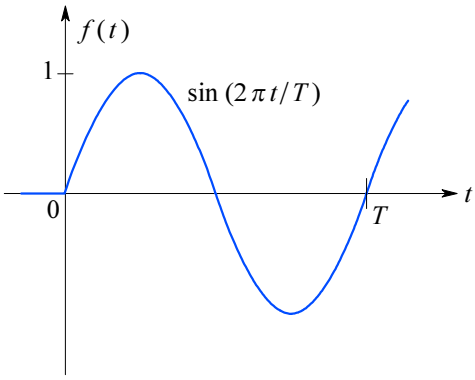


Fig. 1.5.4: Sinusoidal function.

1.5.4 Example 4

We have a sinusoidal function as in Fig. 1.5.4; its corresponding mathematical expression is:

$$f(t) = \sin\omega_1 t$$

where the constant $\omega_1 = 2\pi/T$.
Solution: its $\mathcal{L}$ transform is:

$$F(s) = \int_0^\infty (\sin\omega_1\, t)\,\mathrm{e}^{-st}\,dt \tag{1.5.4}$$

To integrate this function we substitute it using Euler's formula:

$$\sin\omega_1 t = \frac{1}{2\,j}\left(\mathrm{e}^{j\omega_1 t} - \mathrm{e}^{-j\omega_1 t}\right) \tag{1.5.5}$$

Then we have:

$$F(s) = \frac{1}{2\,j}\left(\int_0^\infty \mathrm{e}^{j\omega_1 t}\,\mathrm{e}^{-st}\,dt - \int_0^\infty \mathrm{e}^{-j\omega_1 t}\,\mathrm{e}^{-st}\,dt\right)$$

$$= \frac{1}{2\,j}\left[\int_0^\infty \mathrm{e}^{-(s-j\omega_1)t}\,dt - \int_0^\infty \mathrm{e}^{-(s+j\omega_1)t}\,dt\right] \tag{1.5.6}$$

The solution of this integral is, in a way, similar to that in the previous example:

$$F(s) = \frac{1}{2\,j}\left(\frac{1}{s - j\,\omega_1} - \frac{1}{s + j\,\omega_1}\right)$$

$$= \frac{1}{2\,j} \cdot \frac{s + j\,\omega_1 - s + j\,\omega_1}{s^2 + \omega_1^2} = \frac{\omega_1}{s^2 + \omega_1^2} \tag{1.5.7}$$

This is a typical function of a continuous wave (CW) sinusoidal oscillator, with a frequency ω_1.

1.5.5 Example 5

Here we have the cosine function as in Fig. 1.5.5, expressed as:

$$f(t) = \cos\omega_1\, t$$

Solution: the $\mathcal{L}$ transform of this function is calculated in a similar way as for the sine. According to Euler's formula:

$$\cos\omega_1\, t = \frac{1}{2}\left(\mathrm{e}^{j\,\omega_1 t} + \mathrm{e}^{-j\,\omega_1 t}\right) \tag{1.5.8}$$

Thus we obtain:

$$F(s) = \frac{1}{2}\left[\int_0^\infty \mathrm{e}^{-(s\,-j\,\omega_1)t}\,dt + \int_0^\infty \mathrm{e}^{-(s\,+j\,\omega_1)t}\,dt\right] = \frac{1}{2}\left(\frac{1}{s - j\,\omega_1} + \frac{1}{s + j\,\omega_1}\right)$$

$$= \frac{1}{2} \cdot \frac{s + j\,\omega_1 + s - j\,\omega_1}{s^2 + \omega_1^2} = \frac{s}{s^2 + \omega_1^2} \tag{1.5.9}$$

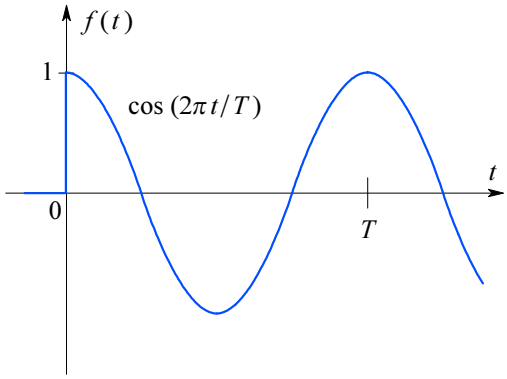


Fig. 1.5.5: Cosine function.

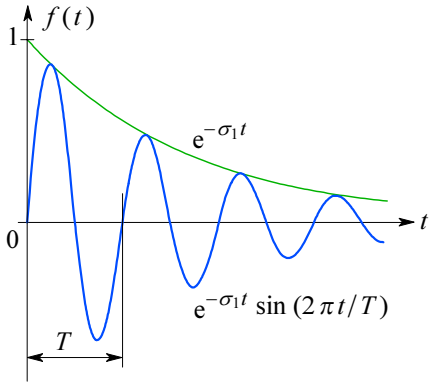


Fig. 1.5.6: Damped oscillations.

1.5.6 Example 6

In Fig. 1.5.6 we have a damped oscillation, expressed by the formula:

$$f(t) = \mathrm{e}^{-\sigma_1 t}\sin\omega_1\, t$$

Solution: we again substitute the sine function, according to Euler's formula:

$$F(s) = \frac{1}{2\,j}\int_0^\infty \mathrm{e}^{-(s+\sigma_1)t}\left(\mathrm{e}^{j\omega_1 t} - \mathrm{e}^{-j\omega_1 t}\right) dt$$

$$= \frac{1}{2\,j}\int_0^\infty \left[\mathrm{e}^{-(s+\sigma_1-j\omega_1)t} - \mathrm{e}^{-(s+\sigma_1+j\omega_1)t}\right] dt$$

$$= \frac{1}{2\,j}\left(\frac{1}{s+\sigma_1-j\omega_1} - \frac{1}{s+\sigma_1+j\omega_1}\right) = \frac{\omega_1}{(s+\sigma_1)^2+\omega_1^2} \qquad (1.5.10)$$

An interesting similarity is found if this formula is compared with the result of Example 4. There, for a CW we have in the denominator s^2 alone, whilst here, because the oscillations are damped, we have $(s+\sigma_1)^2$ instead, and σ_1 is the damping factor.

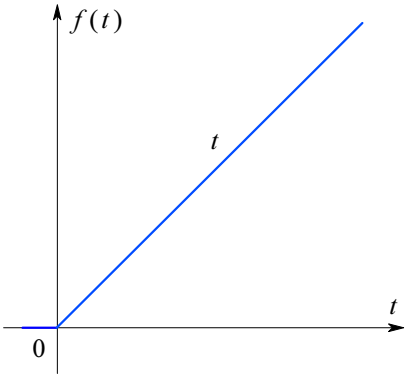


Fig. 1.5.7: Linear ramp $f(t) = t$.

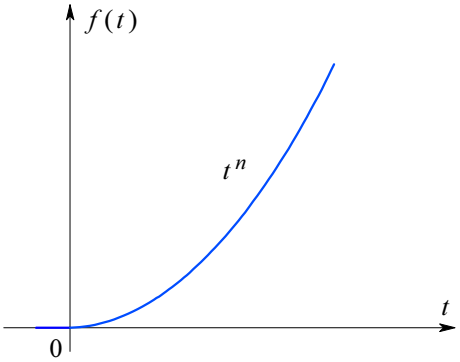


Fig. 1.5.8: Power function $f(t) = t^n$.

1.5.7 Example 7

A linear ramp, as shown in Fig. 1.5.7, is expressed as:

$$f(t) = t$$

Solution: we integrate by parts according to the known relation:

$$\boxed{\int u\,dv = u\,v - \int v\,du}$$

and we assign $t = u$ and $\mathrm{e}^{-s\,t}\,dt = dv$ to obtain:

$$F(s) = \int_0^\infty t\,\mathrm{e}^{-st}\,dt = \frac{t\,\mathrm{e}^{-st}}{-s}\bigg|_{t=0}^{t=\infty} + \frac{1}{s}\int_0^\infty \mathrm{e}^{-st}\,dt$$

$$= 0 - 0 - \frac{1}{s^2}\,\mathrm{e}^{-st}\bigg|_{t=0}^{t=\infty} = \frac{1}{s^2} \qquad (1.5.11)$$

1.5.8 Example 8

Fig. 1.5.8 displays a function which has a general analytical form:

$$f(t) = t^n$$

Solution: again we integrate by parts, decomposing the integrand $t^n\,\mathrm{e}^{-st}$ into:

$$u = t^n \qquad du = n\,t^{n-1}dt \qquad v = \frac{1}{-s}\,\mathrm{e}^{-st} \qquad dv = \mathrm{e}^{-st}\,dt$$

With these substitutions we obtain:

$$F(s) = \int_0^\infty t^n\,\mathrm{e}^{-st}\,dt = \frac{t^n\,\mathrm{e}^{-st}}{-s}\bigg|_{t=0}^{t=\infty} + \frac{n}{s}\int_0^\infty t^{n-1}\,\mathrm{e}^{-st}\,dt$$

$$= \frac{n}{s}\int_0^\infty t^{n-1}\,\mathrm{e}^{-st}\,dt \tag{1.5.12}$$

Again integrating by parts:

$$\frac{n}{s}\int_0^\infty t^{n-1}\,\mathrm{e}^{-st}\,dt = \frac{t^{n-1}\,\mathrm{e}^{-st}}{-s}\bigg|_{t=0}^{t=\infty} + \frac{n\,(n-1)}{s^2}\int_0^\infty t^{n-2}\,\mathrm{e}^{-st}dt$$

$$= \frac{n\,(n-1)}{s^2}\int_0^\infty t^{n-2}\,\mathrm{e}^{-st}\,dt \tag{1.5.13}$$

By repeating this procedure n times we finally arrive at:

$$F(s) = \int_0^\infty t^n\,\mathrm{e}^{-st}\,dt = \frac{n\,(n-1)(n-2)\cdots 3\cdot 2\cdot 1}{s^n}\int_0^\infty t^0\,\mathrm{e}^{-st}\,dt = \frac{n!}{s^{n+1}} \tag{1.5.14}$$

1.5.9 Example 9

The function shown in Fig. 1.5.9 corresponds to the expression:

$$f(t) = t\,\mathrm{e}^{-\sigma_1 t}$$

Solution: by integrating by parts we obtain:

$$F(s) = \int_0^\infty t\,\mathrm{e}^{-\sigma_1 t}\,\mathrm{e}^{-st}\,dt = \int_0^\infty t\,\mathrm{e}^{-(\sigma_1+s)t}\,dt = \frac{1}{(\sigma_1+s)^2} \tag{1.5.15}$$

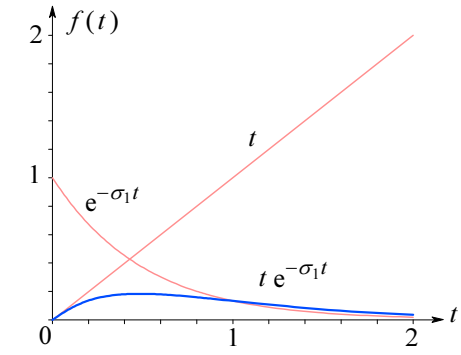
1.5.10 Example 10

Similarly to Example 9, except that here we have $t^{\,n}$, as in Fig. 1.5.10:

$$f(t) = t^{\,n}\,\mathrm{e}^{-\sigma_1 t}$$

Solution: we apply the procedure from Example 8 and Example 9:

$$F(s) = \int_0^\infty t^{\,n}\,\mathrm{e}^{-\sigma_1 t}\,\mathrm{e}^{-st}\,dt = \int_0^\infty t^{\,n}\,\mathrm{e}^{-(\sigma_1+s)t}\,dt = \frac{n!}{(\sigma_1+s)^{n+1}} \tag{1.5.16}$$



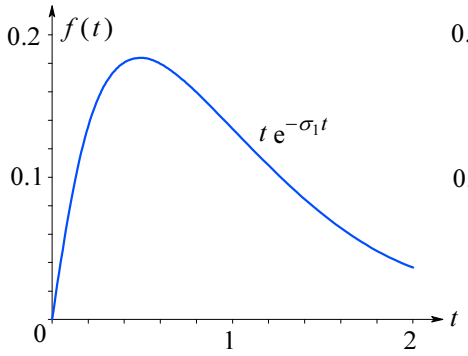
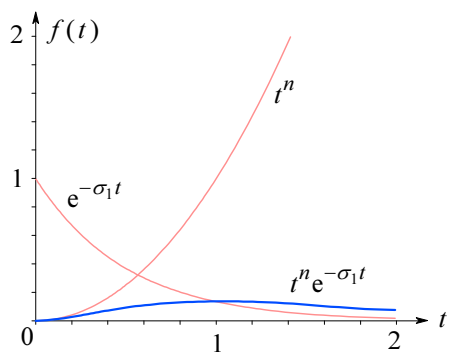


Fig. 1.5.9: Function $f(t) = t\,\mathrm{e}^{-\sigma_1 t}$.



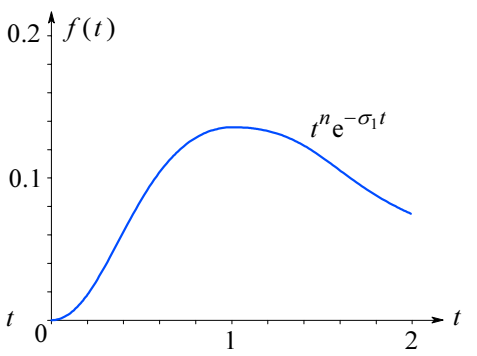


Fig. 1.5.10: Function $f(t) = t^n\,\mathrm{e}^{-\sigma_1 t}$.

These ten examples, which we frequently meet in practice, demonstrate that the calculation of an $\mathcal{L}$ transform is not difficult. Since the results derived are used often, we have collected them in Table 1.5.1.

Table 1.5.1: **Ten frequently met $\mathcal{L}$ transform examples**

No.	$f(t)$	$F(s)$		No.	$f(t)$	$F(s)$
1	$1 \quad (\text{for } t > 0)$	$\dfrac{1}{s}$		6	$\mathrm{e}^{-\sigma_1 t}\sin\omega_1 t$	$\dfrac{\omega_1}{(s^2+\sigma_1^2)+\omega_1^2}$
2	$1 \quad (\text{for } t > a)$	$\dfrac{1}{s}\,\mathrm{e}^{-a\,t}$		7	t	$\dfrac{1}{s^2}$
3	$\mathrm{e}^{-\sigma_1 t}$	$\dfrac{1}{\sigma_1 + s}$		8	t^n	$\dfrac{n!}{s^{n+1}}$
4	$\sin\omega_1 t$	$\dfrac{\omega_1}{s^2+\omega_1^2}$		9	$t\,\mathrm{e}^{-\sigma_1 t}$	$\dfrac{1}{(\sigma_1+s)^2}$
5	$\cos\omega_1 t$	$\dfrac{s}{s^2+\omega_1^2}$		10	$t^n\,\mathrm{e}^{-\sigma_1 t}$	$\dfrac{n!}{(\sigma_1+s)^{n+1}}$

2.1 The Principle of Inductive Peaking

A simple common base transistor amplifier is shown in Fig. 2.1.1. A current step source i_s is connected to the emitter; the time scale has its origin $t = 0$ at the current step transition time and is normalized to the system time constant, RC. The collector is loaded by a resistor R; in addition there is the collector–base capacitance C_{cb}, along with the unavoidable stray capacitance C_s and the load capacitance C_L in parallel. Their sum is denoted by C.

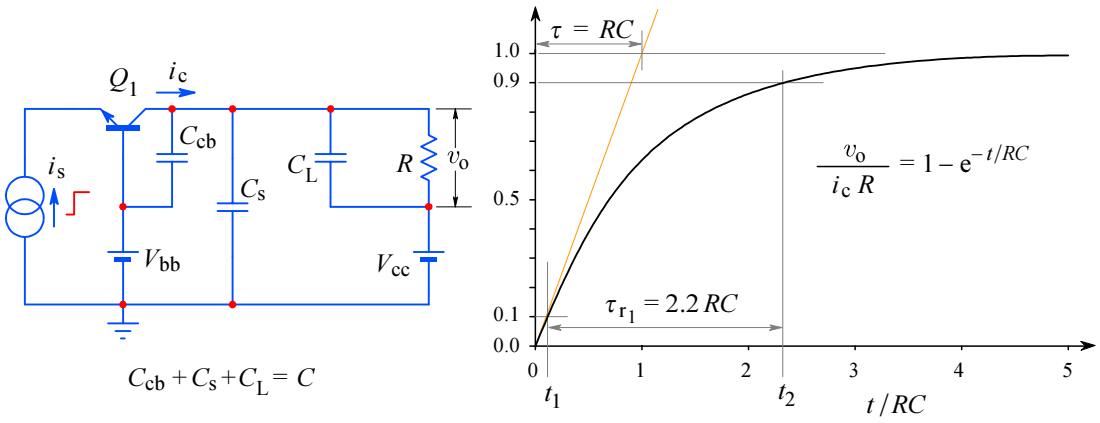


Fig. 2.1.1: A common base amplifier with RC load: the basic circuit and its step response.

Because of these capacitances, the output voltage v_o does not jump suddenly to the value $i_c R$, where i_c is the collector current. Instead this voltage rises exponentially according to the formula (see Part 1, Eq. 1.7.15):

$$v_o = i_c\, R \left(1 - e^{-t/RC}\right) \tag{2.1.1}$$

The time elapsed between 10 % and 90 % of the final output voltage value ($i_c R$), we call the *rise time*, τ_{r_1} (the index '1' indicates that it is the rise time of a single-pole circuit). We calculate it by inserting the 10 % and 90 % levels into the Eq. 2.1.1:

$$0.1\, i_c\, R = i_c\, R \left(1 - e^{-t_1/RC}\right) \quad \Rightarrow \quad t_1 = R\,C \ln 0.9 \tag{2.1.2}$$

Similarly for t_2:

$$0.9\, i_c\, R = i_c\, R \left(1 - e^{-t_2/RC}\right) \quad \Rightarrow \quad t_2 = R\,C \ln 0.1 \tag{2.1.3}$$

The rise time is the difference between these two instants:

$$\tau_{r_1} = t_2 - t_1 = R\,C \ln 0.9 - R\,C \ln 0.1 = R\,C \ln \frac{0.9}{0.1} = \boxed{2.2\, R\,C} \tag{2.1.4}$$

The value $2.2\,RC$ is the reference against which we shall compare the rise time of all other circuits in the following sections of the book.

Since in wideband amplifiers we strive to make the output voltage a replica of the input voltage (except for the amplitude), we want to reduce the rise time of the amplifier as much as possible. As the output voltage rises more current flows through R and less current remains to charge C. Obviously, we would achieve a shorter rise time if we could disconnect R in some way until C is charged to the desired level. To do so let us introduce a switch S between the capacitor C and the load resistor R. This switch is open at time $t = 0$, when the current step starts, but closes at time $t = RC$, as in Fig. 2.1.2. In this way we force all the available current to the capacitor, so it charges linearly to the voltage $i_c R$. When the capacitor has reached this voltage, the switch S is closed, routing all the current to the loading resistor R.

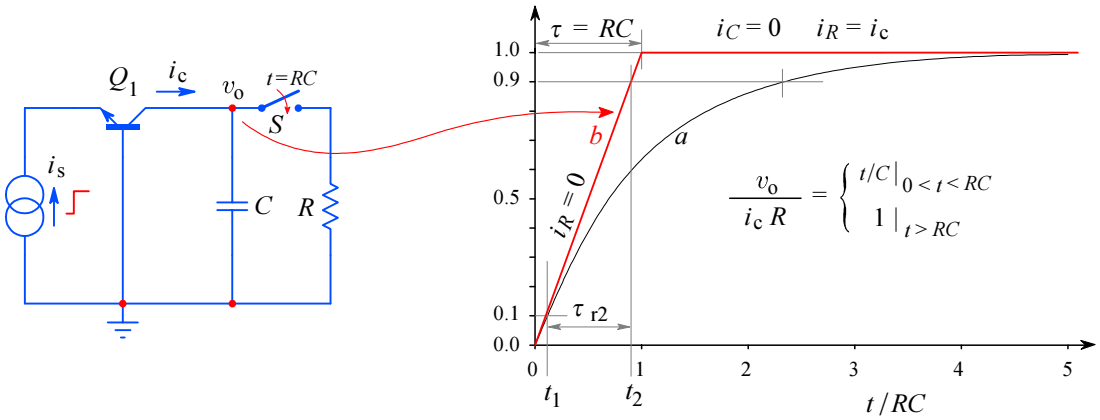


Fig. 2.1.2: A hypothetical ideal rise time circuit. The switch disconnects R from the circuit, so that all of i_c is available to charge C; but after a time $t = RC$ the switch is closed and all i_c flows through R. The resulting output voltage is shown in ***b***, compared with the exponential response in ***a***.

By comparing Fig. 2.1.1 with Fig. 2.1.2, we note a substantial decrease in rise time τ_{r_0}, which we calculate from the output voltage:

$$v_o = \frac{1}{C}\int_0^{\tau} i_c\, dt = \frac{i_c}{C}\, t\Big|_{t=0}^{t=\tau} = i_c R \tag{2.1.5}$$

where $\tau = RC$. Since the charging of the capacitor is linear, as shown in Fig. 2.1.2, the rise time is simply:

$$\tau_{r_0} = 0.9\, RC - 0.1\, RC = 0.8\, RC \tag{2.1.6}$$

In comparison with Fig. 2.1.1, where there was no switch, the improvement factor of the rise time is:

$$\eta_r = \frac{\tau_{r_1}}{\tau_{r_0}} = \frac{2.20\, RC}{0.8\, RC} = 2.75 \tag{2.1.7}$$

It is evident that the **rise time** (Eq. 2.1.6) **is independent of the actual value of the current i_c, but the maximum voltage $i_c R$** (Eq. 2.1.5) **is not**. On the other hand, the smaller the resistor R the smaller is the rise time. Clearly the introduction of the switch S would mean a great improvement. By using a more powerful transistor and a lower value resistor R we could (at least in principle) decrease the rise time at a will (provided that C

remains unchanged). Unfortunately, it is impossible to make a low on-resistance switch, functioning as in Fig. 2.1.2, which would also suitably follow the signal and automatically open and close in nanoseconds or even in microseconds. So it remains only a nice idea.

But instead of a switch we can insert an appropriate inductance L between the capacitor C and resistor R and so **partially** achieve the effect of the switch, as shown in Fig. 2.1.3. Since the current through an inductor can not change instantaneously, more current will be charging C, at least initially. The configuration of the RLC network allows us to take the output voltage either from the resistor R or from the capacitor C. In the first case we have a *series peaking network* whilst in the second case we speak of a *shunt peaking network*. Both types of peaking networks are used in wideband amplifiers.

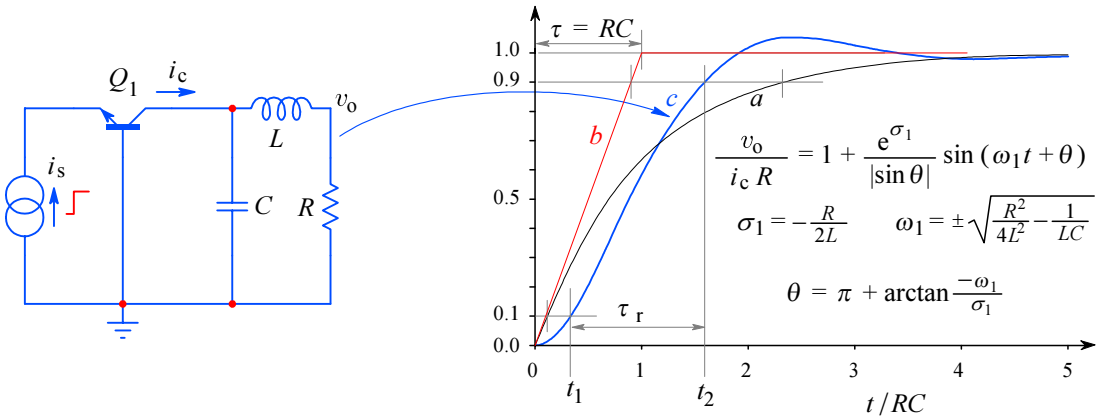


Fig. 2.1.3: A common base amplifier with the series peaking circuit. The output voltage v_o (curve c) is compared with the exponential response (a, $L = 0$) and the response using the ideal switch (b). If we were to take the output voltage from the capacitor C, we would have a shunt peaking circuit (see Sec. 2.7). We have already seen the complete derivation of the procedure for calculating the step response in Part 1, Sec. 1.14. However, the response optimization in accordance with different design criteria is shown in Sec. 2.2 for the series peaking circuit and in Sec. 2.7 for the shunt peaking circuit.

Fig. 2.1.3 is the simplest series peaking circuit. Later, when we discuss T-coil circuits, we shall not just achieve rise time improvements similar to that in Eq. 2.1.7, but in cases in which it is possible (usually it is) to split C into two parts, we shall obtain a substantially greater improvement.

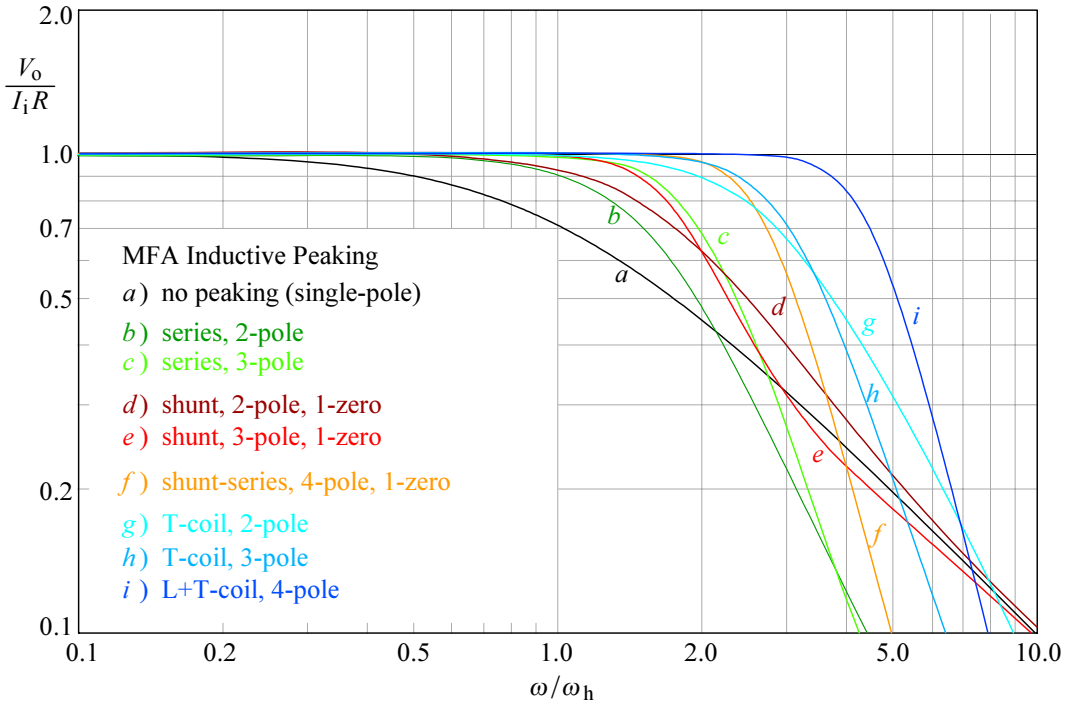


Fig. 2.10.1: MFA frequency responses of all the circuit configurations discussed. By far the 4-pole T-coil response ***i)*** has the largest bandwidth.

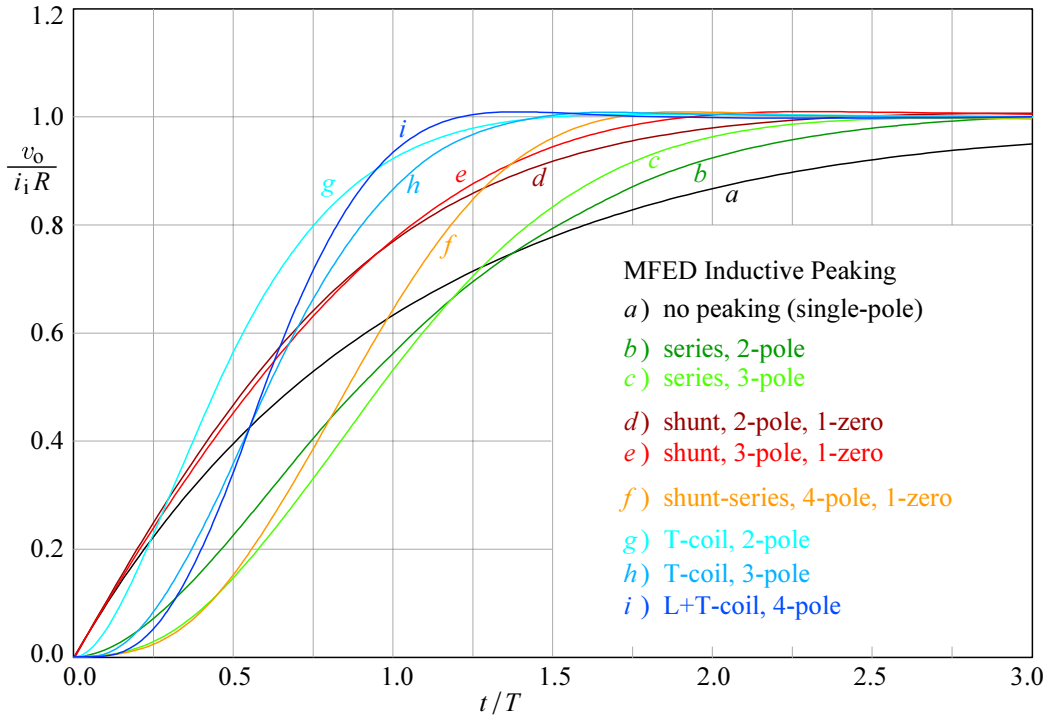


Fig. 2.10.2: MFED step responses of all the circuit configurations discussed. Again, the 4-pole T-coil step response ***i)*** has the steepest slope, but the 3-pole T-coil response ***h)*** is close.

3.2 Transistor as an Impedance Converter

In the previous section we have realized that the amplification factor is frequency dependent, decreasing with frequency above some upper frequency limit (asymptotically to $-20\,\mathrm{dB}/10f$, just like a first-order low pass system). This can help us to derive different impedance transformations from the emitter to the base circuit and back [Ref. 3.1, 3.2]. Knowing the possible transformations is extremely useful in the wideband amplifier design. We shall show how the nature of the impedance changes with these transformations. A capacitive impedance may become inductive and positive impedances may occasionally become negative!

3.2.1 Common base small signal transistor model

As we explained in Sec. 3.1, if the voltage gain is not too high the base emitter capacitance C_π is the dominant cause of the frequency response rolling off at high frequencies. By considering this we can make a simplified small signal high frequency transistor model, as shown in Fig. 3.2.1, for the common base configuration, where i_c, i_e and i_b are the collector-, emitter-, and base-current respectively. For this figure the DC current amplification factor is:

$$\alpha_0 = \frac{I_c}{I_e} \tag{3.2.1}$$

A more correct expression for the mutual conductance is:

$$g_m = \frac{\alpha_0}{r_e} = \frac{\beta_0}{(1+\beta_0)\,r_e} \tag{3.2.2}$$

where β_0 is the common emitter DC current amplification factor. If $\beta_0 \gg 1$ then $\alpha_0 \simeq 1$, so the collector current I_c is almost equal to the emitter current I_e and $g_m \simeq 1/r_e$. This simplification is often used in practice.

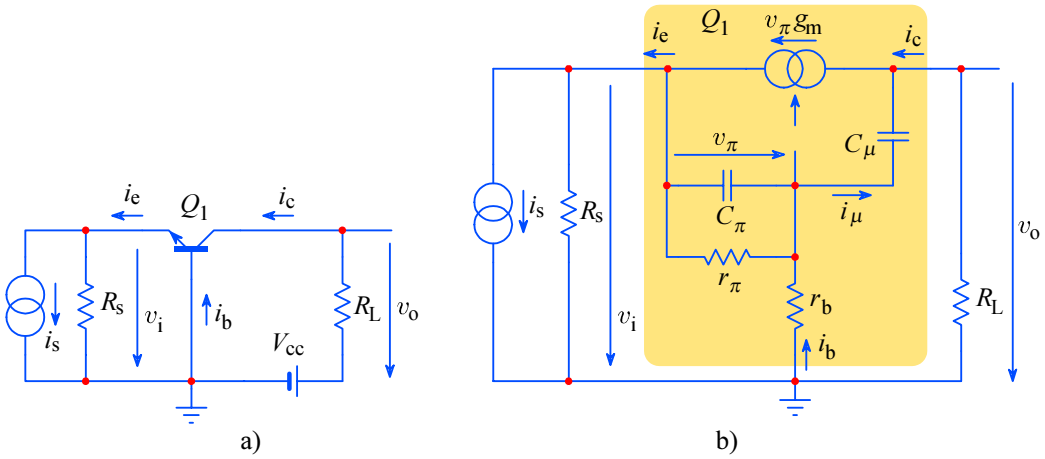


Fig. 3.2.1: The common base amplifier: a) circuit schematic diagram; b) high frequency small signal equivalent circuit.

For the moment let us assume that the base resistance $r_b = 0$ and consider the low frequency relations. The input resistance is:

$$r_\pi = \frac{v_\pi}{i_b} \tag{3.2.3}$$

where v_π is the base to emitter voltage. Since the emitter current is:

$$i_e = i_b + i_c = i_b + \beta_0\, i_b = i_b\,(1 + \beta_0) \tag{3.2.4}$$

then the base current is:

$$i_b = \frac{i_e}{1 + \beta_0} \tag{3.2.5}$$

and consequently:

$$r_\pi = \frac{v_\pi\,(1 + \beta_0)}{i_e} = r_e\,(1 + \beta_0) \approx \beta_0\, r_e \tag{3.2.6}$$

The last simplification is valid if $\beta_0 \gg 1$. To obtain the input impedance at high frequencies the parallel connection of C_π must be taken into account:

$$Z_b = \frac{(1 + \beta_0)\, r_e}{1 + (1 + \beta_0)\, s\, C_\pi\, r_e} \tag{3.2.7}$$

The base current is:

$$i_b = \frac{v_\pi}{Z_b} = v_\pi\, \frac{1 + (1 + \beta_0)\, s\, C_\pi\, r_e}{(1 + \beta_0)\, r_e} \tag{3.2.8}$$

Furthermore it is:

$$v_\pi = i_b\, \frac{(1 + \beta_0)\, r_e}{1 + (1 + \beta_0)\, s\, C_\pi\, r_e} \tag{3.2.9}$$

The collector current is:

$$i_c = g_m\, v_\pi = \frac{\beta_0}{1 + \beta_0} \cdot \frac{1}{r_e}\, v_\pi = \frac{\alpha_0}{r_e}\, v_\pi \tag{3.2.10}$$

If we put Eq. 3.2.9 into Eq. 3.2.10 we obtain:

$$\begin{aligned} i_c &= i_b\, \frac{\beta_0}{1 + \beta_0} \cdot \frac{1}{r_e} \cdot \frac{(1 + \beta_0)\, r_e}{1 + s\,(1 + \beta_0)\, r_e\, C_\pi} \\ &= i_b\, \frac{1}{\dfrac{1}{\beta_0} + s \left(\dfrac{\beta_0 + 1}{\beta_0} \right) r_e\, C_\pi} \\ &\approx i_b\, \frac{1}{\dfrac{1}{\beta_0} + s\, \tau_T} = i_b\, \beta(s) \end{aligned} \tag{3.2.11}$$

In the very last expression we assumed that $\beta_0 \gg 1$ and $\tau_T = r_e\, C_\pi = 1/\omega_T$, where $\omega_T = 2\pi f_T$ is the angular frequency at which the current amplification factor β decreases to unity. The parameter τ_T (and consequently ω_T) depends on the internal configuration and structure of the transistor. Fig. 3.2.2 shows the frequency dependence of β and the equivalent current generator.

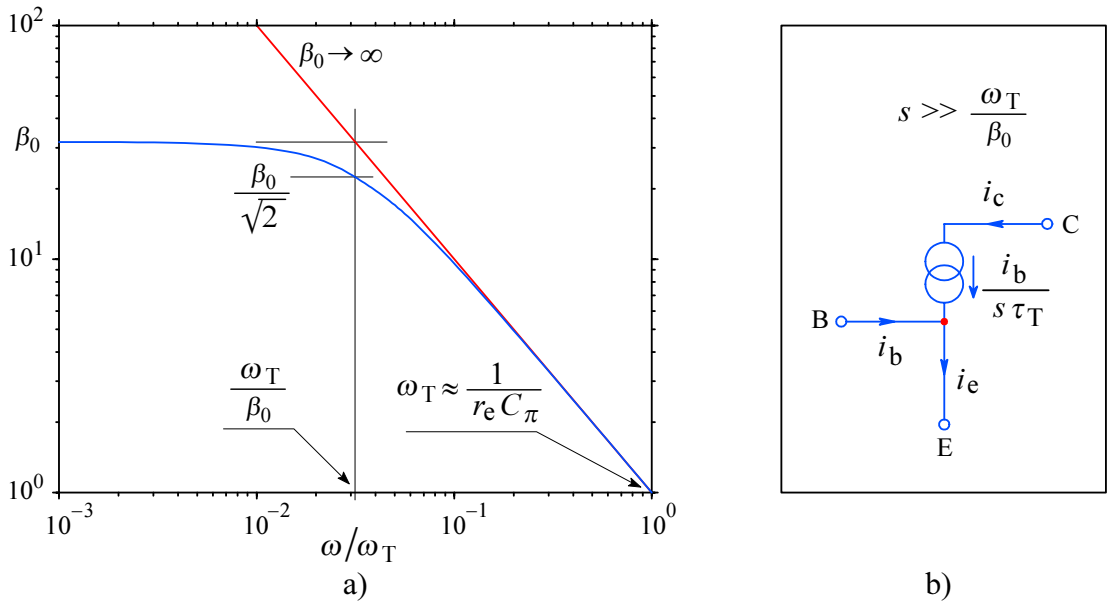


Fig. 3.2.2: a) The transistor gain as a function of frequency, modeled by the Eq. 3.2.11; b) the equivalent HF current generator.

In order to correlate Fig. 3.2.2 with Eq. 3.2.11 we rewrite it as:

$$\frac{i_c}{i_b} \approx \beta_0 \frac{-\frac{\omega_T}{\beta_0}}{s - \left(-\frac{\omega_T}{\beta_0}\right)} = \beta_0 \frac{-s_1}{s - s_1} \tag{3.2.12}$$

where s_1 is the pole at $-\omega_T/\beta_0$. This relation will become useful later, when we shall apply one of the peaking circuits (from Part 2) to the amplifier. At very high frequencies, or if $\beta_0 \gg 1$, the term $s\,\tau_T$ prevails. In this case, from Eq. 3.2.11:

$$\frac{i_c}{i_b} = \beta(s) \approx \frac{1}{s\,\tau_T} = \frac{1}{j\,\omega\,r_e\,C_\pi} \tag{3.2.13}$$

Obviously β is decreasing with frequency. By definition, at $\omega = \omega_T$ the current ratio $i_c/i_b = 1$; then the capacitance C_π can be found as:

$$C_\pi \approx \frac{1}{\omega_T\,r_e} \tag{3.2.14}$$

This simplified relation represents the $-20\,\mathrm{dB}/10f$ asymptote in Fig. 3.2.2a.

3.2.2 The conversion of impedances

We can use the result of Eq. 3.2.11 to transform the transistor internal (and the added external) impedances from the emitter to the base circuitry, and vice versa. Suppose we have the impedance Z_e in the emitter circuit, as displayed in Fig. 3.2.3a, and we are interested in the corresponding base impedance Z_b:

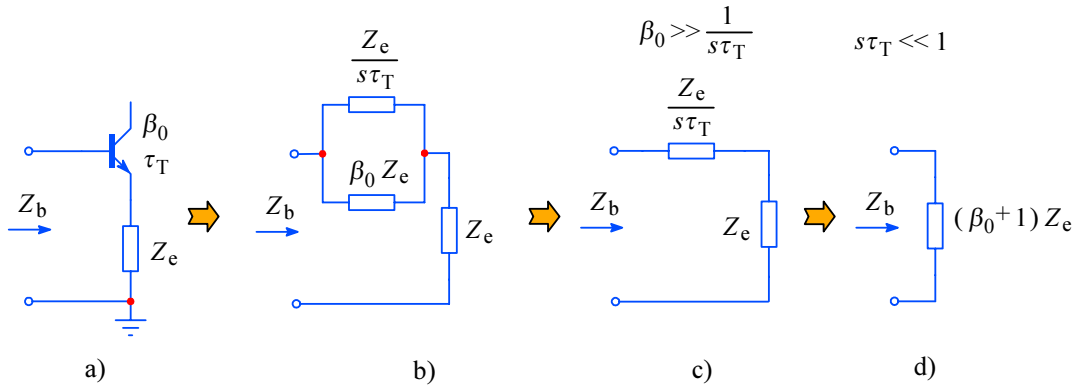


Fig. 3.2.3: Emitter to base impedance conversion: a) schematic; b) equivalent circuit; c) simplified for high β_0; d) simplified for low frequencies.

We know that:

$$Z_b = \beta(s)\, Z_e + Z_e = \left[\beta(s) + 1\right] Z_e \tag{3.2.15}$$

If we insert $\beta(s)$ according to Eq. 3.2.11, we obtain:

$$Z_b = \frac{Z_e}{\dfrac{1}{\beta_0} + s\,\tau_T} + Z_e \tag{3.2.16}$$

The admittance of the first part of this equation is:

$$Y = \frac{\dfrac{1}{\beta_0} + s\,\tau_T}{Z_e} = \frac{1}{\beta_0\, Z_e} + \frac{s\,\tau_T}{Z_e} \tag{3.2.17}$$

and this represents a parallel connection of impedances $\beta_0\, Z_e$ and $Z_e/s\,\tau_T$. By adding the series impedance Z_e, as in Eq. 3.2.16, we obtain the equivalent circuit of Fig. 3.2.3b. At medium frequencies and with a high value of β_0 we can assume that $1/\beta_0 \ll s\,\tau_T$, so we can delete the impedance $\beta_0 Z_e$ and simplify the circuit, as in Fig. 3.2.3c. On the other hand, at low frequencies, where $s\,\tau_T \ll 1$, we can neglect the $Z_e/s\,\tau_T$ component and get a very basic equivalent circuit, displayed in Fig. 3.2.3d.

Eq. 3.2.11 is equally useful when we want to transform the impedance from the base into the emitter circuit as shown in Fig. 3.2.4a. In this case we have:

$$Z_e = \frac{Z_b}{\beta(s) + 1} \tag{3.2.18}$$

Again we calculate the admittance, which is:

$$Y_\mathrm{e} = \frac{\beta(s)+1}{Z_\mathrm{b}} = [\beta(s)+1]\,Y_\mathrm{b} = \beta(s)\,Y_\mathrm{b} + Y_\mathrm{b} \tag{3.2.19}$$

The first part of this admittance is:

$$Y = \frac{\beta(s)}{Z_\mathrm{b}} = \frac{Y_\mathrm{b}}{\frac{1}{\beta_0} + s\,\tau_\mathrm{T}} = \frac{1}{Z_\mathrm{b}} \cdot \frac{1}{\frac{1}{\beta_0} + s\,\tau_\mathrm{T}} \tag{3.2.20}$$

and the impedance is:

$$Z = \frac{Z_\mathrm{b}}{\beta_0} + s\,\tau_\mathrm{T}\,Z_\mathrm{b} \tag{3.2.21}$$

Thus the transformed impedance Z_e is composed of three elements: the series connection of Z_b/β_0 and $s\,\tau_\mathrm{T}\,Z_\mathrm{b}$, in parallel with the impedance Z_b.

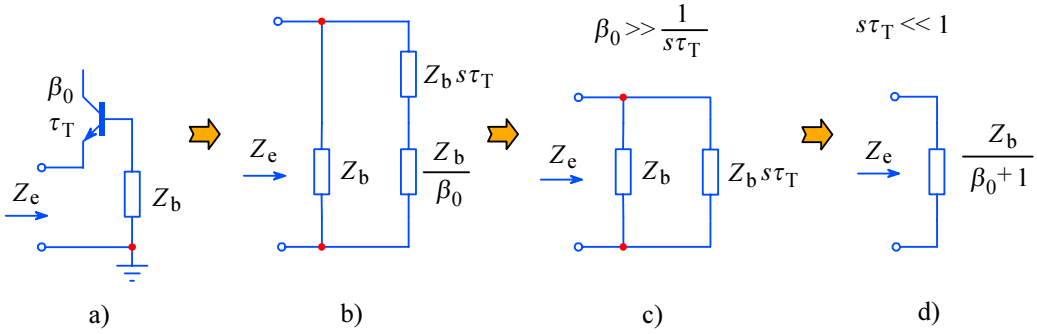


Fig. 3.2.4: Base to emitter impedance conversion: a) schematic; b) equivalent circuit; c) simplified for high β_0 or for $f \simeq f_\mathrm{T}$; d) simplified for low frequencies.

The equivalent emitter impedance is shown in Fig. 3.2.4b.

As in the previous example, for some specific conditions the circuit can be simplified. At medium frequencies and a high β_0, we can assume $\beta_0 \gg 1/s\,\tau_\mathrm{T}$ and therefore neglect the impedance Z_b/β_0, as in Fig. 3.2.4c. At low frequencies, where $s\,\tau_\mathrm{T} \ll 1$, the impedance $Z_\mathrm{b}/(\beta_0+1)$ prevails and we can neglect the parallel impedance Z_b, as in Fig. 3.2.4d.

3.2.3 Examples of impedance transformations

The most interesting examples of impedance transformations are the emitter to base transformation of a capacitive emitter impedance and the base to emitter transformation of an inductive base impedance.

In the first case we have $Z_\mathrm{e} = 1/s\,C$, where C is the emitter to ground capacitance.

To obtain the base impedance we apply Eq. 3.2.5:

$$Z_\mathrm{b} = \frac{\beta(s)+1}{s\,C} = \left[\frac{1}{\dfrac{1}{\beta_0}+s\,\tau_\mathrm{T}}+1\right]\frac{1}{s\,C} = \frac{\dfrac{1}{\beta_0}+s\,\tau_\mathrm{T}+1}{\left(\dfrac{1}{\beta_0}+s\,\tau_\mathrm{T}\right)s\,C}$$

$$= \frac{s\,\tau_\mathrm{T}+\left(1+\dfrac{1}{\beta_0}\right)}{s^2\tau_\mathrm{T}\,C+\dfrac{s\,C}{\beta_0}} \tag{3.2.22}$$

The inverse of Z_b is the admittance:

$$Y_\mathrm{b} = \frac{s^2\tau_\mathrm{T}\,C+\dfrac{s\,C}{\beta_0}}{s\,\tau_\mathrm{T}+\left(1+\dfrac{1}{\beta_0}\right)} \tag{3.2.23}$$

Let us synthesize this expression by a simple continued fraction expansion [Ref. 3.27]:

$$\frac{s^2\tau_\mathrm{T}\,C+\dfrac{s\,C}{\beta_0}}{s\,\tau_\mathrm{T}+\left(1+\dfrac{1}{\beta_0}\right)} = s\,C - \frac{s\,C}{s\,\tau_\mathrm{T}+\left(1+\dfrac{1}{\beta_0}\right)} \tag{3.2.24}$$

The fraction on the right is a negative admittance with the corresponding impedance:

$$Z'_\mathrm{b} = -\frac{s\,\tau_\mathrm{T}+\left(1+\dfrac{1}{\beta_0}\right)}{s\,C} = -\frac{\tau_\mathrm{T}}{C}-\frac{1+\dfrac{1}{\beta_0}}{s\,C} \tag{3.2.25}$$

It is evident that this impedance is a series connection of a negative resistance:

$$R_\mathrm{n} = -\frac{\tau_\mathrm{T}}{C} = -r_\mathrm{e}\,\frac{C_\pi}{C} \tag{3.2.26}$$

and a negative capacitance:

$$C_\mathrm{n} = -\frac{C}{1+\dfrac{1}{\beta_0}} = -\frac{\beta_0}{1+\beta_0}\,C = -\alpha_0\,C \tag{3.2.27}$$

By adding the positive parallel capacitance C, as required by Eq. 3.2.24, we obtain the equivalent circuit which is shown in Fig. 3.2.5. Since we deal with an active circuit (transistor) it is quite normal to encounter negative impedances. The complete base admittance is then:

$$Y_\mathrm{b} = s\,C - \frac{1}{\dfrac{\tau_\mathrm{T}}{C}+\dfrac{1}{s\,\alpha_0\,C}} \tag{3.2.28}$$

By rearranging this expression and substituting $s = j\omega$ we can separate the real and imaginary parts, obtaining:

$$Y_\mathrm{b} = \Re\{Y_\mathrm{b}\} + j\Im\{Y_\mathrm{b}\} = G_\mathrm{b} + j\omega C_\mathrm{b}$$

$$= -\frac{\frac{\tau_\mathrm{T}}{C}}{\tau_\mathrm{T}^2 + \frac{1}{\omega^2\alpha_0^2}} - j\omega C\,\frac{\tau_\mathrm{T}^2 - \frac{\alpha_0 - 1}{\omega^2\alpha_0^2}}{\tau_\mathrm{T}^2 + \frac{1}{\omega^2\alpha_0^2}} \tag{3.2.29}$$

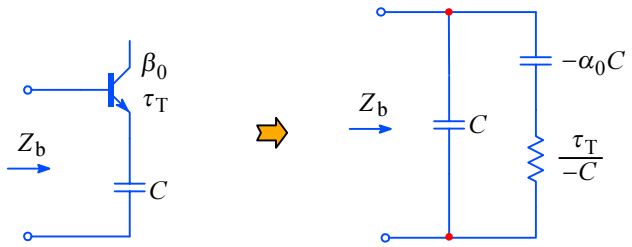


Fig. 3.2.5: A capacitive emitter load is reflected into the base (junction) with **negative** components.

The negative input (base) conductance G_b can cause ringing on steep signals or even continuous oscillations if the signal source impedance has an emphasized inductive component. We shall thoroughly discuss this effect and its compensation later, when we shall analyze the emitter–follower (i.e., common collector) and the JFET source–follower amplifiers.

Now let us derive the emitter impedance Z_e in the case in which the base impedance is inductive ($s\,L$). Here we apply Eq. 3.2.18:

$$Z_\mathrm{e} = \frac{s\,L}{\beta(s)+1} = \frac{s\,L}{\frac{1}{\frac{1}{\beta_0} + s\,\tau_\mathrm{T}} + 1} = \tag{3.2.30}$$

$$= \frac{s\,L\left(\frac{1}{\beta_0} + s\,\tau_\mathrm{T}\right)}{1 + \frac{1}{\beta_0} + s\,\tau_\mathrm{T}} = \frac{s^2 L\,\tau_\mathrm{T} + \frac{s\,L}{\beta_0}}{s\,\tau_\mathrm{T} + \left(1 + \frac{1}{\beta_0}\right)} \tag{3.2.31}$$

By continued fraction expansion we obtain:

$$\frac{s^2 L\,\tau_\mathrm{T} + \frac{s\,L}{\beta_0}}{s\,\tau_\mathrm{T} + \left(1 + \frac{1}{\beta_0}\right)} = s\,L - \frac{s\,L}{s\,\tau_\mathrm{T} + \left(1 + \frac{1}{\beta_0}\right)} \tag{3.2.32}$$

The negative part of the result can be inverted to obtain the admittance:

$$Y'_\mathrm{e} = -\frac{s\,\tau_\mathrm{T} + \left(1 + \frac{1}{\beta_0}\right)}{s\,L} = -\frac{\tau_\mathrm{T}}{L} - \frac{1 + \frac{1}{\beta_0}}{s\,L} \tag{3.2.33}$$

This means we have two parallel impedances. The first one is a **negative** resistance:

$$R_x = -\frac{L}{\tau_T} \tag{3.2.34}$$

and the second one is a **negative** inductance:

$$L_x = -\frac{L}{1+\frac{1}{\beta_0}} = -\frac{\beta_0}{1+\beta_0} L = -\alpha_0 L \tag{3.2.35}$$

As required by Eq. 3.2.32, we must add the inductance L in series, thus arriving at the equivalent emitter impedance shown in the figure below:

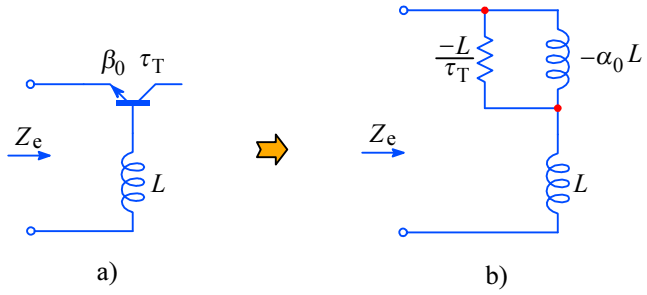


Fig. 3.2.6: The inductive source is reflected into the emitter with **negative** components.

We have just analyzed an important aspect of a common base amplifier, with an inductance (i.e., long lead) between the base and ground. The negative resistance, as given by Eq. 3.2.34, may become the reason for ringing or oscillations if the driving circuit seen by the emitter has a capacitive character. We shall discuss this problem more thoroughly when we shall analyze the cascode circuit.

In a way similar to those used for deriving the previous two results, we can transform other impedance types from emitter to base, and vice versa. The Table 3.2.1 displays the six possible variations and the reader is encouraged to derive the remaining four, which we have not discussed.

Note that all the three transformations for the comon base circuit in the table apply to the base–emitter **junction** to ground only. In order to obtain the correct base **terminal** to ground impedance the transistor base spread resistance r_b must be added in series to the circuits shown in the table.

4.1 A Cascade of Identical, DC Coupled, *RC* Loaded Stages

A multi-stage amplifier with DC coupled, RC loaded stages is shown in Fig. 4.1.1. All stages are assumed to be identical. Junction field effect transistors (JFETs) are being used as active devices, since we want to focus on essential design problems; with bipolar junction transistors (BJTs), we would have to consider the loading effect of a relatively complicated base input impedance [Ref. 4.1].

At each stage load the capacitance C represents the sum of all the stray capacitances along with the $C_{\mathrm{GS}}(1+A_k)$ equivalent capacitance (where C_{GS} is the gate–drain capacitance and A_k is the voltage gain of the individual stage). By doing so we obtain a simple parallel RC load in each stage. The input resistance of a JFET is many orders of magnitude higher than the loading resistor R, so we can neglect it. All loading resistors are of equal value and so are the mutual conductances g_{m} of all the JFETs; therefore all individual gains A_k are equal as well. Consequently, all the half power frequencies $\omega_{\mathrm{h}k}$ and all the rise times $\tau_{\mathrm{r}k}$ of the individual stages are also equal. In order to simplify the circuit further, we have not drawn the the bias voltages and the power supply (which must represent a short circuit for AC signals; obviously, a short at DC is not very useful!).

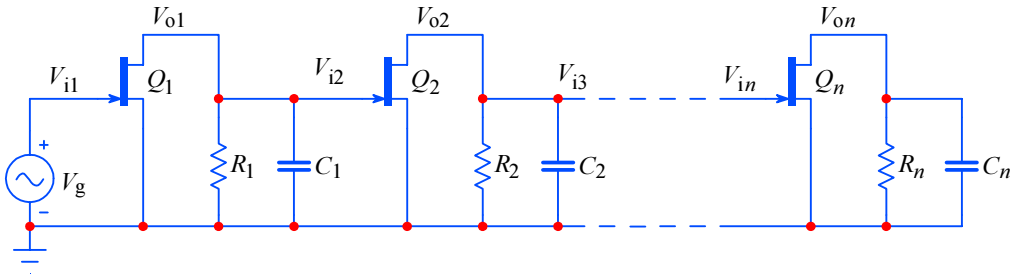


Fig. 4.1.1: A multi-stage amplifier as a cascade of identical, DC coupled, RC loaded stages.

The voltage gain of an individual stage is:

$$A_k = g_{\mathrm{m}}\, R\, \frac{1}{1+j\omega RC} \tag{4.1.1}$$

with the magnitude:

$$|A_k| = \frac{g_{\mathrm{m}}\, R}{\sqrt{1+(\omega/\omega_{\mathrm{h}})^2}} \tag{4.1.2}$$

where:

g_{m} = mutual conductance of the JFET, [S] (siemens, [S] = [1/Ω]);

$\omega_{\mathrm{h}} = 1/RC$ = upper half power frequency of an individual stage, [rad/s].

4.1.1 Frequency Response and the Upper Half Power Frequency

We have n equal stages with equal gains: $A_1 = A_2 = \cdots = A_n = A_k$. The gain of the complete amplifier is then:

$$A = A_1 \cdot A_2 \cdot A_3 \cdots A_n = A_k^n = \left[\frac{g_{\mathrm{m}}\, R}{1+j\,\omega\, RC}\right]^n \tag{4.1.3}$$

The magnitude is:

$$|A| = \left[\frac{g_\mathrm{m}\, R}{\sqrt{1 + (\omega/\omega_\mathrm{h})^2}} \right]^n \tag{4.1.4}$$

To be able to compare the bandwidth of the multi-stage amplifier for different number of stages, we must normalize the magnitude by dividing Eq. 4.1.4 by the system DC gain $A_0 = (g_\mathrm{m}\, R)^n$. The plots are shown in Fig. 4.1.2. It is evident that the system bandwidth, ω_H, shrinks with each additional amplifying stage.

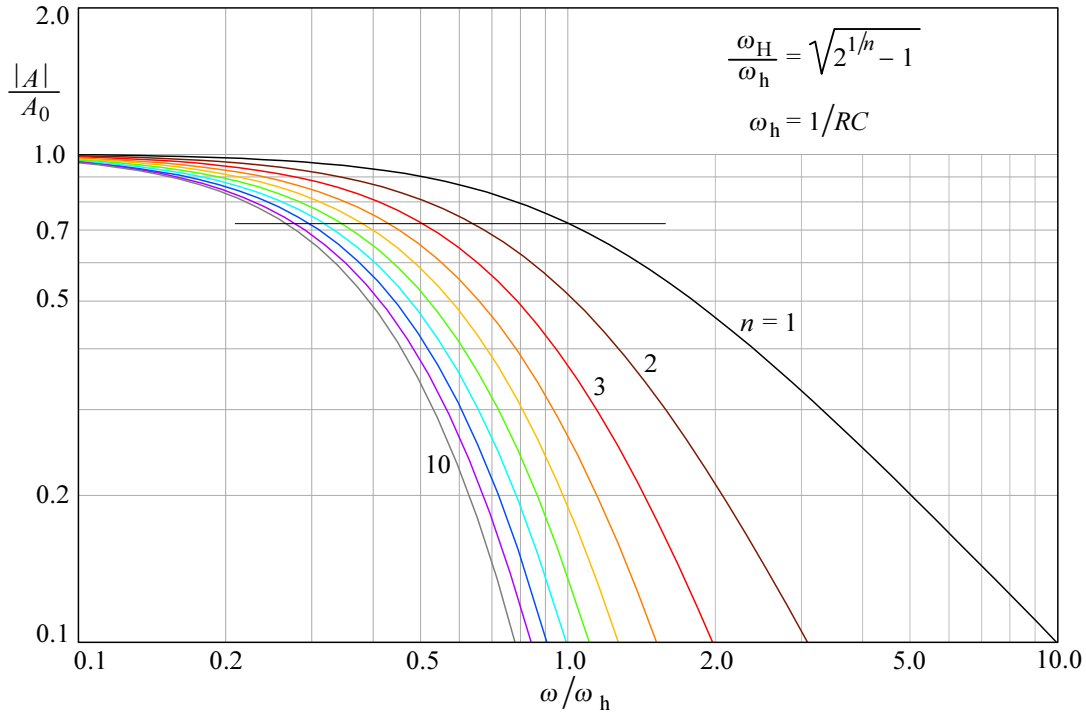


Fig. 4.1.2: Frequency response of an n-stage amplifier ($n = 1, 2, \ldots, 10$). To compare the bandwidth, the gain was normalized, i.e., divided by the system DC gain, $(g_\mathrm{m}\, R)^n$. For each n, the bandwidth (the crossing of the 0.707 level) shrinks by $\sqrt{2^{1/n} - 1}$.

The upper half power frequency of the amplifier can be calculated by a simple relation:

$$\left[\frac{1}{\sqrt{1 + (\omega_\mathrm{H}/\omega_\mathrm{h})^2}} \right]^n = \frac{1}{\sqrt{2}} \tag{4.1.5}$$

By squaring we obtain:

$$\left[1 + (\omega_\mathrm{H}/\omega_\mathrm{h})^2\right]^n = 2 \;\Rightarrow\; \left(\frac{\omega_\mathrm{H}}{\omega_\mathrm{h}} \right)^2 = 2^{1/n} - 1 \tag{4.1.6}$$

The upper half power frequency of the complete n-stage amplifier is:

$$\boxed{\omega_\mathrm{H} = \omega_\mathrm{h} \sqrt{2^{1/n} - 1}} \tag{4.1.7}$$

At high frequencies, the first stage response slope approaches the $-6\,\mathrm{dB/octave}$ asymptote ($-20\,\mathrm{dB/decade}$). The meaning of this slope is explained in Fig. 4.1.3. For the second stage the slope is twice as steep, and for the n^{th} stage it is n times steeper.

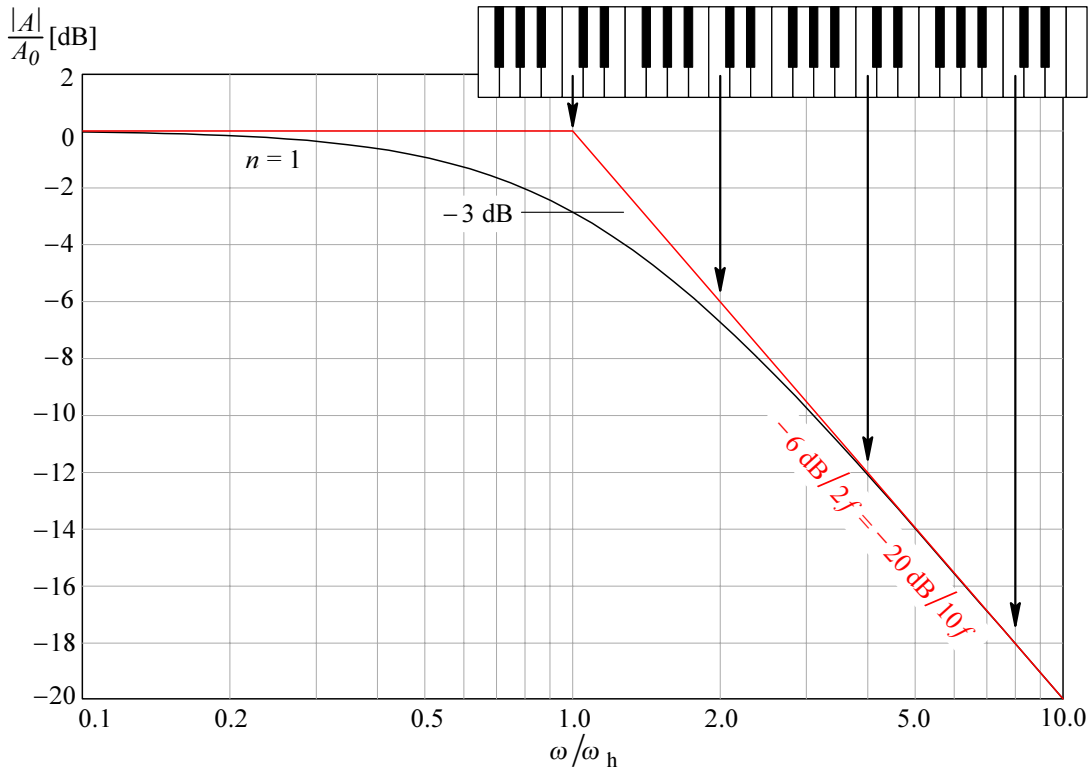


Fig. 4.1.3: The first-order system response and its asymptotes. Below the cut off, the asymptote is the level equal to the system gain at DC (normalized here to 0 dB). Above the cut off, the slope is $-6\,\mathrm{dB/octave}$ (an octave is a frequency span from f to $2f$), which is also equal to $-20\,\mathrm{dB/decade}$ (a frequency decade is a span from f to $10\,f$).

The values of ω_{H} for $n = 1-10$ are reported in Table 4.1.1.

Table 4.1.1

n	1	2	3	4	5	6	7	8	9	10
ω_{H}	1.000	0.644	0.510	0.435	0.386	0.350	0.323	0.301	0.283	0.269

With ten equal stages connected in cascade the bandwidth is reduced to a poor $0.269\,\omega_{\mathrm{h}}$; such an amplifier is definitively not very efficient for wideband amplification.

Alternatively, in order to preserve the bandwidth a n-stage amplifier should have all its capacitors reduced by the same factor, $\sqrt{2^{1/n}-1}$. But in wideband amplifiers we already strive to work with stray capacitances only, so this approach is not a solution.

Nevertheless, the amplifier in Fig. 4.1.1 is the basis for more efficient amplifier configurations, which we shall discuss later.

4.1.2 Phase Response

Each individual stage of the amplifier in Fig. 4.1.1 has a frequency dependent phase angle:

$$\varphi_k = \arctan \frac{\Im\{F(j\omega)\}}{\Re\{F(j\omega)\}} = \arctan(-\omega/\omega_h) \tag{4.1.8}$$

where $F(j\omega)$ is taken from Eq. 4.1.1. For n equal stages the total phase angle is simply n times as much:

$$\boxed{\varphi_n = n \arctan(-\omega/\omega_h)} \tag{4.1.9}$$

The phase responses are plotted in Fig. 4.1.4. Note the high frequency asymptotic phase shift increasing by $\pi/2$ (or 90°) for each n. Also note the shift at $\omega = \omega_h$ being exactly $n\,\pi/4$, in spite of a reduced ω_h for each n.

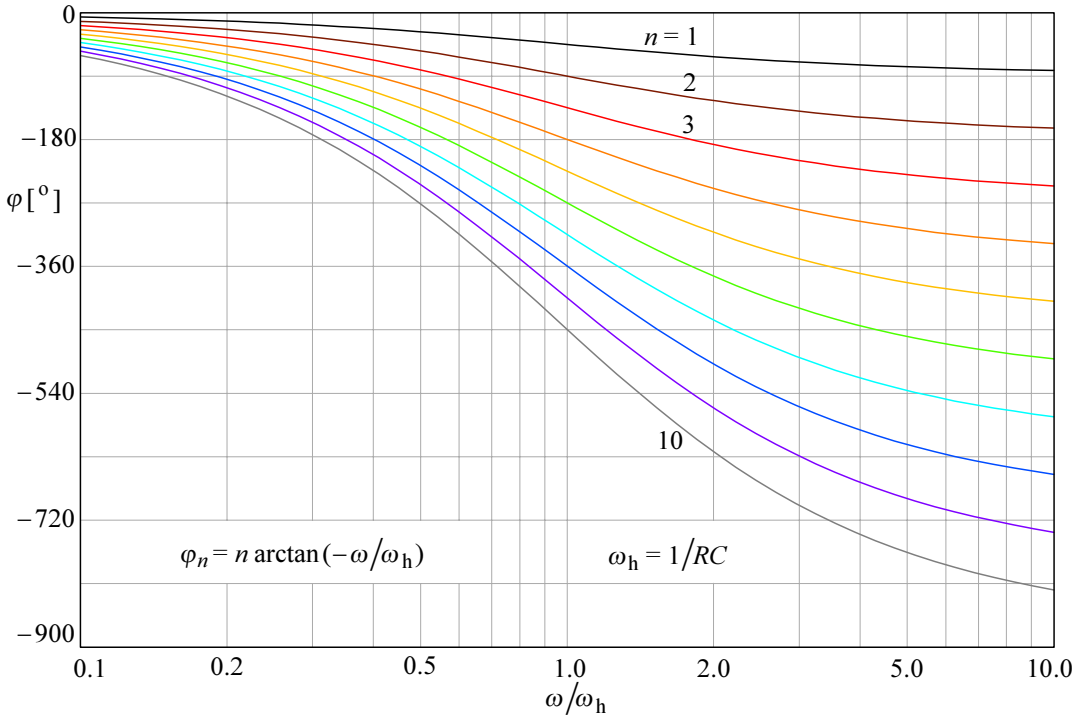


Fig. 4.1.4: Phase angle of the amplifier in Fig. 4.1.1, for n = 1–10 amplifying stages.

4.1.3 Envelope Delay

For a single amplifying stage ($n = 1$) the envelope delay is the frequency derivative of the phase, $\tau_{en} = d\varphi_n/d\omega$ (where φ_n is given by Eq. 4.1.9). The normalized single stage envelope delay is:

$$\tau_e\,\omega_h = \frac{-1}{1 + (\omega/\omega_h)^2} \tag{4.1.10}$$

and for n equal stages:

$$\tau_{en}\,\omega_h = \frac{-n}{1+(\omega/\omega_h)^2} \tag{4.1.11}$$

Fig. 4.1.5 shows the frequency dependent envelope delay for $n = 1$–10. Note the delay at $\omega = \omega_h$ being exactly $1/2$ of the low frequency asymptotic value.

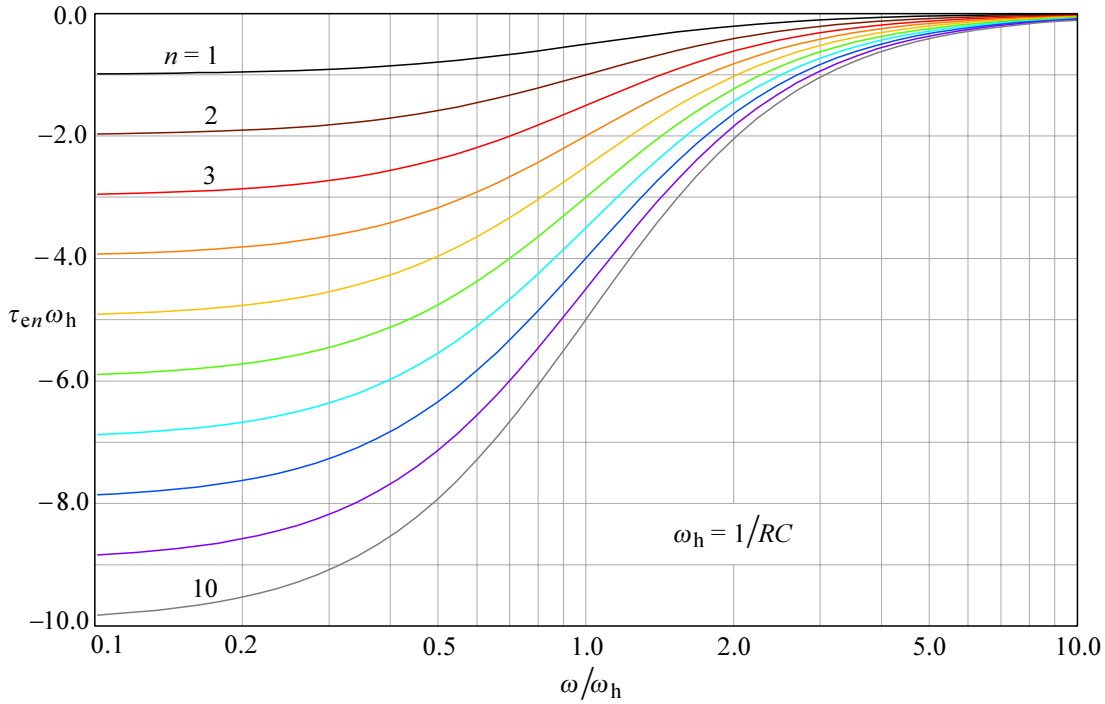


Fig. 4.1.5: Envelope delay of the amplifier in Fig. 4.1.1, for $n = 1$–10 amplifying stages. The delay at $\omega = \omega_h$ is $1/2$ of the low frequency asymptotic value. Note that if we were using f/f_h for the abscissa, we would have to divide the τ_e scale by 2π.

4.1.4 Step Response

To obtain the step response, the amplifier in Fig. 4.1.1 must be driven by the unit step function:

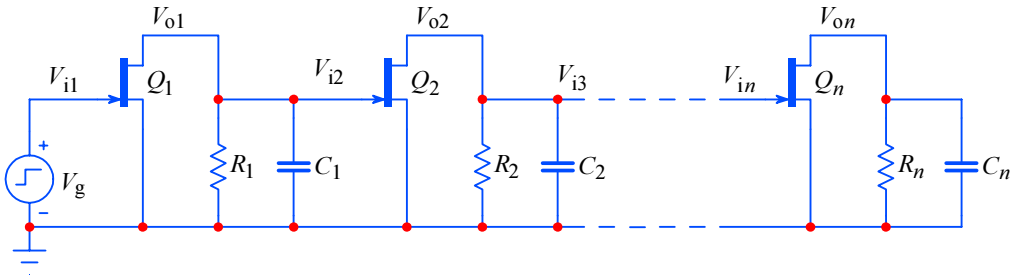


Fig. 4.1.6: Amplifier with n equal DC coupled stages, excited by the unit step

We can derive the step response expression from Eq. 4.1.1 and Eq. 4.1.3. In order to simplify and generalize the expression we shall normalize the magnitude by dividing the transfer function by the DC gain, $g_m R$, and normalize the frequency by setting $\omega_h = 1/RC = 1$. Since we shall use the $\mathcal{L}^{-1}$ transform we shall replace the variable $j\omega$ by the complex variable $s = \sigma + j\omega$.

With all these changes we obtain:

$$F(s) = \frac{1}{(1+s)^n} \tag{4.1.12}$$

The amplifier input is excited by the unit step, therefore we must multiply the above formula by the unit step operator $1/s$:

$$G(s) = \frac{1}{s\,(1+s)^n} \tag{4.1.13}$$

The corresponding function in the time-domain is:

$$g(t) = \mathcal{L}^{-1}\{G(s)\} = \sum \text{res}\, \frac{\mathrm{e}^{s\,t}}{s\,(1+s)^n} \tag{41.14}$$

We have two residues. The first one does not depend of n:

$$\text{res}_0 = \lim_{s \to 0} s \left[\frac{\mathrm{e}^{s\,t}}{s\,(1+s)^n} \right] = 1$$

whilst the second does:

$$\text{res}_1 = \lim_{s \to 1} \frac{1}{(n-1)!} \cdot \frac{d^{(n-1)}}{ds^{(n-1)}} \left[(1+s)^n \, \frac{\mathrm{e}^{s\,t}}{s\,(1+s)^n} \right] =$$

$$= \lim_{s \to 1} \frac{1}{(n-1)!} \cdot \frac{d^{(n-1)}}{ds^{(n-1)}} \left(\frac{\mathrm{e}^{s\,t}}{s} \right) \tag{4.1.15}$$

Since res_1 depends on n, for $n = 1$ we obtain:

$$\text{res}_1 \Big|_{n=1} = -\,\mathrm{e}^{-t} \tag{4.1.16}$$

for $n = 2$:

$$\text{res}_1 \Big|_{n=2} = -\,\mathrm{e}^{-t}\,(1+t) \tag{4.1.17}$$

for $n = 3$:

$$\text{res}_1 \Big|_{n=3} = -\,\mathrm{e}^{-t} \left(1 + t + \frac{t^2}{2} \right) \tag{4.1.18}$$

… etc.

The general expression for the step response for any n is:

$$\boxed{g_n(t) = \mathcal{L}^{-1}\big\{\mathrm{G}(s)\big\} = \text{res}_0 + \text{res}_1 = 1 - \mathrm{e}^{-t} \sum_{k=1}^{n} \frac{t^{k-1}}{(k-1)!}} \tag{4.1.19}$$

Here we must consider that $0! = 1$, by definition.

As an example, by inserting $n = 5$ into Eq. 4.1.19 we obtain:

$$g_5(t) = 1 - \mathrm{e}^{-t} \left(1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} \right) \tag{4.1.20}$$

The step response plots for $n = 1$–10, calculated by Eq. 4.1.19, are shown in Fig. 4.1.7. Note that there is **no overshoot** in any of the curves. Unfortunately, the efficiency of this kind of amplifier in the sense of the 'bandwidth per number of stages' is poor, since it has no peaking networks which would prevent the decrease of bandwidth with n.

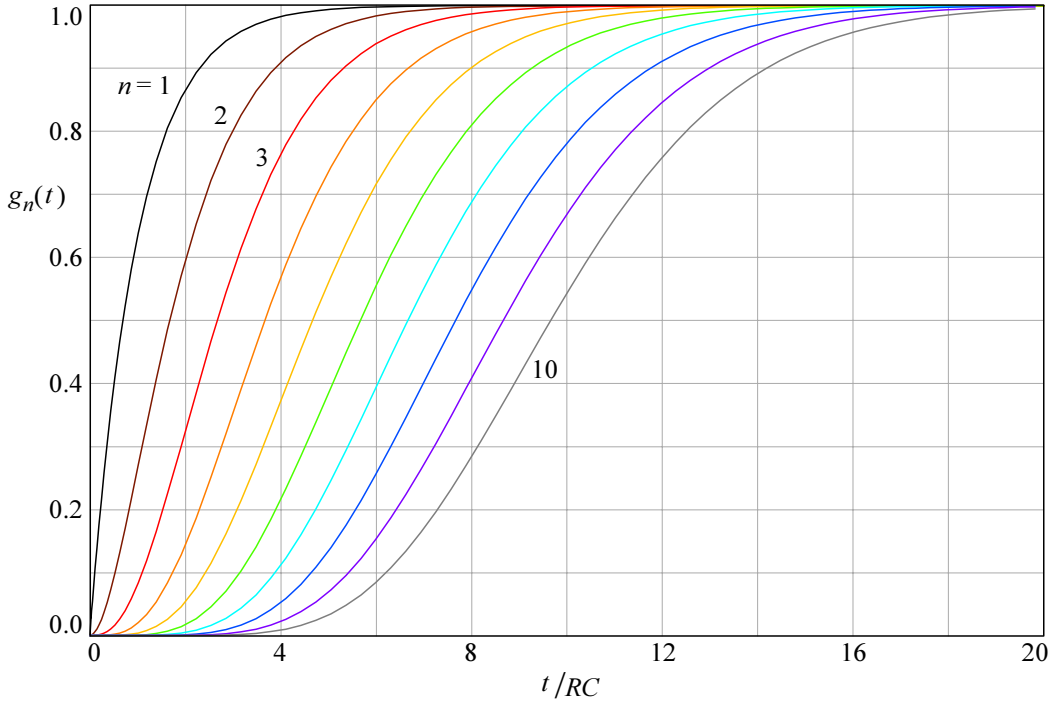


Fig. 4.1.7: Step response of the amplifier in Fig. 4.1.6, for $n = 1$–10 amplifying stages

4.1.5 Rise Time Calculation

In the case of a multi-stage amplifier, where each particular stage has its respective rise time, $\tau_{r1}, \tau_{r2}, \ldots, \tau_{rn}$, we calculate the system's rise time [Ref. 4.2] as:

$$\tau_r = \sqrt{\tau_{r1}^2 + \tau_{r2}^2 + \tau_{r3}^2 + \cdots + \tau_{rn}^2} \tag{4.1.21}$$

In Part 2, Sec. 2.1.1, Eq. 2.1.1–4, we have calculated the rise time of an amplifier with a simple RC load to be $\tau_{r1} = 2.20\,RC$. Since here we have n equal stages the rise time of the complete amplifier is:

$$\boxed{\tau_r = \tau_{r1}\sqrt{n} = 2.20\,RC\sqrt{n}} \tag{4.1.22}$$

Table 4.1.2 shows the rise time increasing with the number of stages:

Table 4.1.2

n	1	2	3	4	5	6	7	8	9	10
τ_{rn}/τ_{r1}	1.00	1.41	1.73	2.00	2.24	2.45	2.65	2.83	3.00	3.16

5.1 Geometrical Synthesis of Inductively Compensated Multi-Stage Amplifiers — A Simple Example

The reader who has patiently followed the discussion presented in previous chapters is probably eager to see all that theory being put into practice.

Before jumping to some more complex amplifier circuits we shall give a relatively simple example of a two-stage differential cascode amplifier, by which we shall illustrate the actual system optimization procedure in some detail, using the previously developed principles in their full potential.

Since we want to grasp the 'big picture' we shall have to leave out some less important topics, such as negative input impedance compensation, cascode damping, etc.; these are important for the optimization of each particular stage which, once optimized, can be idealized to some extent. We have covered that extensively enough in Part 3, so we shall not explicitly draw the associated components in the schematic diagram. But, at the end of our calculations, we shall briefly discuss the influence of those components to final circuit values.

A two-stage amplifier is a 'minimum complexity' system for which the multi-stage design principles still apply. To this we shall add a 3-pole T-coil and a 4-pole L+T-coil peaking networks, discussed in Part 2, as loads to each stage, making a total of 7 poles. There is, however, an additional real pole, owed to the Q_1 input capacitance and the total input and signal source resistance. As we shall see later, this pole can be neglected if its distance from the complex plane's origin is at least twice as large as that of the system real pole set by $-1/R_{\mathrm{a}}C_{\mathrm{a}}$.

Such an amplifier thus represents an elementary example in which everything that we have learned so far can be applied. The reader should, however, be aware that this is by no means the ideal or, worse still, the only possibility. At the end of our calculations, when we shall be able to assess the advantages and limitations offered by our initial choices at each stage, we shall examine a few possibilities of further improvement.

We shall start our calculations from the unavoidable stray capacitances and the desired total voltage gain. Then we shall apply an optimization process, which we like to refer to as the **geometrical synthesis**, by which we shall calculate all the remaining circuit components in such a way that the resulting system will conform to the 7-pole normalized Bessel–Thomson system. The only difference will be that the actual amplifier poles will be larger by a certain factor, proportional (but not equal) to the upper half power frequency ω_{H}. We have already met the geometrical synthesis in its basic form in Part 2, Fig. 2.5.3 when we were discussing the 3-pole T-coil circuit. The name springs from the ability to calculate all the peaking network components from simple geometrical relations which involve the pole real an imaginary components, given, of course, the desired pole pattern and a few key component values which can either be chosen independently or set by other design requirements. Here we are going to see a generalization of those basic relations applied to the whole amplifier.

It must be admitted that the constant and real input impedance of the T-coil network is the main factor which allows us to assign so many poles to only two stages. A cascade of passive 2-pole sections could have been used, but those would load each

other and, as a result, the bandwidth extension factor would suffer. Another possibility would be to use an additional cascode stage to separate the last two peaking sections, but another active stage, whilst adding gain, also adds its own problems to be taken care of. It is, nevertheless, a perfectly valid option.

Let us now take a quick tour of the amplifier schematic, Fig. 5.1.1. We have two differential cascode stages and two current sources, which set both the transistor's transconductance and the maximum current available to the load resistors, R_a and R_b. This limits the voltage range available to the CRT. Since the circuit is differential the total gain is a double of each half. The total DC gain is (approximately):

$$A_0 = 2\,\frac{R_a}{R_{e1}} \cdot \frac{R_b}{R_{e2}} \tag{5.1.1}$$

The values of R_{e1} and R_{e2} set the required capacitive bypass, $C_{e1}/2$ and $C_{e3}/2$, to match the transistor's time constants. In turn, this sets the input capacitance at the base of Q_1 and Q_3, to which we must add the inevitable C_{cb} and some strays.

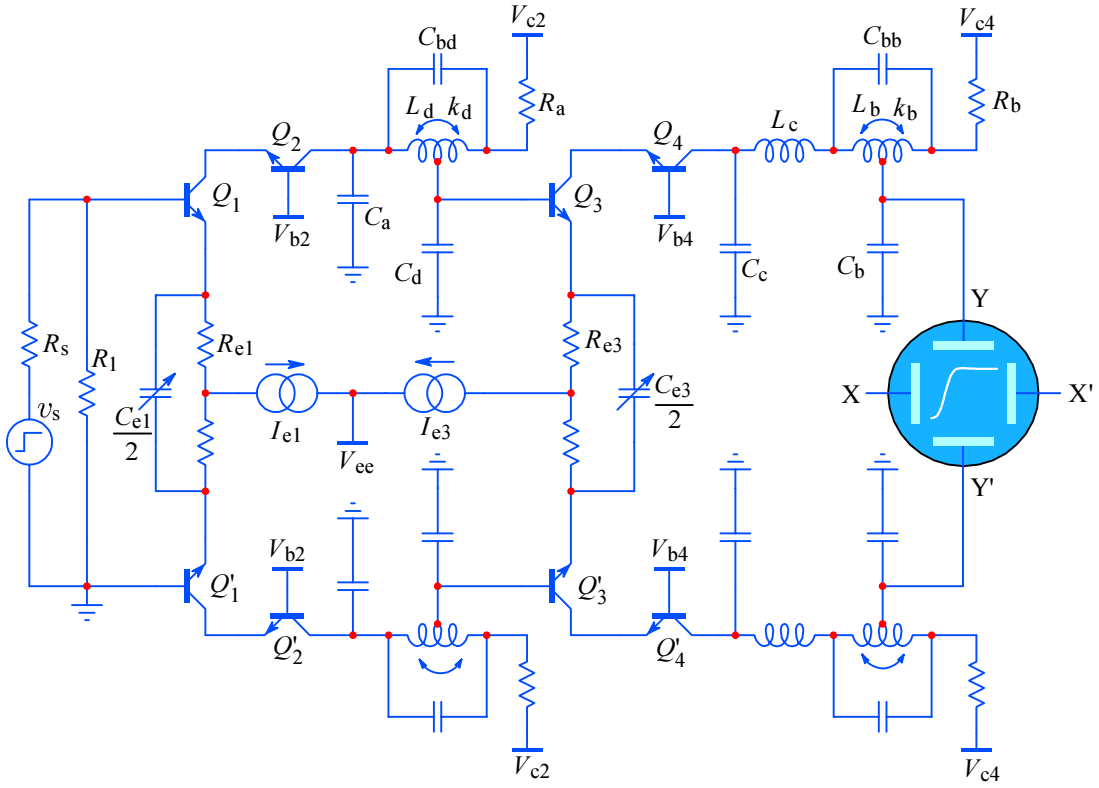


Fig. 5.1.1: A simple 2-stage differential cascode amplifier with a 7-pole peaking system: the 3-pole T-coil inter-stage peaking (between the Q_2 collector and the Q_3 base) and the 4-pole L+T-coil output peaking (between the Q_4 collector and the vertical plates of the cathode ray tube). The schematic was simplified to emphasize the important design aspects — see text.

The capacitance C_d should thus consists of, preferably, only the input capacitance at the base of Q_3. If required by the coil 'tuning', a small capacitance can be added in parallel, but that would also reduce the bandwidth. Note that the associated T-coil L_d will have to be designed as an inter-stage peaking, as we have discussed in Part 3, Sec. 3.6, but we can leave the necessary corrections for the end.

The capacitance C_b, owed almost entirely to the CRT vertical plates, is much larger than C_d, so we expect that R_a and R_b can not be equal. From this it follows that

it might be difficult to apply equal gain to each stage in accordance with the principle explained in Part 4, Eq. 4.1.39. Nevertheless, the difference in gain will not be too high, as we shall see.

Like any other engineering process, geometrical synthesis also starts from some external boundary conditions which set the main design goal. In this case it is the CRT's vertical sensitivity and the available input voltage, from which the total gain is defined. The next condition is the choice of transistors by which the available current is defined. Both the CRT and the transistors set the lower limit of the loading capacitances at various nodes. From these the first circuit component R_b is fixed.

With R_b fixed we arrive at the first 'free' parameter, which can be represented by several circuit components. However, since we would like to maximize the bandwidth this parameter should be attributed to one of the capacitances. By comparing the design equations for the 3-pole T-coil and the 4-pole L+T-coil peaking networks in Part 2, it can be deduced that C_a, the input capacitance of the 3-pole section, is the most critical component.

With these boundaries set let us assume the following component values:

$$\begin{aligned} C_b &= 11\,\text{pF} && (9\,\text{pF of the CRT vertical plates, } 2\,\text{pF stray}) \\ C_a &= 4\,\text{pF} && (3\,\text{pF from the } Q_2\ C_{cb},\ 1\,\text{pF stray}) \\ R_b &= 360\ \Omega && (\text{determined by the desired gain and the available current}) \end{aligned} \qquad (5.1.2)$$

The pole pattern is, in general, another 'free' parameter, but for a smooth, minimum overshoot transient we must apply the Bessel–Thomson arrangement. As can be seen in Fig. 5.1.2, each pole (pair) defines a circle going through the pole and the origin, with the center on the negative real axis.

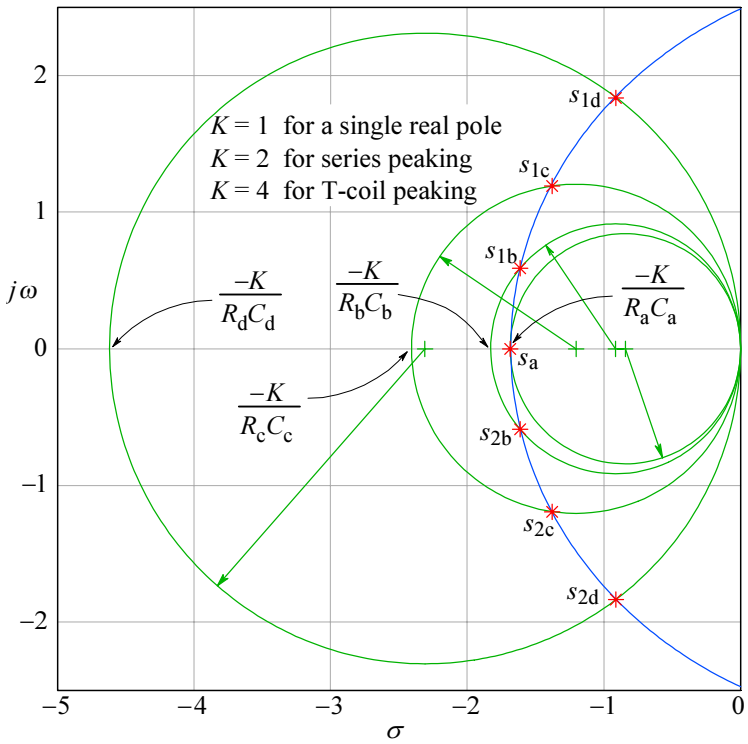


Fig. 5.1.2: The 7 normalized Bessel–Thomson poles. The characteristic circle of each pole (pair) has a diameter determined by the appropriate RC constant and the peaking factor K, which depends on the type of network chosen.

The poles in Fig. 5.1.2 bear the index of the associated circuit components and the reader might wonder why we have chosen precisely that assignment.

In a general case the assignment of a pole (pair) to a particular circuit section is yet another 'free' design parameter. If we were designing a low frequency filter we could indeed have chosen an arbitrary assignment (as long as each complex conjugate pole pair is assigned as a pair, a limitation owed to physics, instead of circuit theory).

If, however, the bandwidth is an issue then we must seek those nodes with the largest capacitances and apply the poles with the lowest imaginary part to those circuit sections. This is because the capacitor impedance (which is dominantly imaginary) is inversely proportional both to the capacitor value and the signal frequency.

In this light the largest capacitance is at the CRT, that is, C_b; thus the pole pair with the lowest imaginary part is assigned to the output T-coil section, formed by L_b and R_b, therefore acquiring the index 'b', $s_{1\mathrm{b}}$ and $s_{2\mathrm{b}}$.

The real pole is the one associated with the 3-pole stage and there it is set by the loading resistor R_a and the input capacitance C_a, becoming s_a.

The remaining two pole pairs should be assigned so that the pair with the larger imaginary part is applied to that peaking network which has a larger bandwidth improvement factor. Here we must consider that $K = 4$ for a T-coil, whilst $K = 2$ for the series peaking L-section (of the 4-pole L+T-scetion). Clearly the pole pair with the larger imaginary part should be assigned to the inter-stage T-coil, L_d, thus they are labeled $s_{1\mathrm{d}}$ and $s_{2\mathrm{d}}$. The L-section then receives the remaining pair, $s_{1\mathrm{c}}$ and $s_{2\mathrm{c}}$.

We have thus arrived at a solution which seems logical, but in order to be sure that we have made the right choice we should check other combinations as well. We are going to do so at the end of the design process.

The poles for the normalized 7th-order Bessel–Thomson system, as taken either from Part 4, Table 4.4.3, or by using the BESTAP (Part 6) routine, along with the associated angles, are:

$$\begin{aligned}
s_\mathrm{a} = \sigma_\mathrm{a} &= -4.9718 & \theta_\mathrm{a} &= 180° \\
s_\mathrm{b} = \sigma_\mathrm{b} \pm j\,\omega_\mathrm{b} &= -4.7583 \pm j\,1.7393 & \theta_\mathrm{b} &= 180° \mp 20.0787° \\
s_\mathrm{c} = \sigma_\mathrm{c} \pm j\,\omega_\mathrm{c} &= -4.0701 \pm j\,3.5172 & \theta_\mathrm{c} &= 180° \mp 40.8316° \\
s_\mathrm{d} = \sigma_\mathrm{d} \pm j\,\omega_\mathrm{d} &= -2.6857 \pm j\,5.4207 & \theta_\mathrm{d} &= 180° \mp 63.6439°
\end{aligned} \qquad (5.1.3)$$

So, let us now express the basic design equations by the assigned poles and the components of the two peaking networks.

For the real pole s_a we have the following familiar proportionality:

$$s_\mathrm{a} = \sigma_\mathrm{a} = D_\mathrm{a} = -4.9718 \;\propto\; \frac{-1}{R_\mathrm{a}\,C_\mathrm{a}} \qquad (5.1.4)$$

At the output T-coil section, according to Part 2, Fig. 2.5.3, we have:

$$D_\mathrm{b} = \frac{\sigma_\mathrm{b}}{\cos^2\theta_\mathrm{b}} = \frac{-4.7583}{0.8821} = -5.3941 \;\propto\; \frac{-4}{R_\mathrm{b}\,C_\mathrm{b}} \qquad (5.1.5)$$

For the L-section of the L+T output network, because the T-coil input impedance is equal to the loading resistor, we have:

$$D_c = \frac{\sigma_c}{\cos^2\theta_c} = \frac{-4.0701}{0.5725} = -7.1094 \quad \propto \quad \frac{-2}{R_b\,C_c} \tag{5.1.6}$$

And finally, for the inter-stage T-coil network:

$$D_d = \frac{\sigma_d}{\cos^2\theta_d} = \frac{-2.6857}{0.1917} = -13.6333 \quad \propto \quad \frac{-4}{R_a\,C_d} \tag{5.1.7}$$

From these relations we can calculate the required values of the remaining capacitances, C_c and C_d. If we divide Eq. 5.1.5 by Eq. 5.1.6, we have the ratio:

$$\frac{D_b}{D_c} = \frac{-\dfrac{4}{R_b\,C_b}}{-\dfrac{2}{R_b\,C_c}} = \frac{2\,C_c}{C_b} \tag{5.1.8}$$

It follows that the capacitance C_c should be:

$$C_c = \frac{C_b}{2}\cdot\frac{D_b}{D_c} = \frac{11}{2}\cdot\frac{-5.3941}{-7.1094} = 4.1730\,\text{pF} \tag{5.1.9}$$

Likewise, if we divide Eq. 5.1.4 by Eq. 5.1.7, we obtain:

$$\frac{D_a}{D_d} = \frac{-\dfrac{1}{R_a\,C_a}}{-\dfrac{4}{R_a\,C_d}} = \frac{C_d}{4\,C_a} \tag{5.1.10}$$

Thus C_d will be:

$$C_d = 4\,C_a\,\frac{D_a}{D_d} = 4\cdot 4\cdot\frac{-4.9718}{-13.6333} = 5.8349\,\text{pF} \tag{5.1.11}$$

Of course, for most practical purposes, the capacitances do not need to be calculated to such precision, a resolution of 0.1 pF should be more than enough. But we would like to check our procedure by recalculating the actual poles from circuit components and for that purpose we shall need this precision.

Now we need to know the value of R_a. This can be readily calculated from the ratio D_a/D_b:

$$\frac{D_a}{D_b} = \frac{-\dfrac{1}{R_a\,C_a}}{-\dfrac{4}{R_b\,C_b}} = \frac{R_b\,C_b}{4\,R_a\,C_a} \tag{5.1.12}$$

resulting in:

$$R_a = \frac{R_b}{4}\cdot\frac{C_b}{C_a}\cdot\frac{D_b}{D_a} = \frac{360}{4}\cdot\frac{11}{4}\cdot\frac{-5.3941}{-4.9718} = 268.5\,\Omega \tag{5.1.13}$$

We are now ready to calculate the inductances L_b, L_c and L_d. For the two T-coils we can use the Eq. 2.4.19:

$$L_b = R_b^2\, C_b = 360^2 \cdot 11 \cdot 10^{-12} = 1.4256\,\mu H \tag{5.1.14}$$

and

$$L_d = R_a^2\, C_d = 268.5^2 \cdot 5.8349 \cdot 10^{-12} = 0.4206\,\mu H \tag{5.1.15}$$

For L_c we use Eq. 2.2.26 to obtain the proportionality factor of the RC constant:

$$L_c = \frac{1+\tan^2\theta_b}{4}\, R_b^2\, C_c = \frac{360^2 \cdot 4.1730 \cdot 10^{-12}}{4 \cdot 0.8821} = 0.1533\,\mu H \tag{5.1.16}$$

The magnetic coupling factors for the two T-coils are calculated by Eq. 2.4.36:

$$k_b = \frac{3-\tan^2\theta_b}{5+\tan^2\theta_b} = \frac{3-0.1336}{5+0.1336} = 0.5584 \tag{5.1.17}$$

and likewise:

$$k_d = \frac{3-\tan^2\theta_d}{5+\tan^2\theta_d} = \frac{3-4.0738}{5+4.738} = -0.1183 \tag{5.1.18}$$

Note that k_d is negative. This means that, instead of the usually negative mutual inductance, we need a positive inductance at the T-coil tap. This can be achieved by simply mounting the two halves of L_d perpendicular to each other, in order to have zero magnetic coupling and then introduce an additional coil, L_e (again perpendicular to both halves of L_d), with a value of the required positive mutual inductance, as can be seen in Fig. 5.1.3. Another possibility would be to wind the two halves of L_d in opposite direction, but then the bridge capacitance C_{bd} might be difficult to realize correctly.

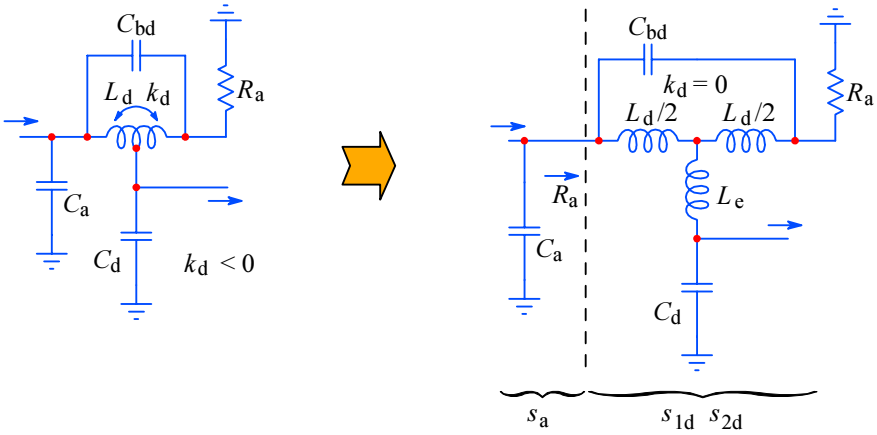


Fig. 5.1.3: With the assigned poles and the resulting particular component values the 3-pole stage magnetic coupling k_d needs to be negative, which forces us to use non-coupled coils and add a positive mutual inductance L_e. Even with a negative k_d the T-coil reflects its resistive load to the network input, greatly simplifying the calculations of component values.

The additional inductance L_e can be calculated from the required mutual inductance given by the negative value of k_d. In Part 2, Eq. 2.4.1–2.4.5 we have defined the T-coil inductance, its two halves, and its mutual inductance by the relations repeated in Eq. 5.1.19 for convenience:

$$L = L_1 + L_2 + 2\,L_\mathrm{M}$$

$$L_1 = L_2 = \frac{L}{2\,(1+k)} \tag{5.1.19}$$

$$L_\mathrm{M} = -\,k\,\sqrt{L_1\,L_2}$$

Thus, if $k = 0$ we have:

$$L_{1\mathrm{d}} = L_{2\mathrm{d}} = \frac{L_\mathrm{d}}{2} = \frac{0.4206}{2} = 0.2103\,\mu\mathrm{H} \tag{5.1.20}$$

and:

$$L_\mathrm{e} = -\,k_\mathrm{d}\,\sqrt{\frac{L_\mathrm{d}}{2}\cdot\frac{L_\mathrm{d}}{2}} = -\,k_\mathrm{d}\,\frac{L_\mathrm{d}}{2} = 0.1183\,\frac{0.4206}{2} = 0.025\,\mu\mathrm{H} \tag{5.1.21}$$

If we were to account for the Q_3 base resistance (discussed in Part 3, Sec. 3.6) we would get k_d even more negative and also $L_{1\mathrm{d}} \neq L_{2\mathrm{d}}$.

The coupling factor k_b, although positive, also poses a problem: since it is greater than 0.5 it might be difficult to realize. As can be noted from the above equations, the value of k depends only on the pole's angle θ. In fact, the 2nd-order Bessel system has the pole angles of $\pm\,150°$, resulting in a $k = 0.5$, representing the limiting case of realizability with conventionally wounded coils. Special shapes, coil overlapping, or other exotic techniques may solve the coupling problem, but, more often than not, they will also impair the bridge capacitance. The other limiting case, when $k = 0$, is reached by the ratio $\Im\{s\}/\Re\{s\} = \sqrt{3}$, a situation occurring when the pole's angle $\theta = 120°$.

In accordance with previous equations we also calculate the value of the two halves of L_b:

$$L_{1\mathrm{b}} = L_{2\mathrm{b}} = \frac{L_\mathrm{b}}{2\,(1+k_\mathrm{b})} = \frac{1.4256}{2\,(1+0.5584)} = 0.4574\,\mu\mathrm{H} \tag{5.1.22}$$

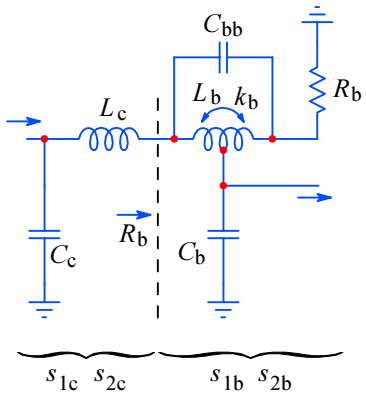


Fig. 5.1.4: The 4-pole output L+T-coil stage and its pole assignment.

The last components to be calculated are the bridge capacitances, C_{bb} and C_{bd}. The relation between the T-coil loading capacitance and the bridge capacitance has been given already in Part 2, Eq. 2.4.31, from which we obtain the following expressions for C_{bb} and C_{bd}:

$$C_{bb} = C_b \frac{1+\tan^2\theta_b}{16} = 11\,\frac{1+0.1336}{16} = 0.7793\,\text{pF} \qquad (5.1.23)$$

and:

$$C_{bd} = C_d \frac{1+\tan^2\theta_d}{16} = 5.8349\,\frac{1+4.0738}{16} = 1.8503\,\text{pF} \qquad (5.1.24)$$

This completes the calculation of amplifier components necessary for the inductive peaking compensation and thus achieving the Bessel–Thomson system response. We would now like to verify the design by recalculating the actual pole values. To do this we return to the relations which we have started from, Eq. 5.1.3 to Eq. 5.1.7 and for the imaginary part using the relations in Part 2, Fig. 2.5.3 . In order not to confuse the actual pole values with the normalized values, from which we started, we add an index 'A' to the actual poles:

$$\sigma_{aA} = -\frac{1}{R_a C_a} = -\frac{1}{268.5\cdot 4\cdot 10^{-12}} = -931.1\cdot 10^6\,\text{rad/s} \qquad (5.1.25)$$

$$\sigma_{bA} = -\frac{4\cos^2\theta_b}{R_b C_b} = -\frac{4\cdot 0.8821}{360\cdot 11\cdot 10^{-12}} = -891.0\cdot 10^6\,\text{rad/s}$$

$$\omega_{bA} = \pm\frac{4\cos\theta_b \sin\theta_b}{R_b C_b} = \pm\frac{4\cdot 0.9392\cdot 0.3433}{360\cdot 11\cdot 10^{-12}} = \pm 325.7\cdot 10^6\,\text{rad/s}$$

$$\sigma_{cA} = -\frac{2\cos^2\theta_c}{R_b C_c} = -\frac{2\cdot 0.5725}{360\cdot 4.1730\cdot 10^{-12}} = -762.2\cdot 10^6\,\text{rad/s}$$

$$\omega_{cA} = \pm\frac{2\cos\theta_c \sin\theta_c}{R_b C_c} = \pm\frac{2\cdot 0.7566\cdot 0.6538}{360\cdot 4.1730\cdot 10^{-12}} = \pm 658.5\cdot 10^6\,\text{rad/s}$$

$$\sigma_{dA} = -\frac{4\cos^2\theta_d}{R_a C_d} = -\frac{4\cdot 0.1917}{268.5\cdot 5.8349\cdot 10^{-12}} = -489.5\cdot 10^6\,\text{rad/s}$$

$$\omega_{dA} = \pm\frac{4\cos\theta_d \sin\theta_d}{R_a C_d} = \pm\frac{4\cdot 0.4439\cdot 0.8961}{268.5\cdot 5.8349\cdot 10^{-12}} = \pm 1015.6\cdot 10^6\,\text{rad/s}$$

If we divide the real amplifier pole by the real normalized pole, we get:

$$\frac{\sigma_{bA}}{\sigma_b} = \frac{-931.1\cdot 10^6}{-4.9718} = 187.3\cdot 10^6 \qquad (5.1.26)$$

and this factor is equal for all other pole components. Unfortunately, from this we cannot calculate the upper half power frequency of the amplifier. The only way to do that (for a Bessel system) is to calculate the response for a range of frequencies around the cut off and then iterate it using the bisection method, until a satisfactory tolerance has been achieved.

Instead of doing it for only a small range of frequencies we shall, rather, do it for a three decade range and compare the resulting response with the one we would get from a non-compensated amplifier (in which all the inductances are zero). Since to this point we were not interested in the actual value of the voltage gain, we shall make the comparison using amplitude normalized responses.

The non-compensated amplifier has two real poles, which are:

$$s_{1N} = -\frac{1}{R_a\,(C_a + C_d)} \quad \text{and} \quad s_{2N} = -\frac{1}{R_b\,(C_b + C_c)} \tag{5.1.27}$$

Consequently, its complex frequency response would then be:

$$F_N(s) = \frac{s_{1N}\,s_{2N}}{(s - s_{1N})(s - s_{2N})} \tag{5.1.28}$$

with the magnitude:

$$\left|F_N(\omega)\right| = \frac{\sqrt{s_{1N}\,s_{2N}}}{\sqrt{(\omega^2 - s_{1N}^2)(\omega^2 - s_{2N}^2)}} \tag{5.1.29}$$

and the step response:

$$\begin{aligned} g(t) &= \mathcal{L}^{-1}\left\{\frac{s_{1N}\,s_{2N}}{s\,(s - s_{1N})(s - s_{2N})}\right\} \\ &= 1 + \frac{s_{2N}}{s_{1N} - s_{2N}}\,e^{s_{1N}t} - \frac{s_{1N}}{s_{1N} - s_{2N}}\,e^{s_{2N}t} \end{aligned} \tag{5.1.30}$$

The rise time is:

$$\tau_r = 2.2\sqrt{\frac{1}{s_{1N}^2} + \frac{1}{s_{2N}^2}} \tag{5.1.31}$$

and the half power frequency:

$$f_h = \frac{\sqrt{s_{1N}\,s_{2N}}}{2\,\pi} \tag{5.1.32}$$

In contrast, the complex frequency response of the 7-pole amplifier is:

$$F_A(s) = A_0\,\frac{-\,s_{aA}\,s_{1bA}\,s_{2bA}\,s_{1cA}\,s_{2cA}\,s_{1dA}\,s_{2cA}}{(s - s_{aA})(s - s_{1bA})(s - s_{2bA})(s - s_{1cA})(s - s_{2cA})(s - s_{1dA})(s - s_{2dA})} \tag{5.1.33}$$

and the step response is the inverse Laplace transform of the product of $F_A(s)$ with the unit step operator $1/s$:

$$g(t) = \mathcal{L}^{-1}\left\{\frac{1}{s}\,F_A(s)\right\} = \sum \text{res}\left(\frac{1}{s}\,F_A(s)\,e^{st}\right) \tag{5.1.34}$$

We shall not attempt to solve either of these functions analytically, since it would take too much space, and, anyway, we have solved them separately for its two parts (3rd- and 4th-order) in Part 2. Because the systems are separated by an amplifier (Q_3, Q_4), the frequency response would be a simple multiplication of the two responses. For the step response we now have 8 residues to sum (7 of the system poles, in addition to the one from the unit step operator). Although lengthy, it is a

relatively simple operation and we leave it as an exercise to the reader. Instead we are going to use the computer routines, the development of which can be found in Part 6.

In Fig. 5.1.5 we have made a polar plot of the poles for the inductively compensated 7-pole system and the non-compensated 2-pole system. As we have learned in Part 1 and Part 2, the farther from origin the smaller is the pole's influence on the system response. It is therefore obvious that the 2-pole system's response will be dominated by the pole closer to the origin and that is the pole of the output stage, s_{2N}. The bandwidth of the 7-pole system is, obviously, much larger.

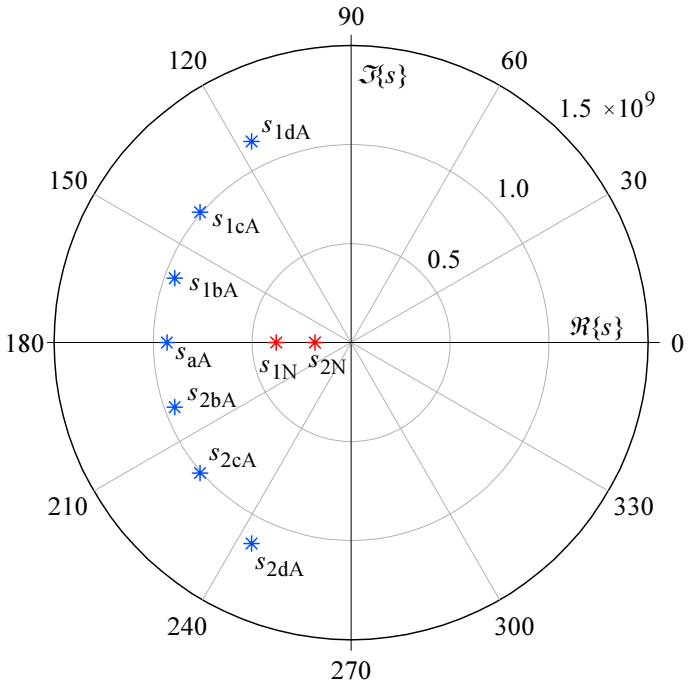


Fig. 5.1.5: The polar plot of the 7-pole compensated system (poles with index 'A') and the 2-pole non-compensated system (index 'N'). The radial scale is $\times 10^9$ rad/s. The angle is in degrees.

The pole layout gives us a convenient indication of the system's performance, but it is the magnitude vs. frequency response that reveals it clearly. As can be seen in Fig. 5.1.6, the non-compensated system has a bandwidth of less than 25 MHz. The compensated amplifier bandwidth is close to 88 MHz, more than 3.5 times larger.

The comparison of step responses in Fig. 5.1.7 reveals the difference in performance even more dramatically. The rise time of the non-compensated system is about 14 ns, whilst for the compensated system it is only 3.8 ns, also a factor of 3.5 times better; in addition, the overshoot is only 0.48 %.

Both comparisons show an impressive improvement in performance. But is it the best that could be obtained from this circuit configuration? After all, in Part 2 we have seen a similar improvement from just the 4-pole L+T-coil section and we expect that the addition of the 3-pole section should yield a slightly better result at least.

One obvious way of extending the bandwidth would be to lower the value of R_b, increase the bias currents, and scale the remaining components accordingly. Then we should increase the input signal amplitude to get the same output. But this is the 'trivial' solution (mathematically, at least; not so when building an actual circuit).

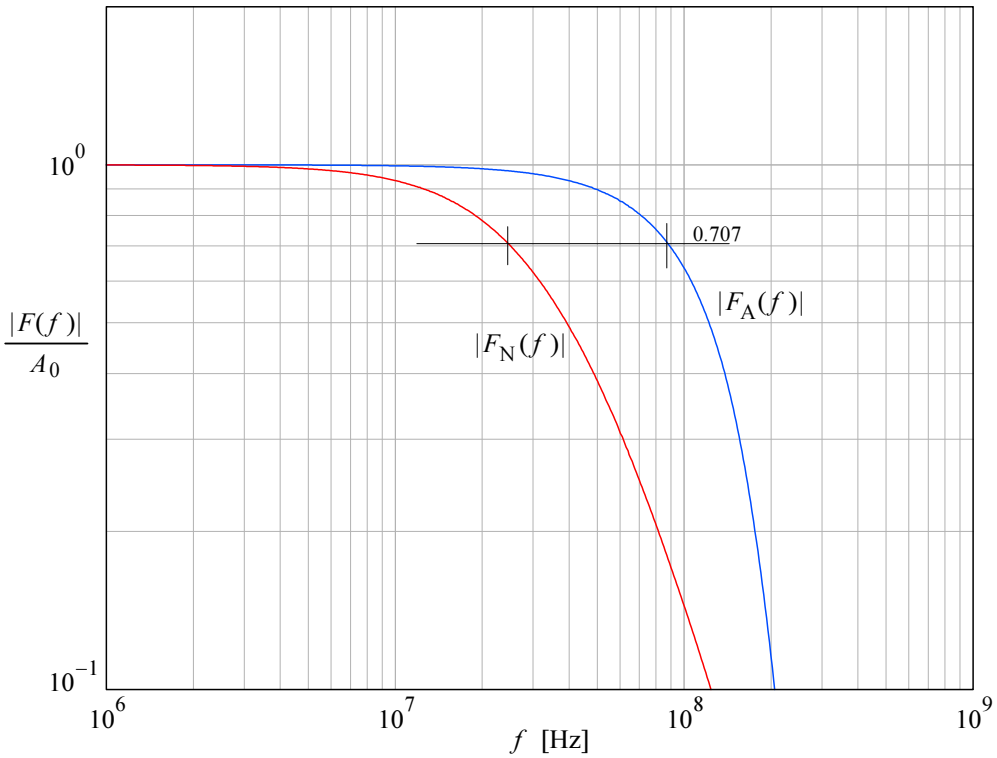


Fig. 5.1.6: The gain normalized magnitude vs. frequency of the 7-pole compensated system $|F_A(f)|$ and the 2-pole non-compensated system, $|F_N(f)|$. The bandwidth of F_N is about 25 MHz and the bandwidth of F_A is about 88 MHz, more than 3.5 times larger.

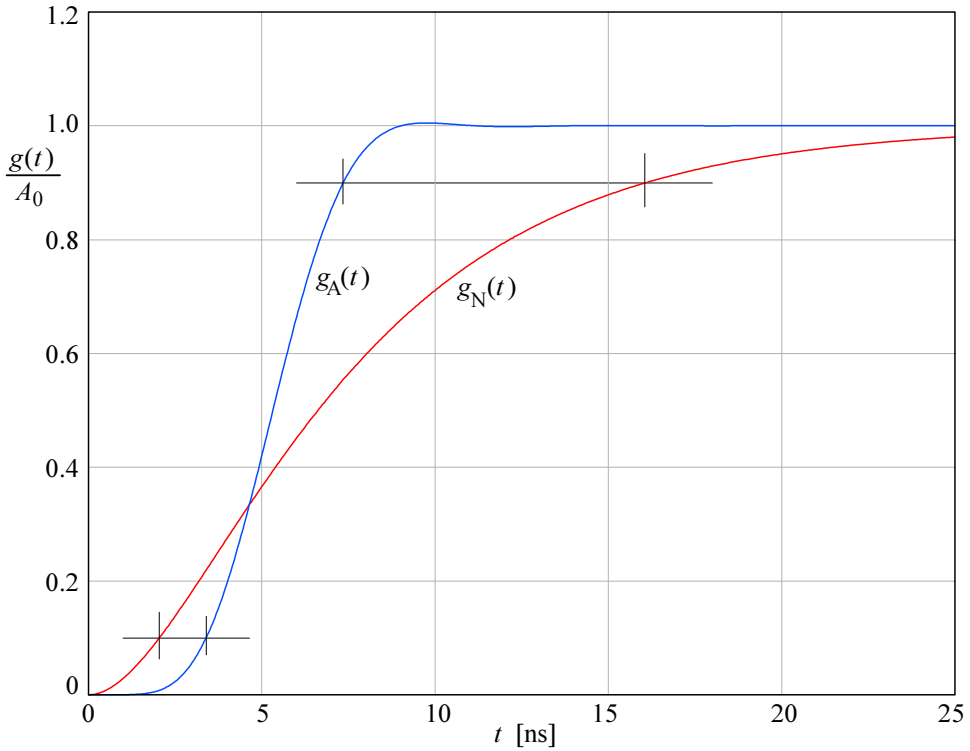


Fig. 5.1.7: The gain normalized step responses of the 7-pole compensated system $g_A(t)$ and the 2-pole non-compensated system $g_N(t)$. The rise time is 14 ns for the $g_N(t)$, but only 3.8 ns for $g_A(t)$. The overshoot of $g_A(t)$ is only 0.48 %.

6.5 Transient Response by Fourier Transform

There are several methods for time domain response calculation. Three of these that are interesting from the system designer's point of view, including the FFT method, were compared for efficiency and accuracy in [Ref. 6.23]. Besides the high execution speed, the main advantage of the FFT method is that we do not even have to know the exact mathematical expression for the system frequency response, but only the graph data (i.e. if we have measured the frequency and phase response of a system). Although the method was described in detail in [Ref. 6.23] we shall repeat here the most important steps, to allow the reader to follow the algorithm development.

There are **five difficulties** associated with the discrete Fourier transform that we shall have to solve:

a) the inability to transform some interesting functions (e.g., the unit step);

b) the correct treatment of the DC level in low pass systems;

c) preserving accuracy with as little spectral information input as possible;

d) find to what extent our result is an approximation owed to finite spectral density;

e) equally important, estimate the error owed to finite spectral length.

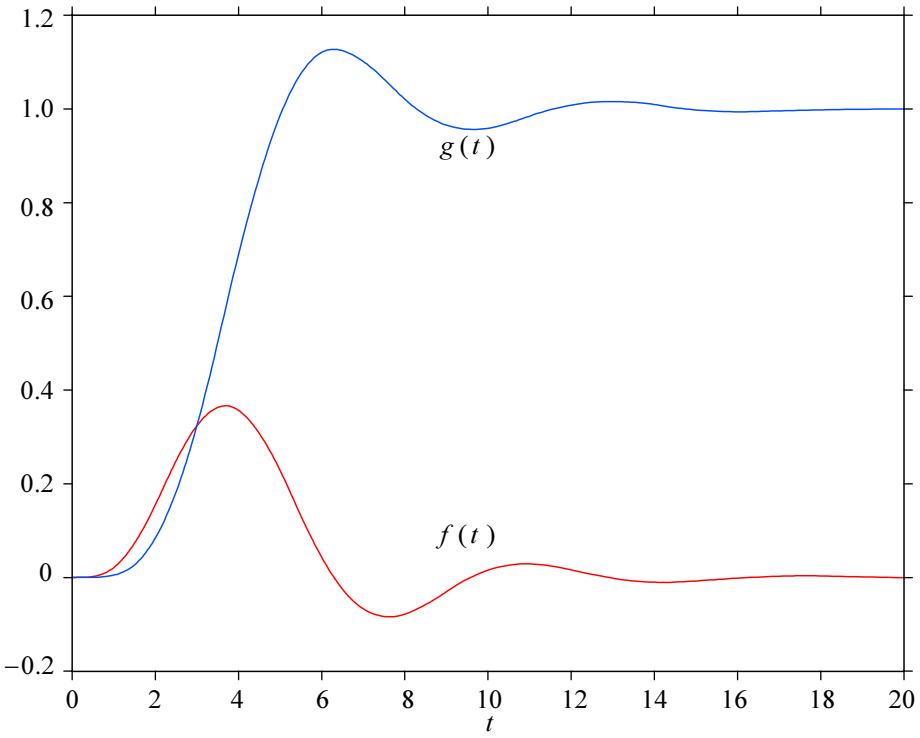


Fig. 6.5.1: The impulse and step response of the 5th-order Butterworth system. The impulse amplitude has been normalized to represent the response to an ideal, infinitely narrow, infinite amplitude input impulse. The impulse response reaches the peak value at the time equal to the envelope delay value at DC; this delay is also the half amplitude delay of the step response. The step response first crosses the final value at the time equal to the envelope delay maximum. Also the step response peak value is reached when the impulse response crosses the zero level for the first time. If the impulse response is normalized to have the area (the sum of all samples) equal to the system DC gain, the step response would be simply a time integral of it.

6.5.1 Impulse Response, Using FFT

The basic idea behind this method is that the Fourier transform is a special case of the more general Laplace transform and the Dirac impulse function is a special type of signal for which the Fourier transform solution always exists. Comparing Eq. 1.3.8 and Eq. 1.4.3 and taking in account that $s = \sigma + j\omega$, we see:

$$\mathcal{L}\{f(t)\} = \mathcal{F}\{f(t)\,\mathrm{e}^{-\sigma t}\} \tag{6.5.1}$$

Since the complex plane variable s is composed of two independent parts (real and imaginary), then $F(s)$ may be treated as a function of two variables, σ and ω. This can be most easily understood by looking at Fig. 6.4.1, in which the complex frequency response (magnitude) of a 5-pole Butterworth function is plotted as a 3D function over the Laplace plane.

In that particular case we had:

$$F(s) = \frac{-s_1 s_2 s_3 s_4 s_5}{(s - s_1)(s - s_2)(s - s_3)(s - s_4)(s - s_5)} \tag{6.5.2}$$

where $s_{1\text{–}5}$ have the same values as in the example at the beginning of Sec. 6.4.1.

When the value of s in Eq. 6.5.2 becomes close to the value of one of the poles, s_i, the magnitude $|F(s)|$ then increases until becoming infinitely large for $s = s_\mathrm{i}$.

Let us now introduce a new variable p such that:

$$p = s\,\Big|_{\sigma\,=\,0} \qquad \text{or:} \qquad p = j\omega \tag{6.5.3}$$

This has the effect of slicing the $|F(s)|$ surface along the imaginary axis, as we did in Fig. 6.4.1, revealing the curve on the surface along the cut, which is $|F(j\omega)|$, or in words: the magnitude $M(\omega)$ of the complex frequency response. As we have indicated in Fig. 6.4.5, we usually show it in a loglog scaled plot. However, for transient response calculation a linear frequency scale is appropriate (as in Fig. 6.4.2), since we need the result of the inverse transform in linear time scale increments.

Now that we have established the connection between the Laplace transformed transfer function and its frequency response we have another point to consider: conventionally, the Fourier transform is used to calculate waveform spectra, so we need to establish the relationship between a frequency response and a spectrum. Also we must explore the effect of taking **discrete values (sampling)** of the time domain and frequency domain functions, and see to what extent we **approximate** our **results** by taking **finite length vectors of finite density sampled data.** Those readers who would like to embed the inverse transform in a microprocessor controlled instrument will have to pay attention to **amplitude quantization (finite word length)** as well, but in Matlab this is not an issue.

We have examined the Dirac function $\delta(t)$ and its spectrum in Part 1, Sec. 1.6.6. Note that the spectral components are separated by $\Delta\omega = 2\pi/T$, where T is the impulse repetition period. If we allow $T \to \infty$ then $\Delta\omega \to 0$. Under these conditions we

can hardly speak of discrete spectral components because the spectrum has become very dense; we rather speak of **spectral density**. Also, instead of individual components' magnitude we speak of **spectral envelope** which for $\delta(t)$ is essentially flat.

However, if we do not have an infinitely dense spectrum, then $\Delta\omega$ is small but not 0, and this merely means that the impulse repeats after a finite period $T = 2\pi/\Delta\omega$ (this is the mathematical equivalent of testing a system by an impulse of a duration much shorter than the smallest system time constant and of a repetition period much larger than the largest system time constant).

Now let us take such an impulse and present it to a system having a selective frequency response. Fig. 6.5.2 shows the results both in the time domain and the frequency domain (magnitude). The time domain response is obviously the system impulse response, and its equivalent in the frequency domain is a spectrum, whose density is equal to the input spectral density, but with the spectral envelope shaped by the system frequency response. The conclusion is that we only have to sample the frequency response at some finite number of frequencies and perform a discrete Fourier transform inversion to obtain the impulse response.

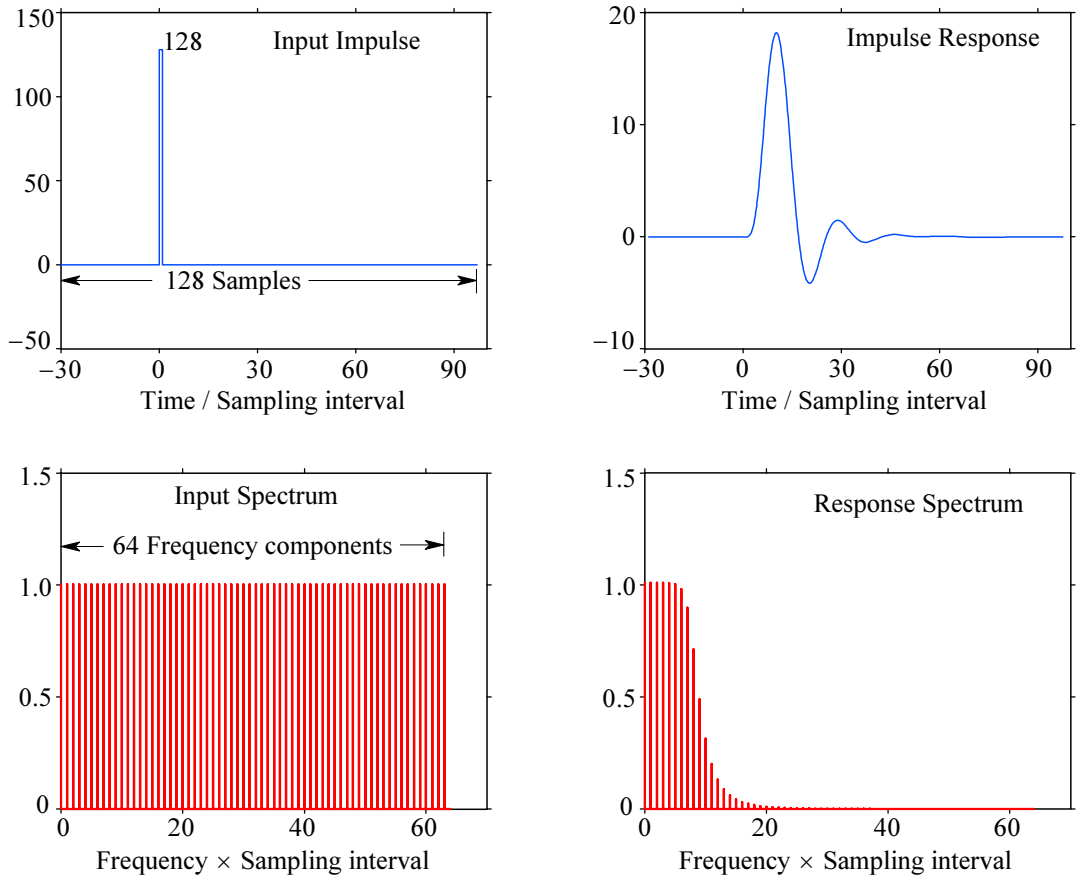


Fig. 6.5.2: Time domain and frequency domain representation of a 5-pole Butterworth system impulse response. The spectral envelope (only the magnitude is shown here) of the output is shaped by the system frequency response, whilst the spectral density remains unchanged. From this fact we conclude that the time domain response can be found from a system frequency response using inverse Fourier transform. The horizontal scale is the number of samples (128 in the time domain and 64 in the frequency domain — see the text for the explanation).

If we know the magnitude and phase response of a system at some finite number of equally spaced frequency points, then each point represents:

$$F_i = M_i \cos(\omega_i t - \varphi_i) \tag{6.5.4}$$

As the contribution of frequencies components which are attenuated by more than, say, 60 dB can be neglected, we do not have to take into account an infinitely large number of frequencies, and the fact that we do not have an infinitely dense spectrum merely means that the input impulse repeats in time. By applying the superposition theorem, the output is then equal to the sum of all the separate frequency components.

Thus for each time point the computer must perform the addition:

$$f(t_k) = \sum_{i(\omega_{\min})}^{i(\omega_{\max})} M_i \cos(\omega_i t_k - \varphi_i) \tag{6.5.5}$$

Eq. 6.5.5 is the **discrete Fourier transform**, with the exponential part expressed in trigonometric form. However, if we were to plot the response calculated after Eq. 6.5.5, we could see that the time axis is reversed, and from the theory of Fourier transform properties (symmetry property, [Ref. 6.14, 6.15, 6.18]), we know that the application of two successive Fourier transforms returns the original function but with the sign of the independent variable reversed:

$$\mathcal{F}\Big\{\mathcal{F}\{f(t)\}\Big\} = \mathcal{F}\{F(j\omega)\} = f(-t) \tag{6.5.6}$$

or more generally:

$$f(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} F(j\omega) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} f(-t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} F(-j\omega) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} f(t) \tag{6.5.7}$$

The main drawback in using Eq. 6.5.5 is the high total number of operations, because there are three input data vectors of equal length (ω, M, φ) and each contributes to every time point result. It seems that greater efficiency might be obtained by using the input frequency response data in the complex form, with the frequency vector represented by the index of the $F(j\omega)$ vector.

Now $F(j\omega)$ in its complex form is a two sided spectrum, as was shown in Fig. 6.4.3, and we are often faced with only a single sided spectrum. It can be shown that a real valued $f(t)$ will always have $F(j\omega)$ symmetrical about the real axis σ. Thus:

$$F_N(j\omega) = F_P^*(-j\omega) \tag{6.5.8}$$

F_N and F_P are the $\omega < 0$ and $\omega > 0$ parts of $F(j\omega)$, with their inverse transforms labeled $f_N(t)$ and $f_P(t)$. Note that F_P^* is the complex conjugate of F_P.

This symmetry property follows from the definition of the **negative frequency concept**: instead of having a single phasor rotating counter-clockwise (positive by definition) in the complex plane, we can always have two half amplitude phasors rotating in opposite directions at the same frequency (as we have already seen drawn in Part 1, Fig. 1.1.1; for vector analysis see Fig. 6.5.3**)**. We can therefore conclude that the

inherent conjugate symmetry of the complex plane allows us to define 'negative frequency' as a clockwise rotating, half amplitude phasor, being the complex conjugate of the usual counter-clockwise (positive by definition) rotating (but now also half amplitude) phasor. And this is not just a fancy way of making simple things complex, but is rather a direct consequence of our dislike of sine–cosine representation and the preference for the complex exponential form, which is much simpler to handle analytically.

One interesting aspect of the negative frequency concept is the ***Shannon* sampling theorem**: for a continuous signal, sampled with a frequency f_s, all the information is contained within the frequency range between 0 and $f_s/2$, because the spectrum from $f_s/2$ to f_s is a mirror image, so the spectrum is symmetrical about $f_s/2$, the *Nyquist* frequency. Therefore a frequency equal to f_s can not be distinguished from a DC level, and any frequency from $f_s/2$ to f_s can not be distinguished from $f_s - f$.

But please, also note that this 'negative frequency' does not necessarily imply 'negative time', since the negative time is defined as the time before some arbitrary instant $t = 0$ at which the signal was applied. In contrast, the negative frequency response is just one half of the full description of the $t \geq 0$ signal.

However, those readers who are going explore the properties of the *Hilbert transform* will learn that this same concept can be extended to the $t < 0$ signal region, but this is beyond the scope of this text.

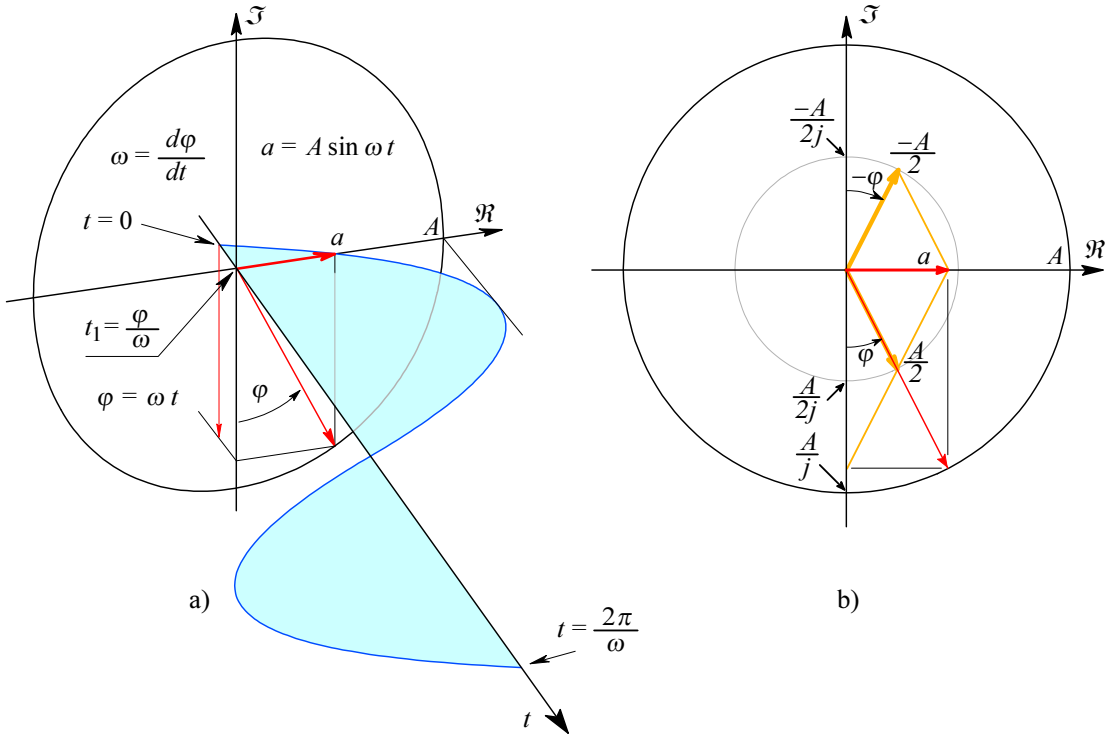


Fig. 6.5.3: As in Part 1, Fig. 1.1.1, but from a slightly different perspective: **a)** the real signal instantaneous amplitude $a(t) = A \sin \varphi$, where $\varphi = \omega t$; **b)** the real part of the instantaneous signal phasor, $\Re\{\vec{0A}\} = \vec{0a} = A \sin \varphi$, can be decomposed into two half amplitude, oppositely rotating, complex conjugate phasors, $-(A/2j) \sin \varphi + (A/2j) \sin(-\varphi)$. The second term has rotated by $-\varphi = -\omega t$ and, since t is obviously positive (see the **a)** graph), the negative sign is attributed to ω; thus, **clockwise rotation is interpreted as a 'negative frequency'**.

Eq. 6.5.8 can thus be used to give:

$$F(j\omega) = F_P(j\omega) + F_P^*(-j\omega) \tag{6.5.9}$$

but from Eq. 6.5.7 we have:

$$\mathcal{F}^{-1}\Big\{F_P^*(-j\omega)\Big\} = f_P^*(t) \tag{6.5.10}$$

hence using Eq. 6.5.9 and Eq. 6.5.10 and taking into account the cancellation of imaginary parts, we obtain:

$$f(t) = f_P(t) + f_P^*(t) = 2\,\Re\Big\{f_P(t)\Big\} \tag{6.5.11}$$

Eq. 6.5.11 means that if $F_P(j\omega)$ is the $\omega \geq 0$ part of the Fourier transformed real valued function $f(t)$, its Fourier transform inversion $f_P(t)$ will be a complex function whose real part is equal to $f(t)/2$. Summing the complex conjugate pair results in a doubled real valued $f(t)$. So by Eq. 6.5.10 and Eq. 6.5.11 **we can calculate the system impulse response from just one half of its complex frequency response using the forward Fourier transform** (not inverse!):

$$\boxed{f(t) = 2\,\Re\left\{\left[\mathcal{F}\Big\{F_P^*(j\omega)\Big\}\right]^*\right\}} \tag{6.5.12}$$

Note that the second (outer) complex conjugate is here only to satisfy the mathematical consistency — in the actual algorithm it can be safely omitted, since only the real part is required.

As the operator $\mathcal{F}\{\}$ in Eq. 6.5.12 implies integration we must use the **discrete Fourier transform (DFT)** for computation. The DFT can be defined by decomposing the Fourier transform integral into a finite sum of N elements:

$$\boxed{F(k) = \frac{1}{N}\sum_{i=0}^{N-1} f(i)\,\mathrm{e}^{-j\frac{2\pi k}{N}}} \tag{6.5.13}$$

That means going again through a large number of operations, comparable to Eq. 6.5.5. Instead we can apply the **Fast Fourier Transform (FFT)** algorithm and, owing to its excellent efficiency, save the computer a great deal of work.

Cooley and *Tukey* [Ref. 6.16] have shown that if $N = 2^B$ and B integer, there is a smart way to use Eq. 6.5.13, owing to the periodical nature of the Fourier transform.

If Eq. 6.5.13 is expressed in a matrix form then the matrix which represents its exponential part can be divided into its even and odd part, and the even part can be assigned to $N/2$ elements. The remaining part can then also be divided as before and the same process can then be repeated over and over again, so that we end up with a number ($\log_2 N$) of individual sub-matrices. Furthermore, it can be shown that these sub-matrices contain only two non-zero elements, one of which is always unity (1 or j).

Therefore multiplying by each of the factor matrices requires only N complex multiplications.

Finally (or firstly, depending on whether we are using the 'decimation in frequency' or the 'decimation in time' algorithm), we rearrange the data, by writing the position of each matrix element in a binary form, and reordering the matrix it by reversing the binary digits (this operation is often refered to as the 'reshuffling').

The total number of multiplications is thus $N \log_2 N$ instead of the N^2 required to multiply in one step. Other operations are simple and fast to execute (addition, change of sign, change of order). Thus in the case of $B = 10$, $N = 1024$, and $N^2 = 1048576$ whilst $N \log_2 N = 10240$, so a reducction of the required number of multiplications by **two orders of magnitude** has been achieved.

Matlab has a command named 'FFT' which uses the 'radix-2' type of algorithm and we shall use it as it is. Those readers who would like to implement the FFT algorithm for themselves can find the detailed treatment in [Ref. 6.16, 6.17 and 6.19].

A property of the FFT algorithm is that it returns the spectrum of a real valued signal as folded about the Nyquist frequency (one half of the frequency at which the signal was sampled). As we have seen in Fig. 6.5.2, if we have taken 128 signal samples, the FFT returns the first 64 spectral components from $\omega = 0, 1, 2, \ldots, 63$ but then the remaining 64 components, which are the complex conjugate of the first ones, are in the reversed order.

This is in contrast to what we were used to in the analytical work, as we expect the complex conjugate part to be on the $\omega < 0$ side of the spectrum. On the other hand, this is equivalent, since the 128 samples taken in the signal time domain window are implicitly assumed to repeat, and consequently the spectrum is also repeating every 128 samples. So if we use the standard inverse FFT procedure we must take into account all 128 spectral components to obtain only 64 samples of the signal back. However, note that the Eq. 6.5.12 **requires only a single–sided spectrum of N points to return N points of the impulse response**. This is, clearly, an **additional two–fold improvement** in algorithm efficiency.

7.1 Using Convolution: Response to Arbitrary Input Waveforms

7.1.1 From Infinitesimal to Discrete Time Integration

The time-domain algorithms that we have developed in Part 6 gave us the system response to two special cases of input signal: the unit-area impulse and the unit-amplitude step. Here we will consider the response to any type of input signal, provided that its application will not exceed neither the input nor the output system capabilities. In technical literature this is known as the BIBO-condition[1]. And, of course, we are still within the constraints of our initial LTIC-conditions[2].

As we have seen in Part 1, Sec. 1.14, the system's time domain response to an arbitrary input signal can be calculated in two ways:

a) by transforming the input signal to complex frequency domain, multiplying it by the system transfer function and transforming the result back to the time domain;

b) directly in time domain by the convolution integral.

A short reminder of the convolution integral definition and the transcription from differential to difference form is in order here. Let $x(t)$ be the time domain signal, presented to the input of a system being characterized by its impulse response $f(t)$. The system output can then be calculated by convolving $x(t)$ with $f(t)$:

$$y(t) = \int_{t_0}^{t_1} f(\tau - t)\, x(t)\, dt \qquad (7.1.1)$$

where τ is a fixed time constant, its value chosen so that $f(t)$ is time reversed. Usually, it is sufficient to make τ large enough to allow the system impulse response $f(t)$ to completely relax and reach the steady state again (not just the first zero-crossing point!).

If $x(t)$ was applied to the system at t_0, then this can be the lower limit of integration. Of course, the time scale can always be renormalized so that $t_0 = 0$. The upper integration limit, labeled t_1, can be wherever needed, depending on how much of the input and output signal we are interested in.

Now, in Eq. 7.1.1 dt is implicitly approaching zero, so there would be an infinite number of samples between t_0 and t_1. Since our computers have a limited amount of memory (and we have a limited amount of time!) we must make a compromise between the sampling rate and the available memory length and adjust them so that we cover the signal of interest with enough resolution in both time and

[1] Bounded input → bounded output. This property is a consequence of our choice of basic mathematical assumptions; since our math tools were designed to handle an infinite amount of infinitesimal quantities, BIBO is the necessary condition for convergence. However, in the real analog world, we are often faced with UBIBO requirements (unbounded input), i.e., our instrumentation inputs must be protected from overdrive. Interestingly, the inverse of BIBO is in widespread use in the computer world, in fact, any digital computer is a GIGO type of device (garbage in → garbage out; unbounded!).

[2] Linearity, Time Invariance, Causality. Although some engineers consider oscillators to be 'acausal', there is always a perfectly reasonable cause why an amplifier oscillates, even if we fail to recognise it at first.

amplitude. So if M is the number of memory bytes reserved for $x(t)$, the required sampling time interval is:

$$\Delta t = \frac{t_1 - t_0}{M} \tag{7.1.2}$$

Then, if Δt replaces dt, the integral in Eq. 7.1.1 transforms into a sum of M elements, $x(t)$ and $y(t)$ become vectors `x(n)` and `y(n)`, where `n` is the index of a signal sample location in memory, and $f(\tau - t)$ becomes `f(m-n)`, with `m=length(f)`, resulting in:

$$\texttt{y(n)} = \sum_{\texttt{n=1}}^{\texttt{M}} \texttt{f(m-n)*x(n)} \tag{7.1.3}$$

Here Δt is implicitly set to 1, since the difference between two adjacent memory locations is a unit integer. Good book-keeping practice, however, recommends the construction of a separate time scale vector, with values from t_0 to t_1, in increments of Δt between adjacent values. All other vectors are then plotted against it, as we have seen done in Part 6.

7.1.2 Numerical Convolution Algorithm

In Part 1 we have seen that solving the convolution integral analytically can be a time consuming task, even for a skilled mathematician. Sometimes, even if $x(t)$ and $f(t)$ are analytic functions, their product need not be elementarily integrable in the general case. In such cases we prefer to take the $\mathcal{L}$ transform route; but this route can sometimes be equally difficult. Fortunately numerical computation of the convolution integral, following Eq. 7.1.3, can be programmed easily:

```
function y=vcon(f,x)
%VCON   Convolution, step-by-step example. See also CONV and FILTER.
%
%      Call :       y=vcon(f,x);
%      where:       x(t) --> the input signal
%                   f(t) --> the system impulse response
%                   y(t) --> the system response to x(t) by convolving
%                            f(t) with x(t).
%      If  length(x)=nx  and  length(f)=nf, then length(y)=nx+nf-1.

%      Erik Margan, 861019, Last editing 890416; Free of copyright!

                          % force f to be the shorter vector :
if length(f) > length(x)
      xx=x; x=f; f=xx;    % exchange x and f via xx
      clear xx
end
nf=length(f);             % get the number of elements in x and f
nx=length(x);
f=f(:).';                 % organize x and f as single-row vectors
x=x(:).';
y=zeros(2,nx+nf-1);       % form a (2)-by-(nx+nf-1) matrix y, all zeros
y(1,1:nx)=f(1)*x;         % first row of y: multiply x by f(1)
for k=2:nf                % second row: multiply and shift (insert 0)
      y(2, k-1:nx+k-1)=[0, f(k)*x];
                          % sum the two rows column-wise and
      y(1,:)=sum(y);      % put result back into first row
end                       % repeat for all remaining elements of f;
y=y(1,:);                 % the result is the first row only.
```

To get a clearer view of what the VCON routine is doing, let us write a short numerical example, using a 6-sample input signal and a 3-sample system impulse response, and display every intermediate result of the matrix y in VCON:

```
 x=[0 1 3 5 6 6];    f=[1 3 -1];    vcon(x,h);

 % initialization - all zeros, 2 rows, 6+3-1 columns:
     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0
 % step 1: multiply x by the first sample of f, f(1)=1 and
 % insert it into the first row:
     0     1     3     5     6     6     0     0
     0     0     0     0     0     0     0     0
 % step 2: multiply x by the second sample of f, f(2)=3,
 % shift it one place to the right by adding a leading zero and
 % insert it into the second row:
     0     1     3     5     6     6     0     0
     0     0     3     9    15    18    18     0
 % step 3: sum both rows vertically, put the result in the first row
     0     1     6    14    21    24    18     0
     0     0     3     9    15    18    18     0
 % iterate steps 2 and 3, each iteration using the next sample of f:
     0     1     6    14    21    24    18     0
     0     0     0    -1    -3    -5    -6    -6

     0     1     6    13    18    19    12    -6
     0     0     0    -1    -3    -5    -6    -6
 % after 2 iterations (because f is only 3 samples long)
 % the result is the first row of y:
     0     1     6    13    18    19    12    -6
 % actually, the result is only the first 6 elements:
     0     1     6    13    18    19
 % since there are only 6 elements in x, the process assumes the rest
 % to be zeros. So the remaining two elements of the result represent
 % the relaxation from the last value (19) to zero by the integration
 % of the system impulse response f.

 % Basically, the process above does the following:
 % (note the reversed sequence of f)

                 0  1  3  5  6  6
          -1  3  1
(↓*) ==>         0                 (→+) ==> 0

                 0  1  3  5  6  6
             -1  3  1
(↓*) ==>         0  1              (→+) ==> 1

                 0  1  3  5  6  6
                -1  3  1
(↓*) ==>         0  3  3           (→+) ==> 6

                 0  1  3  5  6  6
                   -1  3  1
(↓*) ==>         0 -1  9  5        (→+) ==> 13

 % ......... etc.
```

For convolution Matlab has a function named CONV, which uses a built in FILTER command to run substantially faster, but then the process remains hidden from the user; however, the final result is the same as with VCON. Another property of Matlab is the matrix indexing, which starts with 1 (see the lower limit of the sum

symbol in Eq. 7.1.3), in contrast to most programming languages which use memory 'pointers' (base address + offset, the offset of the array's first element being 0).

7.1.3 Numerical Convolution Examples

Let us now use the VCON routine in a real life example. Suppose we have a gated sine wave generator connected to the same 5th-order Butterworth system which we inspected in detail in Part 6. Also, let the Butterworth system's half power bandwidth be 1 kHz, the generator frequency 1.5 kHz, and we turn on the gate in the instant the signal crosses the zero level. From the frequency response calculations, we know that the forced response amplitude (long after the transient) will be:

```
Aout=Ain*abs(freqw(z,p,1500/1000));
```

where `z` are the zeros and `p` are the poles of the normalized 5th-order Butterworth system; the signal frequency is normalized to the system's cut off frequency.

But how will the system respond to the signal's 'turn on' transient? We can simulate this using the algorithms we have developed in Part 6 and VCON:

```
fh=1000;                % system half-power bandwidth, 1kHz
fs=1500;                % input signal frequency, 1.5kHz
t=(0:1:300)/(50*fh);    % time vector, 20us delta-t, 6ms range
nt=length(t);

[z,p]=buttap(5);        % 5th-order Butterworth system
p=2*pi*fh*p;            % denormalized system poles
Ir=atdr(z,p,t,'n');     % system impulse-response

d=25;                   % switch-on delay, 25 time-samples
                        % make the input signal :
x=[zeros(1,d), sin(2*pi*fs*t(1:nt-d))];

y=vcon(Ir,x);           % convolve x with Ir ;

A=nt/(2*pi*fh*max(t)); % denormalize amplitude of Ir for plot

                        % plot the input, the system impulse response
                        % and the convolution result :
plot( t*fh, x, '-g', ...
      t*fh, [zeros(1,d), Ir(1:nt-d)*A],'-r', ...
      t*fh, y(1:nt), '-b')
xlabel('t [ms]')
```

The convolution result, compared to the input signal and the system impulse response, is shown in Fig. 7.1.1 .

Note that we have plotted only the first `nt` samples of the convolution result; however, the total length of `y` is `length(x)+length(Ir)-1`, or one sample less that the sum of the input signal and the system response lengths. The first `length(x)=nt` samples of `y` represent the system's response to `x`, whilst the remaining `length(Ir)-1` samples are the consequence of the system relaxation: since there are no more signal samples in `x` after the last point `x(nt)`, the convolution assumes that the input signal is zero and calculates the system relaxation from the last signal value.

This is equivalent to a response caused by an input step from `x(nt)` to zero. So if we are interested only in the system response to the input signal, we simply limit the response vector to the same length as was the input signal. Also, in the general case the length of the system's impulse response vector, `nr=length(Ir)` , does not have to be equal to the input signal vector length, `nx=nt`. In practice, we often make `nr<<nx` , but `Ir` should be made long enough to include the system relaxation to a level very close to zero, as only then will the sum of all elements of `Ir` not differ much from the system gain.

There is, however, an important difference in the plot and the calculation, that must be explained. The impulse response which we obtained from Butterworth system poles was normalized to represent a unity gain system, since we want to see the frequency dependence on the output amplitude by comparing the input and output signals. Thus our system should either have a gain of unity, or the output should be normalized to the input in some other way (e.g., if the gain was known, we could have divided the output signal by the gain, or multiplied the input signal). But the unity gain normalized impulse response would be too small in amplitude, compared to the input signal, so we have plotted the ideal impulse response.

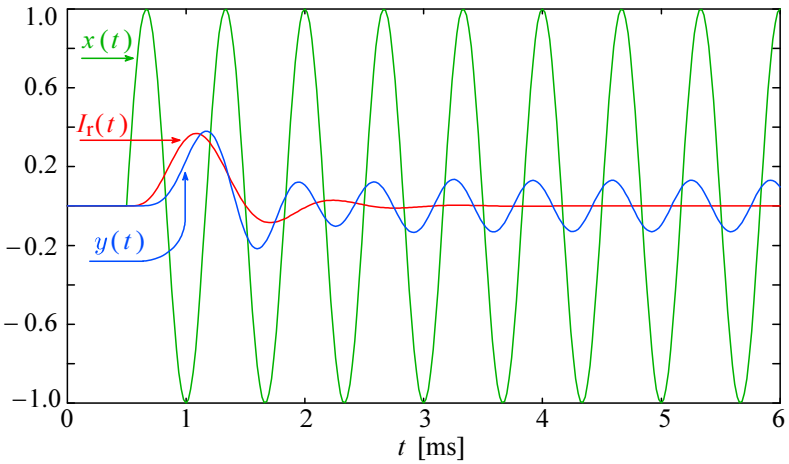


Fig. 7.1.1: Convolution example: response $y(t)$ to a sine wave $x(t)$ switched-on into a 5th-order Butterworth system, whose impulse-response is $I_r(t)$, shown here as the ideal response (instead of unity gain); both are delayed by the same switch–on time (0.5 ms). The system responds by phase shifting and amplitude modulating the first few wave periods, reaching finally the forced ('steady state') response.

Can we check whether our routine works correctly?

Apart from entering some simple number sequences as before, we can do this by entering an input signal for which we have already calculated the result in a different way, say, the unit step (see Fig. 6.5.1, Part 6). By convolving the impulse response with the unit step, instead of the sine wave, we should obtain the now known step response:

```
% continuing from above:
h=[zeros(1:d), ones(1:nt-d)];       % h(t) is the unit step function
y=vcon(Ir,h);                       % convolve h with Ir
plot( t*fh, h, '-g', ...
      t*fh, [zeros(1,d), Ir(1:nt-d)*A],'-r', ...
      t*fh, y(1:nt), '-b')
xlabel('t [ms]')
```

The resulting step response, shown in Fig. 7.1.2, should be identical to that of Fig. 6.5.1, Part 6, neglecting the initial 0.5 ms (25 samples) time delay and the different time scale:

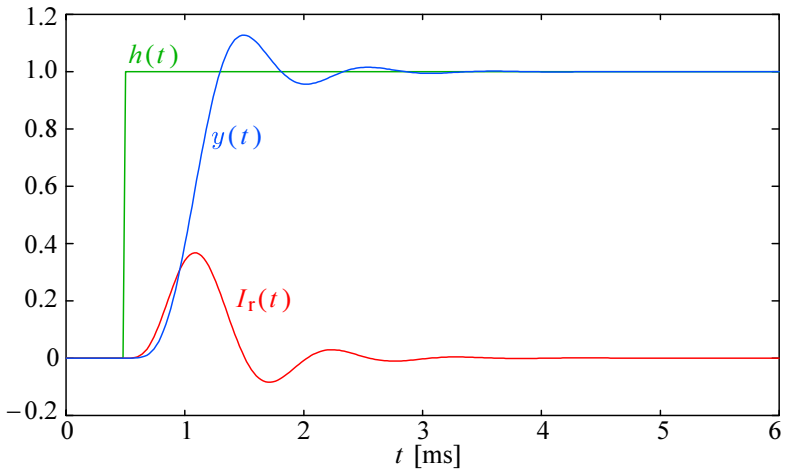


Fig. 7.1.2: Checking convolution: response $y(t)$ of the 5th-order Butterworth system to the unit step $h(t)$. The system's impulse response $I_r(t)$ is also shown, but in its ideal size (not unity gain). Apart from the 0.5 ms (25-samples) time delay and the time scale, the step response is identical to the one shown in Part 6, Fig.6.5.1.

We can now revisit the convolution integral example of Part 1, Sec. 1.15, where we had a unit-step input signal, fed to a two-pole Bessel-Thomson system, whose output was in turn fed to a two-pole Butterworth system. The commands in the following window simulate the process and the final result of Fig. 1.15.1. But this time, let us use the frequency to time domain transform of the TRESP (Part 6) routine. See the result in Fig. 7.1.3 and compare it to Fig. 1.15.1g.

```
[z1,p1]=bestap(2,'t'); % Bessel-Thomson 2nd-order system poles
[z2,p2]=buttap(2);     % Butterworth 2nd-order system poles
N=256;                 % number of samples
m=4;                   % set the bandwidth factor
w=(0:1:N-1)/m;         % frequency vector, w(m+1)=1 ;
F1=freqw(p1,w);        % Bessel-Thomson system frequency response
F2=freqw(p2,w);        % Butterworth system frequency response

[S1,t]=tresp(F1,w,'s');% step-response of the Bessel-Thomson system
I2=tresp(F2,w,'u');    % unity-gain Butterworth inpulse response ;
                       % both have the same normalized time vector
d=max(find(t<=15));    % limit the plot to first 15 time units
I2=I2(1:d);            % limit the I2 vector length to d

     % convolution of Bessel-Thomson system step-response with
     % the first d points of the Butterworth impulse response :
y=vcon(I2,S1);

A=N/(2*pi*m*max(t));   % amplitude denormalization for I2
     % plot first d points of all three responses vs. time :
plot( t(1:d), S1(1:d), '-r',...
      t(1:d), I2(1:d)*A, '-g',...
      t(1:d),  y(1:d), '-b' )
xlabel('Time [s]')
```

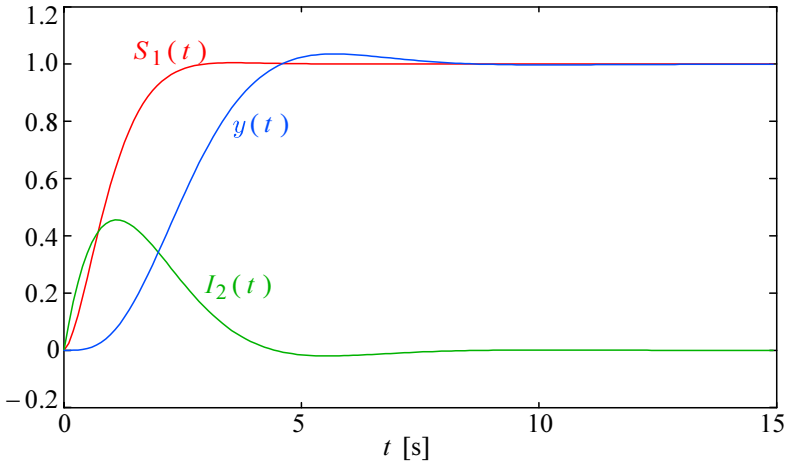


Fig. 7.1.3: Convolution example of Part 1, Sec. 1.15. A Bessel–Thomson 2-pole system step response $S_1(t)$ has been fed to the 2-pole Butterworth system and convolved with its impulse response $I_2(t)$, resulting in the output step response $y(t)$. Compare it with Fig. 1.15.1g.

The VCON function is a lengthy process. On a 12 MHz AT-286 PC, which was used for the first experiments back in 1986–7, it took more than 40 s to complete the example shown above, but even with today's fast computers there is still a noticeable delay. The Matlab CONV routine is much faster.

The reader might question the relevance of the total calculation time, since, for the purpose of a circuit design aid, anything below 1 s should be acceptable (this is comparable with the user's reaction time, when making a simple go/no go assessment of the result). However, imagine an automated optimization program loop, adjusting the values of some 10–20 circuit components. Such a loop might take hundreds or even thousands of executions before reaching satisfactory performance criteria, so a low routine time would be welcome. Moreover, if the routine will eventually be implemented in hardware, acquiring and processing the signal in real time, a low routine time is of vital importance. For example, to continuously process a 16 bit stereo audio stream, divided into 1024 sample chunks, using a 32 word long filter impulse response, the total routine calculation time should be less than 250 μs.

In some cases, particularly with long signal sequences (`N > 1000`), it could be interesting to take the Fourier transform route, numerically.

Here is an example using a signal recorded by a nuclear magnetic resonance imaging (MRI) system. The MRI RF signal is very weak (< 1 mV), so the detection is noisy and there is some interference from another source. We shall try to clean it by using a 5th-order Bessel-Thomson digital filter with a unity gain and a 1 MHz cut off:

```
load R.dat % load the recorded signal from a file "R.dat"
N=length(R);            % total vector length, N=2048 samples
Tr=102.4e-6;            % total record time 102.4 us
dt=Tr/N;                % sampling time interval, 50 ns
t=dt*(0:1:N-1);         % time vector reconstruction

% plot the first 1200 samples of the recorded signal
plot(t(1:1200,R(1:1200),'-g')
xlabel('Time [\mus]')  % input signal R, fist 60us, see Fig.7.1.4
G=fft(R);               % G is the FFT spectrum of R
G=G(1:N/2);             % use only up to the Nyquist freq. ( 10 MHz )
f=(1:1:N/2)/dt;         % frequency vector reconstructed
```

```
[z,p]=bestap(5,'n');    % 5th-order Bessel filter poles
p=p*2*pi*1e+6;          % half-power bandwidth is 1 MHz
F=freqw(z,p,2*pi*f);    % filter frequency response

% multiplication in frequency is equal to convolution in time:
Y=F.*G;                 % output spectrum

x=max(find(f<=8e+6));   % plot spectrum up to 8 MHz
M=max(abs(G));          % normalize the spectrum to its peak value
plot( f(1:x), abs(F(1:x)),    '-r', ...
      f(1:x), abs(G(1:x))/M, '-g', ...
      f(1:x), abs(Y(1:x))/M, '-b' )
xlabel('Frequency [MHz]'), ylabel('Normalized Magnitude')
                        % see Fig.7.1.5

y=2*(real(fft(conj(Y)))-1)/(N/2);  % return to time domain

a=max(find(t<=5e-5));
b=min(find(t>=20e-6));
plot( t(a:b), g(a:b), '-g', t(a:b), y(a:b), '-b' )
xlabel('Time [\mus]')   % see Fig.7.1.6
```

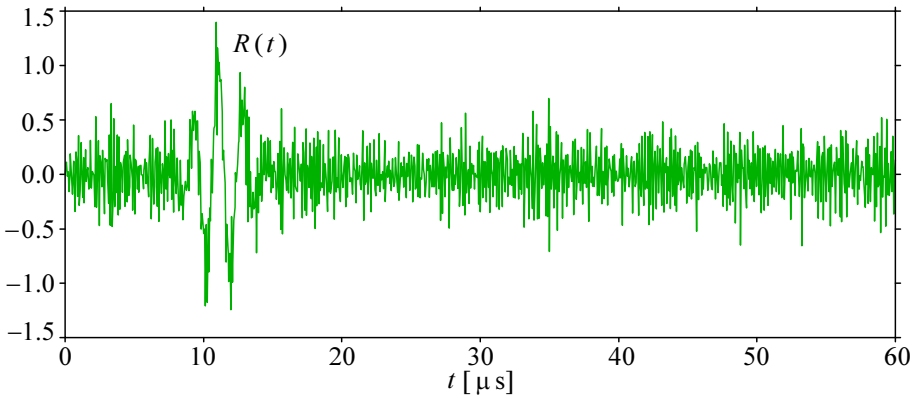


Fig. 7.1.4: Input signal example used for the spectral-domain convolution example (first 1200 samples of the 2048 total record length)

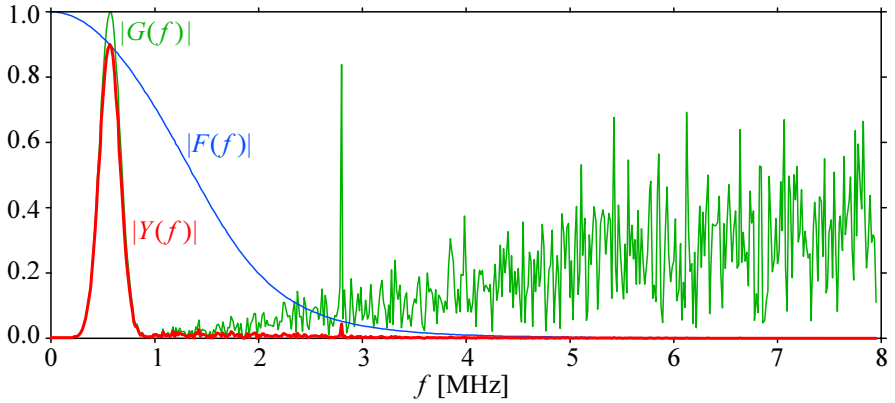


Fig. 7.1.5: The spectrum $G(f)$ of the signal in Fig. 7.1.4a is multiplied by the system's frequency response $F(f)$ to produce the output spectrum $Y(f)$. Along with the modulated signal centered at 560 kHz, there is a strong 2.8 MHz interference from another source and a high level of white noise (rising with frequency), both being substantially reduced by the filter.

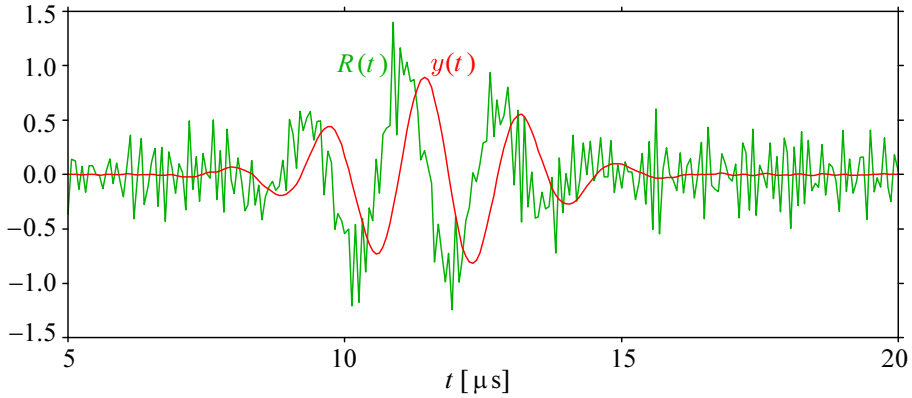


Fig. 7.1.6: The output spectrum is returned to time domain as $y(t)$ and is compared with the input signal $R(t)$, in expanded time scale. Note the small change in amplitude, the reduced noise level and the envelope delay (approx. $1/4$ period time shift), with little change in phase. The time shift is equal to $1/2$ the number of samples of the filter impulse response.

Fig. 7.1.6 illustrates the dramatic improvement in signal quality that can be achieved by using Bessel filters.

In MRI systems the test object is put in a strong static magnetic field. This causes the nucleons of the atoms in the test object to align their magnetic spin to the external field. Then a short RF burst, having a well defined frequency and duration, is applied, tilting the nucleon spin orientation perpendicular to the static field (this happens only to those nucleons whose resonant frequency coincides with that of the RF burst).

After the RF burst has ceased, the nucleons gradually regain their original spin orientation in a top-like precessing motion, radiating away the excess electromagnetic energy. This EM radiation is picked up by the sensing coils and detected by an RF receiver; the detected signal has the same frequency as the excitation frequency, both being the function of the static magnetic field and the type of nucleons. Obviously the intensity of the detected radiation is proportional to the number of nucleons having the same resonant frequency[3].

In addition, since the frequency is field dependent a small field gradient can be added to the static magnetic field, in order to split the response into a broad spectrum. The shape of the response spectral envelope then represents the spatial density of the specific nucleons in the test object. By rotating the gradient around the object the recorded spectra would represent the ‘sliced view’ of the object from different angles. A computer can be used to reconstruct the volumetric distribution of particular atoms through a process called ‘back-projection’ (in effect, a type of spatial convolution).

From this short description of the MRI technique it is clear that the most vital parameter of the filter, applied to smooth the recorded signal, is its group delay flatness. Only a filter with a group delay being flat well into the stop band will be able to faithfully deliver the filtered signal, preserving its shape both in the time and the frequency domain and Bessel–Thomson filters are ideal in this sense. Consequently a sharper image of the test object is obtained.

[3]The 1952 Nobel prize for physics was awarded to *Felix Bloch* and *Edward Mills Purcell* for their work on nuclear magnetic resonance; more info at <http://nobelprize.org/physics/laureates/1952/>.

http://www.springer.com/978-0-387-28340-1

Wideband Amplifiers

Staric, P.; Margan, E.

2006, XVII, 630 p. 360 illus., 313 illus. in color.,
Hardcover
ISBN: 978-0-387-28340-1