Effect Design*

Part 1: Reverberator and Other Filters

JON DATTORRO, *AES Member*

CCRMA, Stanford University, Stanford, CA, USA

The paper is a tutorial intended to serve as a reference in the field of digital audio effects in the electronic music industry for those who are new to this specialization of digital signal processing. The effects presented are those that are demanded most often, hence they will serve as a good toolbox. The algorithms chosen are of such a fundamental nature that they will find application ubiquitously and often.

0 INTRODUCTION

This paper is intended to serve as a point of reference in the field of digital audio effects for the electronic music industry. It is for those who are new to this specialization of digital signal processing so as to advance their skill level at inception. The effects presented herein are those demanded most often; hence they will serve as a good toolbox. The algorithms chosen are of a fundamental nature and therefore will find application ubiquitously and often. They include one for reverberation, two for filtering, two for delay-line interpolation, one for chorus (as well as vibrato and flanging), four for sinusoidal oscillation, and one for noise generation.

It is not necessary to start reading the paper from the beginning. The overall tone of the paper is tutorial, stressing concepts. The supporting mathematics go to some depth in those cases where the algorithms are analyzable. The reader is not required to delve that deeply; in some cases knowledge of the results alone is sufficient. The mathematics serve to develop concepts, to justify conclusions rigorously, and to offer aid when one runs into trouble. Of course, the best way to learn is to try the algorithms and invent one's own.

Our hardware reference standard is a dedicated 24-bit two's complement fixed-point digital signal processor chip [1], typically having 48-bit accumulation of products, but these algorithms will certainly run on any personal computer. Nevertheless, much of the mathematics deals with the impact of finite precision. That will demand consideration when someone complains of too much noise or grit in your output signal.

The two most asked-for effects are chorus and reverberation. Reverberation creates an ambient space in the perception of the listener. The reverberator presented herein is the smallest recursive network we found that meets subjective requirements of good sounding reverberation. This reverberator is not analyzed in great mathematical detail; it is best explored by tinkering, because that is how it was developed. There are few enough knobs so that the sonic impact of each is readily discernible.

Filtering for musical purposes involves somewhat different considerations than what is taught in standard texts on digital signal processing (DSP). The most notable departure is that of the half-power excursion[1] of the magnitude response when regarding audio filters that are typically shallow. Simple and accurate design equations for an easy-to-operate second-order notch filter and resonator are developed from the musician's point of view. A unifying framework for both filter types develops into the Regalia–Mitra topology, which facilitates parametric equalization. We then apply the same simplifying concepts to the musician's popular second-order all-pole filter, which is used for a wide range of purposes, spanning wa-wa to dynamic noise rejection. The musical filtering sections culminate with a unique realization of that popular filter—the versatile and quiet Chamberlin filter topology, the digital analogue to the Moog voltage-

* Manuscript received 1996 March 14; revised 1996 September 14 and 1997 June 28.

[1] A relative as opposed to absolute measure.

controlled filter.

We scrutinize linear interpolation as a means for delay modulation. The modulating delay line forms the basis of many standard audio effects. The inherent filtering artifact of the linear interpolation process is often overlooked, however. We offer an alternative, called allpass interpolation, which avoids the pitfalls in some circumstances and sounds very analog. The chorus effect is well served by this alternative method of interpolation. Chorusing emulates a multiplicity of nearly identical sound sources. When only two sources are emulated (two voices, including the original), we consider that to be the industry-standard chorus effect. This perceptually pleasant effect is hard to describe and must be experienced to be fully apprehended.

Sinusoidal oscillators are found within nearly every audio effect. Although oscillators can generate sound, more often than not they are used to control some modulation process. Delay modulation is a key to successful reverberator design. Writing a few simple instructions, it is easier to design a terse algorithm to generate a sine wave than it is to employ a table lookup. The algorithmic approach also results in a purer sinusoid. We examine several efficient methods of sinusoid generation, and we offer guidelines to aid in the choice.

Noise generation, seemingly the antithesis of sinusoid generation, is discussed. The exceedingly simple maximal-length pseudorandom noise generator is presented as a pleasant and soothing sound source. Not only does this simple circuit produce a pseudorandom bit stream, it also emits a sequence of pseudorandom multibit words, each repeating only once per cycle. The cycle time can easily be designed to exceed the duration over which the human ear can identify patterns. The classical literature on these circuits demonstrates that the autocovariance of the single-bit pseudonoise sequence is a Kronecker delta. Hence the single-bit noise is uncorrelated and spectrally white. In the multibit case we find that to be only approximately true. Short-lived exponential patterns are visible within the pseudonoise sequence, revealing correlation. Hence the power spectrum of a multibit noise realization via this circuit cannot be perfectly white without equalization. We show how the multibit pseudonoise sequence can be precisely modeled as linear FIR filtering of the single-bit sequence. Thus the power spectrum of the uniform amplitude-distribution multibit sequence is known, and we suggest a simple method of equalization.

1 REVERBERATION

Digital reverberators are like paintings. There are zillions of them, all of different colors, as no one wants the same painting in every room. The engineer's pipe dream of the universal reverberator may never be realized. A treatise on artificial reverberation would easily fill volumes. In the past, these networks were so difficult to analyze (like Bach fugues) that they have traditionally been invented through experimentation. The reason for the difficulty is that even the most efficient implementations of reverberators rely on all-pass circuits embedded within very large globally recursive networks. The all-pass circuits themselves have recursive delays measured in hundreds of sample periods, whereas the cumulative delay around a large recursive network can total on the order of tens of thousands of samples.

The early successful commercial inventors of these complicated networks were Griesinger and Blesser. Unfortunately they have written little on this topic. Moorer and Gardner have turned the art more into a science. Moorer elevates the seminal but crude work of Schroeder. Gardner provides a technical chronicle of developments in the art of reverberator design, where he also furnishes a synopsis of his complete translation of the French vanguard Jot.[2] We do not provide sufficient background material to permit the reader to fully understand the development of the reverberation network for plate emulation presented herein. The reader is encouraged to refer to the references [2]–[9, pp. 1–28].

1.1 Simple Reverberation Network[3]

Fig. 1 shows one particular network for producing reverberation. We like this topology for several reasons: 1) It has simple knobs, which easily control particular aspects of the reverberated sound, such as input and decay *diffusion* (decorrelation), decay rate, high-frequency damping, and input signal bandwidth. 2) The style of the topology is more computationally efficient than most others known. 3) It has demonstrated applicability to a broad range of signal sources.

We selected the network in Fig. 1 for presentation because it is the smallest reverberation network we found (in memory and complexity) that is good sounding. We believe that there must be a limitless variety of such networks, however. The question naturally arises as to why the simple digital network shown produces such convincing reverberation. We can answer this only qualitatively.

Consider the plucked string of a violin. Its envelope may be described as having a *coherent* exponential decay. It is this character that is theorized to be one of the primary discriminants of nonreverberated sound. Reverberating this sound, on the other hand, would tend to randomize the string envelope and phase, producing a bumpier, extended, more diffuse and dynamic decay.

This oversimplified qualitative description of the process of reverberation has actually found its way into early commercial products. Long before DSP chips could be integrated into sampler type synthesizers, reverberated sampled sound was simulated by altering the decay characteristics of recorded dry samples by randomizing an overlaid envelope applied at playback. While not absolutely convincing, this kind of aural cue

[2] Jot recently formulated an analytical design method for recursive reverberators. The work is based on a unitary (lossless) feedback loop in a state-space network, where he claims arbitrary time and frequency density

[3] This discussion is adapted from conversations with Barry Blesser and David Griesinger and is supplemented by Appendix 1 in Section 1.5

was enough to cause pioneers [3]–[5] to question the premise of Schroeder's precipitative work with delay lines at Bell Labs during the early 1960s.

One can deduce from Schroeder's work [7] that to achieve the ideal of colorless reverberation, the *eigentone*[4] density of the network needs to approach 3 per hertz. It can also be theorized that the limit on the number of achievable eigentones is proportional to the total delay-line memory [4]. From our current perspective we

[4] An eigentone of a network in this context is a circuit resonance.

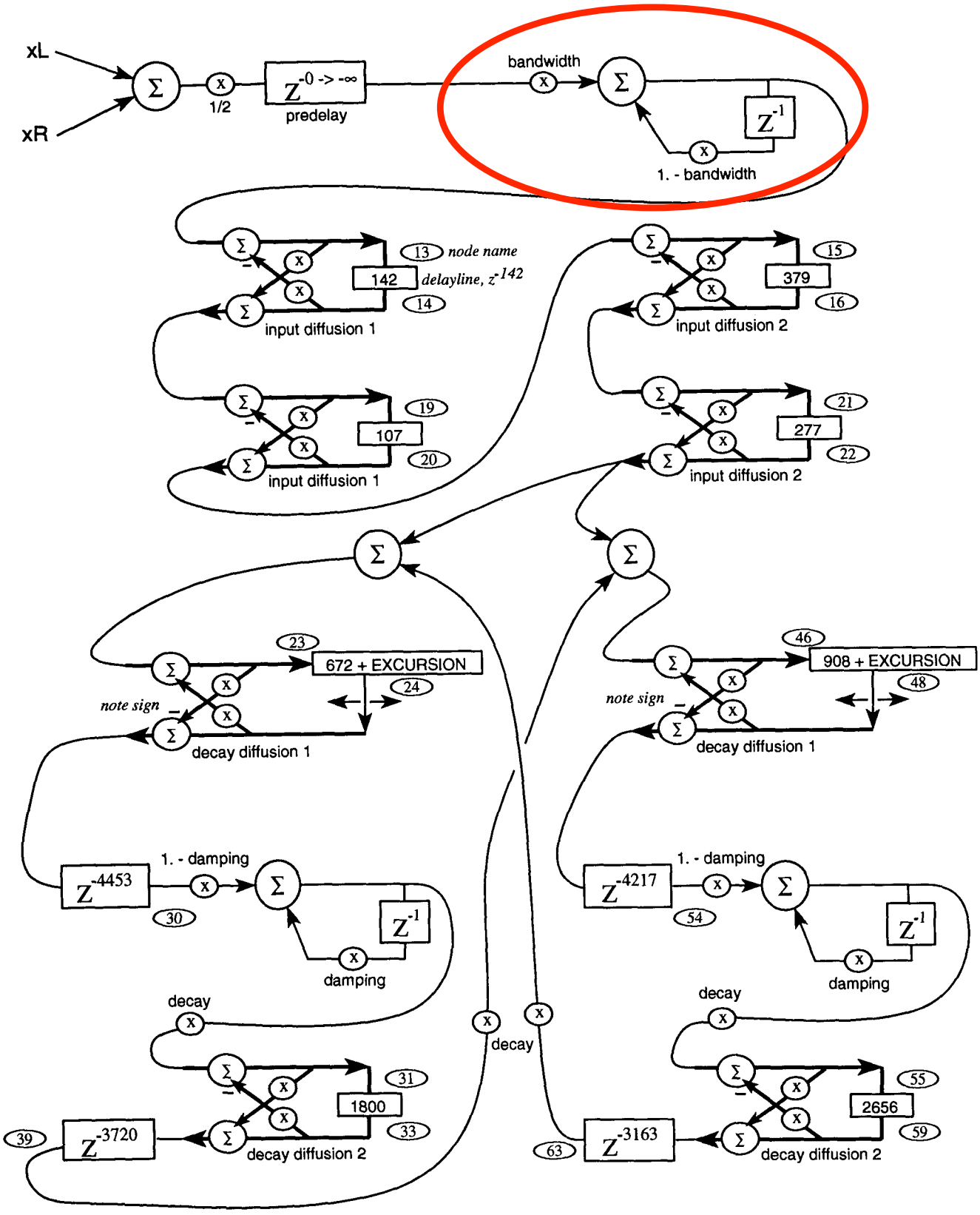

Fig. 1. Simplified plate-class reverberation topology in the style of Griesinger. For output tap structure (y_L, y_R) see Table 2. Delay-line taps at nodes 24 and 48 are modulating.

know that emulation of physical spaces can be convincingly performed using sample rates as low as 20–24 kHz. This is true because of typically rapid acoustical absorption in the high-frequency region, and because the desired output is a mix with the dry input signal. This bandwidth would then require about 30 000 eigentones, hence about 64K words of delay-line memory. In the 1960s, that amount was not economical.[5]

In reverberator design, while a good general rule regarding delay-line memory is certainly "the more the better" [4], the efficient reverberation network shown in Fig. 1 stands as a testimony[6] that Schroeder's eigentone density criterion, predicting about 88K words of memory, is not a hard and fast rule. Of at least equal importance are the decorrelation of the decay and the associated time density of the echoes, that is, one must achieve a balance between eigentone density and echo density.

1.2 Color

On the other hand, our reverberation network's signal response is not colorless. Empirically we find that some of the most sought after commercial reverberators are somewhat colored in their frequency responses. This means that their outputs impose some conspicuous audible resonances upon the input signal. Consequently it is not unusual to find as many musicians and recording engineers who like a particular reverberator as those who do not.

We also find that some recording engineers do *not* want an accurate emulation of a physical space, because the reflection density takes too long to build. Instead, they sometimes want instantaneous high-density reflections with smooth exponential decay of the envelope, having randomization in only the phase trail. This desire most closely describes the plate class of reverberators, which we present here.

1.3 Discussion of the Reverberator

Scrutinizing the reverberation topology in Fig. 1, we can break it down into a cascade set of four input diffusers (the lattices) followed by another set of four tank diffusers, the latter arranged so as to feed back on themselves globally. The first set of diffusers acts to quickly decorrelate the incoming sound somewhat, preparing that sound to be looped indefinitely in the holding tank formed by the second set of diffusers. What we hear comes from a large set of output taps (not shown) located within the tank.

1.3.1 Input Diffusers

All the diffusers are all-pass filters having the topology of a lattice. The purpose of the four input diffusers is to decorrelate the incoming signal quickly before it reaches the tank. The tank recirculation can sometimes become perceptible as strong cyclic events if the input signal is not preconditioned in this manner. This function becomes especially important for the successful reverberation of percussive sounds. One may think of this function as signal-phase randomization, to reduce peakedness and other strong features of the input waveform.

No diffusion corresponds to zero-valued all-pass coefficients, while coefficient magnitudes close to unity produce buzzing that is local to the afflicted all-pass filter. Optimum diffusion for the all-pass filter lies somewhere in a region closer to |0.5| than to the extreme values of the coefficients. The preset values given in Table 1 were determined by trial and error.

1.3.2 Tank

We identify the reverberation tank as the recirculating four lowest diffusers in Fig. 1. We call it a tank because its purpose is to trap the incoming sound by making it recirculate through the global figure eight. The four decay coefficients determine the rate of decay. When the decay coefficients are set very close to 1.0 (and the damping filter within the tank is turned off), the sound will remain held in the tank indefinitely. That in itself is a neat effect, but unless the sound metamorphoses while in the tank, it is easy for us to detect the looping pattern of sound. The purpose of the diffusers within the tank, then, is to eliminate any aural pattern in the recirculation. The tank diffusers are not always successful (being signal dependent), and their settings are critical to achieve an overall exponential decay; everything must be set by ear.

The tank, in summary, is a simple device whose purpose it is to alter the tail of a decaying sound, as mentioned already. The tank diffusers have been further grouped into pairs labeled by the knobs "decay diffusion

[5] The Lexicon model 224 digital reverberation system introduced in 1979 originally possessed only 16K words of memory, operating at a sample rate of 20 kHz. That memory amount doubled shortly thereafter. The Elecktromesstechnik Wilhelm Franz KG EMT-250 digital reverberator distributed in the United States by Gotham Audio Corporation beginning in 1977, operated at a sample rate of 32 kHz, having only 8K words of memory. The precursor to this machine is described in [10, ch. 2].

[6] Given a 30-kHz sample rate and having only 22K words of memory (not including predelay).

Table 1. Reverberation parameters default.

Sample rate F_s = 29761 Hz	
EXCURSION = 16	Maximum peak sample excursion of delay modulation
decay = 0.50	Rate of decay
decay diffusion 1 = 0.70	Controls density of tail
decay diffusion 2 = 0.50	Decorrelates tank signals; decay diffusion 2 = decay + 0.15, floor = 0.25, ceiling = 0.50
input diffusion 1 = 0.750	Decorrelates incoming signal
input diffusion 2 = 0.625	
bandwidth = 0.9995	High-frequency attenuation on input; full bandwidth = 0.9999999
damping = 0.0005	High-frequency damping; no damping = 0.0

1" and "decay diffusion 2." The tank diffusers have overlapping functionality. The dichotomy we make is aurally subtle and pertains to the temporal location of the diffusers in the tank with respect to a stereo tank input, that is, exactly *when* they diffuse the tank signal with respect to the signal onset. The effect of these knobs is best observed using a percussive input, or what Griesinger refers to as a "pink click."[7]

1.3.3 All-Pass Lattice Topology

Each diffuser has been given the topology of a two-multiplier lattice. The eight lattices shown in the reverberator schematic in Fig. 1 are used in this reverberation effect as all-pass filters, each having a long impulse response time.[8] The two coefficients within each individual lattice must remain identical to maintain the all-pass transfer function, which is insensitive to coefficient quantization. The recommended range of these coefficients is from 0.0 to 0.9999999 (q23; see Part 3, Section 9.1, Appendix 7) If the lattice coefficients should exceed 1.0, instability would result. Making them both negative will change the character of the impulse response[9] but does not destroy the all-pass transfer. This change in character is exploited in the lattices having the coefficients called "decay diffusion 1" in the schematic. This character change further enhances the dichotomy between the two pairs of tank diffusers.

All-pass response is the forced (steady-state) response of each lattice output with respect to its own input.[10] Because the impulse response of each individual lattice within the reverberator schematic is so long, in some cases the integration time constant of the human hearing system is exceeded. This means that an all-pass filter output may be perceived as discretized events, that is, not all pass.

This all-pass lattice topology tends to clip prematurely at internal nodes, so the input to each lattice cannot be presented with a full-scale signal at all frequencies. We like this all-pass lattice, however, because it is efficient in its implementation.

1.3.4 Magnitude Truncation

Lattices produce distinct low-level tones, after the input signal has been removed, known as zero-input limit cycles. The origin of these tones stems from ongoing signal quantization in a recursive topology. The spontaneous tones can be eliminated through the use of magnitude truncation (truncation toward zero; see Part 3, Section 9.2, Appendix 8) of the double-precision intermediate results written out to single or lower precision delay-line memory. Magnitude truncation is well known to subdue limit cycles in digital networks composed of ladders and lattices [9].

Only the recursive circuits require magnitude truncation. In Fig. 1 the write to the predelay does not require magnitude truncation. If delay-line memory is 24 bits in width, then the need for magnitude truncation is obviously lessened when compared to having delay-line memory of only 16 bits in width.

Magnitude truncation, in the specific case of reverberator tank topologies employing lattice or ladder all-pass circuits, can reduce the network noise floor by 12–24 dB after the input signal is removed. The reason this is true is that the predominant noise mechanism is zero-input limit-cycle oscillation,[11] a multiplicity of which is perceived as a whooshing ocean noise floor. The magnitude truncation makes the reverberator output eventually go to absolute zero, two's complement. The disadvantage to its use is that the THD + N (total harmonic distortion + noise) of a steady-state sinusoid through the linear reverberator network can be increased by anywhere from 0 to 6 dB.

1.3.5 First-Order Filters

The three single-pole low-pass filters used for input signal bandwidth control and reverberator tank damping will not clip prematurely at any node [11, ch. 11.3], [12, p. 857] when implemented as direct form I. The damping filters cause high frequencies to decay within the tank more quickly than low frequencies. On the input-bandwidth filter, the bandwidth coefficient tracks the cutoff frequency. In contrast, the damping coefficient is high when the damping filter cutoff frequency is low. The recommended range of these coefficients is from 0.0 to 0.9999999 (q23; see Part 3, Section 9.1, Appendix 7).

Because they are all first-order low-pass filters, any low-level zero-input limit cycles they might produce would be at dc, that is, they will *not* produce tones like the lattices [11, ch. 11.5]. Any signal-truncation noise power spectrum generated by the filters themselves will be centered at dc, since it follows the pole frequency. The peak gain of the noise power spectrum is not great because typically the lone pole is relatively far from the unit circle.[12]

1.3.6 Output Tap Points

From the pseudocode note that the delay-line tap structure forming the stereo output signal y_L and y_R is an all wet (reverberated) signal (Table 2). This particular output tap structure is characteristic of the plate emula-

[7] A click source having a pink spectrum.

[8] The impulse response is that of an upsampled first-order all-pass filter. This filter basically has an exponentially decaying impulse response with or without a multiplicative factor of $(-1)^{n-1}$, depending on the sign of the coefficient. The up-sampling factor L is determined by the number of samples in the lone delay line z^{-L} within the lattice. The up-sampling process inserts $L - 1$ zeros between every sample of the impulse response of the corresponding first-order all pass filter.

[9] Via the multiplicative factor $(-1)^{n-1}$ on the impulse response.

[10] The reverberation network as a whole does not have an all-pass transfer function, although we would like that to be the case. Smith [9, pp. 1–28] has found a way to make an entire reverberation network all pass. Smith's method is based on the interconnection of lossless waveguides.

[11] Here we use the term "limit cycle" in the classical DSP sense.

[12] If instead the filters were high pass, limit-cycle tones might be produced at Nyquist while the truncation noise power spectrum would also be concentrated there.

tion class of reverberation networks.[13] Also note that the output tap structure produces a *synthetic* stereo image because the stereo input is converted to a monophonic signal at the reverberator input for this particular topology. Normally, the desired output is a mix of the stereo reverberated signal y_L and y_R with the original (dry, full-bandwidth) stereo input signal x_L and x_R.

1.3.7 Delay Modulation

Linear interpolation or, better yet, all-pass interpolation (as discussed in Part 2, Section 5) can be efficiently employed to modulate slowly[14] the nominal tap point of the two indicated delay lines in the schematic. A slight modulation will introduce undulating pitch change into the tank. For signals with much high-frequency content, such as drum sets, these built-in modulators serve to break up some pretty audible modes, that is, the amount of tank diffusion is effectively increased.

Barring air currents and temperature fluctuations, there is no analogue to this modulation process in a real room (unless the walls are moving). Without the modulation, we may well describe the imaginary space emulated by the given digital network as being enclosed by a picket fence. The slow modulation serves to increase effectively the sheer number of resonances (eigentones, modes of oscillation, picket density) in the tank. The number of resonances in a real room, hall, or plate is probably far beyond what is existent in our little (nonmodulating) reverberation network. In the case of drum sets, the modulation is a godsend. In the case of piano, the modulation, though slight, may be objectionable because of a perceived vibrato.

Ideally, all the delay lines in the tank diffusers should be modulated using different modulation rates and depths. In that case, the diffusion burden becomes more distributed. Hence the required rate and depth of modulation are lessened for each diffuser. When computation time is a constraint, then one should preferentially select the stereo pair of the diffusers appearing earliest in the tank, as we have, to maximize the increase of effective resonances. In this case, the same rate and depth are used for each diffuser in the pair, but we use a quadrature oscillator to decrease the correlation. (Sinusoidal oscillators are discussed in Section 7.) The differing delay-line lengths of all the diffusers also serve to decrease the correlation.

As explained in Section 4; linear interpolation for delay modulation will introduce time-varying low-pass filtering as an artifact, thus supplying some unaccounted damping to the tank. All-pass interpolation overcomes this particular problem and is perfectly applicable to reverberators because the required pitch change is microtonal.[15]

[13] The physical "plate," actually resident in some contemporary recording studios, fills a small room in some embodiments. The best plates are constructed using a solid gold foil. The input signal is typically injected onto the plate via one or two transducers, while each output is the sum of multifarious signal taps, each tap transduced at a different location on the plate.

[14] At a rate on the order of 1 Hz, and at a peak excursion of about 8 samples for a sample rate of about 29.8 kHz.

Table 2. Output taps.

```
/********* left output, all wet *********/
accumulator  = 0.6 X node48_54[266]
accumulator += 0.6 X node48_54[2974]
accumulator -= 0.6 X node55_59[1913]
accumulator += 0.6 X node59_63[1996]
accumulator -= 0.6 X node24_30[1990]
accumulator -= 0.6 X node31_33[187]
y_L = accumulator - 0.6 X node33_39[1066]

/********* right output, all wet *********/
accumulator  = 0.6 X node24_30[353]
accumulator += 0.6 X node24_30[3627]
accumulator -= 0.6 X node31_33[1228]
accumulator += 0.6 X node33_39[2673]
accumulator -= 0.6 X node48_54[2111]
accumulator -= 0.6 X node55_59[335]
y_R = accumulator - 0.6 X node59_63[121]
```

1.4 Conclusion

Choosing a particular reverberator for a particular application is commonplace, and purveyors of such equipment have been known to purchase an audio signal processing box just to acquire one particular algorithm.[16] At some level, the choice of reverberator becomes a matter of taste, much like art. There is no one universal reverberation network that satisfies everyone for each and every application; we speculate that there never will be.

1.5 Appendix 1: Reverberation Recollections

Dear Jon,

What you wrote was fine, but it stimulated my memory of additional snippets. Feel free to use what you want.

I had a personal conversation with Manfred Schroeder in the late 1970s and I asked him the question about what the phrase "maximal incommensurate" delay values meant, as it appeared in one of his reverberation papers. His answer was particularly interesting. This is a paraphrase based on my tired memory:

> We did the electronic reverberation for amusement because we thought it would be fun. Since it took the better part of a day to do 10 seconds of reverberation, we only ran one sample of music. The notion of delay time selections was random in that we just picked a bunch of numbers and there was no mathematical ba-

[15] The sinusoidal low-frequency oscillator driving the modulator must have a rate of update that is the same as the audio sample rate, that is, the two sample rates must be identical. Otherwise, aliasing artifacts will be introduced into the audio signal path.

[16] Much like buying a Compact Disc because one likes the title track.

sis. We just wanted to prove it could be done.

He never related this work to his more profound mathematical and perceptual research, specifically the work on the required 3-eigentone/Hz density and the frequency-phase statistics in a random physical space.

The original EMT reverberator, model 250, operating at a 32-kHz sample rate, used a main memory of 8K words, and the required eigentone density was emulated entirely by randomizing delay lines. Another interesting fact is that colorless reverberation, using all-pass structures, is perceptually not colorless. Even white noise passed through an all pass will not sound like real white noise. When passed through many such all-pass structures, it in fact sounds like a machine shop rather than random noise. It still measures spectrally flat. The reason is that frequency regions get bunched in time. It is very much like a chirped sine wave in radar having a purely flat spectrum but being very different from white noise. The second- and higher order statistical terms out of an all pass are very, very different from a real random process. The utility of an all pass is to pass all frequencies through so that each all pass can see the same spectral density, otherwise comb peaks would align and dominate. Parallel structures of non-all-pass elements achieve a similar issue in that each structure gets fed the full spectrum. All-pass elements are more critical for small delay values. An all pass within a larger loop must be used with great care since it has a sinelike variation in group delay. Hence the effective loop time and reverberation time vary with frequency. After many trips around the loop, the result will be very colored.

Schroeder's had several analyses about reverberation, but his 3-eigentone/Hz theory, which maps to 3 seconds of memory, can be looked at in many ways. His result was empirical, based on listening tests. Consider two eigentones, or poles, separated by 1 Hz and located in the s plane with a real part of -10 Hz. When excited, this will produce two damped exponentially decaying frequencies which differ by 1 Hz. Hence there will be a 1-Hz envelope beat, which is clearly audible. Now add other eigentones, randomly spaced but still at a distance of -10 Hz. Assume 10 such eigentones. All of them will beat with each other, producing a random envelope with a spectrum that is crudely flat from 0 to 10 Hz. One can do this simulation in closed form with variable excitation of each eigentone. Schroeder's result actually depends on the nominal reverberation time since that determines how many eigentones will get excited by a narrow-band input. In the early reverberation boxes, with only 150 ms of reverberation, typically only a few tones would be excited. The envelope had a clear periodicity of 6 Hz on average. It sounds bad. Some regions had only two eigentones excited with a distance of 2 Hz, which was even worse. Development was much more exciting with such limited memory. Today one can use 1 second of DRAM memory. Many simpler structures will thus produce good reverberation.

The perceptual simulations deviated from physical reality in many ways. For example, a natural three-dimensional space has an increasing eigentone density that is proportional to the square of frequency. All electronic simulations tend to have a constant density. The reason is that in a three-dimensional space, the speed of sound along a dimension is proportional to the sine of the wave front direction, whereas in an electronic structure it is always constant.

That is what I remember, so do what you wish with it. Best of luck.

Sincerely yours,

BARRY BLESSER
Blesser Associates
Electronics & Software Consultants
Belmont, MA 02178, USA

2 MUSICAL FILTERING

The first-order recursive filter is by far the safest and most economical choice. Low in noise, it should be used wherever possible, and in cascade if necessary. For the design of shelving filters, which are conventionally first order, refer to [13]. When a filter having a steeper transition band is desired, it is usually sufficient to employ a second-order filter. Musical filters do not often see orders higher than that.[17]

In this section we discuss filtering requirements for musicians whose criteria are quite different from those of the electronics engineer. Our treatment of filtering will consider *only* the second-order case and predominantly all-pass topologies. The applications of these filters are broad; we note a particular suitability to parametric equalization. The more involved topic of truncation noise recirculation is not discussed in this section, although we do discuss limit cycles and internal signal overflow. The more curious reader is referred to [12] to find remedies for truncation noise within the direct form I topology.

For those readers new to digital filtering, the eminent theoretician, practitioner, and mentor of DSP and electronic music, Julius O. Smith, presents a splendid introduction to classical digital filter theory in [15, ch. 2], requiring only basic mathematical skills. Strawn's audio signal processing book [15] is written from the musician's standpoint, hence it is highly recommended.

2.1 Filter (Q) Selectivity

Electronics engineers are accustomed to think of digital filters analytically in terms of pole–zero constellation and locus, cutoff frequency, passband ripple, transition band or slope, stopband attenuation, and so on. Musicians and recording engineers are more comfortable thinking in terms of filter parameters—gain or cut, center frequency, filter Q (selectivity) or bandwidth. Formally, filter Q is defined as the positive quantity

$$Q = \frac{\omega_c}{\Delta\omega} = \frac{\omega_c}{\omega_2 - \omega_1} \tag{1}$$

[17] When a filter that is steeper than second order is required, it is advisable to construct it as a cascade of second-order sections. That will mitigate any coefficient sensitivity or truncation noise problems [11, ch. 11.4–11.6], [14].

that is, the center frequency divided by the bandwidth. The bandwidth is determined from the particular definition of the cutoff frequencies ω_1 and ω_2 (in radians). Traditionally the cutoff frequency coincides with an absolute half-power level. In the archetypal case of a steep unity-gain (0-dB) low-pass filter we recall this level as corresponding to the frequency at which the magnitude-squared response reaches -3.01 dB $[=10 \log_{10}(1/2)]$. But shallow audio filters may not have an absolute half-power level. So we must refine the definition of cutoff frequency in terms of half-power *excursion*, *not* an absolute level [16].

Take, for example, the cut filter magnitude-squared response shown in Fig. 2. This example response has a Q of 2. We define the two musical cutoff frequencies as corresponding to the level at which

$$\frac{1 - |H_c(e^{j\omega})|^2}{1 - |H_c(e^{j\omega_c})|^2} = \frac{1}{2}. \tag{2}$$

We must solve this equation for ω. There are two solutions, ω_1 and ω_2. Referring to Fig. 2, this equation instructs us to measure the bandwidth halfway down the trough of the magnitude-*squared* response. This makes intuitive sense. We cannot use the traditional definition of cutoff frequency for this example because the trough is not deep enough. But note that if $|H_c(e^{j\omega_c})|^2 = 0$ (notch filter), then the solution to Eq. (2) would correspond to the traditional definition of cutoff frequency.

The situation is pretty much the same for the resonator. Whereas the cut filter approaches unity asymptotically at $z = \pm 1$, the resonator is loosely defined as a second-order peaked filter having a peak gain normalized to unity at its center frequency. The resonator is easily formulated such that its magnitude-squared response is an exact flip of the corresponding cut filter about the horizontal half-power excursion line, that is, it is symmetrical with the cut filter. We shall shortly see how. This is the reason why many of the numbers are exactly the same in both Figs. 3 and 2.

For the resonator (rather, the normalized boost filter) we acquire the two musical cutoff frequencies ω_1 and ω_2, solving the slightly different equation

$$\frac{1 - |H_{b_{norm}}(e^{j\omega})|^2}{1 - |H_{b_{norm}}(\pm 1)|^2} = \frac{1}{2}. \tag{3}$$

Like before, the bandwidth is measured halfway *up* the peak of the magnitude-squared response in Fig. 3. Again we note that if $|H_{b_{norm}}(\pm 1)|^2 = 0$ (perfect resonator), then the solution to Eq. (3) corresponds to the traditional definition of cutoff frequency.

Having gained an understanding of musical filter Q, we begin with two unique and musically useful digital filter transfer functions, which precisely fit our definition of filter selectivity.

2.2 Cut Filter

When constructing a notch filter, we expect there to be an absolute zero of transmission at some selected frequency in its transfer function. If we use a filter that only has zeros (that is, no poles), we can indeed make a notch. The problem with this approach is that the rest of the magnitude response would not be very flat, as we might like it to be. We might also like a "surgical" notch, one that has high selectivity. Fig. 4 is an example showing the magnitude transfer of a badly designed notch filter evaluated along the unit circle in the z plane. The zero radius is $R = 1$, whereas the zero angle is $\theta = 1$ rad. This transfer function has two trivial poles at $z = 0$.

The magnitude response shown in Fig. 4 would pretty much obliterate a musical signal, especially because of the gain at high frequencies. Note that if the zero were

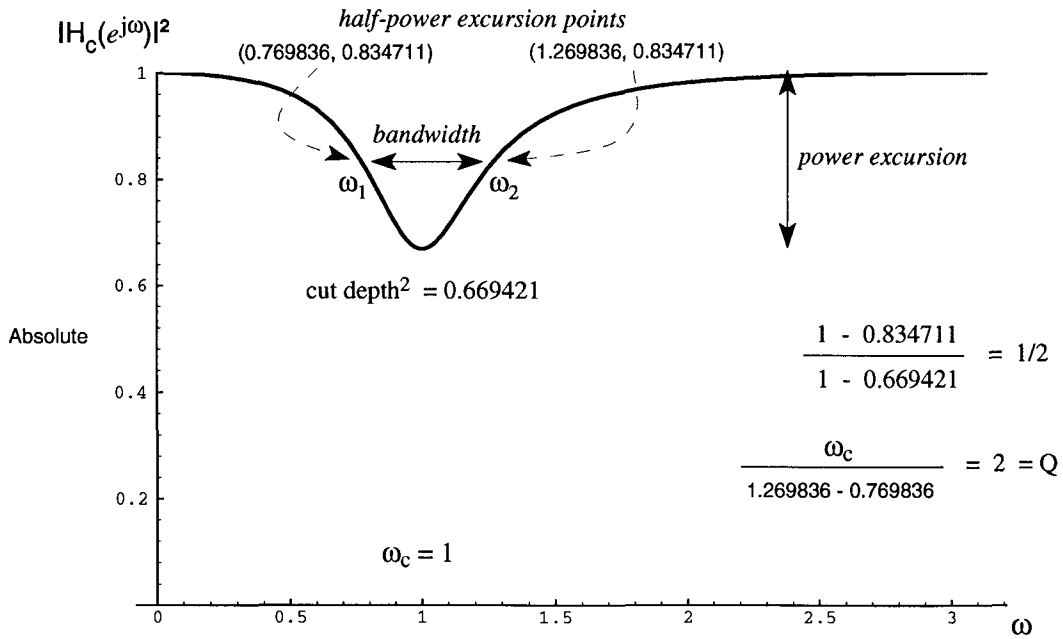

Fig. 2. Cut filter excursion ≈ 1.7 dB.

moved to a new fixed frequency, the rest of the magnitude response would change its shape in an undesirable way. Hence this particular notch filter is not very useful for surgical filtering.

In [13] it was shown how to make the passband portions of the notch filter flat, and how to achieve high selectivity. This is accomplished by adding nontrivial poles. The result is illustrated in Fig. 5 and expressed in the transfer function

$$H_n(z) = \frac{1}{2}(1 + \beta)\frac{1 + 2\gamma z^{-1} + z^{-2}}{1 + \gamma(1 + \beta)z^{-1} + \beta z^{-2}}. \quad (4)$$

This notch filter [Eq. (4)] has an absolute zero having controllable selectivity there at its center frequency, while its magnitude at dc and Nyquist is *always* 1, regardless of the center frequency. We must determine how to obtain a trough of arbitrary depth while maintaining the other attributes. This would be called a *parametric* cut filter. Before we do that, however, we look at the resonator, which is an exact powerwise flip of this notch filter about a horizontal.

2.3 Resonator

One use of a perfect resonator in electronic music is to synthesize ping sounds via impulsive excitation. We discuss a more general resonator for use as a musical filter. It is easy to construct a simple resonator using only poles. But such an approach has problems similar to those we encountered with the all-zero notch filter, especially with regard to shape, selectivity, and magnitude. In particular, the peak magnitude will vary as the center frequency is changed to new fixed values.

In [11, ch. 4.3] it was shown how to normalize the height of the resonator peak magnitude as the center frequency changes, namely, by adding two zeros, one at Nyquist and the other at dc. This musically useful result is illustrated in Fig. 6 and expressed by the equation

$$H_r(z) = \frac{1}{2}(1 - \beta)\frac{1 - z^{-2}}{1 + \gamma(1 + \beta)z^{-1} + \beta z^{-2}}. \quad (5)$$

This filter has a peak gain that is *always* precisely 1, regardless of the center frequency. This is characteristic of a resonator. The two zeros make the skirts of the

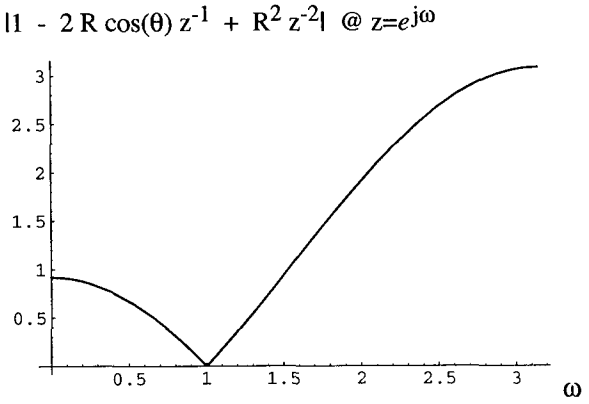

Fig. 4. Poor notch filter; $R = 1$, $\theta = 1$.

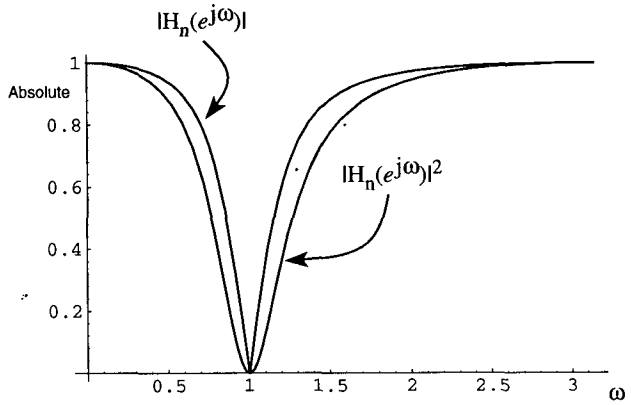

Fig. 5. Notch filter; $Q = 2$, $\omega_c = 1$.

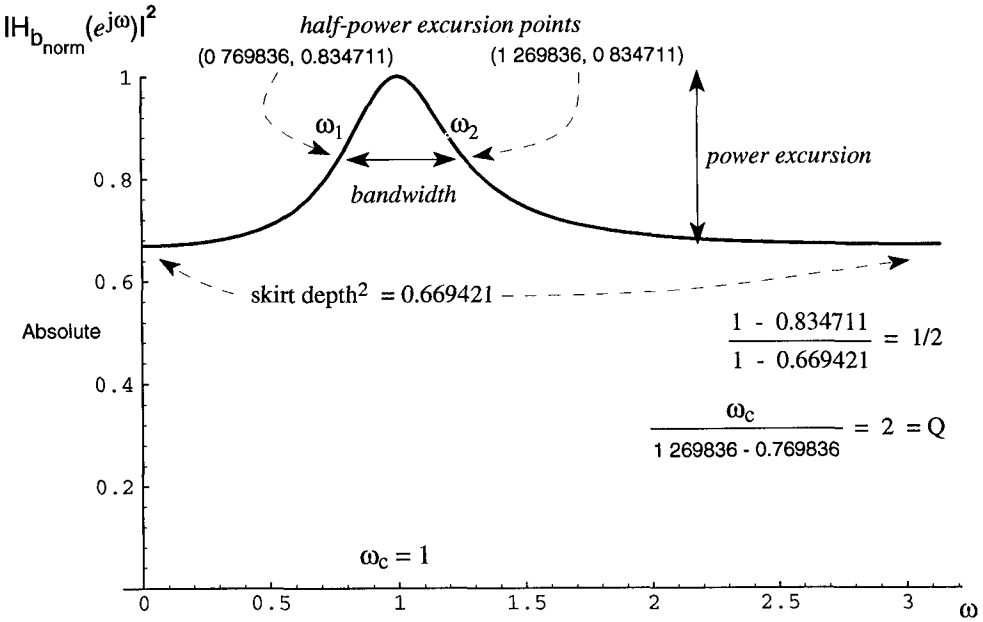

Fig. 3. Resonator excursion ≈ 1.7 dB.

magnitude-squared response go to zero at the extremities. When the extremities reach zero, we call this the *perfect* resonator [Eq. (5)]. We must determine how to make skirts of arbitrary depth—the resonator. We must also determine how to place the skirts at absolute magnitude 1 while achieving arbitrary peak heights; that would be called a parametric boost filter. We have yet to define γ and β.

2.4 Musical Filter Topology

The two transfer functions $H_n(z)$ and $H_r(z)$ have some desirable theoretical and practical properties. First, there is a strong bond between Eqs. (4) and (5). Because their denominators are identical, there exists one circuit that can generate both. Second, there is a simple relationship between the coefficients β and γ and the musical filter parameters ω_c and Q.

Consider the all-pass lattice topology shown in Fig. 7. It has the all-pass transfer function

$$A(z) = \frac{Y(z)}{X(z)} = \frac{\beta + \gamma(1 + \beta)z^{-1} + z^{-2}}{1 + \gamma(1 + \beta)z^{-1} + \beta z^{-2}}. \quad (6)$$

Some characteristics of the all pass filter are summarized in Fig. 8 and the equations

$$|A(e^{j\omega})| = 1, \quad A(\pm 1) = 1, \quad A(e^{j\omega_c}) = -1. \quad (7)$$

It is interesting that the non-minimum-phase all-pass filter will shortly become integral to a parametric filter that is indeed a minimum-phase design. We also note in passing that the transfer to $D_r(z)$ from the input comprises only the denominator (the poles) of $A(z)$,

$$D_r(z) = \frac{X(z)}{1 + \gamma(1 + \beta)z^{-1} + \beta z^{-2}}.$$

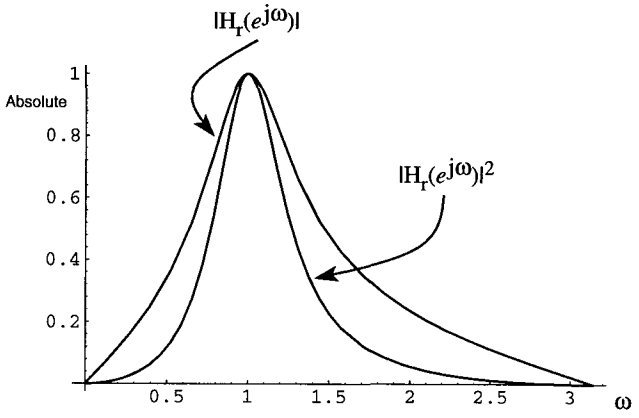

Fig. 6. Perfect resonator; $Q = 2$, $\omega_c = 1$.

Back to the problem at hand, it is easily proven that

$$H_r(z) = \frac{1 - A(z)}{2} \quad (8)$$

$$H_n(z) = \frac{1 + A(z)}{2}. \quad (9)$$

Substituting Eq. (6) into Eqs. (8) and (9), we can derive Eqs. (5) and (4). This means that we can construct notch and perfect resonant filters from an all-pass filter. We only have left to show that using the all-pass filter we can construct cut, resonant, and boost filters as well. We will use the fact [Eq. (7)] that at the critical frequency

$$\omega_c = \arccos(-\gamma) \quad (10)$$

the all-pass filter output is 180° out of phase with respect to a steady-state sinusoid at its input. This critical frequency becomes the normalized center frequency $\omega_c = 2\pi f_c T$ (with T being the sample period) for all filter types employing the topology shown in Fig. 9.

In Fig. 9 we have introduced a new control coefficient k. When $k = 0$, the network in Fig. 9 implements the notch [Eq. (9)] exactly, and when $k = 2$, this same network implements the perfect resonator [Eq. (8)] exactly. Within these bounds, this control (for $k < 1$) gives us the ability to specify the depth of the cut, leaving the magnitude at the extremal frequencies equal to 1. Using the same network for the resonator, we can control the depth of the skirts (when $k > 1$), leaving the absolute

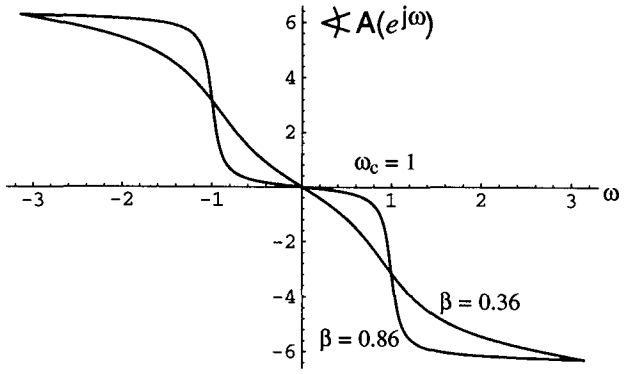

Fig. 8. All-pass radian phase responses.

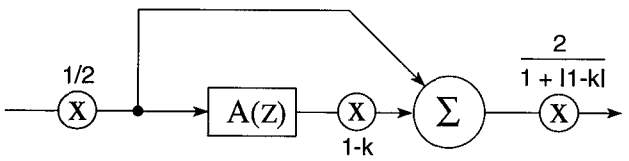

Fig. 9. Cut, notch, or resonator type filter.

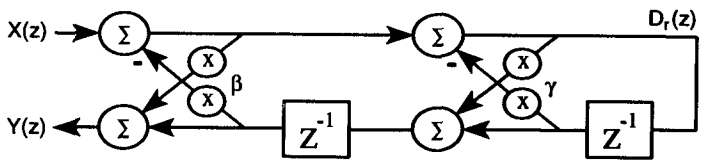

Fig. 7. Lattice second-order all-pass filter.

peak frequency magnitude at precisely 1. These actions explain the unusual looking normalization coefficient at the output.

The action of the control coefficient k is characterized in Table 3 and illustrated in the magnitude-squared responses of Figs. 10 and 11. The cut depth [Eq. (11)] and the skirt depth [Eq. (12)] must be squared to resolve with Figs. 10 and 11, respectively,

$$\text{absolute cut depth} = \frac{1-(1-k)}{1+|1-k|} = \frac{k}{2-k} \quad ; 0 \leq k < 1 \tag{11}$$

$$\text{resonator absolute skirt depth} = \frac{1+(1-k)}{1+|1-k|} = \frac{2-k}{k} \quad ; 1 < k \leq 2 \tag{12}$$

Table 3. Control coefficient k for Fig. 9.

$k = 0$		Notch, $H_n(z)$
$0 < k < 1$		Cut, $H_c(z)$
$k = 1$	Yields	Input signal
$1 < k < 2$		Resonator, $H_{b_{norm}}(z)$
$k = 2$		Perfect resonator, $H_r(z)$

These depth equations are easily deduced from Eqs. (10) and (7) and are independent of the center frequency.

The center frequency is unequivocally determined by Eq. (10) for these cut, notch, resonator, and perfect resonator filters shown in Figs. 10 and 11. This center frequency corresponds to the peak or trough extremum of the magnitude transfer evaluated on the unit circle in the z plane.

The ordinate axis is drawn at the lower half-power excursion frequency ω_l for the plots in Figs. 10 and 11. The half-power excursion frequencies (the two musical cutoff frequencies) are given for the cut, notch, resona-

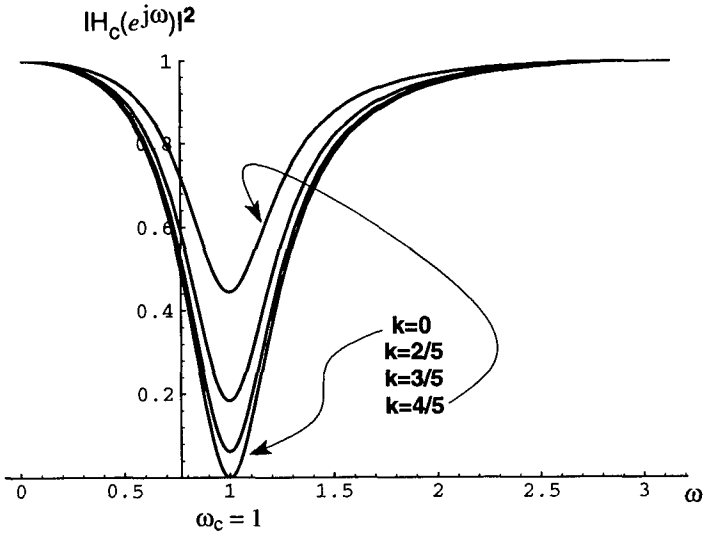

Fig. 10. Cut filter magnitude-squared responses for various values of k; $Q = 2$.

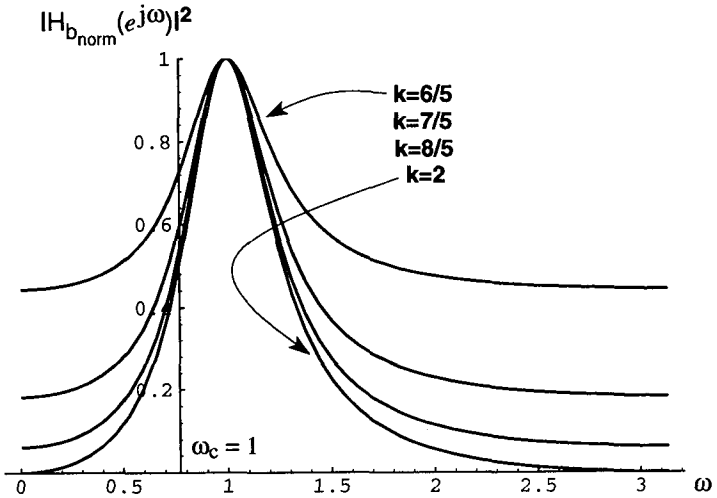

Fig. 11. Resonator magnitude-squared responses for various values of k; $Q = 2$.

tor, and perfect resonators by the equations

$$\cos(\omega_{2,1}) = \frac{(1+\beta)^2\cos(\omega_c) +,- (\beta-1)\sqrt{2(1+\beta^2) - (1+\beta)^2\cos^2(\omega_c)}}{2(1+\beta^2)} \quad (13)$$

$$\beta = \frac{1-\tan(\omega_c/(2Q))}{1+\tan(\omega_c/(2Q))} = \frac{1-\tan(\Delta\omega/2)}{1+\tan(\Delta\omega/2)}. \quad (14)$$

Given a particular center frequency, the all-pass lattice coefficient β [Eq. (14)] precisely controls *selectivity* (the filter Q [Eq. (1)]) for these cut, notch, resonator, and perfect resonator filters.[18] Whereas the lattice coefficient γ is a function only of ω_c as we see from an inspection of Eq. (10), here we see that β is a function of both ω_c and Q as per our new definition of musical cutoff frequency, Eqs. (3) and (2).

On the one hand, it is very good that we have discovered closed-form mathematical relationships describing how to modify the two lattice coefficients to control the musical filter parameters. But from a control standpoint, we would like to have a way to decouple the filter coefficients so that only one of them governs the center frequency whereas the other governs only the selectivity parameter. (We almost have that in γ.) Later on we will see the Chamberlin filter topology, which nearly reaches that ideal.

2.4.1 Regalia k Coefficients

In [13] it was understood that a simple algebraic change in variable would result in a new design which substitutes the parametric boost filter for the resonator, hence incurring the loss of the resonator and the perfect resonator. Employing the same topology as before, the coefficients in Fig. 12(a) are derived from those in Fig. 9 via the substitution

$$1 - k \rightarrow \frac{1-k}{1+k} \quad (15)$$

[18] This equation for β is exact in terms of the selectivity definition, Eq. (1).

and via a scaling by the boost factor k on the output, but only when $k > 1$. The transfer function of the circuit in Fig. 12(b) is identical to that in Fig. 12(a). By pushing the output coefficient forward, we simplify the other coefficient.

The action of the control coefficient k is now characterized in Table 4 and illustrated in the magnitude-squared responses of Fig. 13. The cut depth [Eq. (16)] and the boost [Eq. (17)] must each be squared to resolve with Fig. 13,

$$\text{Regalia absolute cut depth} = k \quad ; 0 \leq k < 1 \quad (16)$$

$$\text{Regalia absolute boost} = k \quad ; 1 < k < \infty. \quad (17)$$

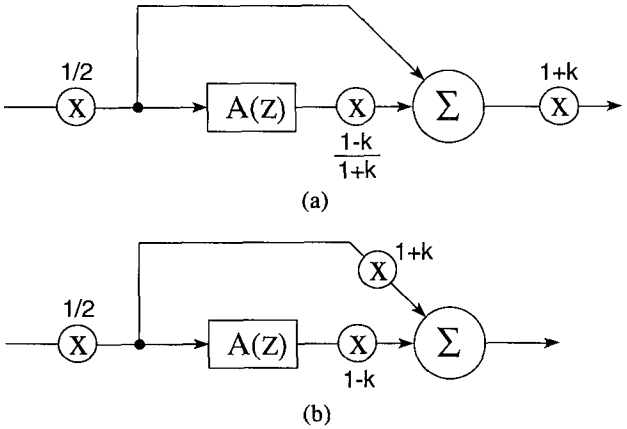

Fig. 12. Cut, notch, or boost type filter. Transfer functions in (a) and (b) are identical.

Table 4. Control coefficient k for Figs. 12 and 14.

$k = 0$	Yields	Notch, $H_n(z)$
$0 < k < 1$		Cut, $H_c(z)$
$k = 1$		Input signal
$1 < k < \infty$		Boost, $H_b(z)$

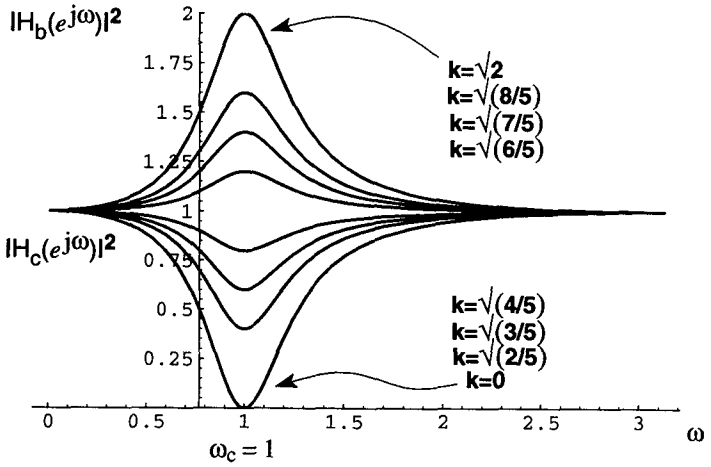

Fig. 13. Cut and boost responses for various values of k; $Q = 2$.

These results can be derived by subtituting Eq. (15) into Eqs. (11) and (12). As before, these results are independent of the center frequency. The combined plot in Fig. 13 highlights the symmetry of the cut with the boost filters, hence, the symmetry of Q.

In general, the parametric filters of Figs. 9 and 12 are minimum phase. From the all-pass characteristics [Eq. (7)] it can be deduced that, independent of center frequency,

Regalia absolute skirt depth of boost filter = 1

$$; 1 < k < \infty . \qquad (18)$$

The ordinate axis is again drawn at the lower half-power excursion frequency (the lower musical cutoff frequency ω_1) in Fig. 13. The two musical cutoff frequencies $\omega_{2,1}$ for the boost filter are derived using a slightly different Eq. (19) as compared to that for the resonator [Eq. (3)]. But it can be shown that the results are the same as before, that is, Eq. (13) remains valid under

$$\frac{|H_b(e^{j\omega})|^2 - 1}{|H_b(e^{j\omega_c})|^2 - 1} = \frac{1}{2} . \qquad (19)$$

Unlike Eq. (3), Eq. (19) does *not* reduce to the traditional definition of cutoff frequency because $H_b(e^{j\omega_c})$ is, by definition, never zero. Because we were able to derive the Regalia k coefficients in Fig. 12 from the network in Fig. 9 while retaining the same topology, all the equations thus far remain applicable, that is, for β, γ, ω_c, and $\omega_{2,1}$.

2.4.2 Regalia–Mitra Topology

Fig. 14 shows the parametric filter topology originally presented[19] in [13].[20] Although our development led us to Fig. 12, the transfer function of the circuit in Fig. 14 is identical.

[19] We moved to the input the originally internal scaling by ½ in order to avoid subsequent overflow in a fixed-point implementation.

[20] Regalia and Mitra [13] also formulate the construction of first-order shelving filters using the same topology. In the audio industry, shelving filters are typically first-order designs. They are used to uniformly boost or cut selected portions of the high- or low-frequency region. They are called "shelves" because their magnitude responses approach unity asymptotically.

2.4.3 Lattice Topology in Practice

The foregoing filter topologies constructed from the all-pass lattice suffer two drawbacks to their implementation: 1) the lattice produces spontaneous low-level audible zero-input limit-cycle tones, and 2) the lattice topology is prone to signal overflow at internal nodes before the all-pass output has reached full scale.[21]

The first problem is solved by magnitude truncation of[22] all lattice memory elements [9]. The second problem is solved by scaling the lattice input, as already shown in Figs. 9, 12, and 14. When premature internal overflow persists (which is more likely for high Q, cut *or* boost), it becomes necessary to provide a user-controlled input-signal-level adjustment.[23] Compensation will be required at the filter output, under separate user control. Keep in mind that the cost of output amplification is the concomitant amplification of the filter's internal signal-truncation noise floor. So this input scaling process should be limited.

As an alternative to the use of the all-pass lattice, we recall that the direct form I filter topology does not suffer from internal signal overflow. That is because its single accumulator has *infinite headroom* when used in a fixed-point implementation [12, pp. 857, 875][24], [11, ch. 11.3], [14, ch. 6.7.1]. In Fig. 15 we show an implementation of the second-order all-pass filter [Eq. (6)] comprising embedded direct form I first-order all-pass sections.[25] This topology retains the dichotomy of the musical filter coefficients as in the lattice, while employing the exact same coefficients, and avoids internal overflow.

Empirically we observe that the limit-cycle tones produced by the direct form I are much quieter than those produced by the corresponding lattice in Fig. 7, in general.[26] Truncation error feedback (not shown) in the di-

[21] Conditional saturation is helpful but does not solve the problem because internal clipping sounds bad. In [11, ch. 4.3] a novel topology for the perfect resonator is shown.

[22] See Part 3, Section 9.2, Appendix 8.

[23] This knob is probably required anyway to counterbalance boosts at the filter output.

[24] Overflow is not always a bad thing.

[25] Conversion to direct form II using Rossum's technique [17] would eliminate some memory elements while providing automatic input scaling. The scaling is necessary to prevent internal overflow in that topology.

[26] The direct form I may require error feedback [12] to be truncation-noise competitive with the lattice, however.

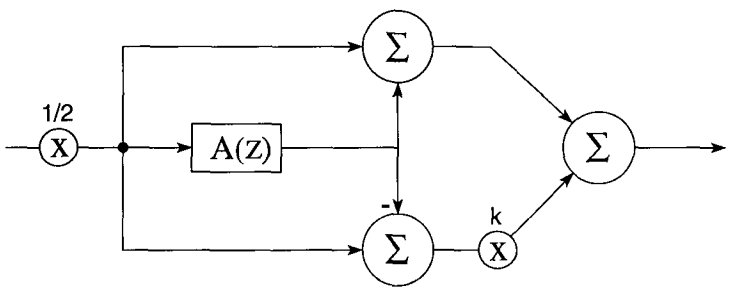

Fig. 14. Regalia–Mitra topology.

rect form is known to further minimize limit-cycle oscillation [18], thus providing an alternative to magnitude truncation as a remedy. For both the lattice and the embedded direct form, stability is assured by $|\gamma| < 1$ and $|\beta| < 1$.

2.5 Appendix 2: Filter Errata in the Literature

A mistake has been perpetuated regarding the center frequency of the second-order digital filter. The polar representation of complex conjugate filter poles is often found, correctly written, as

$$z_{pole} = Re^{\pm j\theta} . \tag{20}$$

The erroneous hypothesis can be recognized wherever the filter's normalized center radian frequency ω_c is ascribed to the radian pole angle θ. Hence, the distinction between *center* frequency and *pole* frequency is obscured in the literature.[27] There it is argued that for high selectivity, this distinction is of little practical importance, but that tenuous assumption of practical equivalence has, consequently, promulgated specious theoretical conclusions within the audio community.

One such erroneous conclusion is that the perfect resonator transfer function [Eq. (5)], for arbitrary center frequency, does *not* have a peak magnitude exactly equal to 1 when evaluated on the unit circle in the z plane. The errant proof evaluates Eq. (5) at the resonant frequency (at $z = e^{j\theta}$), in complete disregard of the pole radius R. Evaluation at the true center frequency (at $z = e^{j\omega_c}$) given by Eq. (10) shows that conclusion to be false, that is, it *is* true that the perfect resonator as given by Eq. (5) always has a peak gain of exactly unity.

We can establish a correspondence between pole and center frequencies by equating the general denominator of a second-order transfer function, written in terms of the pole radius and angle [Eq. (20)] [11, ch. 4.3], [12, Eq. (27)], to the perfect resonator [Eq. (5)],[28]

$$H_r(z) = \frac{\frac{1}{2}(1-\beta)(1-z^{-2})}{1+\gamma(1+\beta)z^{-1}+\beta z^{-2}} = \frac{\frac{1}{2}(1-\beta)(1-z^{-2})}{1-2R\cos(\theta)z^{-1}+R^2z^{-2}} . \tag{5}$$

Using Eq. (10), we can easily deduce the following identifications;

$$\beta = R^2$$

$$\gamma = \frac{-2R\cos(\theta)}{1+R^2} = -\cos(\omega_c) .$$

This proves that the only instance where the center frequency ω_c would be the same as the second-order pole (or resonant) frequency θ is for conjugate poles right on the unit circle ($R = 1$). But in that circumstance one has an oscillator, not a filter.

These results can be extended to the resonator in general. Similar conclusions can be drawn from an examination of the second-order all-pole transfer function [see Eq. (23)], and from the second-order all-zero transfer function, such as the one in Fig. 4.

An instance where the center frequency *is* identical to the pole frequency is for the case of the first-order resonator. The equivalence is independent of pole radius R, unlike the second-order case. This instance may be the reason for the propagation of the erratum regarding the second-order case. The transfer function of the first-order resonator is

$$F_r(z) = \frac{1-R}{1-Re^{j\theta}z^{-1}} .$$

This filter has only one pole. But notice that the one filter

[27] The *resonant* frequency is that frequency at which a filter rings when excited by an impulse. The resonant frequency is the pole frequency, which is the same as the pole angle θ in the z plane. The center frequency is the frequency at peak magnitude response in the steady state, when a filter is excited by a sinusoid of infinite duration. In general, center and resonant frequencies are not identical [19, ch. 5.5].

[28] The poles occur in complex conjugate pairs when the filter coefficients of z (that is, γ and β) are real. When the filter coefficients are real, then it is easily shown that the filter's impulse response must also be real.

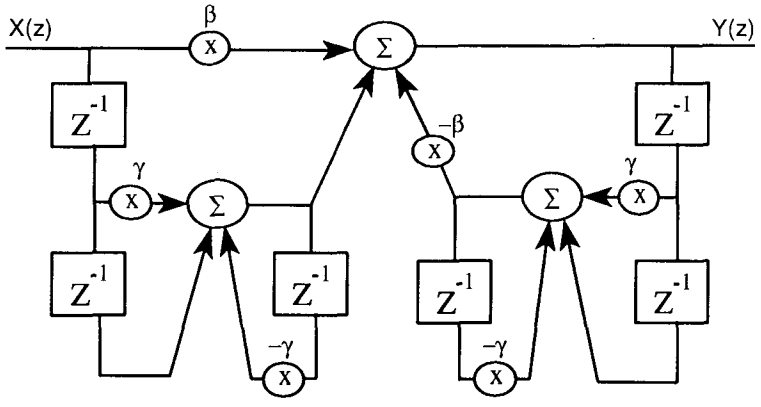

Fig. 15. Embedded direct form I, second-order all-pass filter.

coefficient is complex, in general. Hence the impulse response of this filter cannot be real. One may surmise that there must be some interaction among multiple poles in the z plane, which destroys radial symmetry.

3 CHAMBERLIN FILTER TOPOLOGY

Next we consider high-fidelity musical filtering using a different topology and the musician's all-pole low-pass filter type. We apply our earlier redefinition of cutoff frequency, in terms of half-power excursion, to this new construction, which establishes a tie to our previous work.

The musician's all-pole filter has antecedents in the electronic music industry,[29] appearing in currently renowned and vintage music synthesizers [22]. The filters we considered previously had zeros in the transfer function. We were concerned about the control of those filters as a musician might like to control them. Here we present an additional goal; namely, to come up with filter coefficients where each will control individually only center frequency or selectivity (filter Q). To do so, we rederive the Chamberlin [23] all-pole (two-pole) low-pass filter topology entirely from the perspective of the discrete-time domain.[30]

We argue that the truncation noise performance of the Chamberlin filter topology is very good, although practitioners have known that for years.[31] In so doing we introduce a new more musical and conservative measure of noise performance that we call "transparency," and which we denote criterion 1. Using a more traditional approach, denoted criterion 2, we compare the truncation noise power observed at the Chamberlin filter output to the input-signal quantization noise power, which can be construed as the *noise gain*. We discover that for the Chamberlin topology, the worst noise gain is the same as the peak gain squared of the whole filter acting upon the input signal. That turns out to be the reason why the noise performance is so good.

3.1 Shape of the Musician's Low-Pass Filter

The electronics engineer's low pass has zeros in the stopband and is very flat in the passband.[32] The stopband zeros serve to provide high attenuation there. In contrast, musicians have a taste for peaked filters, even when the desired filter is of the low-pass variety. Because the musician's peak-center frequency is typically quite low (requiring poles closer to the unit circle), zeros are largely unnecessary due to the relatively high attenuation at frequencies far away from the low-frequency poles. When the peak-center frequency is high, on the other hand, the stopband excursion of the all-pole filter magnitude response may not span 3 dB. In fact, when the peak-center frequency reaches $\pi/2$, the all-pole low-pass filter ceases being low pass because the magnitude response at π starts to exceed the response at dc.

Due to the fact that the Chamberlin filter is all pole, there is little control over the rate of transition from passband to stopband. To increase the transition rate of the low-pass filter, the accepted solution is to cascade an identical all-pole filter. This works in practice because the musician's working range of the low-pass peak-center frequency is much less than $\pi/2$ for reasonable sample rates. Zeros placed at the Nyquist frequency, for example, would have little impact over the musician's working range. Therefore the cascade is preferred to zeros at Nyquist. Zeros elsewhere in the stopband region would entail more computation, hence they are undesirable. In this development, we will consider only a single filter section.

We expect some kind of boosting response, as shown in Fig. 16. The corresponding transfer function must be at least second-order to get the peak center away from dc. Notice that the filter is normalized to unity at dc.[33] Once again, we must refine our notion of cutoff frequency by relating it to half-power excursion, as before. For this filter type, we define the passband excursion from the value of the magnitude-squared response at dc to the peak value of the response. Reminding ourselves that this magnitude-squared response is periodic in 2π, we then similarly define the stopband excursion from the peak to the value at Nyquist.[34] In Fig. 16 the half-power excursion points are indicated, defining the musician's bandwidth of the all-pole low-pass filter.

We find the frequencies of the half-power excursion points (the musical cutoff frequencies) here much like we did before: the passband half-power excursion frequency is found solving Eq. (21) for ω,

$$\frac{|H_{\text{chx}}(e^{j\omega})|^2 - 1}{|H_{\text{chx}}(e^{j\omega_c})|^2 - 1} = \frac{1}{2}. \tag{21}$$

We call this frequency ω_1. Similarly, we call ω_2, the solution to Eq. (22) for the half-power excursion in

[29] The classic Moog analog synthesizers, for example, employed fourth-order all-pole voltage-controlled filters (VCFs). His constant-Q design was also known as the Moog ladder, after the appearance of the schematic [20]. A cascade of two Chamberlin filters can be considered as the digital counterpart to the Moog VCF because some of the same characteristics are shared. They are both all-pole constant-Q designs tuned by a single sweepable parameter. Rossum [21] of Emu considers essential nonlinear ingredients to make digital filters sound more "analog."

[30] This filter was originally derived from an analog state-variable filter by application of the impulse-invariant transformation.

[31] The Chamberlin all-pole design is a reputed resident within the contemporary digital synthesizers by Peavey and Kurzweil.

[32] The Butterworth filter, for example (which is a good choice for audio with regard to minimal ringing), has all its zeros at Nyquist.

[33] To bring the boost at the peak-center frequency ω_c down to unity, additional scaling is required beyond what we recommend here.

[34] The electronics engineer's transition band and stopband are merged in this development. Because of the lack of zeros here, the electronics engineer's boundaries are not as clear. Also, the electronics engineer would measure bandwidth from dc, unlike our measurement.

the stopband,

$$\frac{|H_{chx}(e^{j\omega})|^2 - |H_{chx}(-1)|^2}{|H_{chx}(e^{j\omega_c})|^2 - |H_{chx}(-1)|^2} = \frac{1}{2}. \quad (22)$$

Neither of these two cutoff frequency definitions [Eqs. (21) and (22)] reduce to the traditional definition because none of the terms can go to zero in this all-pole design.

3.2 Transfer Function Development

We begin with a simpler second-order transfer function having no zeros, so we can expect some of the previously discovered equations to be different,

$$H_{chx}(z) = \frac{\alpha}{1 + \lambda z^{-1} + \beta z^{-2}}. \quad (23)$$

We seek the relationship of the ideal coefficients λ and β to the peak-center frequency ω_c and selectivity Q,

$$\omega_c = \arccos\left[\frac{-(1+\beta)\lambda}{4\beta}\right]. \quad (24)$$

If we express λ as

$$\lambda = \frac{4\beta\gamma}{1+\beta}$$

then we find the simpler expression for peak-center frequency,

$$\omega_c = \arccos(-\gamma). \quad (10)$$

This equation for the center frequency is the same as before, and both Eqs. (10) and (24) are exact.

At the peak-center frequency the magnitude-squared response reaches its peak height. Exactly,

$$\max_{\omega}[|H_{chx}(e^{j\omega})|^2] = \frac{4\alpha^2\beta}{(\beta-1)^2(4\beta-\lambda^2)} = \frac{\alpha^2(1+\beta)^2}{(\beta-1)^2[1+\beta^2-2\beta\cos(2\omega_c)]}. \quad (25)$$

For the low-pass filter we normalize the transfer function to unity at dc, so α becomes

$$\alpha = 1 + \lambda + \beta.$$

The two musical cutoff frequencies were determined exactly, using *Mathematica* [24],[35] as

$$\cos(\omega_{2,1}) = \cos(\omega_c) +,- \frac{\cos^2,\sin^2(\omega_c/2)(\beta-1)\sqrt{2[1+\beta^2-2\beta\cos(2\omega_c)]}}{\sqrt{1+\beta+8\beta^2+\beta^3+\beta^4 +,- 4\beta(1+\beta)^2\cos(\omega_c) - \beta(1-6\beta+\beta^2)\cos(2\omega_c)}} \quad (26)$$

We were not able to determine an exact expression for the β coefficient in terms of ω_c and Q as we did for γ, but the following guess turns out to be a good approximation:

$$\beta \approx \frac{1 - \sin(\omega_c/(2Q))}{1 + \sin(\omega_c/(2Q))}. \quad (27)$$

The plot in Fig. 17 shows that our expression for β is good over the recommended operating peak-center frequency range of $\omega_c = 0$ to $\pi/2$. To make this plot, we substitute the desired Q into the exact equations for ω_2 and ω_1 [Eq. (26)] using the approximation to β [Eq. (27)], and then we sweep over ω_c. If we had the exact expression for β, then the sheet would be a taut plane having unit slope with respect to the desired Q. But the

[35] The extensions of *Mathematica* [25] to analog and digital signal processing are highly recommended.

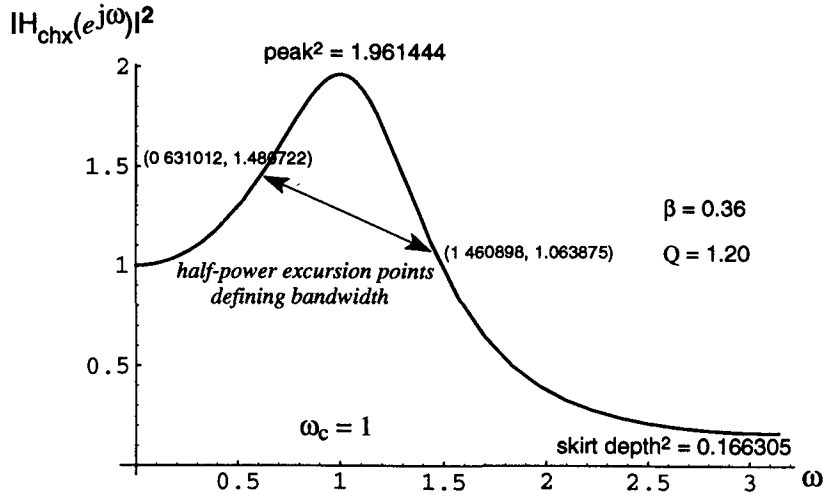

Fig. 16. All-pole low-pass magnitude-squared response.

approximation Eq. (27) is far better than some others in the literature [26]–[28], [15, ch. 2-III, p. 123]. The largest percentage errors in the recommended center frequency range are for a desired Q of 1, having error maxima of 21.8 to 27.6%.

3.2.1 Approximations

To achieve our stated goal of obtaining filter coefficients that control center frequency or filter selectivity individually, we now make series approximations to our expressions for the ideal filter coefficients we have found thus far. Using the good approximation Eq. (27), we find that the first several terms of the equivalent Maclaurin series for the z^{-1} coefficient in Eq. (23) are

$$\lambda \approx -\left(2 - \frac{1}{Q}\omega_c - \omega_c^2 + \left(\frac{1}{2Q} + \frac{1}{24Q^3}\right)\omega_c^3 + \frac{1}{12}\omega_c^4 - \left(\frac{1}{24Q} + \frac{1}{48Q^3} + \frac{1}{1920Q^5}\right)\omega_c^5 + \cdots\right)$$

and the first several terms of the Maclaurin series for the z^{-2} coefficient in Eq. (23) are

$$\beta \approx 1 - \frac{1}{Q}\omega_c + \frac{1}{2Q^2}\omega_c^2 - \frac{5}{24Q^3}\omega_c^3 + \frac{1}{12Q^4}\omega_c^4 - \frac{61}{1920Q^5}\omega_c^5 + \cdots.$$

These series are hard to predict. The *Mathematica* script used to generate them is

```
beta = (1 − Sin [wc/(2 Q)]) / (1 + Sin [wc/(2 Q)]) ;
Simplify [Series [Simplify [Factor [−4 beta Cos [wc]/(1 + beta)]], {wc, 0, 5}]]
Series [beta, {wc, 0, 5}]
```

Fig. 18 shows the musical filter topology[36] that implements a truncated series approximation to the ideal filter coefficients, hence decoupling somewhat the control of ω_c and Q. Fig. 18 incorporates the first three terms from the λ series and the first two terms from the β series. Thus the coefficient F_c is identified with ω_c while the coefficient Q_c is identified with $1/Q$. Because the circuit implements so few terms from the β and λ series which

[36] We adopt Chamberlin's nomenclaure [23]. Chamberlin points out that this filter topology simultaneously possesses a high-pass and a band-pass output at the nodes labeled hp and bp, respectively. We discuss only the low-pass filter function lp of this circuit in detail here.

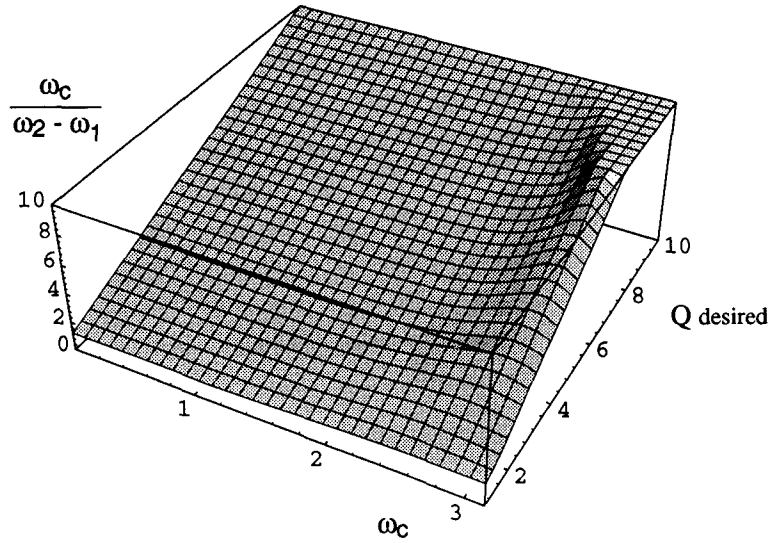

Fig. 17. Actual all-pole filter Q as a function of center frequency and desired Q.

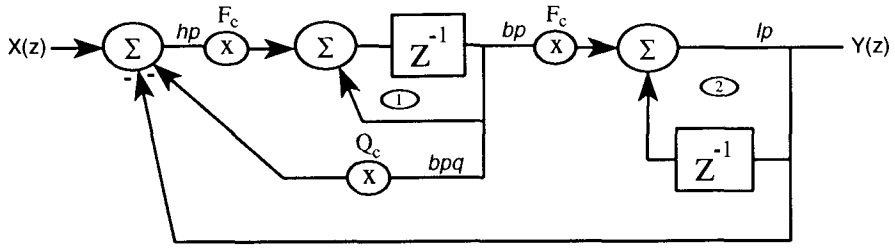

Fig. 18. Chamberlin topology, second-order all-pole filter. Input scaling by ½ and output compensation not shown (see Section 3.3.6).

are themselves approximations due to Eq. (27), these stated identities are crude. We refine the approximate relation of F_c to ω_c in Eq. (29), but we will leave the circuit in Fig. 18 as it is. After we characterize the circuit a little more, we will find that the filter coefficients in the figure provide sufficiently autonomous control for musical purposes.

The all-pole low-pass transfer function of this further approximation to Eq. (23) in Fig. 18 is

$$H_{ch}(z) = \frac{Y(z)}{X(z)} = \frac{F_c^2 z^{-1}}{1 - (2 - F_cQ_c - F_c^2)z^{-1} + (1 - F_cQ_c)z^{-2}} \approx z^{-1}H_{chx}(z) \tag{28}$$

redefining

$$\lambda \equiv -(2 - F_cQ_c - F_c^2), \qquad \beta \equiv 1 - F_cQ_c \tag{29}$$

where

$$F_c \approx 2\sin\left(\frac{\omega_c}{2}\right) \quad \text{rad}, \qquad Q_c \approx \frac{1}{Q}.$$

The transmogrified numerator of Eq. (23) now shows a delay operator in Eq. (28). This comes about because of the need to eliminate an otherwise delay-free loop in the circuit of Fig. 18.[37] The numerator coefficient α has become F_c^2 to force Eq. (28) to unity at dc ($z = 1$), as stipulated. This refined approximation to F_c in Eq. (29) is from [23], [26]; it becomes more exact for high Q. When the peak-center frequency ω_c is low and Q is high, F_c in Eq. (29) reduces to ω_c. [An exact expression for F_c is given by Eq. (31).]

3.2.2 Stability/Parameter Decoupling

The stability of complex conjugate poles demands the constraint

$$0 \leq (1 - F_cQ_c) < 1 .$$

Restated,

$$0 < F_c \leq \frac{1}{Q_c}. \tag{30}$$

This condition is ascertained from Eq. (28) by demanding a pole radius[38] of magnitude less than 1.

From previous considerations we presume that the tuning range for the all-pole low-pass filter is

$$\frac{\pi}{2} > \omega_c \geq 0 .$$

If we substitute Eq. (29) into the equation for the actual peak-center frequency ω_c [Eq. (24)] in this range, we discover in Eq. (31) that F_c and Q_c are not completely decoupled,[39] except for very high selectivity ($Q_c \approx 0$),

$$0 < \cos(\omega_c) = \frac{4(1 - F_cQ_c) - F_c^2(2 - Q_c^2) + F_c^3Q_c}{4(1 - F_cQ_c)} \leq 1 . \tag{31}$$

The identity in Eq. (31) is exact. Further, we find on the right-hand side inequality that

$$F_c \leq \frac{2}{Q_c} - Q_c$$

and on the left-hand side,

$$F_c < \frac{-Q_c + \sqrt{8 + Q_c^2}}{2}.$$

From the stability condition Eq. (30), the minimum value of F_c is zero. This is achieved for the right-hand side inequality of Eq. (31) when Q_c reaches $\sqrt{2}$. Thus we have an upper bound on Q_c to keep the actual peak-center frequency within the prescribed tuning range,

$$0 < Q_c < \sqrt{2} .$$

To maintain complex conjugate poles in Eq. (28),[40] we find that the following inequality holds:

$$0 \leq (F_c + Q_c) \leq 2 .$$

This is found by rooting the denominator of Eq. (28),

$$H_{ch}(z) = \frac{F_c^2 z^{-1}}{(1 - az^{-1})(1 - a^*z^{-1})} \tag{32}$$

where

$$a = 1 - \frac{F_c(F_c + Q_c)}{2} + jF_c\sqrt{1 - \frac{(F_c + Q_c)^2}{4}}.$$

Using the upper bound we found for Q_c, we see that there will always be some finite range of F_c over which the poles will be complex conjugate.

[37] It is remarkable that the delay-free loop is eliminated without compromise to the digital filter coefficients, because delay-free loops can be troublesome when it is desired to maintain autonomy of the coefficients in an analog-to-digital filter transformation.

[38] See Eq. (5) in Section 2.5, Appendix 2 to find the pole radius.

[39] A similar conclusion can be reached by solving Eq. (27) for Q in terms of F_c and Q_c via Eq. (29); we leave that for the reader. But note that Eq. (27) is an approximation whereas Eq. (31) is exact.

[40] That is, for peak-center frequency away from but asymptotically including dc.

Combining all three criteria, we finally conclude that to maintain a stable low-pass filter in the form of Eq. (28), having complex conjugate poles conforming to the prescribed tuning range, then the constraint must hold:

$$0 < F_c < \min\left[\frac{1}{Q_c}, 2 - Q_c, \frac{2}{Q_c} - Q_c, \frac{-Q_c + \sqrt{8 + Q_c^2}}{2}\right]. \quad (33)$$

We learn from Eq. (33) that an artificial upper bound on the value of Q_c equal to 1 yields a universal upper bound on F_c equal to 1 as well (corresponding to ω_c of about $\pi/3$). We conclude that we can guarantee stability of complex conjugate poles for any value of either filter coefficient as long as they individually remain within the range of $0 \rightarrow 1$.

3.2.3 Peak Gain

We examined the actual peak gain of $H_{ch}(e^{j\omega})$ (evaluated at ω_c) over the prescribed ranges of F_c and Q_c (both $0 \rightarrow 1$) substituting the truncated series approximation coefficients [Eq. (29)] into Eq. (25), then taking the square root. We found the peak gain to be greater than but approximately equal to $1/Q_c$. The largest excess beyond this estimate occurs for low center frequency and low Q, or for high center frequency and high Q. At $F_c = 0.000001$ and $Q_c = 1$ we found the greatest excess at about 15.5% more than $1/Q_c$.

There is no separate control over peak gain in the Chamberlin topology; it is controlled indirectly through Q_c. We recommend a maximum peak gain of 24 dB for musical purposes, corresponding to a minimum Q_c of about 0.0625 (filter $Q = 16$).

3.3 Performance of the Chamberlin Filter

Now we wish to know whether our approximations are good enough. To do this, an engineer might calculate the root locus of the Chamberlin poles to see how closely their trajectory matches that of a second-order constant-Q filter. Instead we will repeat the musical analysis, as in Fig. 17, relating Q and center frequency; but this time we will *not* use the ideal coefficients. Rather we use the actual filter coefficients given by the truncated series approximations in Eq. (29).

Specifically, the radian frequencies ω_c, ω_2, and ω_1 in Fig. 19 are calculated using the actual filter coefficients, evaluating Eq. (31) to get ω_c, and by substituting the truncated series approximation for β [Eq. (29)] into Eq. (26). Fig. 19 tells the whole story by relating actual filter Q [Eq. (1)] to the filter coefficients. Ideally, we are looking for a planar relationship. Nonetheless, the sheet is fairly unwrinkled up to $F_c \approx 1$, corresponding to a tuning range of peak-center frequencies up to $\pi/3$. Further, it appears that for our purposes the selectivity parameter control Q_c is sufficiently decoupled from the tuning frequency control F_c. Hence we can expect good agreement between theory and practice in that region.[41]

3.3.1 Integrator Analysis

Generally speaking, it is *not* a good idea to implement an ideal digital integrator unless it can be guaranteed that there exists a zero of transmission across it at dc. This is certainly the case for integrator 1 in Fig. 18, which has the required zero across it, but integrator 2 has no such zero. In that case one must then prove that there can exist no signal from any source having dc content upon arrival at the input to the integrator under scrutiny. Audio signals normally enter the digital circuit at the designated input node, but noise having dc content is routinely generated in any practical implementation at every node where signal truncation occurs. These noise sources often appear in contemporary DSP chips at the input to each multiplier because multiplier inputs cannot accommodate double-precision operands the way the ac-

[41] At a sample rate of 44.1 kHz, $\pi/3$ corresponds to a bandwidth of 7350 Hz. Considering that the topmost note of the pianoforte reaches only 4186 Hz, that tuning range is good enough for musical purposes.

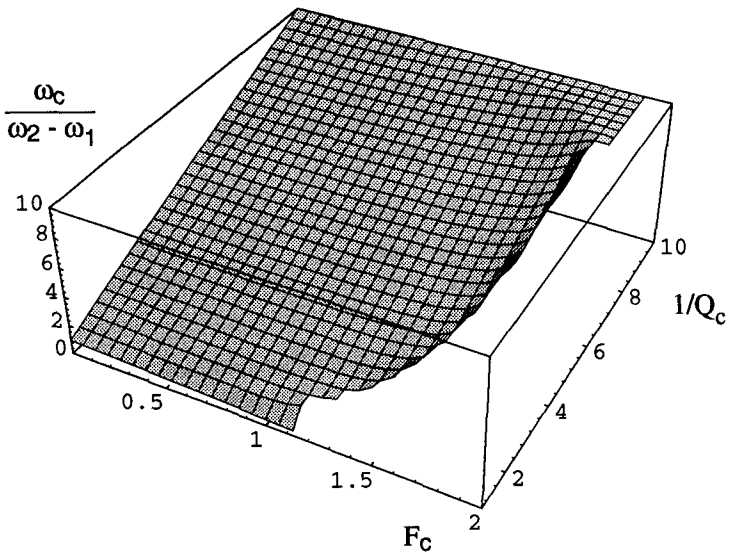

Fig. 19. Actual Chamberlin filter Q as a function of F_c and $1/Q_c$.

cumulators can.[42] The high-rate noise is accurately modeled as a deterministic source [12, Eq. (6)], input to a fictitious adder resting in front of the multiplier. Fig. 20 demonstrates the application of the noise model to one of the noise sources (e_2) on its way to integrator 2.

For the Chamberlin topology we have the remarkable result that the input to integrator 2 never sees any signal having dc content. For verification, we now look at the most interesting signal, which is the noise source at the input to the multiplier at node bp. There we have

$$\frac{I_2(z)}{e_2(z)} = -F_c \frac{1 - (2 - F_cQ_c)z^{-1} + (1 - F_cQ_c)z^{-2}}{\Delta} \tag{34}$$

where Δ is the denominator of Eq. (28). The transfer function (34) has a zero of transmission at dc, which can be proven by substituting $z = 1$. The three other possible sources (at nodes hp, bpq, and the filter input) acquire a simple zero of transfer at dc in the form $1 - z^{-1}$ by the time they arrive at I_2.

3.3.2 Truncation Noise; Spectral Criterion

The object of our noise analysis is to find the internal truncation noise generated by the circuit itself, which then appears at the observed output. The design goal of fidelity might be stated as follows:

Criterion 1: It is desired that the filter circuit generate no truncation noise appearing at the filter output which would fall above the spectral noise floor due to quantization of the original input signal.

Under this criterion it would be a violation for any portion of the truncation noise spectrum created by the filter itself to fall above the presumed spectrally white quantization noise floor of the input signal. Hence this criterion is the more conservative of the two criteria for fidelity that we will consider. Under this musical and spectral interpretation of fidelity, the filter is then termed *transparent* to the input signal.

Because the noise we are studying is deterministic (the sources are known), when it occurs early in a filter topology it undergoes filtering as does the signal itself. Using the truncation noise model described for the integrator analysis, we then need to know the transfer function from each of the internal noise sources to the low-pass output. Having obtained this information, we can predict the frequency-dependent amplification of each presumably wide-bandwidth noise source.

We assume that all internal truncation occurs at the same bit level. Suppose, for example, that internal noise due to truncation were being generated at the 20-bit spectral level. Then any given noise magnitude response would need to possess at least a 24-dB boost beyond unity (in any frequency region) before criterion 1 were violated, assuming 16-bit input signal fidelity and a spectrally flat signal-quantization noise floor.[43]

Like the one shown in Fig. 20, there are three noise sources arising due to truncation at each of the respective multiplier inputs. These nodes are labeled hp, bpq, and bp in Fig. 18.

1) *hp*: The noise transfer function from hp to the output is $-H_{ch}(z)$. This means that the noise transfer from hp is the same as the signal transfer (with sign inversion), which is certainly not a bad situation. This source is the largest of the noise contributors when observed from the filter's low-pass output.

2) *bpq*: The noise transfer from bpq to the output is $Q_cH_{ch}(z)$. We know that the peak gain of $H_{ch}(e^{j\omega})$ is approximately $1/Q_c$, which means that the peak of this noise magnitude response from bpq is about 1. This is good.

3) *bp*: Lastly, we consider the noise transfer from bp to the output, which can be written

$$\left(\frac{1 - z^{-1}}{-F_cz^{-1}} - Q_c\right) H_{ch}(z) .$$

When filter Q is high, the first term introduces a zero at dc (but only to the noise), which is exactly what happens when we employ first-order truncation error feedback [12]. The zero squelches the noise in the low-frequency region, but boosts it in the high-frequency region. Using Eq. (29) we find that the crossover point (the frequency at which $|1 - z^{-1}|/F_c = 1$), above which the $1 - z^{-1}$ term begins to boost, is approximately ω_c. In this particular circumstance, the second-order low-pass filter $H_{ch}(z)$ kicks in above ω_c to remove the boosted noise. Hence the center frequency and peak gain of this noise source are about the same as that of hp, for high Q.

Summarizing the three noise sources, the noise source at hp would demand 1 extra bit in the filter signal path for every 6 dB of peak gain at the filter output (Section 3.5, Appendix 3), because node hp sees the same gain

[42] We presume *no* truncation post-accumulation for these integrators. This presumption is justified based on the alternative, which is a leaky integrator, requiring a multiply in its loop.

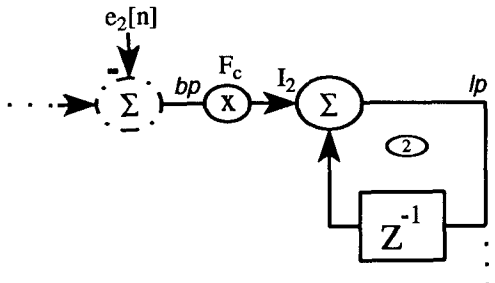

Fig. 20. Truncation noise source model (Appendix 8).

[43] (20 − 16 bits) × 6 dB/bit = 24 dB. Another way of looking at this example would be to say that under fidelity criterion 1, four extra bits beyond 16 would be required in the filter signal path to maintain filter transparency if any one of the noise magnitude responses were capable of a 24-dB gain in any frequency region, that is, 1 bit for every 6 dB of gain beyond unity magnitude response. Refer to Appendix 3 in Section 3.5 for supporting noise concepts.

as does the input signal. The noise source at bpq is insignificant when compared with the other two. Node bp is much like node hp. We finally determine the required total number of extra bits as an uncorrelated sum of the respective peak gains;

$$\approx \log_2\sqrt{1 + (16^2 + 1^2 + 16^2)} = 4.5 \text{ bit}.$$

So for a recommended maximum peak gain of 24 dB (16 absolute) across the filter, under criterion 1, five more bits would be required in the single-precision signal path.

3.3.3 Truncation Noise; Power Criterion

What we have thus far are the deterministic noise transfer functions, which were of interest because of the way that the first criterion for fidelity was stated. The design goal of fidelity has a second, quantitative (statistical) interpretation which is stated as follows:

Criterion 2: It is desired that the truncation noise power generated by the filter circuit, observed at the filter output, be less than the noise power due to quantization of the original input signal.

Here we compare internally generated truncation noise power observed at the filter output to the signal-quantization noise power observed at the input. This criterion is a classical high-rate interpretation,[44] ubiquitously ascribed as the −6-dB average noise power per individual bit [14, ch. 3.7.3]. We begin by assuming that the number of bits in the filter's single-precision signal path is the same as the number of bits representing the input signal. We then determine the number of extra bits required in the signal path to maintain input signal fidelity using this criterion. To do so we must calculate what is commonly termed the noise gain [29, ch. 9.2.2]. This is essentially an estimate of the total power boost of the internally generated, presumed spectrally white noise. Any boost beyond unity is bad news, whereas any gain of unity or less is good.[45] The calculation of noise gain is often performed in the frequency domain exploiting the Parseval energy relation [Eq. (35)], which integrates the magnitude-squared response of a noise transfer function,

$$G = \sum_{n=-\infty}^{\infty} |h[n]|^2 = \frac{1}{2\pi}\int_0^{2\pi} |H(e^{j\omega})|^2 \, d\omega \,. \tag{35}$$

Hence the result of the calculation is unitless since the integrand is a ratio. From the noise transfers we know that the worst offender is due to the noise source at hp in Fig. 18. Substituting the poles of that transfer [Eq. (32)] into the integration results from [14, ch. 4.7, p. 187; ch. 6.9.1, p. 357], we calculate the noise power gain at hp as

$$G_{hp} = \frac{F_c^4\{a[1 - (a^*)^2] - a^*(1 - a^2)\}}{(a - a^*)(1 - a^2)[1 - (a^*)^2](1 - aa^*)} \,. \tag{36}$$

The vertical axis in Fig. 21 represents G_{hp} evaluated over the recommended operating ranges of F_c ($0 \rightarrow 1$) and Q_c ($1 \rightarrow 0.0625$). There is zero noise gain for $F_c = 0$, because the filter is shut down at that point. We see that the worst noise power boost occurs at hp, for high center frequency and high Q, where G_{hp} reaches 10.7826, which translates to 1.8 bit,

$$\frac{10 \log (1 + G_{hp})}{20 \log 2} = \log_2\sqrt{1 + G_{hp}}$$

$$= \log_2\sqrt{1 + 10.7826} \text{ bit}.$$

[44] The term "rate" here refers to the number of bits per sample.

[45] A noise gain of 1 says that the noise transfer function will not increase the amount of noise that it passes, but says nothing about the spectral distribution of the noise passed.

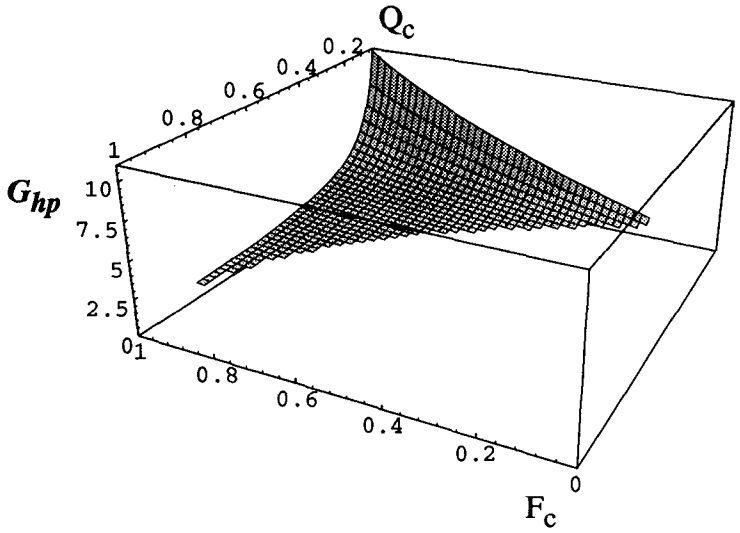

Fig. 21. Showing G_{hp} only where it exceeds unity.

G_{bp} looks much like G_{hp}. The aggregate of the noise-power gain contributions, corresponding to all the presumably uncorrelated noise sources in the circuit of Fig. 18, is calculated simply as the sum $G_{hp} + G_{bpq} + G_{bp}$ when all signal truncations occur at the same bit level. This sum demands 2.2 bit, worst case as above:

$$\log_2 \sqrt{1 + G_{hp} + G_{bpq} + G_{bp}}$$

$$= \log_2 \sqrt{1 + 10.7826 + 0.0421196 + 10.4769}\ \text{bit}.$$

We conclude that the aggregate requires a total of three extra bits (beyond the desired fidelity) in the filter signal path to maintain fidelity in accord with criterion 2.

3.3.4 Truncation Noise Summary

From the perspective of either criterion, the all-pole low-pass Chamberlin topology looks very good from the standpoint of truncation noise performance. This is true because the pole gain, which is the determinant of noise gain in general, does not exceed the desired filter peak gain for the Chamberlin topology; the maximum pole gain *is* the maximum peak gain, which we recommend to be 24 dB.

3.3.5 Limit-Cycle Oscillation

Zero-input limit cycling arises due to ongoing signal quantization within a recursive topology [11]—a nonlinear operation in an otherwise discrete linear system.[46] The quantized filter coefficients are parameters to limit cycles, but are not the cause. Limit cycles manifest themselves as annoying low-level tones at a circuit's outputs after input signal has been removed. Signal quantization in most modern DSP chips often takes place at the single-precision multiplier inputs where double-precision operands cannot be accepted and so must be truncated.[47] The limit-cycle tones can therefore be visualized to enter the topology at the same places as the truncation noise. One such input port is shown in Fig. 20. Similarly to the truncation noise sources, if limit-cycle tones occur early in a filter topology they will be filtered just like the signal (at the same point of entry) itself.

In the decade just passed, we have learned that limit-cycle oscillation is minimized by truncation error feedback [18], which was devised to minimize the amplification of truncation noise [12]. Essentially, error feedback introduces zeros strategically placed on the unit circle into the noise transfer function, but leaves the signal transfer function alone. Therefore, a reasonable hypothesis is that with or without error feedback, if the noise transfer from a quantizer to the low-pass output has a term $1 - z^{-1}$, then it provides some immunity to limit cycles as well as some squelching of truncation noise, both artifacts caused by that same quantizer.[48] From our truncation noise analysis of the Chamberlin topology we see that only one of the truncation noise transfers (bp) has such a term. Hence limit-cycle tones cannot be completely ruled out, although empirically we do not recall overt problems with the circuit behavior in that regard.

We have relaxed the mathematical rigor in this section because limit-cycle analyses for arbitrary topologies are analytically difficult, in general. The question remains as to whether limit cycles are a serious problem for the Chamberlin topology. Further analysis is certainly called for. In the meantime we design for the worst case. So our recourse is to minimize the potential tones' amplitudes of oscillation by providing internal signal truncation at lower bit levels. That is tantamount to providing a higher precision signal path. A good rule of thumb is that there exist about 6 dB of limit-cycle suppression for each appended bit of precision.

3.3.6 Signal Overflow Analysis

The study of overflow is concerned with the observation of the signal magnitude at sensitive nodes. Typically, one or several sensitive internal nodes may overflow (or underflow) sooner than the output. A saturation nonlinearity clips (appropriately full-scale positive or negative) the overflowed node, as this is highly preferable to a two's complement wraparound nonlinearity. The audible consequence of clipping at internal nodes is much more objectionable than clipping at the filter output, however, so it must be precluded completely. The sensitive internal nodes are, once again, the multiplier inputs because they typically cannot accept overflowed inputs like the accumulators can.[49] These are labeled hp and bp in Fig. 18.

In our overflow analysis, what we are really interested in is the relationship of the sensitive nodes to the output. So we form a ratio R of transfers to the sensitive nodes with respect to the transfer to the output node.

1) *Node hp*: Formulate [Eqs. (28) and (29)]

$$R_{hp}(z) = \frac{hp(z)/X(z)}{H_{ch}(z)} = \frac{(1 - z^{-1})^2}{F_c^2 z^{-1}}$$

$$|R_{hp}(e^{j\omega})| = \frac{\sin^2(\omega/2)}{\sin^2(\omega_c/2)}.$$

This describes a boost over the output at high frequencies. The worst case of overflow comes at the highest frequency ($z = -1$) and for a peak-center frequency ω_c at the top of its utmost recommended range ($\pi/2$), for

[46] Signal quantization converts a discrete-time system to a digital system.

[47] Multipliers then produce double-precision results, which are usually fed to accumulators that can accept double-precision inputs.

[48] In fact, Laakso et al. [30] show that any zero in the noise transfer function provides some limit-cycle immunity. The common solution to both artifacts suggests that the two phenomena are homologous.

[49] Recall that most contemporary fixed-point accumulators are designed to tolerate infinite intermediate output overflow simply by virtue of nonsaturating adders [11, ch. 11.3]. So saturation at an accumulator output (when necessary) is never performed upon intermediate accumulated results.

there the boost over the unity output is by the factor 2.

2) *Node bp*: Formulate

$$R_{bp}(z) = \frac{bp(z)/X(z)}{H_{ch}(z)} = \frac{(1 - z^{-1})}{F_c}$$

$$|R_{bp}(e^{j\omega})| = \frac{\sin(\omega/2)}{\sin(\omega_c/2)}.$$

Similarly, the worst-case boost over the output is absolute $\sqrt{2}$.

Based upon this analysis, a simple technique to eliminate internal signal overflow, which we have found works quite well, is to precede the Chamberlin topology with a fixed input level attenuation of ½ and to follow with a compensation factor of 2 at the output. The output compensation amplifies the filter's internally generated truncation noise, however, that is, the ratio of the input signal power to the filter's own noise power degrades by 6 dB.

As shown in Fig. 16, the Chamberlin low-pass filter is a boosting filter. For some input signals the low-pass output may overflow. But overflow at the output will not always occur because the filtered signal may not have significant energy in the frequency region of the boost. Further, some small amount of clipping at the output is not offensive to the musician. Hence is it *not* desirable to automatically normalize the filter peak gain for the musician by attenuating the input signal because there will be an objectionable loss in perceived volume. For then, the musician would demand a knob for output compensation. Such a knob is emphatically discouraged because of the consequent amplification of internally generated truncation noise.

The most viable solution to the output overflow problem is to provide a filter input signal level user control. The user then determines at what input level any clipping at the output becomes offensive. When an input level user control exists for a boosting filter, it becomes unnecessary to provide user-controlled output compensation to maximize the output signal level.

3.4 Estimate of Signal Path Width

The 21-bit path-width estimate given in Table 5 maintains input signal fidelity of 16 bits at the Chamberlin low-pass filter output under criterion 2. The filter internal signal path width can be minimized by reducing the maximum peak gain or by compromising the bit-fidelity requirement. Note that under the more conservative fidelity criterion 1, the estimate of the required total number of bits in the filter signal path becomes 23 ($n = 5$, $r = 0$).

The integrating accumulators must retain double precision feedback to maintain stability.

3.5 Appendix 3: Truncation Noise Spectral Level versus Noise Power

We seek the relationship of the truncation noise power spectral level $|\eta(e^{j\omega})|^2/M$ to the noise power N because we wish to prove that for every additional 6 dB of S/N, the average noise spectral level drops by the same amount. The analysis of truncation noise is much like that of quantization noise [14, ch. 6.9.1, p. 353]. It is interesting that the classical high-rate estimation of quantization noise [14, ch. 3.7.3] is statistical in nature, hence does not include the actual sample rate F_s in its quantification. We therefore expect our final result to reflect this.

Here we regard truncation noise as deterministic so that it has an integrable spectrum, and we note that a discrete unity-level complex sinusoid of any duration and frequency has finite power $S = 1$. We then pose the problem: given $10 \log (S/N) = 96$ dB,[50] input signal duration equal to M samples, and input signal power $S = 1$, find the average truncation noise power spectral level η^2/M in relation to the noise power N.

We easily find the noise power, solving

$$10 \log (N) = -96$$

where

$$N = T\int_0^{1/T} \frac{|\eta(e^{j2\pi fT})|^2}{M}\,df = \frac{1}{2\pi}\int_0^{2\pi} \frac{|\eta(e^{j\omega})|^2}{M}\,d\omega \tag{37}$$

where $\omega = 2\pi fT$ throughout this paper. Eq. (37)[51] is a statement of noise power versus noise power spectral level, where $T = 1/F_s$ and M is the number of samples

[50] The approximate expected signal-to-noise ratio for a 16-bit fidelity signal is 96 dB (6 dB per bit) [14, ch. 3.7.3].

[51] Scaling of the Parseval energy relation [Eq. (35)] by $1/M$ to yield power in Eq. (37) is discussed in [31].

Table 5. Estimate of minimum required internal signal path width at 24-dB maximum peak gain.

Bit Budget	Attribute
$N = 16$	Output fidelity, assumed input signal quantization
$n = 3$	Truncation noise immunity (criterion 2)
$o = 1$	Internal signal overflow prevention
$m = 1$	Limit-cycle suppression (6 dB per bit)
$r = \max[0, (24/6) - n - m]$	Signal path LSBs for user-controlled input level attenuation*
$N + n + o + m + r = 21$ bits	Total

* Assuming that the filter internal signal path resolution ultimately exceeds the input signal resolution, then no input signal information will be lost through the use of an input level control, provided that the attenuation is limited to the difference in resolution (6 dB per bit, 24 dB recommended).

in the original data record. If we designate η^2/M to be the average noise power spectral level (the average of $|\eta(e^{j\omega})|^2/M$ over frequency), then solving Eq. (37) we find that

$$\frac{\eta^2}{M} = N \tag{38}$$

which is independent of the sample rate, as expected.

This result [Eq. (38)] indicates that the noise power spectral level is proportional to the noise power. This means that if the noise power drops by 6 dB, then so will its average spectral level (assuming M fixed). We needed to know this to justify the claim (supporting fidelity criterion 1 in Section 3.3.2) that the average difference in truncation noise spectral level between a 16-bit and a 20-bit quantized signal is 24 dB [$=(20 - 16) \times 6$].

4 REFERENCES

[1] D. C. Andreas, J. Dattorro, and J. W. Mauchly, "Digital Signal Processor for Audio Applications," U.S. patent 5,517,436 (1996 May 14).

[2] W. G. Gardner, "Reverberation Algorithms," in *Applications of Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. (Kluwer Academic, Norwell, MA, 1997).

[3] J. A. Moorer, "About this Reverberation Business," *Computer Music J.*, vol. 3, pp. 13–28 (1979 June), also in *Foundations of Computer Music*, C. Roads and J. Strawn, Eds. (MIT Press, Cambridge, MA, 1988).

[4] D. Griesinger, "Practical Processors and Programs for Digital Reverberation," in *Audio in Digital Times, Proc. Audio Eng. Soc. 7th Int. Conf.* (Toronto, Ont., Canada, 1989 May 14–17), pp. 187–195.

[5] B. A. Blesser and K. O. Bader, "Electric Reverberation Apparatus," U.S. patent 4,181,820 (1980 Jan. 1).

[6] D. Griesinger, "Room Impression, Reverberance, and Warmth in Rooms and Halls," presented at the 93rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 1064 (1992 Dec.), preprint 3383.

[7] M. Schroeder, "Natural Sounding Artificial Reverberation," *J. Audio Eng. Soc.*, vol. 10, p. 219 (1962 July).

[8] J. M. Jot and A. Chaigne, "Digital Delay Networks for Designing Artificial Reverberators," presented at the 90th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 39, p. 383 (1991 May), preprint 3030.

[9] J. O. Smith, "Elimination of Limit Cycles and Overflow Oscillations in Time-Varying Lattice and Ladder Digital Filters," in *Music Applications of Digital Waveguides*, Rep. STAN-M-39, Center for Computer Research in Music and Acoustics (CCRMA), Dept. of Music, Stanford University, Stanford, CA (1987 May).

[10] A. V. Oppenheim, Ed., *Applications of Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1978).

[11] L. B. Jackson, *Digital Filters and Signal Processing*, 3rd ed. (Kluwer Academic, Norwell, MA, 1996).

[12] J. Dattorro, "The Implementation of Recursive Digital Filters for High-Fidelity Audio," *J. Audio Eng. Soc.*, vol. 36, pp. 851–878 (1988 Nov.); Comments, ibid. (*Letters to the Editor*), vol. 37, p. 486 (1989 June); Comments, ibid. (*Letters to the Editor*), vol. 38, pp. 149–151 (1990 Mar.).

[13] P. A. Regalia and S. K. Mitra, "Tunable Digital Frequency Response Equalization Filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 118–120 (1987 Jan.).

[14] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[15] J. Strawn, Ed., *Digital Audio Signal Processing* (A-R Editions, Madison, WI, 1985).

[16] J. A. Moorer, "The Manifold Joys of Conformal Mapping: Applications to Digital Filtering in the Studio," *J. Audio Eng. Soc.*, vol. 31, pp. 826–841 (1983 Nov.).

[17] D. P. Rossum, "Dynamic Digital IIR Audio Filter and Method which Provides Dynamic Digital Filtering for Audio Signals," U.S. patent 5,170,369 (1992 Dec.).

[18] T. I. Laakso, "Error Feedback for Reduction of Quantization Errors due to Arithmetic Operations in Recursive Digital Filters," D. Tech. thesis, Rep. 9, Laboratory of Signal Processing and Computer Technology, Helsinki University of Technology, Espoo, Finland (1991).

[19] K. Steiglitz, *A Digital Signal Processing Primer* (Addison-Wesley, Menlo Park, CA, 1996).

[20] *Moog Voltage-Controlled Ladder Filter*, Internet: http://www.sara.nl/Rick.Jansen/Emusic/Moog/moogvcf_schematic.gif and Internet: http://www-ccrma.stanford.edu/~stilti/papers/moogvcf.pdf.

[21] D. Rossum, "Making Digital Filters Sound 'Analog,'" in *Proc. Int. Computer Music Conf.* (San Jose, CA, 1992), pp. 30–33.

[22] "CEM3328 Four Pole Low Pass VCF," Curtis Electromusic Specialties, Los Gatos, CA (1983).

[23] H. Chamberlin, *Musical Applications of Microprocessors* (Hayden, Indianapolis, IN, 1980).

[24] *Mathematica*, version 2, Wolfram Research, Champaign, IL (1994).

[25] B. L. Evans, L. J. Karam, K. A. West, and J. H. McClellan, "Learning Signals and Systems with *Mathematica*," *IEEE Trans. Education*, vol. 36, pp. 72–78 (1993 Feb.).

[26] K. W. Martin and M. T. Sun, "Adaptive Filters Suitable for Real-Time Spectral Analysis," *IEEE Trans. Circuits Sys.*, vol. CAS-33, pp. 218–229 (1986 Feb.); also *IEEE J. Solid-State Circuits*, vol. SC-21, no. 1 (1986 Feb.), pp. 108–119.

[27] M. R. Petraglia, S. K. Mitra, and J. Szczupak, "Adaptive Sinusoid Detection Using IIR Notch Filters

and Multirate Techniques," *IEEE Trans. Circuits Sys. II*, vol. 41, pp. 709–717 (1994 Nov.).

[28] T. Kwan and K. Martin, "Adaptive Detection and Enhancement of Multiple Sinusoids Using a Cascade IIR Filter," *IEEE Trans. Circuits Sys.*, vol. 36, pp. 937–947 (1989 July).

[29] P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[30] T. I. Laakso, P. S. R. Diniz, I. Hartimo, and T. C. Macedo, Jr., "Elimination of Zero-Input and Constant-Input Limit Cycles in Single-Quantizer Recursive Filter Structures," *IEEE Trans. Circuits Sys. II*, vol. 39, pp. 638–646 (1992 Sept.).

[31] A. Luthra, "Extension of Parseval's Relation to Nonuniform Sampling," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36 (1988 Dec.), pp. 1909–1911.

THE AUTHOR

Jon Dattorro is from Providence, RI. He trained as a classical pianist, attended the New England Conservatory of Music where he studied composition and electronic music, and performed as soloist with Myron Romanul and the Boston Symphony Orchestra for Children's Concerts at Symphony Hall. His scores include a ballet and a piano concerto.

Mr. Dattorro received a B.S.E.E. with highest distinction from the University of Rhode Island in 1981, where he was a student of Leland B. Jackson. In 1984 he received an M.S.E.E. from Purdue University, specializing in digital signal processing under S. C. Bass. He is currently working towards a Ph.D. in electrical engineering at Stanford University.

He designed the Lexicon Inc. model 2400 Time Compressor with Charles Bagnaschi and Francis F. Lee in 1986, and he designed most of the audio effects from Ensoniq Corp. between 1987 and 1995. He shares two patents in digital signal processing chip design with David C. Andreas, J. William Mauchly, and Albert J. Charpentier. Personal mentors are Salvatore J. Fransosi, Pozzi Escot, and Chae T. Goh.