

June 5, 2013

## **CONTROLLING A LOW-PASS PEAK GAIN**

### **INTRODUCTION**

It is usually easy enough to design a fixed active filter. That is, as long as the parameters of the filter (usually cutoff frequency, damping, and gain) need not be variable, we envision few problems. If it is the case that the actual filter we build does not approximate the desired response closely enough, we usually can “tweak” it to adjust for tolerances of the passive components and/or bandwidth limitations of any active components [1].

Problems can arise when we need a tunable filter. Typically it is the cutoff frequency that we need to adjust. If the filter is of low order (such as 2) we only need a dual pot to do this. As long as the frequency range needed is fairly low (say under 50 kHz) we expect success. If asked to design a 2<sup>nd</sup>-order Butterworth with a cutoff that varies between 200 Hz and 20 kHz, we might well take out a sheet of paper and draw the circuit on the spot.

If we were asked to change the gain or the damping however, we would be forced in general to consider certain trade-offs. Not that we can't get what we want, but these two parameters will likely interact. Often a higher Q (lower damping) means that the gain increases. If we were looking for a fixed peak gain, we would need an attenuation.

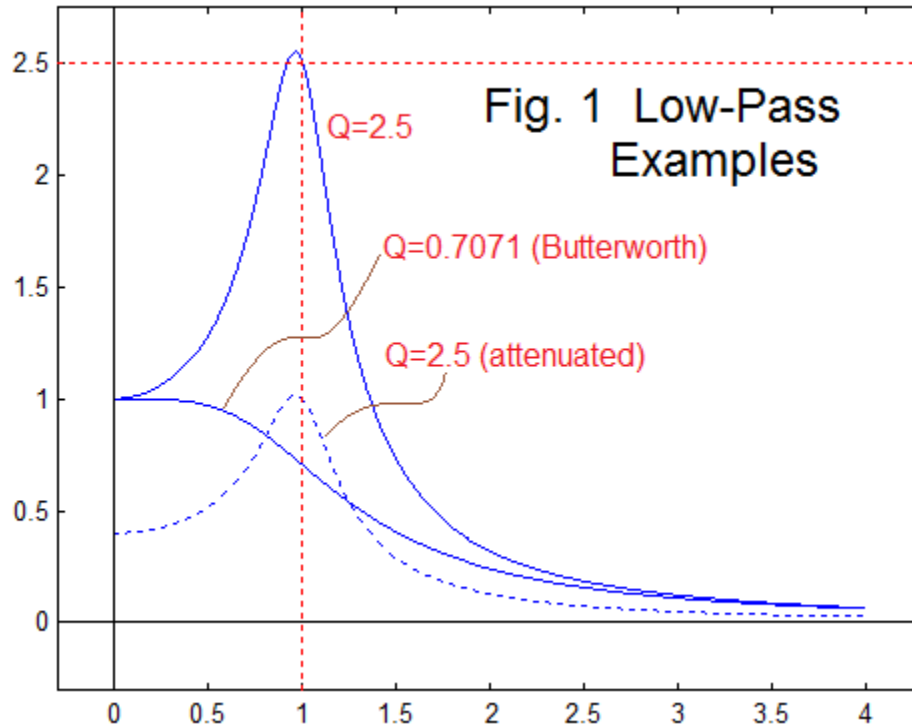
### **WHAT IS THE DC GAIN AND WHAT IS THE PEAK GAIN?**

Here we concentrate on a 2<sup>nd</sup>-order filter section, and assume a normalized pole radius of 1. This should sufficiently illustrate the problems and potential solutions. Accordingly a 2<sup>nd</sup>-order low-pass transfer function would be simply:

$$T(s) = \frac{1}{s^2 + Ds + 1} \quad (1)$$

Here D is the “damping” or the inverse of “Q”. It is trivial to see that the DC gain of this section is 1. All we need to do is substitute s=0 into equation (1). That's the easy part of course.

Let's jump ahead and look at some example frequency responses (Fig. 1).



Where did all this come from? First of all, Fig. 1 shows two filters (solid blue) one for a Q of  $1/\sqrt{2}=0.7071$ , a “Butterworth” response, and the other for a Q of 2.5, which shows resonance. The actual plots were generated with the Matlab **freqs** function. Lacking this, it is always true that the magnitude of the frequency response of  $T(s)$  is:

$$|T(s)| = [T(j\omega)T(-j\omega)]^{1/2} \quad (2)$$

which we can calculate and solve for  $\omega=0$  to 4 here. We have also plotted the  $Q=2.5$  result divided by Q (the dashed blue line) for comparison – note that this peaks close to 1. The gain of the Butterworth at  $\omega=1$  is 0.7071 or almost exactly -3db. At  $\omega=1$  note that the gain of the  $Q=2.5$  is NOT the peak (but close to it) and the peak gain is not Q (but close to it). If we wanted the more resonate low-pass ( $Q=2.5$ ) to have a peak near 1, it is necessary that the DC gain is reduced (dashed blue).

What is at issue here is illustrated by the difference between the solid blue curve and the dashed blue curve for  $Q=2.5$ . Why would we care much – they are just scaled versions of each other? Well, first of all, the gain near the cutoff frequency might mean that a signal at that frequency, input with high gain, could well clip against the power supply limits. Alternatively, frequencies below this frequency are going to be attenuated.

Consider a broadbanded signal at the input. For example, it might be a narrow pulse with a frequency 0.1, so there would be a whole bunch of harmonics like 0.1, 0.2, 0.3, and so on, of which we would expect a Butterworth low-pass to substantially pass the first 10 or so, and tend to reject those above. The low-pass with a Q of 2.5, solid blue, would treat the signal much the same, except perhaps the harmonics at say from 0.8 to 1.2, which would be amplified (with the possibility of clipping as mentioned). The dashed curve with the same frequency response would be less loud overall. But the issue is complicated as upper harmonics can have a much stronger effect on the impression of loudness than the lower ones. For example, a sinewave at a certain amplitude might sound relatively soft, but if we make it instead a square wave at the same amplitude, it can sound many times louder, subjectively. For similar reasons, when using a VCF for music synthesis, a change of Q that results in a gain increase might result in some clipping as well, and this adds to the perceived spectral modifications, often in a musically valid way, but to the embarrassment of the engineer!

Keep in mind that the use of low-pass sections with Q's much greater than Butterworth, or much less than Butterworth for that matter, are mainly to allow the construction of higher-order, but still relatively flat overall responses [2]. In our music synthesis work, however, we have a lot of use for sharper filters with moving characteristics in creating dynamically varying spectra for our sounds.

Above we noted that for the case of Q=2.5 that the peak was close to a frequency of 1, and that the peak amplitude was close to Q. If Q is as high as, say 10, then this is a virtual identity. The exact frequency of the peak and the exact amplitude of the peak will be calculated below in an appendix. But right here note that if we accept that the peak is close to a frequency of 1 (that is,  $s=j$ ), it is clear from just equation (1) that the magnitude is Q, since:

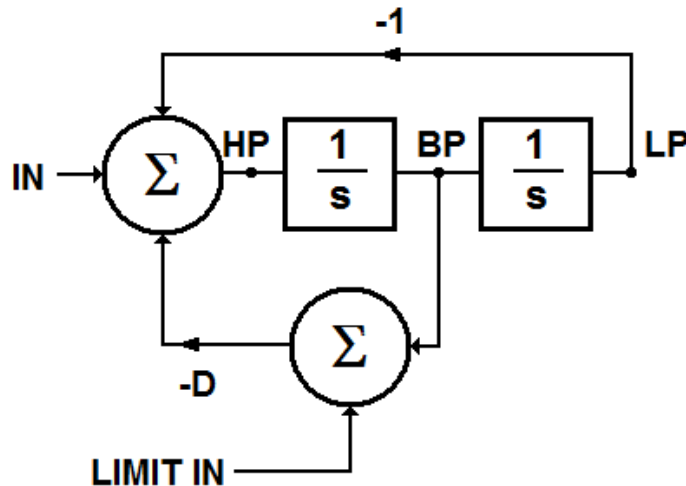
$$|T(s = j)| = \left| \frac{1}{j^2 + Dj + 1} \right| = \frac{1}{D} = Q \quad (3)$$

So the magnitude at a frequency 1 is always Q, and for large Q, this is also the peak magnitude.

So the finding is that for most values of Q that result in significant peaking, the peak is of magnitude approximately equal to Q, and that correcting this peak gain back to 1 is a matter of multiplying by  $D = 1/Q$ . That is, as in going from the solid blue curve for Q=2.5 in Fig. 1 to the dashed blue curve. It turns out that with the familiar state-variable filter, this can offer us a convenient alternative input point where the peak gain remains 1 for a varying Q. The idea is not new.

## THE STATE-VARIABLE WITH A "LIMIT" INPUT

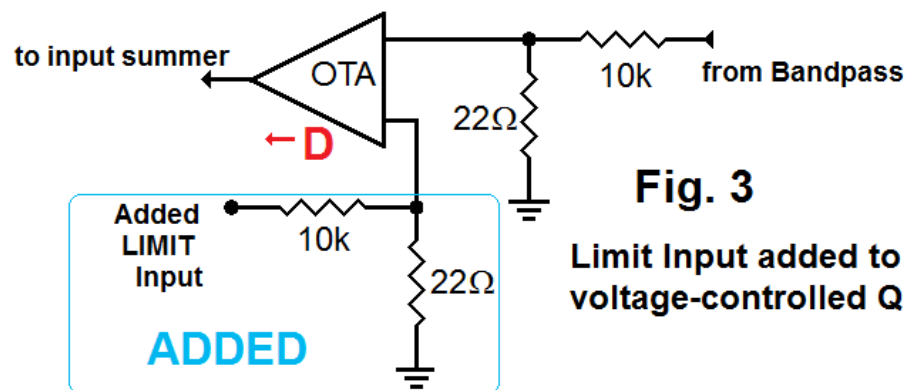
Fig. 2 shows the modified state-variable. We have simply added the lower summer as an alternative place to put the input. Because the lower summer cascades into the normal upper one, this is the the same thing, except note that the output of the lower summer is fed through a multiplier  $-D$ . This is what we need to set the  $Q$  of the filter, but at the same time, conveniently, the input is also multiplied by the exact same  $D$ , achieving the attenuation we were looking for.



**Fig 2**

**State-Variable  
Showing the  
"LIMIT" Input**

As stated, the idea is not new. Unfortunately I am not that clear on exactly where it came from. As an explicit design feature, it appeared in 1976 in one of our VCF options for the ENS-76 series [3], in EN#71, and is also described in another article [4] in the same EN#71. This in turn leads back, as many of our best ideas have, to Terry Mikulic's work [5] in January of 1974. It also appeared in a somewhat later app note in 1979 [6]. It may appear in some of the earliest state-variable literature. What was slightly different in the music synthesis examples was the ease with which we could add this once we had decided to use a voltage-controlled  $Q$ . Essentially it was just a single resistor (or two) to an input terminal of the OTA used for the controlled  $Q$  (Fig. 3)



**Fig. 3**

**Limit Input added to  
voltage-controlled  $Q$**

## APPENDIX – MORE EXACT CALCULATIONS

Although the simplified calculations discussed above are probably adequate for many purposes, here we will do the exact calculations, which have been published in our app notes series long ago [7,8]. We begin with equation (1) above. Completing the calculation suggested by equation (2), but leaving it in a squared-magnitude form, we have:

$$|T(s)|^2 = [T(j\omega)T(-j\omega)] = \frac{1}{\omega^4 + (D^2 - 2)\omega^2 + 1} \quad (4)$$

To find the maximum of  $|T(s)|$  it is convenient to minimize the denominator of equation (4) and we can call this denominator  $P(\omega)$ :

$$P(\omega) = \omega^4 + (D^2 - 2)\omega^2 + 1 \quad (5)$$

and then

$$\frac{dP(\omega)}{d\omega} = 4\omega^3 + 2(D^2 - 2)\omega = 0 \quad (6)$$

gives us (in addition to  $\omega=0$ ):

$$\omega_{max} = \sqrt{1 - D^2/2} \quad (7)$$

Using equation 4 (taking the square root for  $|T(s)|$ ) and plugging in  $\omega_{max}$  we arrive at the peak value of the response as:

$$|T(s)|_{max} = \left(\frac{2}{D}\right) \sqrt{\frac{1}{4 - D^2}} \quad (8)$$

If we have not made any math errors, the results of equations (7) and (8) are correct. We note first of all that equation (7) gives a real result only when  $D < \sqrt{2}$ . That is, only when the Q exceeds  $1/\sqrt{2}$  (Q greater than Butterworth) - where there a peaking. For higher dampings, the roll-off of the low-pass is monotonic from 0. In the other direction, when D becomes small,  $\omega_{max}$  approaches 1, and  $|T(s)|_{max}$  in equation (8) approaches  $1/D = Q$ , agreeing with equation (3). So, for the region we are most interested, everything looks good. Note that equation (8) gives a real result only as long as D is less than 2. The condition on equation (7) that  $D < \sqrt{2}$ . certainly meets the condition  $D < 2$ . But what's going on here: why two conditions on D?

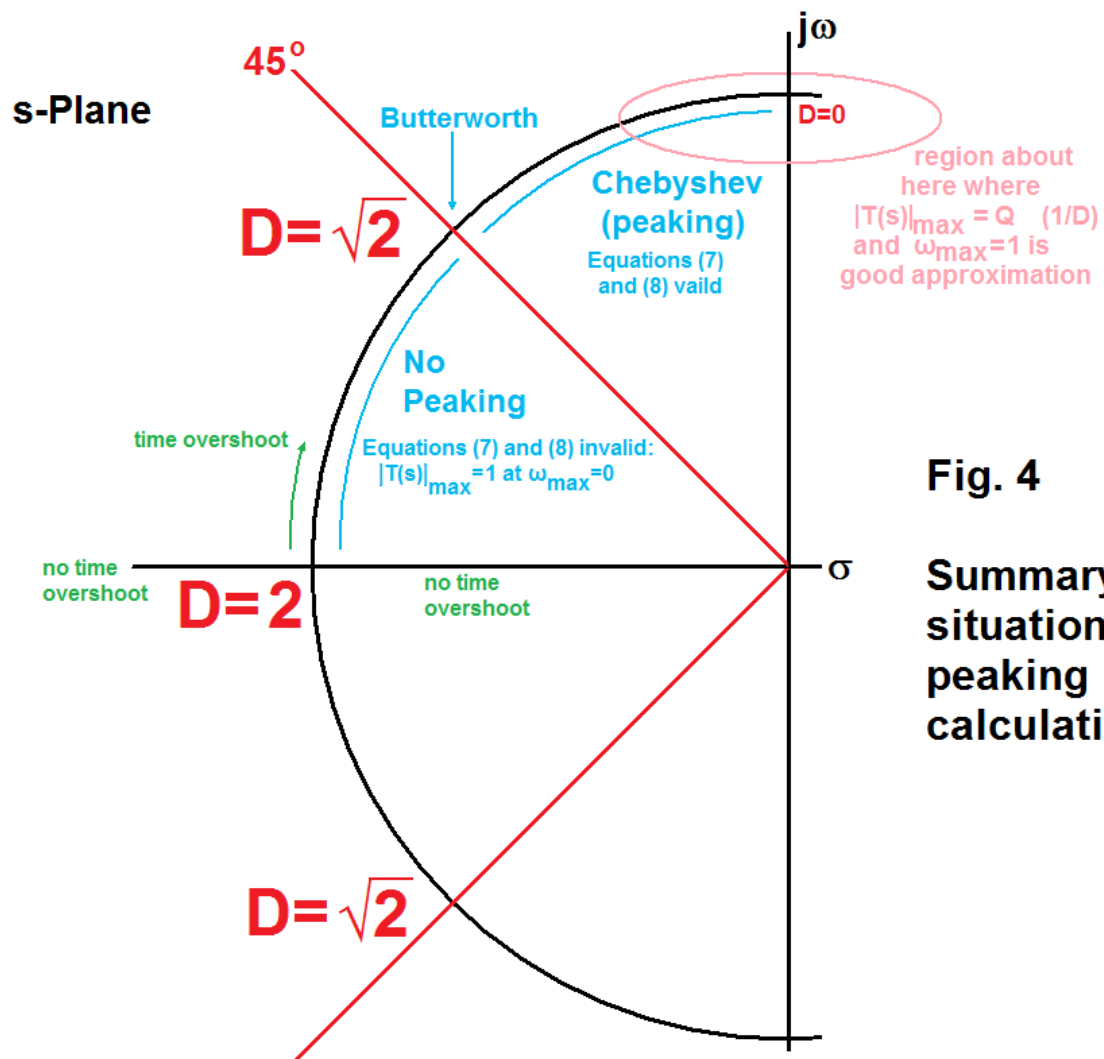
Here is the situation for the interested reader. Equation (6) is third order. This equation is to be solved for relative max/min values of  $\omega$ . One root is clearly  $\omega=0$ . This corresponds to the minimum or maximum at  $\omega=0$  or course. In the cases where there is no peaking (monotonic decrease)  $\omega=0$  is a maximum. This is true for all cases where damping is greater than Butterworth ( $D > \sqrt{2}$ ). In the case for damping less than Butterworth ( $D < \sqrt{2}$ ), there is a relative minimum at  $\omega=0$ , and the quadratic remains of equation (6) gives us the real results (negative as well as positive frequency) of equation (7). Note that with the damping greater than Butterworth, ( $D > \sqrt{2}$ ), equation (7) gave us imaginary frequencies as the “other” two solutions to equation (6). So the case where equation (7) gives a real result is really retelling the occurrence of the beginning of peaking in the frequency response – going beyond Butterworth to Chebyshev responses.

Now for equation (8). Ouch! This equation is telling us to watch out for the cases where  $D$  goes to 2 and beyond. If  $D=2$ , equation (8) blows up. If  $D > 2$ , then the result is imaginary. [Note that when  $D = \sqrt{2}$ , the Butterworth case, well short of  $D=2$ , equation (8) gives  $|T(s)|_{\max} = 1$ . Remember that right at  $D = \sqrt{2}$  we have the last legitimate case to use equation (8). Beyond this, as  $D$  exceeds  $\sqrt{2}$ ,  $\omega_{\max}$  becomes imaginary, and  $|T(s)|_{\max}$  goes back up above 1, a bogus result.] Thus equation (8) fails not where it blows up or becomes imaginary (although it would do so) but where equation (7) becomes imaginary.

We note that the case of  $D=2$  is a significant value in terms of the network. For dampings of  $D < 2$ , the poles of the network become imaginary. This is what we want. The poles for  $D=2$  are a pair of two real poles at  $s = -1$ . For smaller dampings, the complex conjugate poles move on a circle or radius 1 in the  $s$ -plane, arriving at  $\pm j$  when  $D$  goes to 0 (infinite  $Q$ ). One significance of the complex poles is that we need these for any sort of reasonable filtering. Indeed, 2<sup>nd</sup>-order Butterworth consists of a complex pair at  $-\sqrt{2}/2 \pm j\sqrt{2}/2$ . A second significance is that the time response shows overshoot starting at  $D=2$  and increases as  $D$  gets smaller.

To summarize, equations (7) and (8) are limited to values of  $D = \sqrt{2}$  and smaller, and these were the cases we were interested in anyway. Also, keep in mind that time overshoot begins as  $D$  gets smaller than 2, and “corner peaking” begins as  $D$  gets smaller than  $\sqrt{2}$ . See Fig. 4 for summary.

To actually verify (not prove) our results, we can do some calculations. This is conveniently done with Matlab, and the code that we used to test the equations (for various arbitrary choices of  $Q$ ) is printed at the end. The code is the best documentation of the calculations. [If there is ever a doubt about interpreting a written equation, the code is usually unambiguous.] Similar code produced Fig. 1. The results agree to four decimal places.



**Fig. 4**  
**Summary situation for peaking calculations**

## **REFERENCES**

- [1] B. Hutchins, "Analog Signal Processing, Chapter 7: Passive and Active Sensitivity", *Electronotes*, Vol. 20, No. 195, July 2000  
<http://electronotes.netfirms.com/EN195.pdf>
- [2] A good classic reference to many practical active filter issues is: S-016: Lab Problems and Examples in Active, Voltage-Controlled, and Delay Line Methods  
<http://electronotes.netfirms.com/S016.pdf>
- [3] B. Hutchins, "The ENS-76 Home Build Synthesizer System, Part 5, VCF Options" *Electronotes*, Vol. 8, No. 71, Nov. 1976 (see Fig. 8 of this article).
- [4] B. Hutchins, "The Bandpass Filter Response and Electronic Music Applications" *Electronotes*, Vol. 8, No. 71, Nov. 1976 (see Fig. 7 of this article).
- [5] T. Mikulic, "Reader's Equipment: Four Analog Circuits from Terry Mikulic," *Electronotes*, Vol. 5, No. 34, Jan 25, 1974
- [6] B. Hutchins, "A State-Variable Configuration With Constant Peak Response," Electronotes Application Note. AN-121, Feb 5, 1979  
<http://electronotes.netfirms.com/AN121.PDF>
- [7] B. Hutchins, "Resonant Peaking in Second-Order Filters," Electronotes Application Note. AN-90, June 11, 1978
- [8] B. Hutchins, "Second-Order Frequency Response," Electronotes Application Note. AN-156, Dec. 5, 1979



## Typical Matlab Code Used in Calculations

```
% an396Ver

Q=2.5

w=0.9:.0001:1.1;
T=freqs(1,[1 1/Q 1],w);

figure(1)
plot(w,abs(T))
hold on
plot([-1 5],[Q Q], 'r:')
plot([1 1],[-1 20], 'r:')
plot([0 0],[-1 50], 'k')
plot([-1 5],[0 0], 'k')
axis([0.9 1.1 .8 1.1*Q])
hold off
figure(1)

% examine calculated data
[M,ix]=max(abs(T));
fmax=0.9+(ix-1)*.0001
M

% equations (7) and (8)
D=1/Q;
fmaxth=sqrt(1-D^2/2)
Mtheory=(2/D)*sqrt(1/(4-D^2))
```