ELECTRONOTES 225

Newsletter of the Musical Engineering Group

1016 Hanshaw Road, Ithaca, New York 14850

Volume 23, Number 225 September 2015

FILTERING WITH "RINGS" OF POLES/ZEROS

-by Bernie Hutchins

INTRODUCTION

For many practitioners of the digital filter design art, the exposure to the essential idea was by a specific example rather than by general principals. This was often extremely useful as it showed by example WHY things worked. Typically, an engineer with some knowledge of ordinary analog filtering heard the term "digital filter" and supposed it had something to do with the quantization of the signal samples. A long trip through the mathematics likely resolved this issue in favor of the samples being, rather, discrete IN TIME. A comfortable short trip was apparent from real examples, however obtained.

In the early days, when it got to the serious matter of actually designing digital filters with respectable characteristics, two factors came into play which favored an "Infinite Impulse Response" (IIR) approach. First, the IIR methods: Bilinear-z (BZ) and Impulse Invariant (IIV), etc.) were based on familiar analog filters which were usually considered "prototypes" to be emulated – numerical (trapezoidal) integration for BZ and sampling of an analog impulse response for IIV. Secondly, implementation resources (mostly hardware multiplies) were then very limited, so a high-order such as might be required for a "Finite Impulse Response" (FIR) seemed unlikely to be of practical value for at least a while.

It should be understood that IIR theory was (and, encompassing so much of analog filtering, still is) more difficult than FIR theory – at least at the point where the mathematics was first developed. Arguably today's FIR design procedures are harder to understand (having advanced with history) than IIR methods which are largely unchanged after 50 years. (Both are also largely a matter of plugging specifications into canned programs.)

So there was a former period of time for which FIR filters were not remotely practical while the best of the FIR design procedures (least-squares, equi-ripple, etc.) were not even developed yet. Such was a time of "tricks" where a simple FIR structure could be proposed and analyzed, often as very valuable teaching examples. The "moving-sum" or "moving-average" was such a popular "toy" to consider. It fully illustrated an obvious result coming out of an obvious mechanism, and the discrete time mathematics of the z-transform approach was cooperative. We could easily see the "why" of the calculated responses.

The case of the moving-average can be reviewed. For example, if we were running a store and wanted to keep track of how well business was progressing, but knew at the same time that there were significant day-by-day random fluctuations, we automatically think of averaging over several days. The last few days likely. How many days? Well perhaps three. But what if our best days were, for reasons we can understand, Friday, Saturday, and Sunday. At the same time perhaps we close on Tuesday. We are not surprised (or alarmed) if such a three-day moving average peaks every Sunday and has low values for and three days that include Tuesday. Soon enough, we determine that a 7-day (weekly) moving average is likely quite reasonable. And keep in mind that we are talking about the last 7 days. If it's Friday, we add today's income to the average while subtracting the previous Friday. To the extent that Friday's are similar, we get a similar value. Note that this is a running-average and not an average.

This is perhaps a good point to recall that normally a moving average is done by taking the sum of N consecutive values and dividing by N. The individual terms are not individually weighted, so there is only one "multiply" (by 1/N) at the end. In cases where N is very large, it is usually computationally advantageous to keep the previous sum, subtract the oldest sample and add the newest sample, and then divide by N, and so on – often called a "head-tail" method. The math has to be equivalent.

In describing this as an era of tricks I have in mind simply the fact that we chose FIR impulse responses that were simple enough so that the series the z-Transform gave us were well-known geometric series which we could sum. It was surprising to me later that anyone actually memorized these geometric summation results as plug-in formulas, as it was so easy to just do it on the fly and the details were flexible – although the right answers always resulted. Given a length-N impulse response h(n) for n=0 to n=N-1, the z-Transform is the series:

$$H(z) = \sum_{n=0}^{N-1} h(n) z^{-n} \qquad (1)$$

where H(z) is the transfer function. This sometimes offered a simple path to finding the zeros of the filter, avoiding any numerical root-finding. Another important from of equation (1) was to evaluate it at $z = e^{j\omega T}$ (the unit circle in the z-plane) where ω is the radial frequency (ω=2πf with f ordinary frequency in Hz) and T is the sampling interval ($T = 1/f_s$ where f_s is the sampling frequency in Hz). Thus:

$$H\left(e^{j\omega T}\right) = \sum_{n=0}^{N-1} h(n) e^{-jn\omega T} \tag{2}$$

and $H\left(e^{j\omega T}\right)$ is the frequency response. Taking $|H\left(e^{j\omega T}\right)|$ we have the most familiar notion of a filter specification (the magnitude of the frequency response without regard to phase). $H\left(e^{j\omega T}\right)$ is properly called the "Discrete-Time Fourier Transform" (DTFT). The DTFT is in turn simply the Fourier Series of a periodic function of frequency, with the h(n) being the "Fourier Series coefficients" (reversing the usual roles of time and frequency).

As one would expect, the "analysis" of a frequency response, equation (2), is often used for filter design. For one thing, we most often start with some notion of $H\left(e^{j\omega T}\right)$, a "specification", and can directly invert equation (2) (invert the Fourier Series) or obtain h(n) by least squares, or by frequency sampling [1]. In addition, in an iterative process such as the Parks-McClellan equiripple algorithm, equation (2) lets us know how thing improve with successive iterations.

All this is given by way of a quick review and reminder.

RING OF ZEROS: MOVING AVERAGE AND COMB

In the case here of the moving average (Fig. 1a) and the comb (Fig. 1b), we are directly specifying h(n). The point of this report will be that in many cases where we start with such simple impulse responses h(n), we get "rings of zeros" which may be perfectly symmetrical or just usefully non-symmetrical. There are variations on these basic structures. For example, the multiplier on the right side of Fig. 1b might be -1 instead of +1. Or we might have the case where N is just 1. And these structures may well be parts of more involved schemes. The unifying ideas here are that we can expect to do a design without a complicated procedure, do our analysis with simple mathematics, and achieve an easy-to-understand utility as a result.

Here for specificity we will set N=16 (sixteen zeros), which is a length-17 sum. Applying equation (1) to the moving average of Fig. 1a, we have:

$$H_{MA}(z) = \frac{\left[1 + z^{-1} + z^{-2} + \ .\ .\ .\ . + z^{-15} + z^{-16}\right]}{17} \tag{3}$$

It is not obvious where the roots of this polynomial are, although the procedures for simplifying it are well-known. We remember that we are to multiply both sides by (1-z). Or is it $(1\text{-}z^{-1})$? Well, either works equally well. Let's use $(1\text{-}z^{-1})$.

$$\begin{aligned}(1 - z^{-1})H_{MA}(z) &= [\,1 + z^{-1} + z^{-2} + \ .\ .\ .\ . + z^{-15} + z^{-16}]/17 \\ &\quad - \quad [\,z^{-1} + z^{-2} + \ .\ .\ .\ .\ + z^{-16} + z^{-17}]/17 \\ &= \ [1 - z^{-17}]\,/\ 17\end{aligned} \tag{4}$$

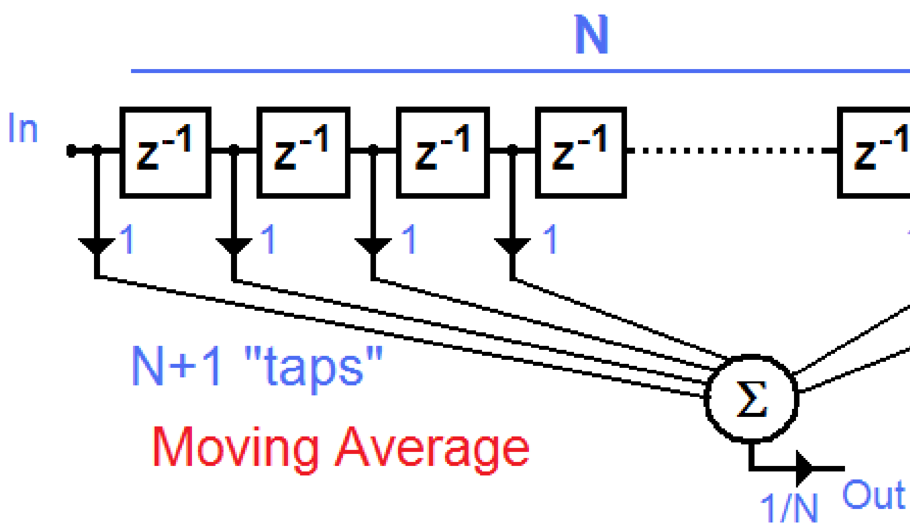


Fig. 1a

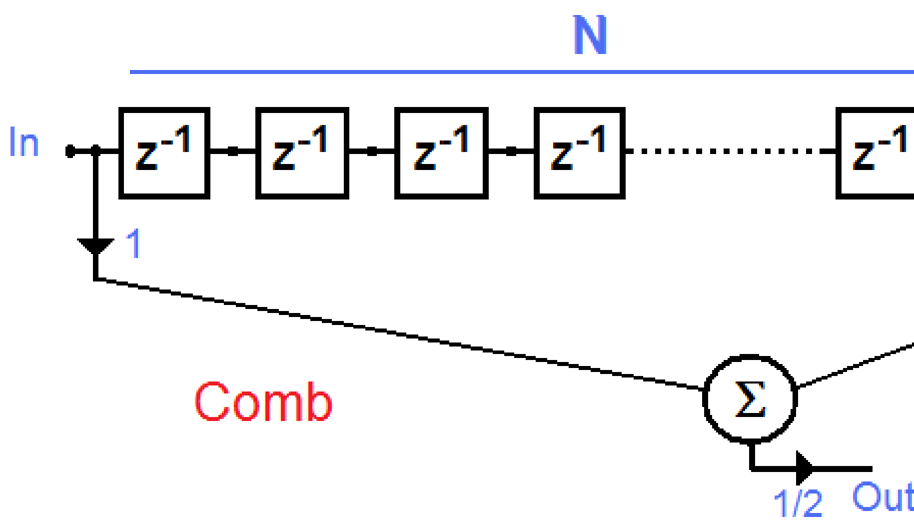


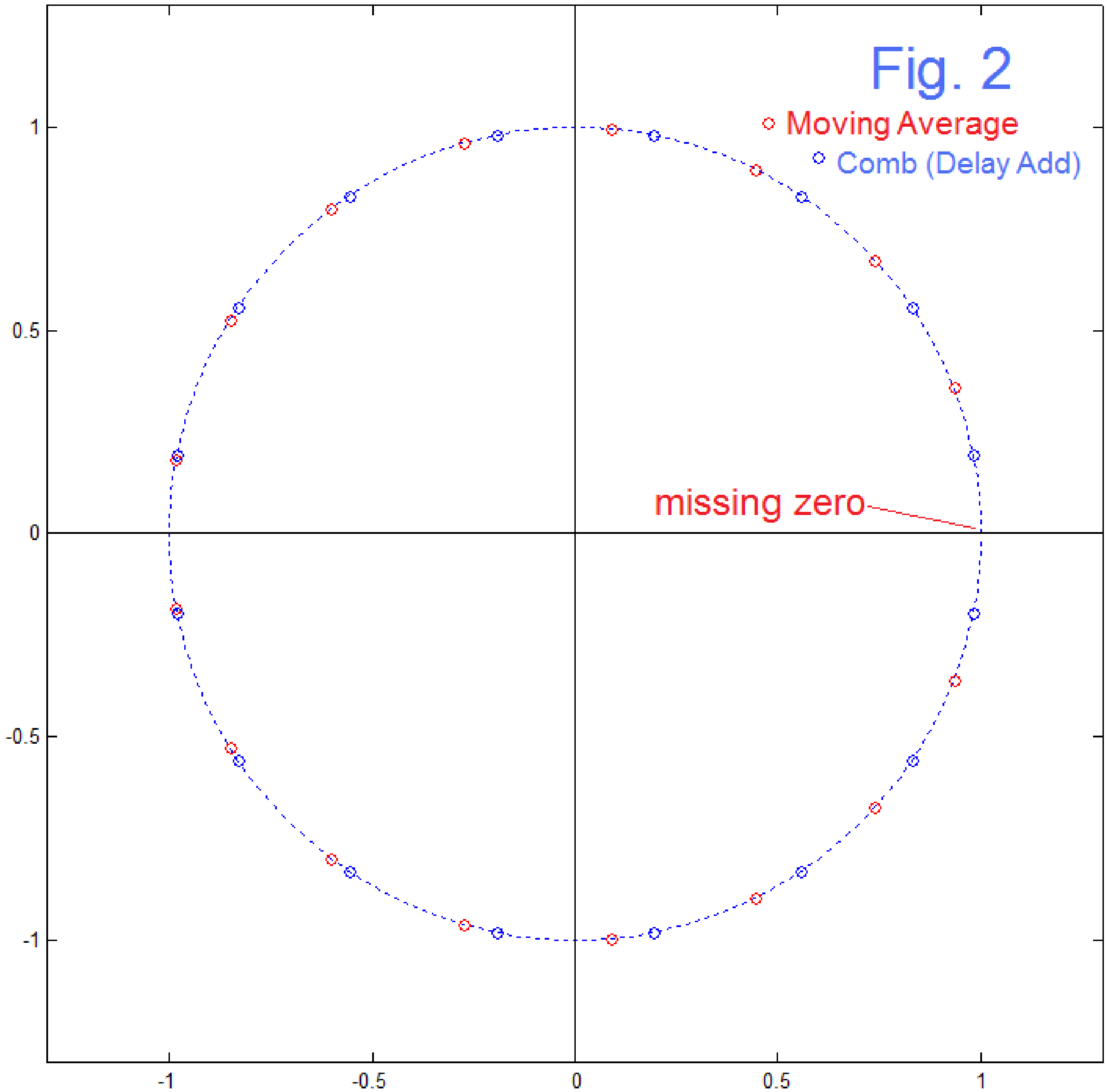
Fig. 1b

Thus $H_{MA}(z)$ is given by:

$$H_{MA}(z) \;=\frac{[1-z^{-17}]}{17(1-z^{-1})} \qquad (5)$$

Normally we don't expect to be able to solve for the roots of the 17th-order polynomial in closed form, but we can solve this one because it is really a 1st-order polynomial in z^{-17}. The roots (zeros), of the numerator, are the 17 numbers which when rotated by themselves 17 times end up at z=1. These are often called the "roots of unity". Now, here $h_{MA}(n)$ is just 17 sequential "taps" of value 1/17. And 17 taps should have just 16 zeros, not 17. [As a reminder, it takes two taps (a cancelation) to get the first zero.] So what's going on?

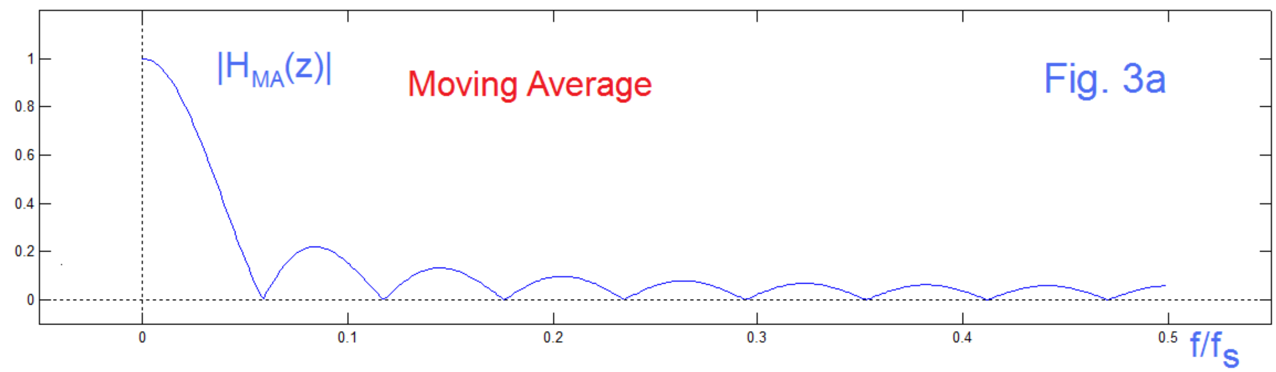
The other curious thing about equation (5) is that it has a pole at z=1. Being an FIR filter, it should have no poles at all. Clearly, the numerator $1-z^{-17}$ has one of its 17 zeros at z=+1 and this is cancelled by the pole of the denominator when we look at the entire $H_{MA}(z)$. So we end up (net) with just the 16 zeros, equally spaced at 17 points around the unit circle (red in Fig. 2), except the zero at z=+1 is missing.

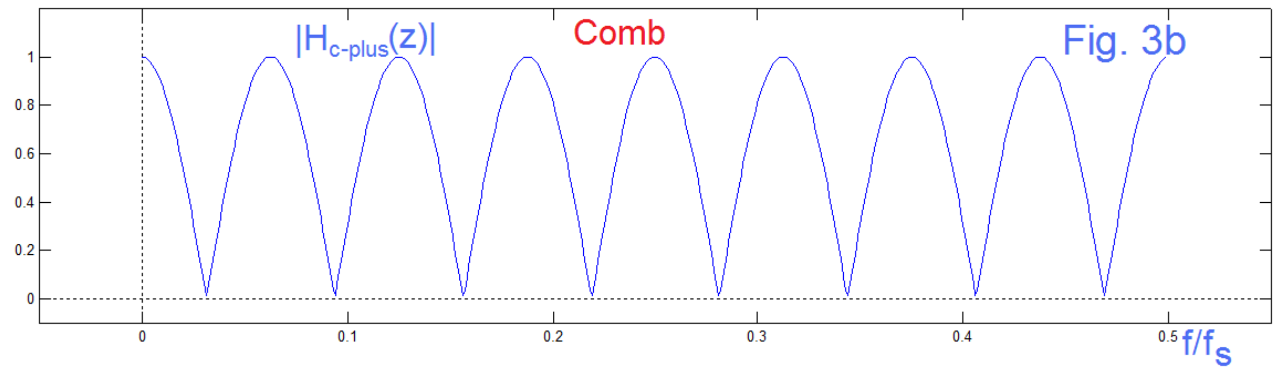


An excellent comparison example to the moving average is the so-called comb (Fig. 1b) which is also based on a delay line but adds only the first and last tap (sum divided by 2). This has a simple transfer function:

$$H_{c-plus}(z) = \left(\frac{1}{2}\right)[1 + z^{-16}] \qquad (6)$$

The zeros of equation (6) are at z=(-1)$^{1/16}$ which (again related to roots of unity) are at 16 equally spaced points about the unit circle (again – blue in Fig. 2). We are looking for 16 rotations here to end up at z=-1. The zeros are separated by 360°/16 = 22.5°. Starting counter-clockwise from z=1, the first zero is at 11.25°. Rotating 16 times by 11.25° is 180° or z=-1, and so on. This sort of result is "simple" but it is very easy to become confused, and there are a lot of variations on the comb.

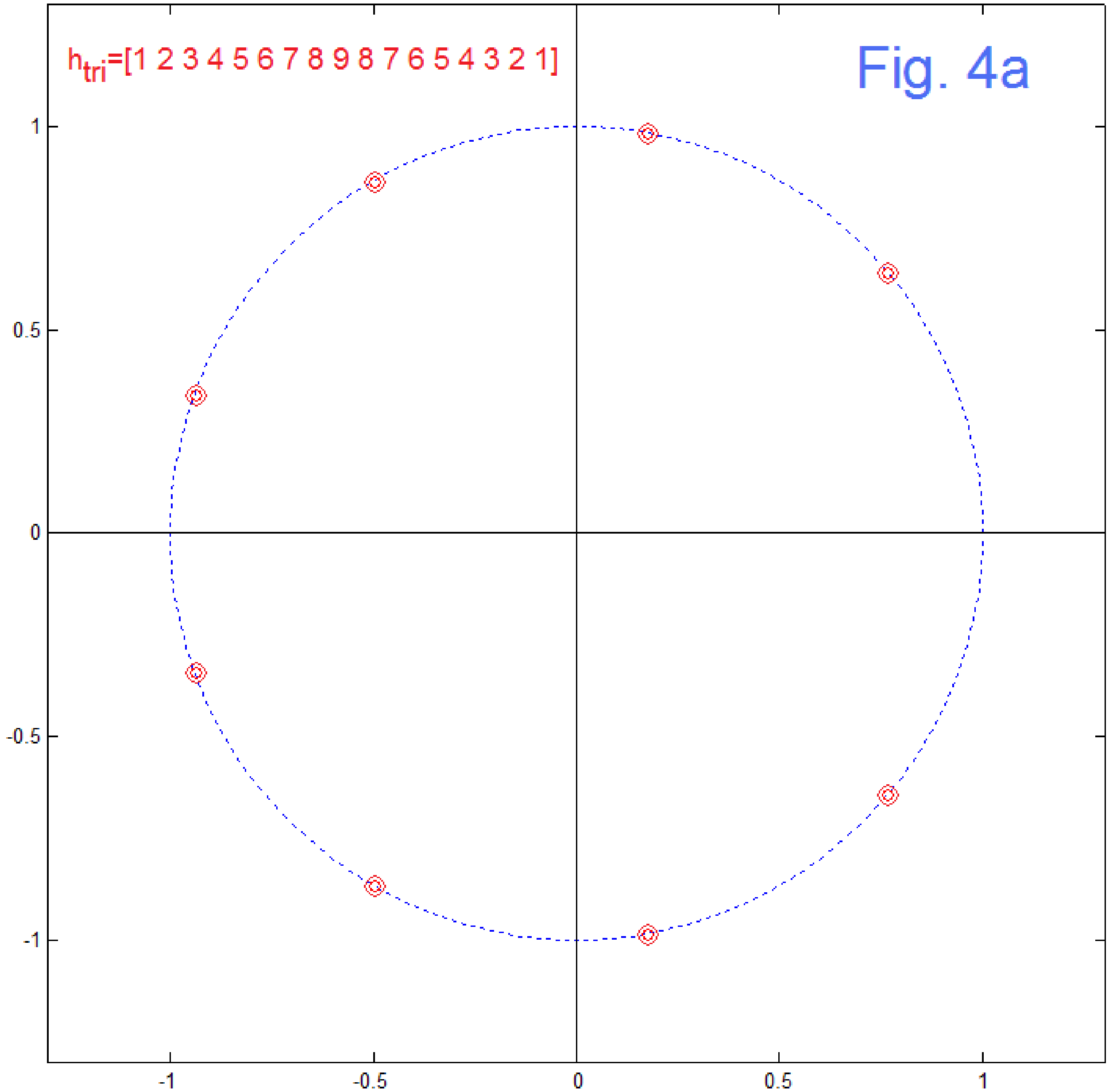


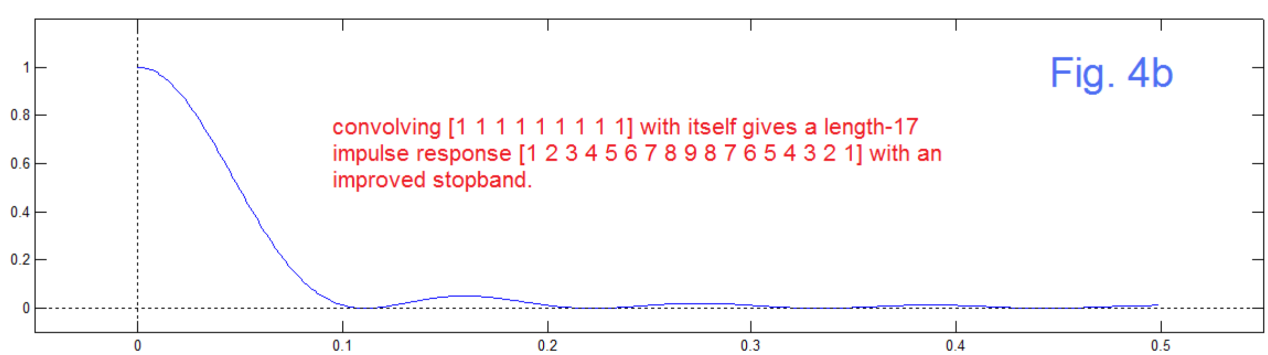


While we enjoy looking at the positions of the roots of the various transfer function we examine, we likely are most familiar with frequency responses, and the magnitude responses for the moving average and the comb are plotted in Fig. 3. While the zero plots did not seem drastically different, we see that the frequency responses are. Likely we recognize the basic low-pass nature of the moving average, and the periodic nulls which give the comb its name. Recall that the moving average is far from a low-pass that we would (aside from its simplicity) brag about. Recall also that a companion to the comb filter we use here would be the delay-subtract, perhaps $H_{c\text{-minus}}$, where the end tap is subtracted and the response is shifted so that a null occurs at DC. All this is a reminder that we need to look carefully at what we are dealing with in its particulars.

SIMPLE VARIATIONS

Simple variations on the themes here involve things like changing lengths and perhaps the signs of tap weights, as well as using familiar convolution and modulation processes that relate to time series (impulse responses) and the corresponding transfer functions. For example we might try changing the sign of every other tap (multiplying by -1^n) to reflect the frequency response. Or we might convolve the impulse response to square the frequency response – a common analysis shortcut for a triangular impulse response (Fig. 4). For Fig. 4 we have convolved [1 1 1 1 1 1 1 1 1] (length 9) with itself for a length 17 impulse response. This has second-ordered zeros (Fig. 4a) and a reduced stopband (with a less sharp transition band), as in Fig. 4b. An expected trade-off.





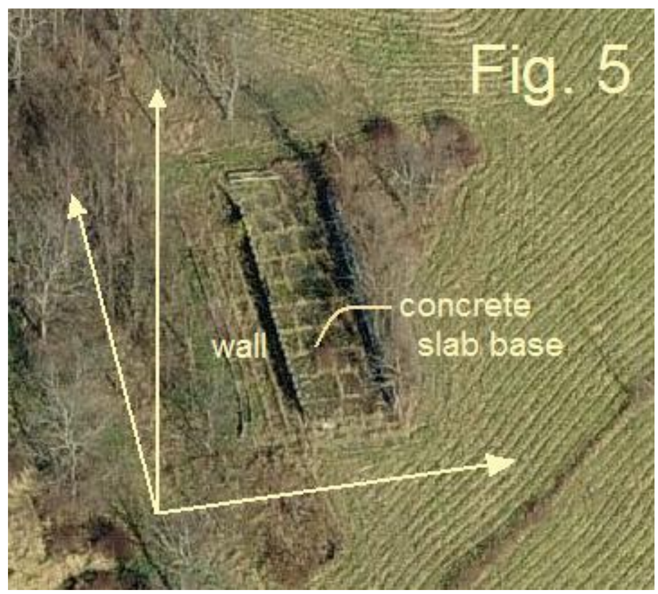
Another simple trick (which we shall not complete here) would be to represent the comb as the sum of two moving averages. This we would do by forming a second moving average from the first by shortening it by two, shifting it one to the right, and subtracting.

MORE COMPLEX TRICKS

There is actually a measure of incentive for what we are doing here. First, there is a sort of "blank slate" notion of starting with something that is the most basic. Not just the notion of starting an analysis with the first mathematical steps, but rather with a simple geometric item such as a "ring of zeros" which seems intuitively neutral. Thus a ring is neutral (a comb) which is modified by messing with just one of the many zeros giving a moving average. Secondly, we can think in terms of certain fundamental building blocks which we modify and manipulate (as suggested above) into something different, dragging along the fundamental understanding.

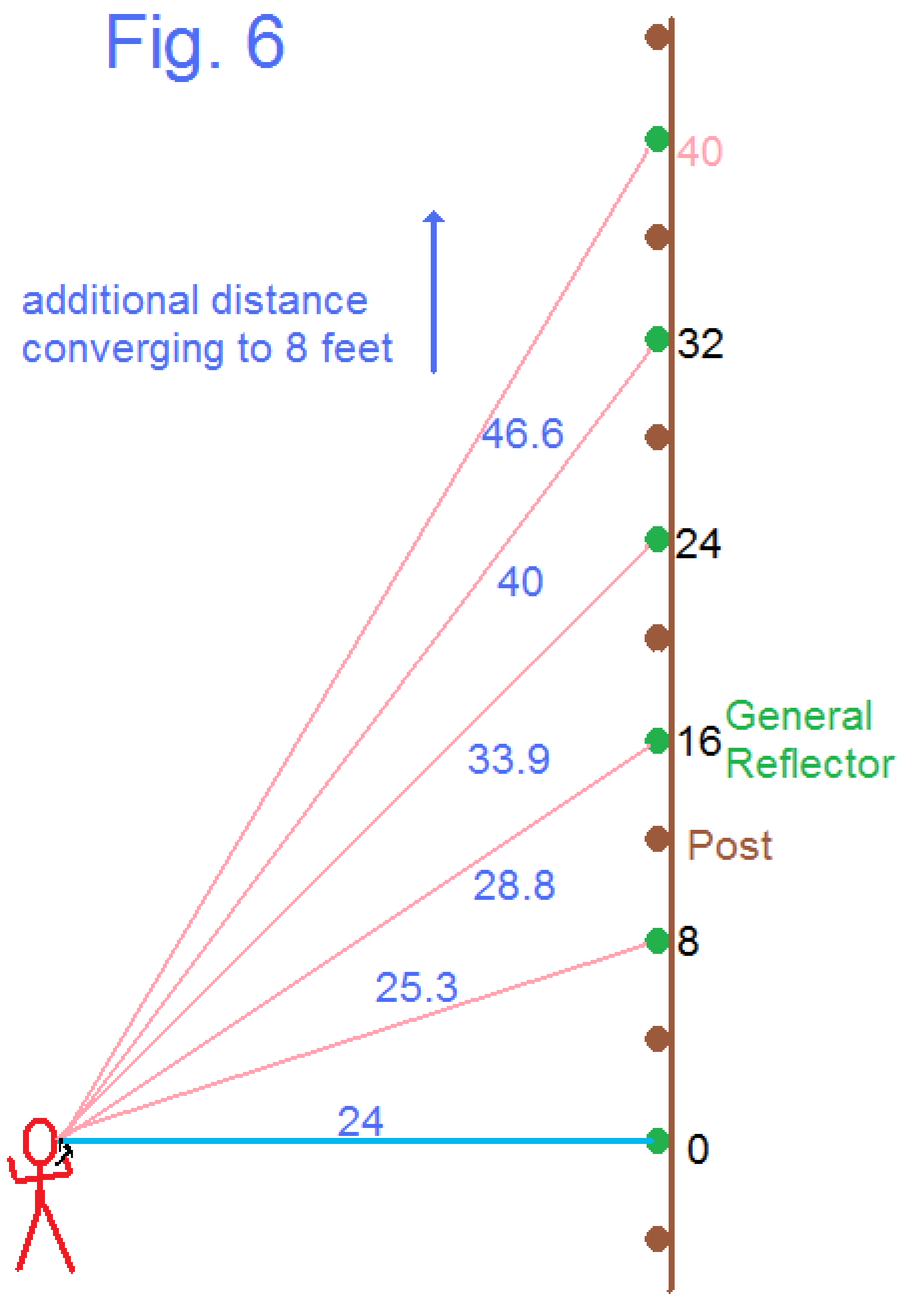
A third reason is exemplified by a recent application note [2] where we were considering something that was arguably related to a physical process that really had to be physically simple. This was the case of a multitude of acoustic (or similar) paths that by their natures had different lengths yet which recombined later. Often such physical cases have to be a recombination by addition. There is no process I can imagine that can "flip" (subtract) two sounds. OF COURSE the two sounds can "destructively interfere" if they arrive back together out of phase, <u>but they add</u> (as in the nulls of the comb above).

As a second example of delay-add, consider the image in Fig. 5. The farmers among our readers may recognize this as an abandoned trench silo, or as in the above-ground design shown, it is often called a "bunker-silo". This one is seen as the center rectangle of photo, a ground-level concrete slab which evidently had cracked into composite sections after 50 years or so. The walls were made of pressure-treated lumber with heavy posts, perhaps 8 feet apart and sticking above the ground perhaps 8 feet. The actual walls were PT 2x8 tongue/groove. My family built it in the 1960s with the notion that it was to store silage for the winter feeding of our cattle. Clearly the silo lasted considerably longer than our enthusiasm for raising cattle!



My interest here is in the construction phase of the operation. There were at one time about 30 posts on each wall (you can still count them). And we were pounding nails. I don't much recall what I heard initially, but as the wall went up, there was a delightful echo that developed. I knew enough about sound and physics to understand pretty much what was going on. Apparently the post by themselves scattered sound too widely and/or too little to make much impression. But the whole sequence of side panels was an excellent reflector with repeating structures. The posts more or less interrupted (and thus structured) the echo. As I think about it today, I can imagine many secondary factors.

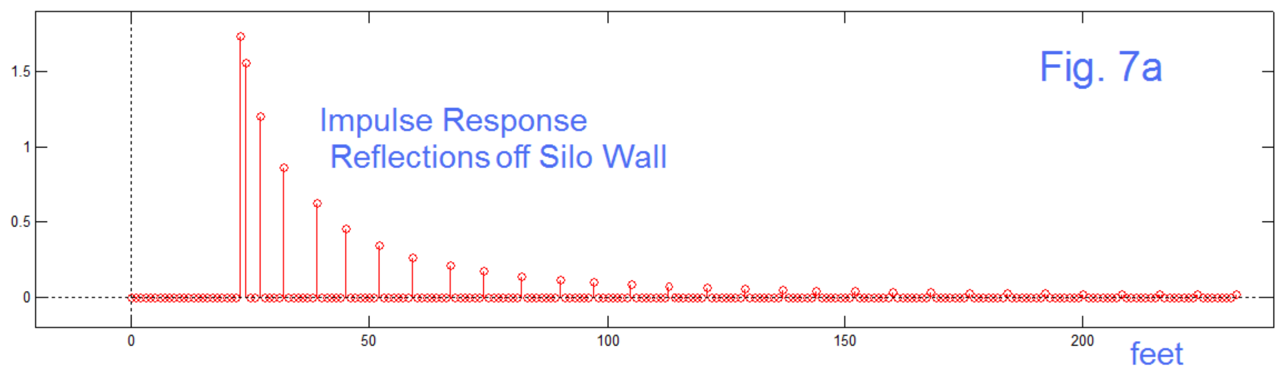
Fig. 6 suggests a possible geometry showing posts and the wall in brown, and suggesting that each panel is in some sense a reflector (green – shown in the middle of each panel). In some manner, it is clear that each panel/post constitutes some sort of sound-trap/reflector or else there would be nothing to hear. Only 6 of something like 30 panels are shown. Note that the reflectors are 8 feet apart. If in a perfect line, this would be 1100/8 = 137.5 Hz, a low but audible pitch.



The person shown with the hammer is of course in general standing back away from the wall (here shown as 24 feet from a reflector). In this case, note that the second reflector is at 25.3 feet – not much difference. Thus the second reflection arrives quite quickly. Subsequent reflections space further apart and the differences approach that of 8 feet. Thus we would suppose that the pitch should start high and drop – or would it. The whole "tone" can't be longer than about 1/4 second (30 posts) so there is not a lot of time to establish a pitch even if it were constant. THIS IS NOT A SIMPLE PROBLEM. Sure it is delay-add. But the delays change, as do the tap weights. Shall we look at this further?

When we reach a point where we think something interesting can be examined, there is the approach of jumping right in and a second approach of systematically doing our study one step at a time.   Here we had a pretty good idea how to write down an impulse response of reflections coming off an array, and the empirical fact that an interesting sound was already observed (although many years ago). So why not just go for it?

We know the distances and they seem large enough that we can represent them as an integer number of feet. This gives us delays in units of 1/(1100 ft/sec) or about 0.91 milliseconds/foot. Relatively few of the delays have a non-zero path to the summer. The non-zero paths have multipliers set to 0.   The rest of the paths we will take to be 1/distance^2. That's the basic idea.



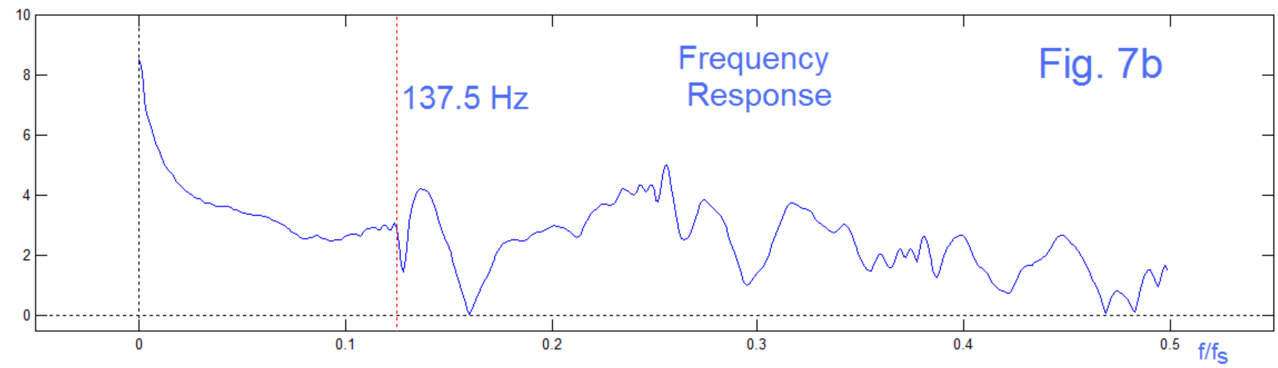


Fig. 7a is not a surprise at all, based on Fig. 6. But Fig. 7b is quite unexpected. We likely expected to see some sort of frequency response with a preference for frequencies around 137.5 Hz. Okay - we might argue that there is a peak slightly above that, and we could blame it on the fact mentioned that the first reflections are going to be substantially sharp. Probably true.   The zero plot (showing the major zeros) is in Fig. 8 and that doesn't help much.   However, we do see that it clearly shows that the current problem is appropriate to our "ring of zeros" theme.   But clearly we have taken too big a step.

What happens if we simplify - just look at every 8th tap as being non-zero and keep the amplitudes all the same (Fig. 9a)?   This seems better controlled, even though it departs from our physical notion of the actual silo in important ways.   But we do need to regain a firm footing. We see From Fig. 9b that we now get something expected: a nicely formed frequency response and one that does indeed have a peak about 137.5 Hz, as we hoped for. Likewise, the zero plot of Fig. 10 looks promising.

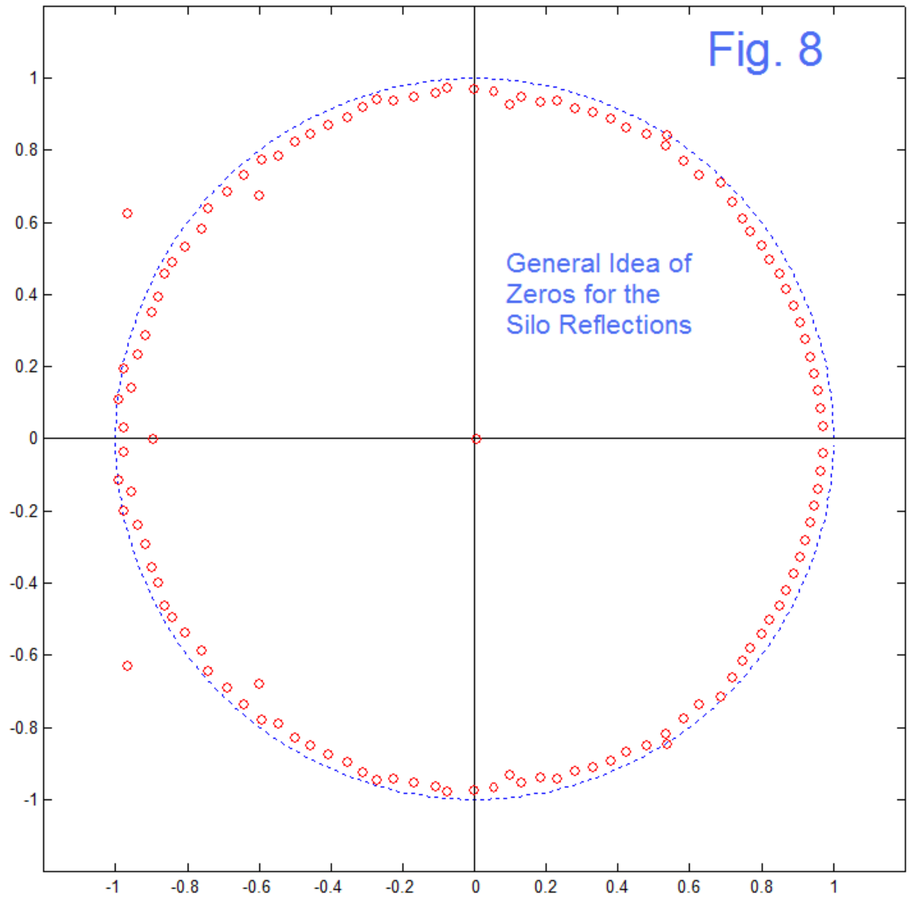
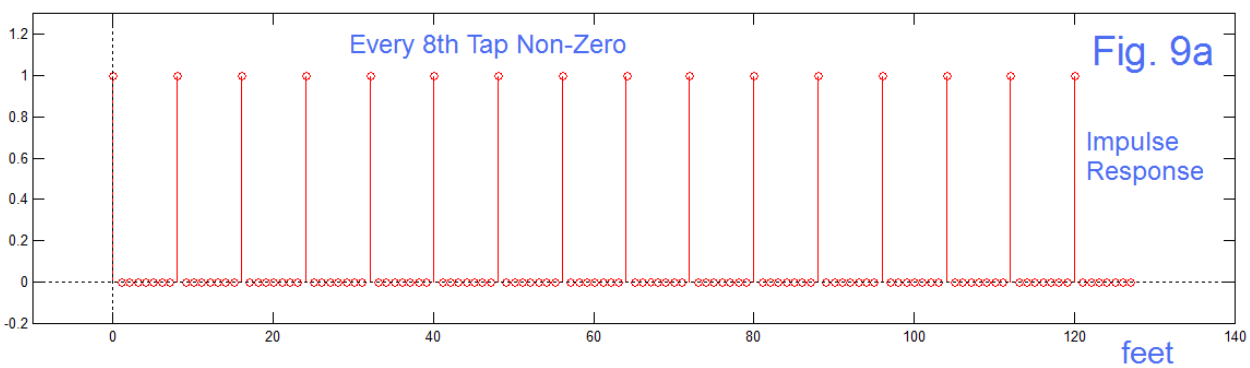
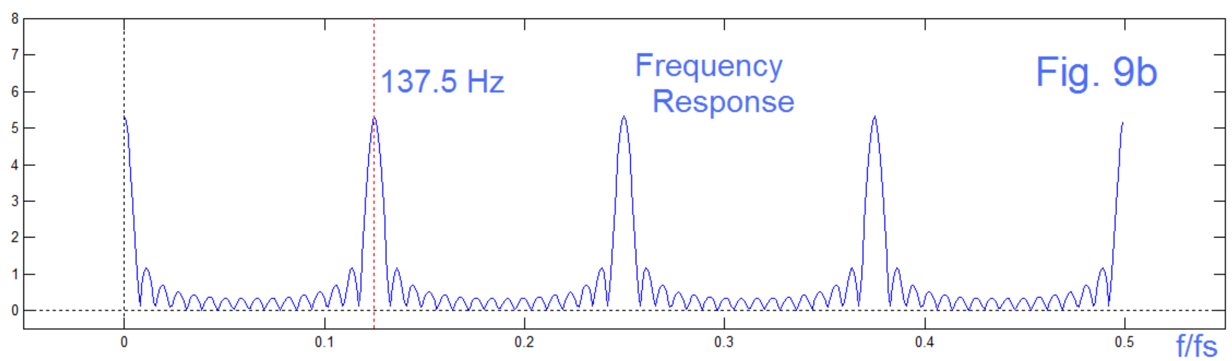
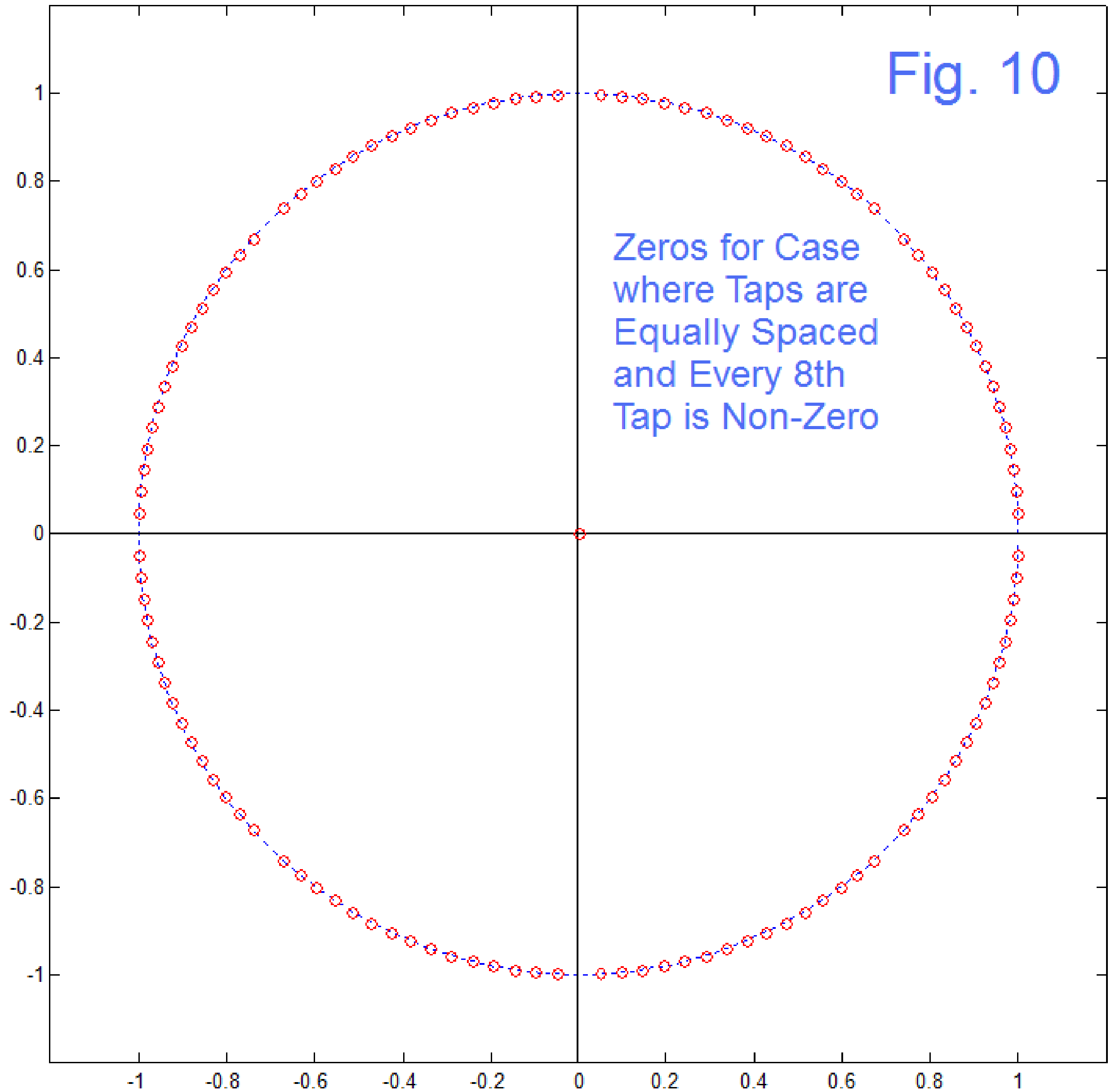


Fig. 8
General Idea of
Zeros for the
Silo Reflections


Every 8th Tap Non-Zero
Fig. 9a
Impulse
Response
feet


137.5 Hz
Frequency
Response
Fig. 9b
f/fs



Is it clear that the zero plot of Fig. 10 is correct? Here is a chance to exercise our notion of combining known results. The simplified impulse of Fig. 9a is just 16 repetitions of the sequence [1 0 0 0 0 0 0 0] so its transfer function is:

$$H(z) = 1 + z^{-8} + z^{-16} + \ldots\ldots + z^{-112} + z^{-120} \tag{7a}$$

Using the familiar trick we multiply both sides by $1 - z^{-8}$:

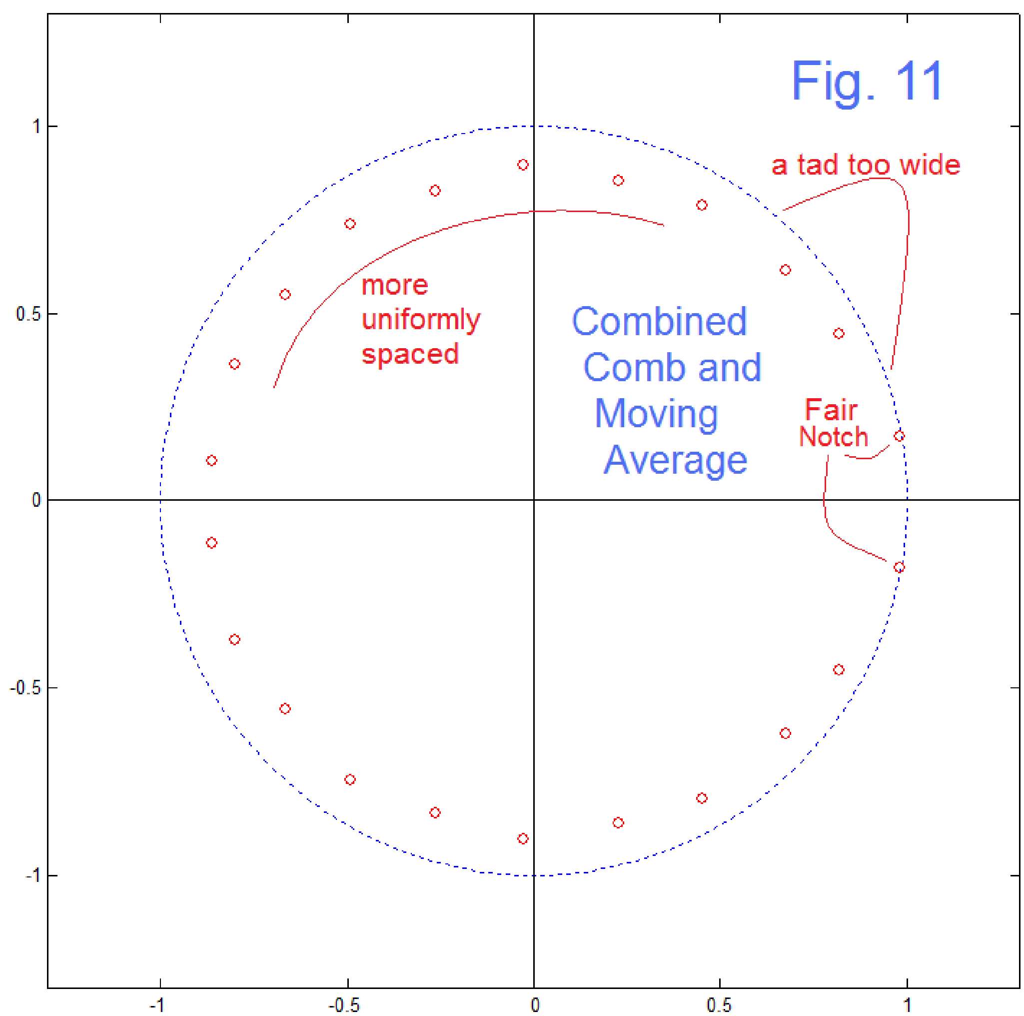
$$(1 - z^{-8})H(z) = 1 - z^{-128} \tag{7b}$$

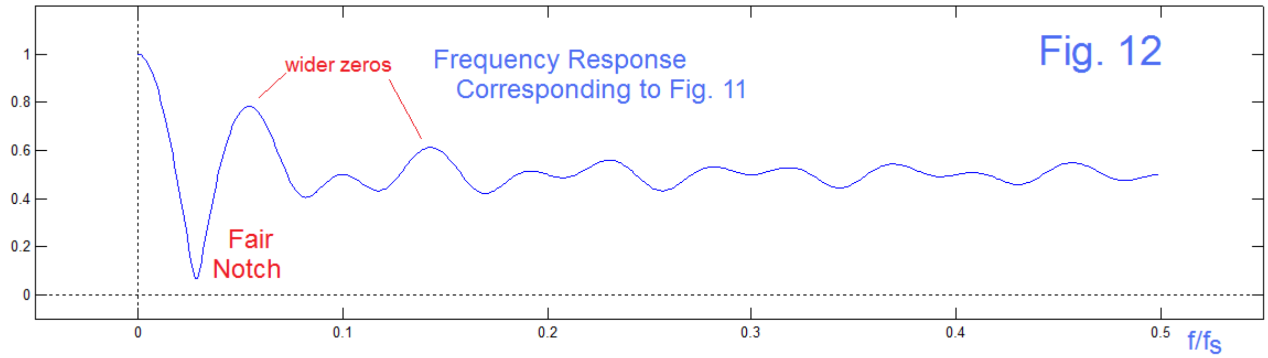
so H(z) has zeros at 128 equally spaced points about the unit circle with eight poles cancelling eight of these zeros (Fig. 10). [It is easy to get fooled here. Initially I had been surprised that it was every 16^{th} zero and not every 8^{th} zero that was cancelled.] The corresponding frequency response is a respectable multiple bandpass (we can call it a comb if we wish) which has a peak at 137.5 Hz. So this works. The more complicated estimate of the impulse response (Fig. 7a) while more realistic was clearly too big a jump to start with. Audibly, Fig. 9a is a reasonable pitched buzz and a "thump" combined. (Fig. 7a is just the thump.) A similar "buzz" occurs with railroad tracks [3].

We will in a moment go on to talk about arrays of poles, but we need to tie up the earlier example of a combination of a comb and a moving average. Here we have in mind a case where a signal is applied at an input and then delayed and added (a comb). However instead of having just one echo at the far end, we disperse this contribution over a range of delays, such that the sum of the delays is still one. Here is our example in Matlab:

```
htest=[1 0 0 0 0 0 0 0 0 0 0  0 0 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1]
```

Accordingly we have an input summed with a delayed and attenuated moving average.





The notion here [2] was that we had a group of combs where all the individuals had their first zero at about the same frequency (fair notch), as in Fig. 11, but that due to the differences in the actual delays, the upper notches of the comb would increasingly "wash" each other out and that some sort of a filter looking like a single notch could result (Fig. 12). That is, a sharpened notch can result without the usual pole pair. Purely FIR. Clearly we also have here an entry under our "ring of zeros" theme.

Likely the reader is familiar with the idea that we can have a ring of zeros and that by removing a pair (or pairs) from the circular array have a result that looks a lot like a pole. This is true of the moving average and much of what we saw above.

RINGS OF POLES

Above we have looked at the various ways we can use rings of zeros to form various frequency responses and/or approach various physical situations. Setting zeros with a particular numerator of a transfer function has the obvious counterpart in setting poles with the denominator. The differences are (1) we will need to use feedback to achieve poles, and (2) we need our poles to be stable; inside the unit circle, not outside or even ON the unit circle. So a lot of applications above, where we have found so many unit circle zeros to be a good solution, need to be modified – at least.

{ We have somewhere in the past remarked on why, so many filters we design, do have zeros that fall exactly on the unit circle! After encountering these so often, we tend to ask the question, and the answer is likely apparent: it's because so many filters we use in practice have stopbands. That is, our goals in filtering very often involve one or more bands where we would ideally want a zero response throughout the band. In as much as the magnitude response is always a sum of complex exponentials, equation (2), we at best expect to realize a small response that hovers around zero, and this means that we will have a unit circle zero (pair usually) each time the response crosses zero. }

Fig. 13 shows the simplest case where feedback is used. Here the output is delayed and then multiplied by some constant g and summed with the input. That is:

$$H(z) = \frac{Out(z)}{\mathrm{In(z)}} = \frac{1}{1-\mathrm{g}z^{-\mathrm{N}}} \tag{8}$$

So we have poles when the denominator is zero if

$$z = \left| g^{1/N} \right| 1^{1/N} \qquad (9)$$

which we would normally write as just $g^{1/N}$ (the N^{th} roots of g). Here we emphasize that this is a result that scales the "roots of unity". For example, if g=0.5 and N=8 the poles are at 0.917 times the roots of unity. They are perhaps further out that one might suppose from the value chosen for g. But they <u>are on a ring</u>, equally spaced.

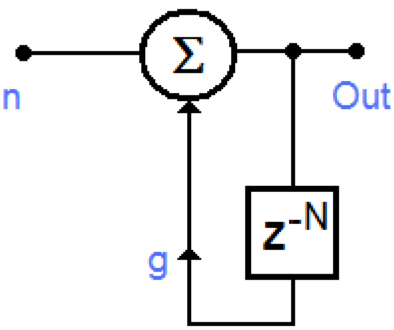


Fig. 13

Feedback Produces Poles

For example, we might choose H(z) to have a denominator:

```
den=[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.7]
```

so g = 0.7. This gives 16 poles as in Fig. 14a, a radius of $0.7^{1/16}$ = 0.9780 such that 16 rotations end up at 0.9780. A nice, complete, ring of stable poles. The corresponding frequency response is shown in Fig. 15a. The response has uniform peaks since all poles are equally spaced. Notice that this is a comb filter, and that as the poles are moved closer to the unit circle, we expect a very nice sharp (bandpass) comb. This is a <u>very useful</u> compliment to the notch comb such as in Fig. 3b.

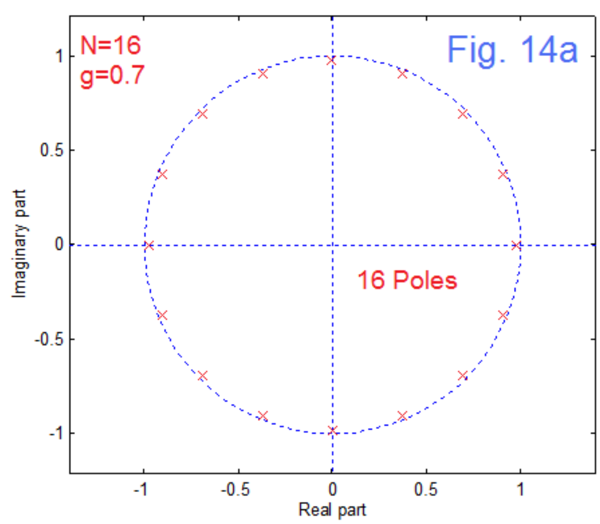


Fig. 14a

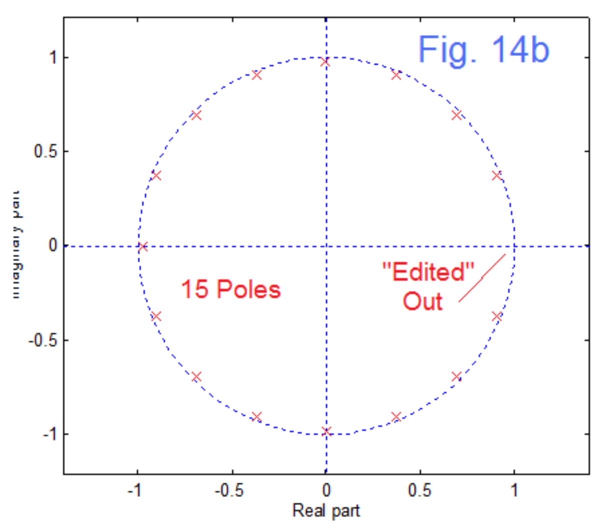
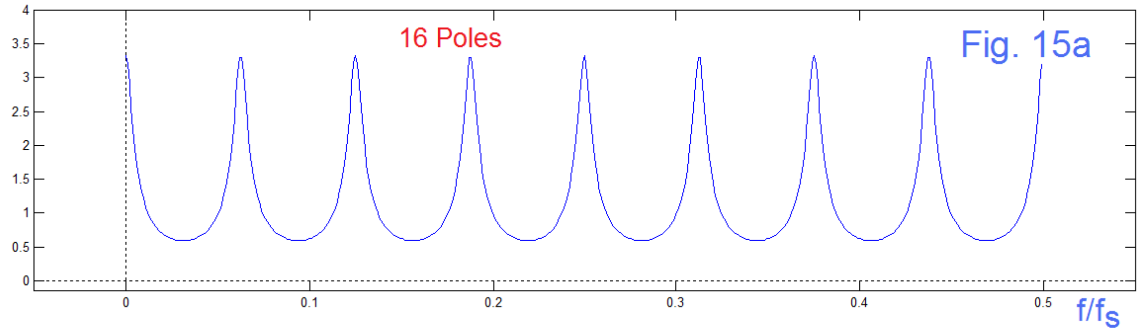
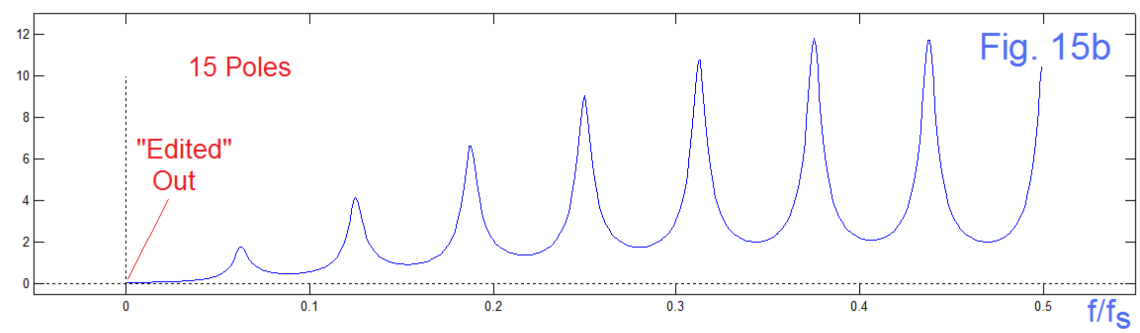


Fig. 14b

We saw in the case of rings of zeros that some simple modification led to interesting alternative responses. In the case of poles, the ideas are similar but slightly harder to implement. For example, suppose we decide to drop from 16 poles to just 15 poles by removing the pole closest to z=1 (Fig. 14a goes to Fig. 14b). This is going to be a radically different denominator (for example, using Matlab's ***poly*** on the 15 kept poles).



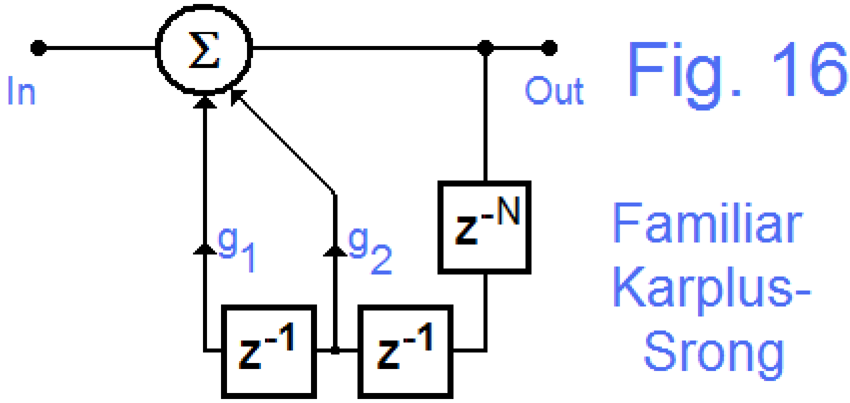



```
newden =  1.0000 0.9780 0.9564 0.9353 0.9147 0.8945 0.8748 0.8555
           0.8367 0.8182 0.8002 0.7825 0.7653 0.7484 0.7319 0.7158
```

which is length 16, corresponding to the 15 poles. If we were going to realize this directly, we would need 15 taps. We could of course also just use the single feedback tap of 0.2 and get the original 16 poles, and punch out the poles we don't want with a zeros at 0.9780, using a numerator term of [1.0000 -0.9780]. So this is not terribly easy. More to the point, perhaps, the resulting frequency response (Fig. 15b) is not obviously useful.

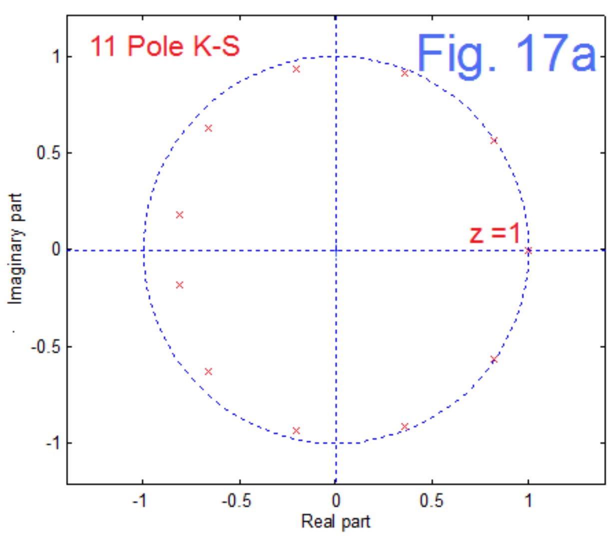
Recall however that for Fig. 11 we found it interesting to replace a tap near the end with a combination of taps, and in this way we found a reasonably good notch coming from "washed out" combs. In effect, we added a moving average to a usual comb. Can we add a moving average (or similar) to the feedback loop? That is, can be put some sort of filter in a feedback loop from a tap at the end before feeding back? Fig. 16 shows what we have in mind and likely it is familiar to our readers as the simplest version of the Karplus-Strong (K-S) plucked string algorithm [4]!

Here we have added to the delay line two additional delays (or just moved taps around) and arranged for some combination of two taps to replace the single tap of Fig. 13. We recognized immediately that in order to have stable poles in Fig. 13, it was necessary that $|g|<1$. Here we need to be more careful.

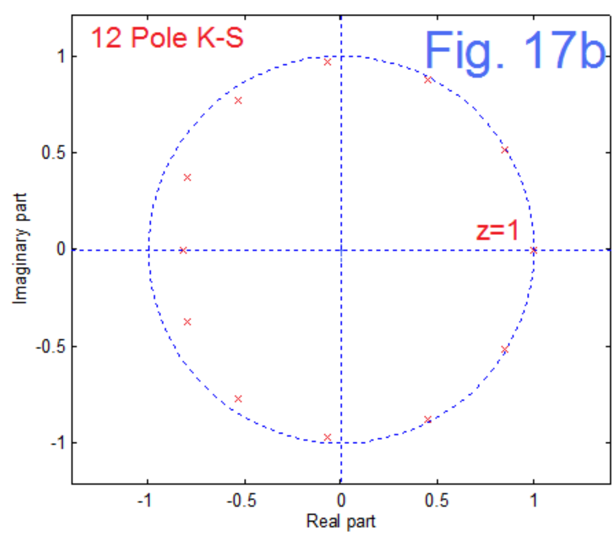


Indeed, we are going to require some aid to calculation: like the Matlab ***roots*** function. Clearly if g_1=0, we would need $|g_2|$<1 and if g_2=0, we would need $|g_1|$<1, and could easily calculate the pole positions. To this we add the fact that in the original K-S paper, the authors used $g_1 = g_2$ =1/2. This is a strange choice, perhaps, if one is thinking of filtering and poles. But K-S involved this structure not so much as a filter (or any notion that that was the stability consideration) but of a self-modifying wavetable. That is, the structure of Fig. 13 with g=1 and the input zeroed, would circulate a waveform as though the samples of the waveshape were being read (periodically) from a table. This sequence would not change. If however, we tried, for example, g_1=g_2=1/2, the sequence would change each full cycle. Further, it is likely intuitively clear that if the values of g_1 and g_2 are small (relative to 1) that the circulating sequence would rapidly decay, and further, that if g_1 and g_2 were individually (or in some combination), large (again relative to 1), we could expect some sort of blow-up. So how good was the guess?

To answer this question, it is a simple matter of choosing g_1=g_2=1/2 for some N. In the examples using the original K-S taps of 1/2, 1/2, we have chosen an 11-pole example (Fig. 17a) and a 12-pole example (Fig. 17b), and for clarity, we simply give the coefficients of the denominator below the figures [don't forget the (–) signs are due to the feedback].



[1 0 0 0 0 0 0 0 0 0 -1/2 -1/2]



[1 0 0 0 0 0 0 0 0 0 0 -1/2 -1/2]

We notice three things immediately. First, we get some sort of a ring of poles (our theme here). Second, the ring is shifted off center to the right. Thirdly, both cases have a pole at exactly z=1 (checked from the printout).

The fantastic thing about the first two observations (ring, and offsets ring) is that this is an array of poles (a set of resonators "ringing" in parallel) that happens to resemble what we would likely try to implement by a much more laborious scheme. That is, the ring of approximately equally spaced pole frequencies resembles what we would suppose a plucked string would involve (string stiffness causing the upper harmonics to run sharp). The offset also would shift the perfect harmonics, but here the principal result of interest is that the higher harmonics decay more rapidly, as would be true of a physical string.

The pole at z=1 (DC) is unstable, or at least is the boundary between stable and unstable. It is not clear if the original authors realized this problem. However, it is likely the case that it made little or no difference – since it was a component at DC and was thus inaudible. Had the pole been slightly further to the right (say z = 1.1) then possibly the output signal could have progressed to clipping. Almost certainly the choice of 1/2, 1/2 as feedback taps was a matter of needing to do a multiply (division actually) by using a bit shift. This is viewed as a machines restriction of the then-available processors. Today, we don't hesitate to use, perhaps, taps of 0.45 and 0.45.

Here we can look at a generalization of the K-S scheme as being the addition of a filter g(z) in the feedback loop of Fig. 13 replacing the fixed value of g. The condition |g(z)|<1 for stability is the same. As such, the taps of 1/2, 1/2 constitute a two-tap moving average which we know to have a gain of 1 at DC and gains less than 1 for all other frequencies in the allowable bandwidth. Hence the pole at z=1. Nothing prevents us from trying other filters. [One of the most "annoying" features of K-S is the number of options that can take up an afternoon of (enjoyable) exploration.]

Let's do a complete example. Choose a length-25 denominator d:

```
d=[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -.19 -.19 -.19 -.19 -.19]
```

which has the 24 poles

```
pt=roots(d)
   0.9977
   0.9537+ 0.2799i        0.9537- 0.2799i
   0.8266+ 0.5299i        0.8266- 0.5299i
   0.6313+ 0.7211i        0.6313- 0.7211i
   0.4032+ 0.8230i        0.4032- 0.8230i
   0.2178+ 0.8730i        0.2178- 0.8730i
  -0.0058+ 0.9291i       -0.0058- 0.9291i
  -0.2635+ 0.8993i       -0.2635- 0.8993i
  -0.4946+ 0.7819i       -0.4946- 0.7819i
  -0.9280
  -0.8829+ 0.2520i       -0.8829- 0.2520i
  -0.7658+ 0.4514i       -0.7658- 0.4514i
  -0.6548+ 0.6077i       -0.6548- 0.6077i
```

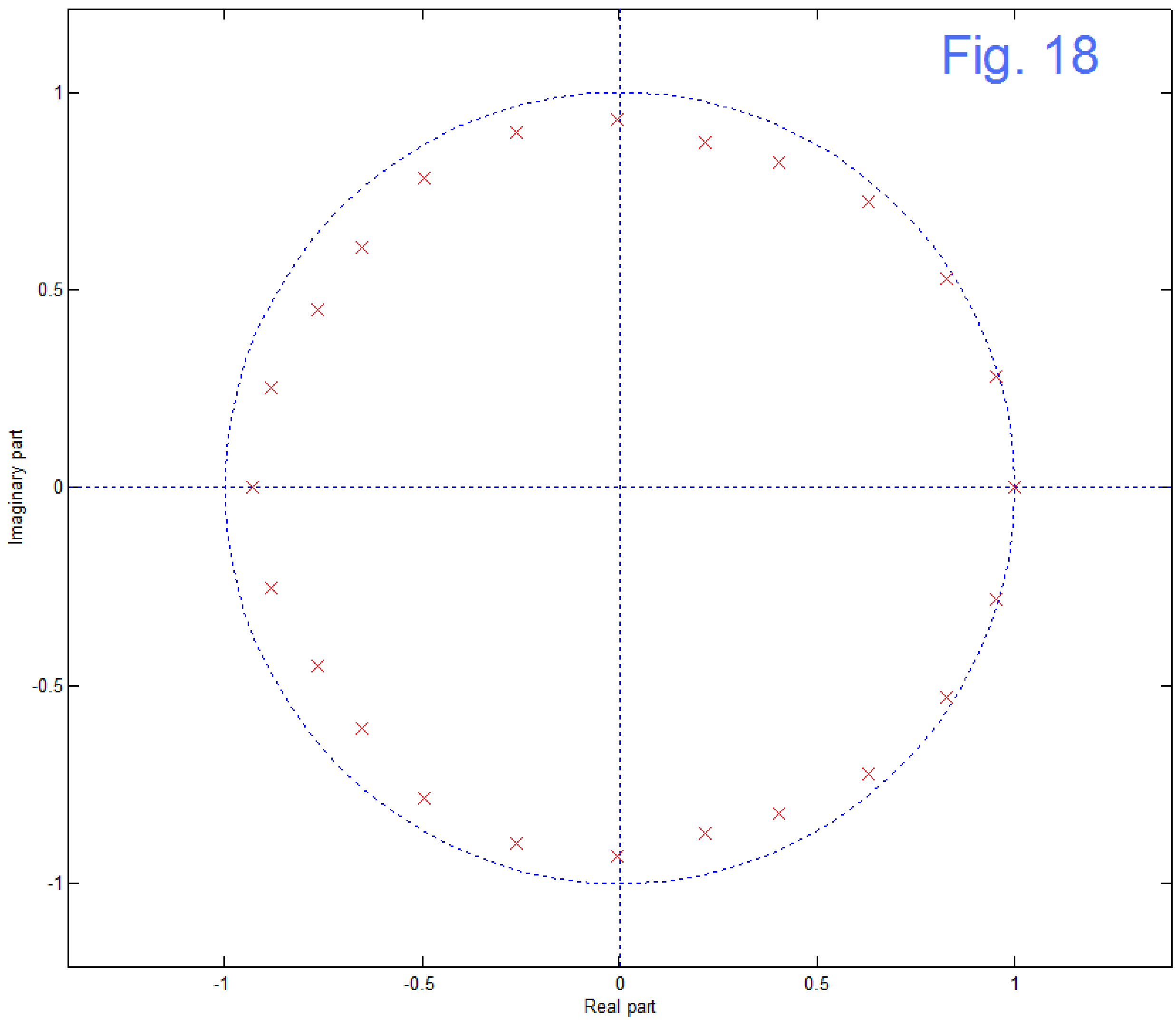
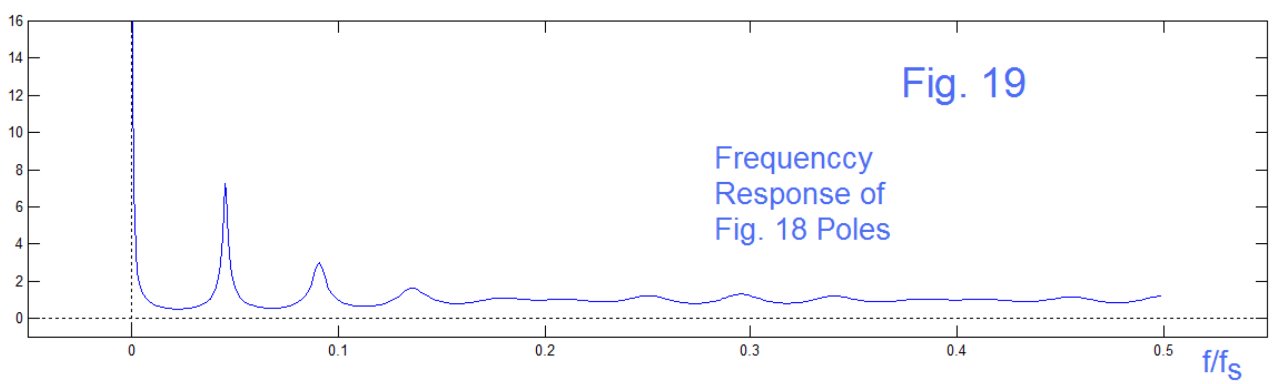
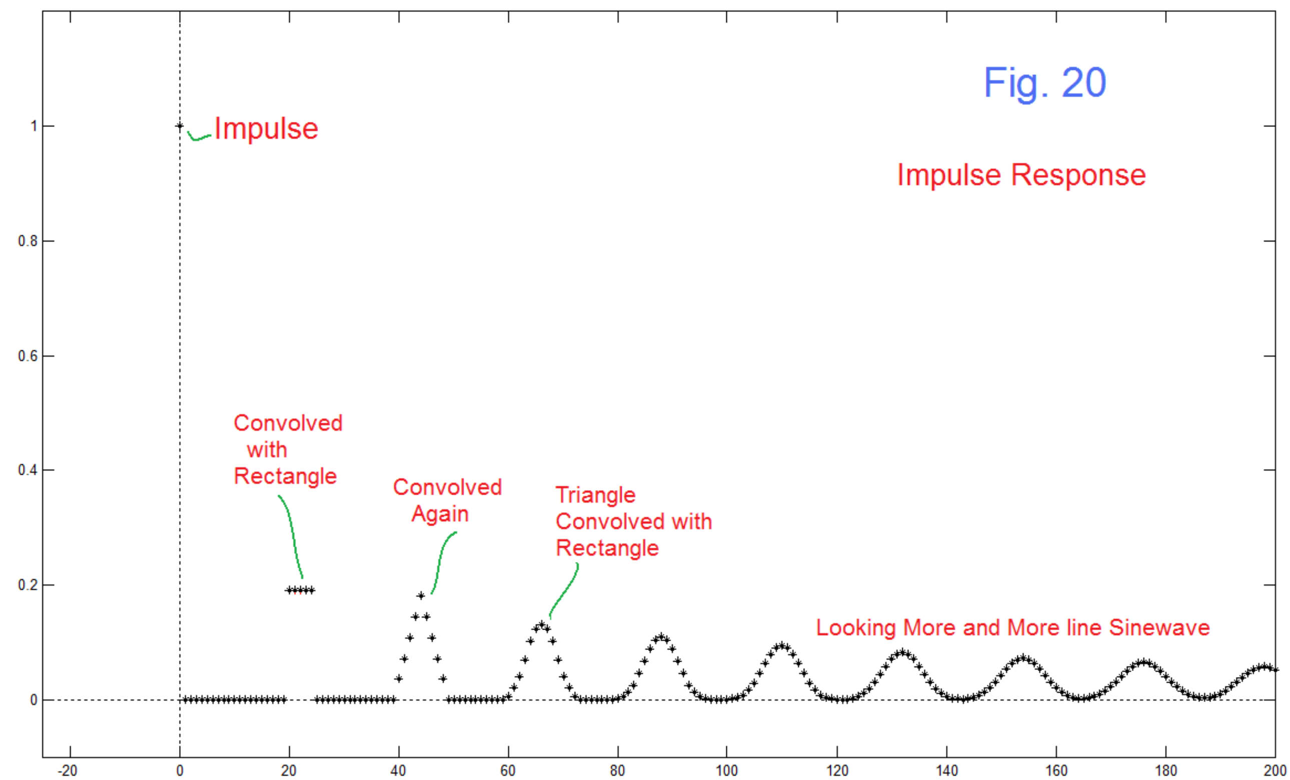


Figure 18 is a plot of these 24 poles, and we note the roughly circular, equally-spaced but right offset ring structure. From the graph, we can not be sure there is not a pole at z=1. Note from the tabulation that the pole is at 0.9977. Note that we show here all five of the rectangular taps at the end as 0.19; if they had been 0.20, the pole would be at z=1. Other combinations of taps could well give poles outside the unit circle. Fig 19 shows the corresponding frequency response.





Possibly what is the <u>most revealing</u> is to look at our example <u>in the time domain</u>, and this is the lovely plot of Fig. 20. Adding to the “all afternoon” game we postulated is the additional possibility of using different inputs (or alternative preloadings of the delay line). Most basically, we likely think in terms of inputting an impulse (plucking a string). Fig. 20 shows the output for the poles of Fig. 18.

At the far left, we see the input impulse. After the impulse moves along the untapped portion of the delay line, it reaches the five rectangular taps (each 0.19) at the end. In consequence, five output samples of value 0.19 come out (a length-5 rectangle). This goes back down the delay line and when it again encounters the rectangular taps, we get a length-9 triangle. Once more around, we get a triangle convolved with the same rectangle (length 13).

<u>So, in the time-domain view, all that is happening here is that any current cycle is being repeatedly convolved with whatever filter we have in the feedback loop</u>. Correspondingly in the frequency domain the signal is being repeatedly filtered by the feedback filter. Here the filter is low-pass (rectangular taps) so we can understand how the output tends to look sinusoidal at the largest resonance (period 25). Everything fits.

Three more examples will finish what we are up to here. First, we want to look at a system that is not <u>just</u> marginally unstable (pole at z=1), but unstable:

d=[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -.22 -.22 -.27 -.22 -.22]

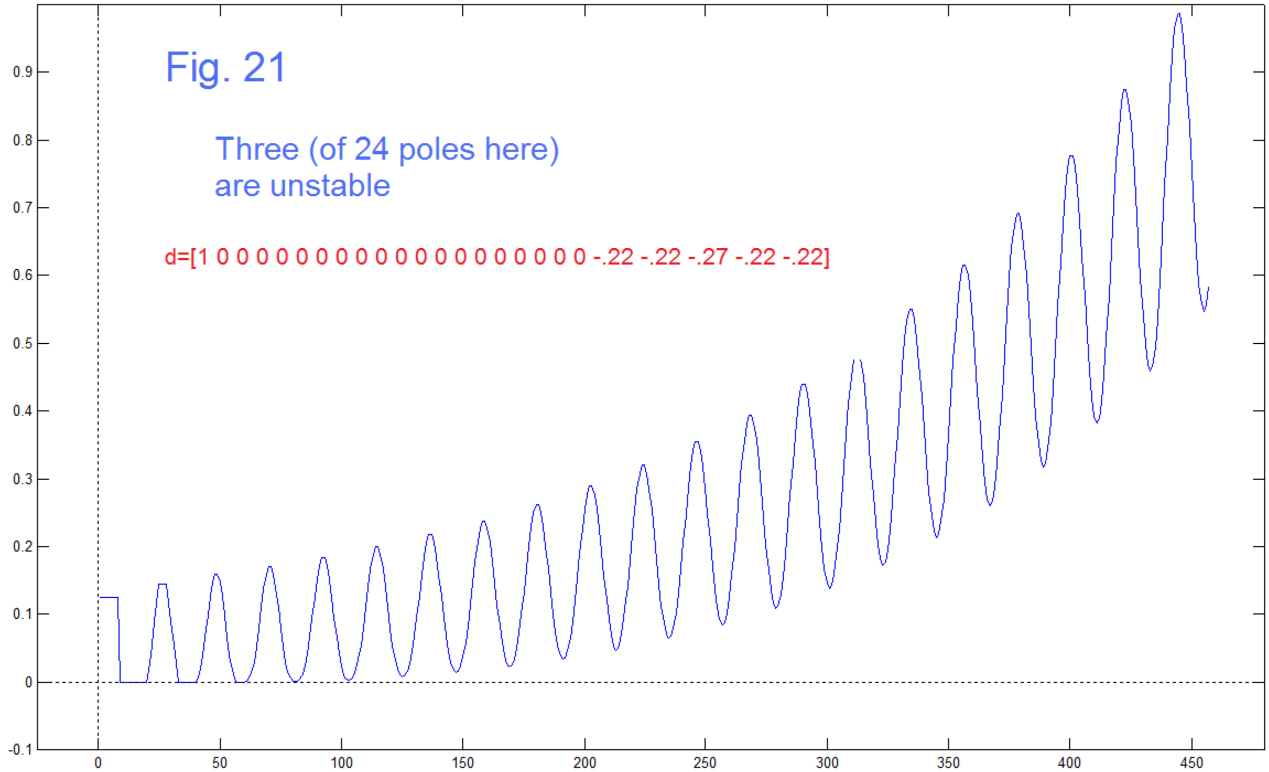
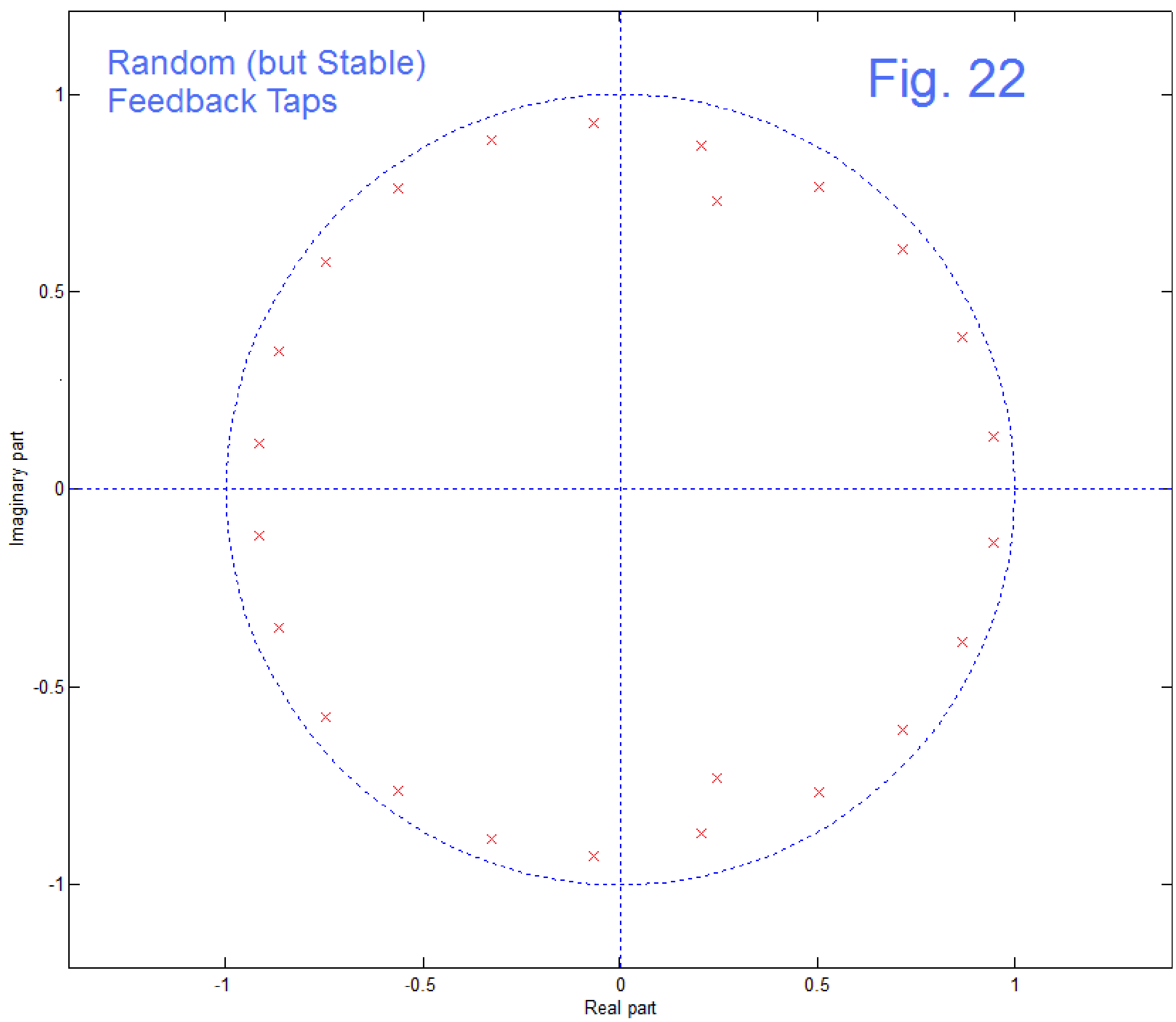
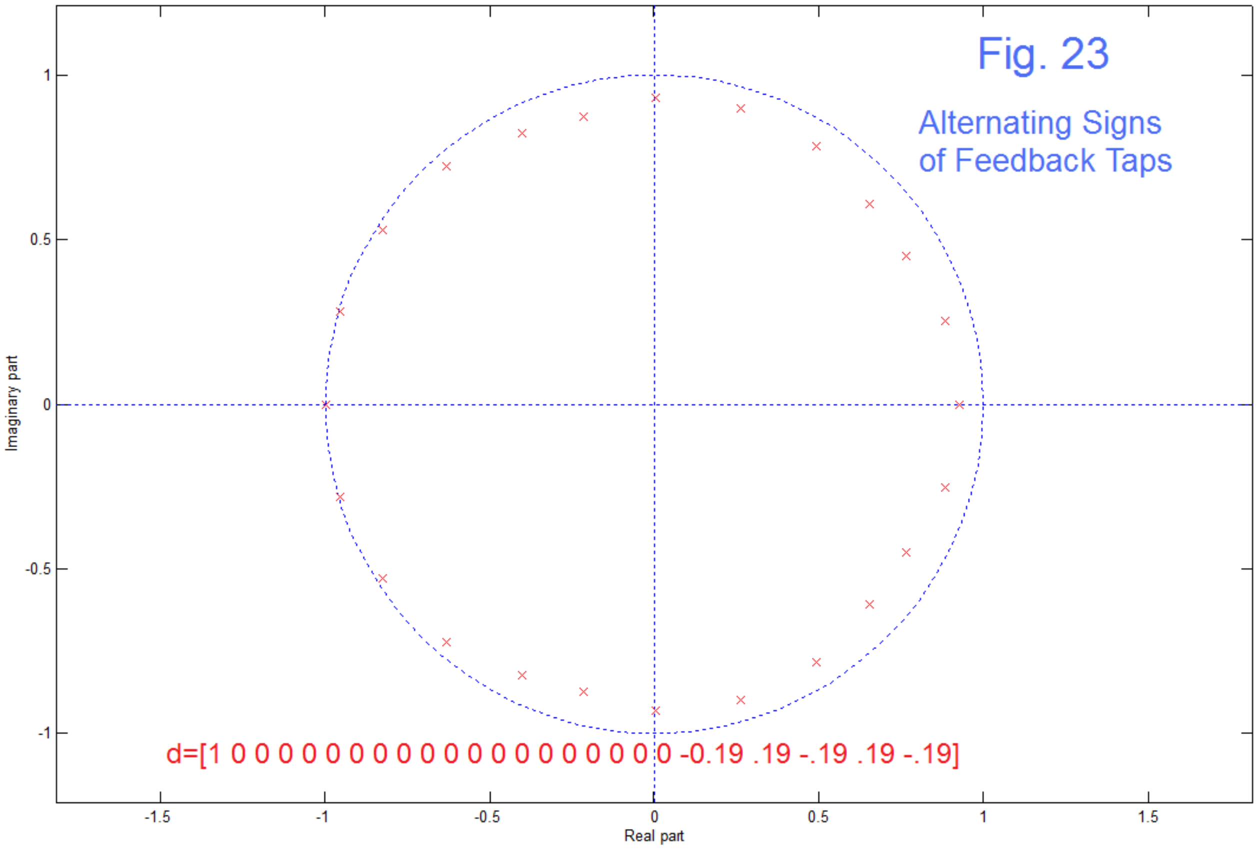


Fig. 21
Three (of 24 poles here)
are unstable
d=[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -.22 -.22 -.27 -.22 -.22]
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
50
100
150
200
250
300
350
400
450



Random (but Stable)
Feedback Taps
Fig. 22
1
0.5
0
-0.5
-1
Imaginary part
-1
-0.5
0
0.5
1
Real part



In fact, this (impulse response as shown in Fig. 21 – the K-S output) has three unstable poles (of the 24) and its impulse response blows up as we expect. Audibly there is little objectionable effect until the sequence ends at which time there is the expected "thump" of the change of the DC level from that at the end back to zero.

Secondly. Possibly we think of ourselves as being in a rut with regard to ring-like arrays of poles. [After all however, that was the theme here!] It might seem that we could get something much more general by trying radically different feedback filters. We have been very conservative in choosing the feedback taps. What if we chose the five taps on the end (for example) as random numbers. Likely we might guess that the pole array would scatter much better throughout the unit circle. We would however want to restrict ourselves to stable systems. Choosing five taps at random would of course not guarantee a stable system. We do know that the feedback filter g(z) would need to have a maximum gain of 1: |g(z)|<1. So we could check random trials for this. As usefully (more so actually) we can compute the magnitude of g(z) and normalize to something (perhaps like 0.98). That is, the taps remain random but they are all adjusted by some constant (the same for all five) to give stable but "high Q" simulation. What we get is – A RING-LIKE ARRAY OF POLES! (Fig. 22 typical). It won't go away. Why?

From Fig. 13 we immediately saw that a single feedback tap at the far end gave a circle of poles. Now, consider two taps, as in Fig. 16, (or a relatively small number of taps all way out at the far end). We have a phenomenon where there is just "something way out there" and the details are "unfocused". MOSTLY we have that zero-tap weight middle as a defining feature, leading to the corresponding ring-like structure of the poles. We are restricted to categories. We do NOT get general arrays because we have not allowed a general impulse response.

One final note. We have noted that a filter can be "modulated" by multiplying the impulse response by (-1^n) reversing low-pass/high-pass. This applies of course to g(z), the feedback filter. This is known with the K-S procedure where the skewing of the ring of poles can be to the left rather than to the right (Fig. 23). This results in an interesting but "less musical" tone as it has mainly higher harmonics (note the higher Q poles are on the left). [This is not a new finding for K-S.] Note that the signs alternate. It is not a matter of making the right side taps negative – although it may seem to be that way if there is only one tap as we have in Fig. 13. Easy to be misled!

REFERENCES:

[1] B. Hutchins, "Basic Elements of Digital Signal Processing – Filter Element, Part 1" Electronotes, Vol. 20, No. 197, April 2001. (See also additional parts in EN#198 and EN#199)
http://electronotes.netfirms.com/EN197.pdf

[2] ELECTRONOTES APPLICATION NOTE NO. 422, Mar 6, 2015, "NOTCH FILTER AS A 'WASHED-OUT' COMB"
http://electronotes.netfirms.com/AN422.pdf

[3] https://www.youtube.com/watch?v=5uxsFglz2ig
Here workers are welding a rail and then hammering the loose material. Pounding begins at about 3:00 of the video. Try to ignore the sound of the hammer blows and listen for the following "buzz" reflections.

[4] Karplus, K., and A. Strong, "Digital synthesis of plucked-string and drum timbres", Computer Music Journal Vol. 7, No. 2, pp 43-55, Summer 1983
http://users.soe.ucsc.edu/~karplus/papers/digitar.pdf