Analog Filter Banks with Low Intermodulation Distortion

James A. Cherry and W. Martin Snelgrove
Department of Electronics, Carleton University, Ottawa, Canada K1S 5B6
jac@doe.carleton.ca, snelgar@doe.carleton.ca

Abstract

It is demonstrated that the filter-bank structure of [1] can be designed to significantly reduce the effect of intermodulation distortion. This gives it an important advantage over other structures in radio systems. The principle is explained, and supported with Volterra analysis and simulation.

I. Introduction

Filters are key components of a radio, often dominating selectivity and dynamic range. The rapid development of cellular telephony has recently spurred increased research on all aspects of radio, and because off-chip filters often limit size, flexibility, and cost, there is a strong motivation to replace them with on-chip filtering. A superheterodyne radio uses filters at every stage — radio-frequency (RF), intermediate-frequency/frequencies (IF) and baseband, and so needs many types. Off-chip filters are usually passive and use high-Q inductors, surface acoustic wave technology, or crystal or ceramic resonators — technologies not available on standard silicon.

Passive filters have quite poor selectivity on chip, because inductors have low Q. They also only have useful Q (in the range 5 to 10) above 1GHz [2] and so can contribute mostly to the RF section. Active-RC filters are quite slow, and useful only at baseband or for a very low IF; switched-C filters can be pushed to an IF of a few MHz [3]; and transconductance-C filters are faster (up to 400MHz and beyond [4]) but have poor linearity when designed for high speed and low power; A/D conversion at the IF or baseband can be used to move some filtering to the digital domain [5-7], but some is still needed on the analog side to reduce the linearity requirements on the converter. The Q of an on-chip inductor can be improved by active techniques [8,9] but again at the cost of linearity.

Linearity is vital in the design of a radio receiver [10], because it may have to select a signal that is very close in frequency to much stronger interferers — and the presence of a small nonlinearity can cause nonlinear mixing products of the interferers to produce signals that overlap the spectrum of the desired signal. For example, the base station for a cell-phone system can have interferers 80dB larger than the desired signal separated by only 60kHz at a carrier of nearly 900MHz [11].

The literature on active filtering has typically looked for structures that have low sensitivities to errors in their components [12] or low noise gains [13]. For radio, it may be much more important to find structures very tolerant of the nonlinearities of their amplifiers.

II. Modeling nonlinearities in radio amplifiers

For memoryless systems such as amplifiers and mixers a Taylor series expansion is often a good description of behaviour, because nonlinear terms are small by design and so the series converges rapidly. Also by design, a balanced circuit amplifier gain has odd symmetry so that

$$y = a_1 x + a_3 x^3 \tag{1}$$

is often a good representation of input-output behaviour, with the cubic term generally small compared to the linear one. For large enough signals the cubic would eventually dominate, and the input level x at which the two terms are equal is referred to as the "third order intercept" $\mathrm{IP}_3 = \sqrt{a_1 / a_3}$. A radio is designed to operate with signals well below IP_3.

When the input contains a desired signal at ω_0 and an interferer at ω_1

$$x = V_0 e^{j\omega_0 t} + V_1 e^{j\omega_1 t} + \overline{V_0} e^{j(-\omega_0)t} + \overline{V_1} e^{j(-\omega_1)t} + \ldots \tag{2}$$

there will be output terms

$$y = 3a_3 V_0^2 \overline{V_0} e^{j\omega_0 t} + 3a_3 V_0 V_1 \overline{V_1} e^{j\omega_0 t} + \ldots \tag{3}$$

at frequencies

$$\begin{aligned} \omega_0 &= \omega_0 + (\omega_0 - \omega_0) \\ \omega_0 &= \omega_0 + (\omega_1 - \omega_1) \end{aligned} \tag{4}$$

Then $y = a_1 V_0 e^{j\omega_0 t} + \ldots$, the linear-gain term at the desired frequency, has superimposed on it two third-order terms, the first due to gain compression, the second due to "desensitization" by the interferer. The latter becomes a problem when its amplitude $3a_3 V_0 V_1 \overline{V_1}$ approaches that of the linear-gain term $a_1 V_0$, i.e., when

$$V_1 \to \frac{IP_3}{\sqrt{3}} \tag{5}$$

Hence the need for interferers to be small compared with IP_3.

When the input contains two interferers at ω_1 and ω_2

$$x = V_1 e^{j\omega_1 t} + V_2 e^{j\omega_2 t} + \overline{V_1} e^{j(-\omega_1)t} + \overline{V_2} e^{j(-\omega_2)t} + \ldots \tag{6}$$

the output has desensitization terms as well as a term like

$$y = 3a_3 \overline{V_1} V_2^2 e^{j\omega_0 r} + \ldots \tag{7}$$

at the frequency

$$\omega_0 = 2\omega_2 - \omega_1 = \omega_2 + (\omega_2 - \omega_1) \tag{8}$$

Since channels in a radio system are evenly spaced, this third-order intermodulation product falls on top of a channel — possibly the desired channel. This becomes a problem when the distortion term amplitude $3a_3 \overline{V_1} V_2^2$ approaches the linear-gain amplitude $a_1 V_0$ from a small input $x = V_0 e^{j\omega_0 t}$, i.e., when

$$V_1 V_2^2 \rightarrow \frac{\mathrm{IP}_3^2}{3} V_0 \qquad (9)$$

Again, interferers must be small compared with IP_3.

III. Nonlinearities in filters

The Taylor series can only be applied to memoryless nonlinearities, and we are interested in the design of highly selective active filters. We are accustomed to using Laplace techniques, but these only apply to linear systems. Within an active filter circuit there are many nonlinearities — the current-voltage characteristics of a transistor are nonlinear as are almost all its parasitic resistances and capacitances. These nonlinearities drive the filtering components of the nominal circuit and are further filtered by parasitics. Analysis of a system simultaneously containing weak (Taylor-type) nonlinearities and memory generally involves Volterra kernels [14, 15] $g_i(t_1, t_2, \ldots, t_i)$ which are convolved with all possible products of delayed input signals in a form with the structure of both the Taylor and linear convolution equations:

$$\begin{aligned} y(t) &= \int d\tau g_1(\tau) x(t-\tau) \\ &+ \int d\tau_1 \int d\tau_2 g_2(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) \\ &+ \int d\tau_1 \int d\tau_2 \int d\tau_3 g_3(\tau_1, \tau_2, \tau_3) x(t-\tau_1) x(t-\tau_2) x(t-\tau_3) \\ &+ \ldots \end{aligned} \qquad (10)$$

A frequency-domain version of this generalized convolution can be had by taking the Fourier transform, and reads

$$\begin{aligned} Y(\omega) &= G_1(\omega) X(\omega) + \\ &\int d\Omega_1 G_2(\Omega_1, \omega - \Omega_1) X(\Omega_1) X(\omega - \Omega_1) + \\ &\int d\Omega_1 \int d\Omega_2 G_3(\Omega_1, \Omega_2, \omega - \Omega_1 - \Omega_2) X(\Omega_1) X(\Omega_2) X(\omega - \Omega_1 - \Omega_2) \\ &+ \ldots \end{aligned} \qquad (11)$$

meaning that the output power at a given frequency ω can be obtained by convolving a kernel G_i with all combinations of input frequencies that add to ω. The kernels may be computed for simple systems by positing an input that is a sum of incommensurate complex exponentials $\sum e^{j\omega_i t}$, and for a larger system by mathematically manipulating the kernels of its components [15].

IV. Distortion in a Gm-C biquad

If the transconductors of the simple biquad of Fig. 1 are linear, it has a transfer function

$$\frac{V_o}{V_i} = \frac{\frac{g_{m,\,in}}{C_1} s}{s^2 + \frac{1}{R_1 C_1} s + \frac{g_{m,\,f1} g_{m,\,f2}}{C_1 C_2}} \qquad (12)$$

If they each have the weak cubic nonlinearity of (1) with coefficients $a_{1,\,\mathrm{in}}$, $a_{1,\,f1}$ and $a_{1,\,f2}$ etc. (adding a subscript to

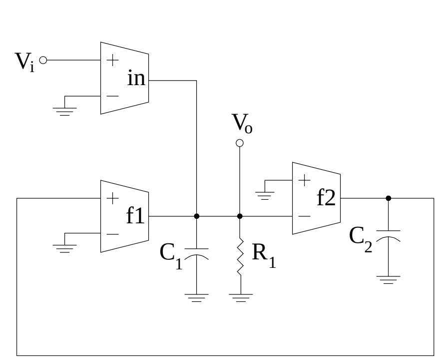

Fig. 1: A transconductance-C biquad

name the transconductor modeled), then the third-order Volterra kernel becomes [17]

$$\begin{aligned} G_3(\Omega_0, \Omega_1, \Omega_2) &= \left[C_1 \sum_l j\Omega_l + \frac{1}{R_1} + \frac{a_{1,\,f1} a_{1,\,f2}}{C_2 \sum_l j\Omega_l} \right]^{-1} \\ &\times \left(6a_{3,\,in} - 6\left[\prod_l G_1(\Omega_l)\right] \left[\frac{a_{3,\,f2} a_{1,\,f1}}{C_2 \sum_l j\Omega_l} + \frac{a_{3,\,f1} a_{1,\,f2}}{C_2^3 \prod_l j\Omega_l} \right] \right) \end{aligned} \qquad (13)$$

where $j = \sqrt{-1}$ and

$$G_1(\Omega_1) = \frac{\frac{a_{1,\,in}}{C_1} j\Omega_1}{(j\Omega)^2 + \frac{1}{R_1 C_1} j\Omega_1 + \frac{a_{1,\,f1} a_{1,\,f2}}{C_1 C_2}} \qquad (14)$$

These expressions assist in biquad design in the presence of interferers: for example allowing us to evaluate the effect of changing Q while keeping other parameters (peak gain, input levels and input frequencies, and transconductor IP_3) constant. Fig. 1 shows the intermodulation distortion term amplitude with peak gain 20dB, 100µV inputs, desired tone 100MHz at bandcenter, interferers at 99 and 98 MHz, and IP_3 of 10mV. The desired tone's linear output amplitude is a constant -60dBV.

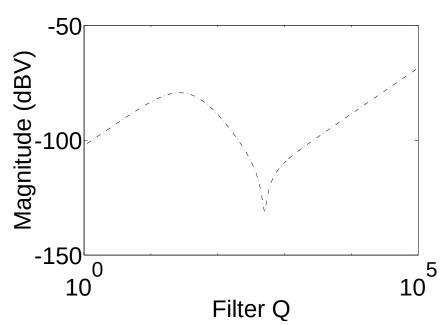

Fig. 2: Intermodulation at the filter output versus Q

The shape of the curve allows for an intuitive explanation. First of all, increasing Q generally increases the gain to the output of an intermodulation product at the center frequency, so the curve tends upwards. At low Q all three tones are in-band, and the circuit behaves more or less as an amplifier would. As Q increases, though, the interferers move out of band and excite the nonlinearity less strongly, improving the situation. Finally, at a high enough Q the gain enhancement for intermodulation products again starts to dominate and distortion starts to rise. This suggests that there is an optimum Q at which to design, though unfortunately other constraints (such as the finite bandwidth of the actual radio input, or limited accuracy in a Q-tuning circuit) may make this point unreachable.

V. The Filter Bank

The curve of Fig. 1 suggests that, with a given amplifier, it would be desirable to notch out interferers as soon as possible so that they could not excite G_3. A typical biquad notch filter does not do this, though, since the notch is formed only for one amplifier and the others are still driven by interferers. Thus structures like cascades and ladder simulations [16] cannot be expected to gain much benefit from their notches, at least in the stages up to and including the one forming the notch.

The filter bank structure of [1], though, has the property we want. It consists of a number of infinite-Q resonators (three in the diagram), each tuned to a different frequency and with overall feedback providing damping. The signal $E(\omega)$ ideally has notches at the resonator frequencies, because it is followed by a block with infinite gain at these frequencies. Designers accustomed to op-amps will recognize this as a sort of “virtual ground”.

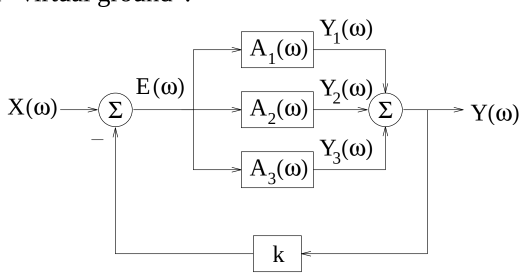

Fig. 3: A filter bank

Linear analysis of Fig. 1 shows that each of the three resonators has an output

$$Y_i(\omega) = \frac{A_i(\omega)}{1 + k\sum A_i(\omega)} X(\omega) \tag{15}$$

If one resonator is tuned to the desired signal and the others to interferer frequencies, then the interferers are nulled *at the input* of the biquads because

$$E(\omega) = \frac{1}{1 + k\sum A_i(\omega)} X(\omega) \tag{16}$$

and an A_i term goes to infinity at each signal frequency — so the relevant third-order distortion term should be nulled.

Fig. 1 shows the transfer function for a filter bank and compares it to two possible competing biquads: a low-Q one with similar passband performance and a high-Q one with similar stopband performance. The notches allow us to get a good steep stopband without using a high Q, which we know would be dangerous (recall Fig. 1).

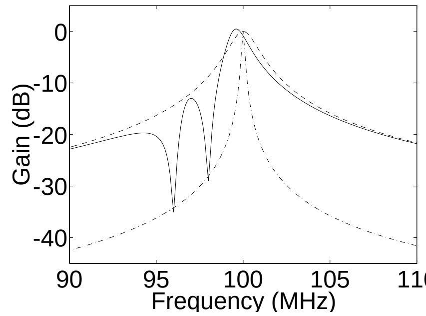

Fig. 4: Filter-bank transfer function compared to biquads

While (15) predicts infinitely deep notches, a practical filter will not have its resonators tuned exactly to infinite Q, and so will have finite depth notches. Fig. 1 was drawn for resonators with Q=630. This will set a practical limit to the improvement we can expect in distortion performance, and to how close the notches can be to the bandcenter.

VI. Volterra Analysis of the Filter Bank

The filter bank is a feedback structure containing biquads, which individually have the Volterra representation of (13) and (14). Let the k^{th} Volterra kernel of the A_i blocks be $A_{i,k}$ (so that $A_{i,1}$ is just the familiar transfer function). All the components have odd symmetry, and combining them with additions and subtractions doesn’t change that, so the overall system will only have odd-order Volterra kernels. Feedback can generate higher-order kernels, but we’ll assume that nonlinearities are weak enough that they can be ignored. Calling the third-order kernel modeling overall gain from input to output i $G_{i,3}$ it can be shown that

$$\begin{aligned} G_{i,3}(\Omega_0, \Omega_1, \Omega_2) &= \left(\prod_{j=0}^{2} \frac{1}{1 + k\sum_l A_{l,1}(\Omega_j)} \right) \\ &\times \left(A_{i,3}(\Omega_0, \Omega_1, \Omega_2) - \frac{A_{i,1}\left(\sum_l \Omega_l\right) \sum_j A_{j,3}(\Omega_0, \Omega_1, \Omega_2)}{1 + k\sum_i A_{j,1}\left(\sum_l \Omega_l\right)} \right) \end{aligned} \tag{17}$$

where the first bracketed factor represents the effect that the loop has of reducing the input signals to the biquads at each ω_l and the second one represents the effect of loop feedback at the frequency $\sum \Omega$ on reducing the error induced by intermodulation distortion. In the $2f_2 - f_1$ case that we are interested in, where $\Omega_0 = \Omega_1 = \omega_2$ and $\Omega_2 = -\omega_1$, all four denominator terms are large and help to reduce intermodulation from the open-loop $A_{i,3}$. Even if the biquads provide only 20dB of loop gain at each frequency we can expect more than 60dB of linearization!

VII. Simulation

The block diagram of Fig. 1 with individual filters modeled as in Fig. 1 with nonlinear transconductors was simulated using a Runge-Kutta numerical integration program. The desired tone was at ω_0=100MHz while the interferers were at ω_1=98MHz and ω_2=96MHz. A method of numerically extracting Volterra series terms was applied [17] and a comparison of the distortion terms for the three filters in Fig. 1 is presented in Table 1. All three architectures have 0dB linear gain at bandcenter yet the filter bank shows a marked improvement over the single-filter implementations.

Table 1: Simple filter bank distortion term improvements.

Third-order term	Filter bank win over single low-Q filter (dB)	Filter bank win over single high-Q filter (dB)
Compression	2.9	22.9
ω_1 desens.	32.9	52.6
ω_2 desens.	44.5	64.5
ω_1, ω_2 intermod	33.0	53.4

A more realistic simulation was done using SPICE models of a resonator patterned after [4] in which inductors of Q=5 are Q-enhanced to Q=90 for a low-Q single filter and Q=800 for a high-Q single filter, then tuned to a frequency of ω_0=1.83GHz. A filter bank made up of three high-Q filters with notches at ω_1=1.81GHz and ω_2=1.79GHz was compared with the two single filters. Table 2 shows the improvements to be expected. The intermodulation wins are lower than might be hoped, but this can be shown [17] to be due to the high bandcenter gain, 40dB, of each filter. Actual filters would be designed with much smaller gains, and this can be shown to lower the filter bank distortion considerably; realistic filters with smaller gains were not readily available for simulation.

Table 2: Realistic filter bank improvements.

Third-order term	Filter bank win over single low-Q filter (dB)	Filter bank win over single high-Q filter (dB)
Compression	-5.6	47.4
ω_1 desens.	15.0	34.0
ω_2 desens.	16.5	32.9
ω_1, ω_2 intermod	4.6	5.7

VIII. Conclusions

By selecting a filter structure properly it is possible to radically improve the linearity of active filters, and in particular the parallel-resonator filter-bank structure of Fig. 1 appears very good. This may open the way to replacing some of the passive filters of a conventional radio with active ones based on Gm-C or active-LC circuits.

This improvement requires that resonators be tuned to key interferers, and there are two situations in which this may be practical.

- when probable locations for interferers are known at the system level — such as at the “alternate channels” or in nearby bands allocated to other purposes.
- when the other biquads can be tuned directly to interferers actually present, as in the design of a cellular-telephony base station when several channels must be received anyway. In this situation all of the filter bank outputs would be used for reception, each also providing notches for the others.

Acknowledgments

James Cherry’s work was supported by an NSERC postgraduate scholarship. NT and BNR also contribute, along with NSERC and other members of the Ottawa-Carleton Research Institute, to Snelgrove’s Industrial Research Chair, which paid for the equipment used. The Gm-C filter work was initiated with funding from Micronet, which in turn is funded under NSERC’s NCE programme.

References

[1] M. Padmanabhan and K. W. Martin, “Resonator-Based Filter-Banks for Frequency-Domain Applications,” *IEEE Trans. Circ. Syst.*, Oct. 1991, pp. 1145-1159.

[2] N. M. Ngyuen and R. G. Meyer, “Si IC-Compatible Inductors and LC Passive Filters,” *IEEE Journal SS Circ.*, Aug. 1990, pp. 1028-1031.

[3] B. S. Song, “A 10.7MHz Switched-Capacitor Bandpass Filter,” *IEEE Journal SS Circ.*, Apr. 1989, pp. 320-334.

[4] W. M. Snelgrove and A. Shoval, “A Balanced 0.9um CMOS Transconductance-C Filter Tunable over the VHF Range,” *IEEE Journal SS Circ.*, Mar. 1992, pp. 314-323.

[5] T. H. Pearce and A. C. Baker, “Analogue to Digital Converter Requirements for HF Radio Receivers,” *Proc. IEE Colloquium on System Aspects and Applications of ADCs for Radar, Sonar, and Communications*, 1987.

[6] S. A. Jantzi, W. M. Snelgrove, and P. F. Ferguson Jr., “A Fourth-Order Bandpass Sigma-Delta Modulator”, 1992 *Custom Integrated Circuits Conference*, pp. 16.5.1-16.5.4.

[7] F. W. Singor and W. M. Snelgrove, “10.7MHz Bandpass Sigma-Delta A/D Modulator,” 1994 *Custom Integrated Circuits Conference*, pp. 163-166.

[8] R. A. Duncan, K. W. Martin, and A. S. Sedra, “A Q-enhanced active-RLC Bandpass Filter,” 1993 *International Symposium on Circuits and Systems*, pp. 1416-1419.

[9] S. Pilipos and Y. Tsividis, “Design of Active RLC Integrated Filters with Applications in the GHz Range,” 1994 *International Symposium on Circuits and Systems*, pp. 645-649.

[10] H. L. Kraus, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*. New York: Wiley, 1980.

[11] R. E. Fisher, “A Subscriber Set for the Equipment Test,” *Bell Syst. Tech. J.*, Jan. 1979, p. 133.

[12] H. J. Orchard, “Inductorless Filters,” *Electron. Lett.*, June 1966, pp. 224-225.

[13] W. M. Snelgrove and A. S. Sedra, “Synthesis and Analysis of State-Space Active Filters Using Intermediate Transfer Functions,” *IEEE Trans. Circ. Syst.*, Mar. 1986, pp. 287-301.

[14] V. Volterra, *Theory of Functionals and of Integro and Integro-Differential Equations*. New York: Dover, 1959.

[15] E. Bedrosian and S. O. Rice, “The Output Properties of Volterra Systems (Nonlinear Systems with Memory) Driven by Harmonic and Gaussian Inputs,” *Proc. IEEE*, Dec. 1971, pp. 1688-1707.

[16] R. Schaumann, M. S. Ghausi, and K. R. Laker, *Design of Analog Filters*. Englewood Cliffs: Prentice-Hall, 1990.

[17] J. A. Cherry, “Distortion Analysis of Weakly Nonlinear Filters Using Volterra Series,” M.Eng. thesis, Carleton University, Ottawa, Canada, 1994.