

On the Characterization and Reduction of Distortion in Bandpass Filters

James A. Cherry, *Student Member, IEEE*, and W. Martin Snelgrove, *Member, IEEE*

Abstract— In this paper we consider the nonlinear mixing products that arise and cause distortion in weakly nonlinear bandpass filters that might be used in a radio receiver [1], [2]. Volterra series analysis is applied to a G_m - C filter with weak cubic nonlinearities in the transconductance amplifiers so that explicit formulae for the amplitudes of the mixing products can be derived and an architecture for the reduction of distortion proposed. The analysis is used to aid in the characterization of distortion in practice, both in SPICE simulation and laboratory work. It is believed that this is the first time that measurements of the *phase* of Volterra transfer functions in the laboratory have been published in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.

Index Terms— Bandpass filters, harmonic distortion, volterra series.

I. INTRODUCTION

FILTERS are key components of a radio, often dominating selectivity and dynamic range. With the advent of cellular telephony, there has been a strong push to move filters on-chip, thereby reducing size and cost while improving flexibility. A superheterodyne radio uses filters at every stage—radio frequency (RF), intermediate frequency (IF), and baseband—and so needs many types.

Currently there exist several different architectures of on-chip filters, each with their own advantages and disadvantages. On-chip passive filters have quite poor selectivity because of the low Q of on-chip inductors; this can be improved by active techniques [3], [4] at the expense of linearity. Their useful Q is only for frequencies above 1 GHz [5], and thus passive filters contribute mostly to the RF stage. Active-RC filters are quite slow and useful only at baseband or a low IF, while switched-C filters can be pushed to an IF of a few MHz [6]. Transconductance-C (G_m - C) filters are faster (up to 400 MHz and beyond [7]), but have poor linearity when designed for high speed and low power.

Indeed, linearity is vital in the design of a radio receiver because it may have to select a signal that is very close in frequency to much stronger interferers. For example, an AMPS base station can have interferers 80 dB larger than the desired signal separated by only 60 kHz at a carrier of nearly 900 MHz [8]. The presence of even a weak nonlinearity can cause the nonlinear mixing products of the interferers to produce

Manuscript received May 5, 1997; revised August 27, 1997. This work was supported by the Natural Science and Engineering Research Council of Canada. This paper was recommended by Associate Editor D. A. Johns.

The authors are with the Department of Electronics, Carleton University, Ottawa, Ont., Canada K1S 5B6 (e-mail: jac@doe.carleton.ca; snelgar@doe.carleton.ca).

Publisher Item Identifier S 1057-7122(98)02541-0.

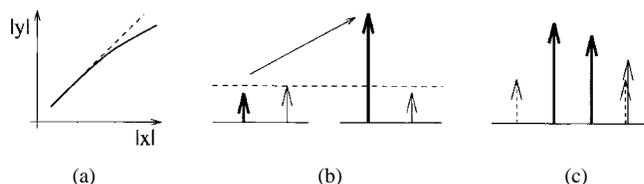


Fig. 1. Graphical illustration of (a) compression, (b) desensitization, and (c) intermodulation.

signals that overlap the spectrum of the desired signal, thereby distorting it and rendering it unrecoverable.

This paper examines how distortion is affected by the weak nonlinearities that are present in typical on-chip bandpass filters. It is organized as follows. Section II defines and classifies various types of distortion that occur in radios. Section III introduces Volterra series, explains why they are useful for characterizing distortion, and applies them for this purpose to a balanced G_m - C bandpass filter. Section IV builds on this to show why a bank of filters with feedback can reduce distortion under certain conditions. Section V discusses a measurement technique for finding distortion from the transient analysis of a circuit in SPICE, thereby confirming the calculated results presented in the previous sections. Section VI applies the technique to an actual G_m - C filter in the laboratory, demonstrating its practical viability. Lastly, Section VII draws conclusions about the work presented here.

II. DISTORTION

In this paper we shall define *distortion* as the degradation of the desired signal due to nonlinearity. We classify distortion into three types, illustrated in Fig. 1. In Fig. 1(a), the gain at a particular frequency does not remain constant as the input amplitude is increased; it eventually decreases, and this is called *gain compression*. In Fig. 1(b), it is shown that increasing the amplitude of an interferer can reduce the gain at the desired signal frequency. This is called the *desensitization* of the desired signal by the interferer. Lastly, Fig. 1(c) shows that two interferers intermodulate to produce intermodulation products (the dashed arrows) at the output of a nonlinear system. If the desired signal overlaps an intermodulation product, it will be degraded. The three types of distortion, then, are compression, desensitization, and intermodulation.

Distortion as defined here, therefore, arises when an intermodulation product (either by a tone with itself or between two tones) due to nonlinearity occurs at the same frequency as the desired tone. Modulated signals also distort, but the resulting

behavior is more complicated to analyze [9]. This paper treats signals as sinusoids for simplicity, but the general results still apply to modulated signals.

III. MODELING NONLINEARITY TO CALCULATE DISTORTION

A. Memoryless Systems

For memoryless systems, a Taylor series is often sufficient to characterize the input–output behavior. Consider the case of an amplifier with input x and output y . The equation

$$y = a_1x + a_3x^3 \quad (1)$$

is usually an adequate model for the input–output behavior because, by design, the nonlinearities are small and so higher order terms may be neglected. As well, a balanced design has odd symmetry, and thus the x^2 term can be neglected. The *input third-order intercept point* IIP_3 [10] is defined as the point where the linear and cubic terms are equal, i.e., where $a_1x = a_3x^3$, or

$$x = \text{IIP}_3 = \sqrt{\frac{a_1}{a_3}}. \quad (2)$$

Amplifiers are designed to operate with signal levels well below IIP_3 .

We can express the distortion terms defined in Section II mathematically for such an amplifier as follows. Suppose the input contains a desired signal at ω_0 and an interferer at ω_1 :

$$x = V_0 \exp(j\omega_0 t) + \bar{V}_0 \exp(-j\omega_0 t) + V_1 \exp(j\omega_1 t) + \bar{V}_1 \exp(-j\omega_1 t). \quad (3)$$

The output is found from putting (3) in (1). Along with the desired linear gain term at ω_0 , there will be output terms

$$y = 3a_3V_0^2\bar{V}_0 \exp(j\omega_0 t) + 3a_3V_0V_1\bar{V}_1 \exp(j\omega_0 t) + \dots \quad (4)$$

at frequencies

$$\begin{aligned} \omega_0 &= \omega_0 + (\omega_0 - \omega_0) \\ \omega_0 &= \omega_0 + (\omega_1 - \omega_1). \end{aligned} \quad (5)$$

From (5), we see that the first term of (4) is due to compression, whereby the nonlinearity causes the signal to “fold back on” (i.e., distort) itself, and the second term of (4) is due to desensitization, where the interferer interacts with the desired signal and produces an overlapping (i.e., distorting) tone.

Similarly, when the input contains two interferers at ω_1 and ω_2

$$x = V_1 \exp(j\omega_1 t) + \bar{V}_1 \exp(-j\omega_1 t) + V_2 \exp(j\omega_2 t) + \bar{V}_2 \exp(-j\omega_2 t) \quad (6)$$

the output will contain a term like

$$y = 3a_3\bar{V}_1V_2^2 \exp(j\omega_0 t) + \dots \quad (7)$$

at frequency

$$\omega_0 = \omega_2 + (\omega_2 - \omega_1). \quad (8)$$

If the desired signal is at frequency ω_0 in (8), intermodulation distortion results.

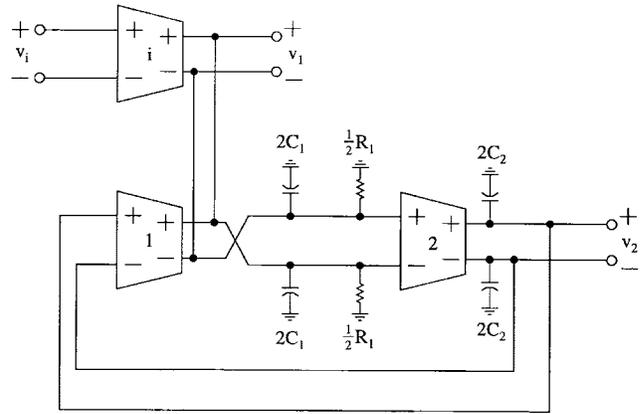


Fig. 2. A bandpass biquadratic G_m - C filter.

IIP_3 is sufficient for characterizing nonlinearity in an amplifier, or in a filter with an input tone that is in-band. However, we are interested in the effect of possibly out-of-band interfering tones on the in-band tone in a filter. Thus, a different approach will be required.

B. Systems with Memory

For the design of highly-selective active filters, we are accustomed to Laplace techniques. These, however, only work for linear systems, and real filters contain weak nonlinearities—for example, transistors have nonlinear i - v characteristics and parasitics that drive the filtering components and are themselves filtered. Systems with memory and weak nonlinearities are amenable to analysis using Volterra series [11], [12]. Confronted with a nonlinear system, perhaps most designers would opt for harmonic balance techniques to aid them. We shall contrast harmonic balance and Volterra series shortly.

A Volterra series is an infinite sum of n -fold convolution integrals [13]

$$\begin{aligned} y(t) &= \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n g_n(t_1, \dots, t_n) \\ &\quad \times \prod_{r=1}^n x(t - t_r) \\ &= \frac{1}{1!} \int_{-\infty}^{\infty} dt_1 g_1(t_1) x(t - t_1) \\ &\quad + \frac{1}{2!} \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 g_2(t_1, t_2) \\ &\quad \times x(t - t_1) x(t - t_2) \\ &\quad + \frac{1}{3!} \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} dt_3 g_3(t_1, t_2, t_3) \\ &\quad \times x(t - t_1) x(t - t_2) x(t - t_3) \\ &\quad + \dots \end{aligned} \quad (9)$$

with the $g_n(t_1, \dots, t_n)$, denoted the *Volterra kernels*, serving to characterize the various orders of nonlinearity. Equation (9) can be transformed into the frequency domain using the

TABLE I
 VOLTERRA COEFFICIENTS FOR BANDPASS BIQUAD

Term	Coefficient for general n
$C_1 \frac{dy}{dt} + \frac{1}{R_1} y + \frac{g_{m1}g_{m2}}{C_2} \int y dt$	$(C_1 \sum j\omega_i + \frac{1}{R_1} + \frac{g_{m1}g_{m2}}{C_2 \sum j\omega_i}) M_n$
$g_{mi} x + \epsilon_i x^3$	$\begin{cases} g_{mi}, & n = 1 \\ 6\epsilon_i, & n = 3 \\ 0, & \text{otherwise} \end{cases}$
$\frac{\epsilon_2 g_{m1}}{C_2^3} [\int y^3 dt]$	$\frac{\epsilon_2 g_{m1}}{C_2 \sum j\omega_i} M_n^{(3)}$
$\frac{\epsilon_1 g_{m2}}{C_2^3} [\int y dt]^3$	$\frac{\epsilon_1 g_{m2}^3}{C_2^3} 3! \sum_{(v;3,n)} \sum'_{N'} \frac{M_{v_1} M_{v_2} M_{v_3}}{\sum_{v_1} j\omega_i \sum_{v_2} j\omega_i \sum_{v_3} j\omega_i}$
$\frac{3g_{m2}^2 \epsilon_1 \epsilon_2}{C_2^3} [\int y dt]^2 [\int y^3 dt]$	$\frac{3g_{m2}^2 \epsilon_1 \epsilon_2}{C_2^3} \sum_{(v;3,n)} \left(\sum'_{N_{12}} \frac{2! M_{v_1} M_{v_2}}{\sum_{v_1} j\omega_i \sum_{v_2} j\omega_i} \right) \left(\frac{M_{v_3}^{(3)}}{\sum_{v_3} j\omega_i} \right)$
$\frac{3g_{m2} \epsilon_1 \epsilon_2^2}{C_2^3} [\int y dt] [\int y^3 dt]^2$	$\frac{3g_{m2} \epsilon_1 \epsilon_2^2}{C_2^3} \sum_{(v;3,n)} \left(\sum'_{N_1} \frac{M_{v_1}}{\sum_{v_1} j\omega_i} \right) \left(\sum'_{N_{23}} \frac{2! M_{v_2}^{(3)} M_{v_3}^{(3)}}{\sum_{v_2} j\omega_i \sum_{v_3} j\omega_i} \right)$
$\frac{\epsilon_1 \epsilon_2^3}{C_2^3} [\int y^3 dt]^3$	$\frac{\epsilon_1 \epsilon_2^3}{C_2^3} \sum_{(v;3,n)} \sum'_{N'} \frac{3! M_{v_1}^{(3)} M_{v_2}^{(3)} M_{v_3}^{(3)}}{\sum_{v_1} j\omega_i \sum_{v_2} j\omega_i \sum_{v_3} j\omega_i}$

n -fold Fourier transform

$$G_n(\omega_1, \dots, \omega_n) = \int_{-\infty}^{\infty} du_1 \dots \int_{-\infty}^{\infty} du_n g_n(t_1, \dots, t_n) \times e^{-j\omega_1 u_1} \dots e^{-j\omega_n u_n} \quad (10)$$

leading to

$$Y(\omega) = \frac{1}{1!} G_1(\omega) X(\omega) + \frac{1}{2!} \int_{-\infty}^{\infty} d\omega_1 G_2(\omega_1, \omega - \omega_1) \times X(\omega_1) X(\omega - \omega_1) + \frac{1}{3!} \int_{-\infty}^{\infty} d\omega_1 d\omega_2 G_3(\omega_1, \omega_2, \omega - \omega_1 - \omega_2) \times X(\omega_1) X(\omega_2) X(\omega - \omega_1 - \omega_2) + \dots \quad (11)$$

where the (complex) $G_n(\omega_1, \dots, \omega_n)$ are denoted the *Volterra transfer functions*. Here it is assumed the $g_n(t_1, \dots, t_n)$ and $G_n(\omega_1, \dots, \omega_n)$ are *symmetric* which means permuting their arguments does not change their values. Unsymmetric kernels can always be made symmetric [12].

When the system equations are known, there exist methods to determine the kernels g_n and transfer functions G_n . These are outlined in Appendix A. Even when the system equations are unknown, we can still determine G_n numerically as explained in Sections V and VI. Once G_n are found, the output spectrum may be expressed in terms of them. For example, for a single input tone $x(t) = V_0 \cos(\omega_0 t)$, $\omega_0 = 2\pi f_0$, the output spectrum is

$$y(t) = \sum_{n=1}^{\infty} \left(\frac{V_0}{2} \right)^n \sum_{k=0}^n \frac{\exp[j(2k-n)\omega_0 t]}{k!(n-k)!} G_{k, n-k}(\omega_0) \quad (12)$$

where $G_{k, n-k}(\omega_0)$ denotes $G_n(\omega_1, \dots, \omega_n)$ with the first k of the ω_i equal to $+\omega_0$ and the remaining $n-k$ equal to $-\omega_0$.

Expanding (12) for a few different values of $(2k-n)$ leads to

$$y(t) = \left[\frac{V_0^2}{2} G_2(\omega_0, -\omega_0) + \dots \right] + e^{j\omega_0 t} \left[\frac{V_0}{2} G_1(\omega_0) + \frac{V_0^3}{16} G_3(\omega_0, \omega_0, -\omega_0) + \dots \right] + e^{j2\omega_0 t} \left[\frac{V_0^2}{8} G_2(\omega_0, \omega_0) + \dots \right] + e^{j3\omega_0 t} \left[\frac{V_0^3}{48} G_3(\omega_0, \omega_0, \omega_0) + \dots \right] + \dots + e^{-j\omega_0 t} \left[\frac{V_0}{2} G_1(-\omega_0) + \frac{V_0^3}{16} G_3(-\omega_0, -\omega_0, \omega_0) + \dots \right] + e^{-j2\omega_0 t} \left[\frac{V_0^2}{8} G_2(-\omega_0, -\omega_0) + \dots \right] + e^{-j3\omega_0 t} \left[\frac{V_0^3}{48} G_3(-\omega_0, -\omega_0, -\omega_0) + \dots \right] + \dots \quad (13)$$

Thus the magnitude and phase at each frequency can be found from an infinite sum of Volterra transfer functions with the proper frequency arguments. Just as with Taylor series, Volterra series work best when the nonlinearities are weak—that is, when the terms inside the square brackets in (13) diminish rapidly with increasing order. Unfortunately, *a priori* analytic determination of “weakness” is likely impractical on a real problem; usually, one must find G_n and evaluate them numerically to determine if they vanish for large n . The formulae become less reliable as nonlinearity gets stronger, though quantifying how much less reliable is difficult. Nonetheless, it is fairly safe to say that Volterra series work anywhere a Taylor series would.

For an input $x(t) = V_0 \cos(\omega_0 t) + V_1 \cos(\omega_1 t)$ with two tones, the term at frequency $N\omega_0 + M\omega_1$, $N, M \geq 0$ is given by

$$e^{j(N\omega_0 + M\omega_1)t} \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \frac{(V_0/2)^{2l+N} (V_1/2)^{2k+M}}{(N+l)! l! (M+k)! k!} \times G_{N+l, l; M+k, k}(\omega_0, \omega_1) \quad (14)$$

where $G_{N+l, l, M+k, k}(\omega_0, \omega_1)$ is $G_n(\omega_1, \dots, \omega_n)$ with

$$N + 2l + M + 2k = n$$

the first $N + l$ of ω_i equal to $+\omega_0$, the next l of ω_i equal to $-\omega_0$, the next $M + k$ of ω_i equal to $+\omega_1$, and the last k of ω_i equal to $-\omega_1$. These output spectra, along with several others for different types of inputs, can be found in [13].¹

C. Distortion in a Band Pass Biquad

Consider the simple G_m - C filter shown in Fig. 2. The active devices are transconductance amplifiers (TA's) that turn the voltage difference at their input terminals into a current at the output, i.e., $i = f(v_+ - v_-)$. The time-domain system equations are

$$\begin{aligned} \frac{dv_1}{dt} &= \frac{1}{C_1} \left[-\frac{v_1}{R_1} + f_i(v_i) + f_1(v_2) \right] \\ \frac{dv_2}{dt} &= \frac{1}{C_2} f_2(-v_1). \end{aligned} \quad (15)$$

If we assume the TA's are linear, then $i = g_m(v_+ - v_-)$ where g_m is the transconductance. Substituting in (15), taking Laplace transforms, and solving gives

$$\frac{V_1}{V_i} = \frac{\frac{g_m i}{C_1} s}{s^2 + \frac{1}{R_1 C_1} s + \frac{g_{m1} g_{m2}}{C_1 C_2}} \quad (16)$$

$$\frac{V_2}{V_i} = \frac{\frac{-g_{m1} g_{m2}}{C_1 C_2}}{s^2 + \frac{1}{R_1 C_1} s + \frac{g_{m1} g_{m2}}{C_1 C_2}} \quad (17)$$

v_1 and v_2 will be recognized as bandpass and low-pass outputs, respectively. The important parameters in (16) are

$$\omega_0 = \sqrt{\frac{g_{m1} g_{m2}}{C_1 C_2}}, \quad A_0 = g_{mi} R_1, \quad Q = \omega_0 C_1 R_1 \quad (18)$$

where ω_0 is the center frequency, A_0 is the gain at ω_0 , and Q is the quality factor.

Let us now assume a weak cubic nonlinearity in the dc transfer curve of the TA's, and that the transfer curve does not change significantly at the frequencies at which the filter operates:

$$i = g_m(v_+ - v_-) + \epsilon(v_+ - v_-)^3. \quad (19)$$

Assuming a weak nonlinearity is justified because, by design, the TA's will be close to linear; furthermore, no second-order term will be present since the circuit is balanced. Lastly, the frequency-independent assumption is reasonable because it is unlikely a practical TA would be operated at frequencies close to the switching frequency of the transistors. Substituting (19)

¹One reviewer quite rightly mourned the complexity of the notation introduced thus far. Regrettably, the authors find that this seems to be the norm: a survey of papers that use Volterra series will reveal many differing notations, all exhibiting considerable complexity.

in (15), eliminating v_2 , and writing $v_i = x$ and $v_1 = y$ leads to

$$\begin{aligned} C_1 \frac{dy}{dt} + \frac{y}{R_1} + \frac{g_{m1} g_{m2}}{C_2} \int y dt \\ = g_{mi} x + \epsilon_i x^3 - \frac{\epsilon_2 g_{m1}}{C_2} \left[\int y^3 dt \right] \\ - \frac{\epsilon_1 g_{m2}^3}{C_2^3} \left[\int y dt \right]^3 - \frac{\epsilon_1 \epsilon_2^3}{C_2^3} \left[\int y^3 dt \right]^3 \\ - \frac{3\epsilon_1 \epsilon_2 g_{m2}^2}{C_2^2} \left[\int y dt \right]^2 \left[\int y^3 dt \right] \\ - \frac{3\epsilon_1 \epsilon_2^2 g_{m2}}{C_2^2} \left[\int y dt \right] \left[\int y^3 dt \right]^2. \end{aligned} \quad (20)$$

The equation is rather complicated, but the Volterra transfer functions from x to y [which we shall label $M_n(\omega_1, \dots, \omega_n)$] can still be derived. The calculation for general n is shown in Appendix B.

Using the results summarized in Table I, we could write explicit formulae for all M_n ; let us do so for M_1 and M_3 only. For $n = 1$, only the first two rows of the table are nonzero:

$$\left[C_1 j\omega_1 + \frac{1}{R_1} + \frac{g_{m1} g_{m2}}{C_2 j\omega_1} \right] M_1(\omega_1) = g_{mi}. \quad (21)$$

Solving for M_1 , identifying $j\omega_1$ with s , and rearranging gives

$$M_1(\omega_1) = \frac{\frac{g_{mi}}{C_1} s}{s^2 + \frac{1}{R_1 C_1} s + \frac{g_{m1} g_{m2}}{C_1 C_2}} \quad (22)$$

which is the same as (16). This is the expected result: the Volterra transfer function M_1 is the same as the linear transfer function obtained by standard techniques. For $n = 2$, the coefficients of the right-hand side (RHS) terms of (20) are zero; a simple proof by induction shows $M_n = 0$ for all even n . For $n = 3$, the first four rows of the table are nonzero:

$$\begin{aligned} \left[C_1 \sum_3 j\omega_i + \frac{1}{R_1} + \frac{g_{m1} g_{m2}}{C_2 \sum_3 j\omega_i} \right] M_3(\omega_1, \omega_2, \omega_3) \\ = 6\epsilon_i - \frac{\epsilon_2 g_{m1}}{C_2 \sum_3 j\omega_i} M_3^{(3)}(\omega_1, \omega_2, \omega_3) \\ - \frac{\epsilon_1 g_{m2}^3}{C_2^3} \sum_N' \frac{3! M_1(\omega_1) M_1(\omega_2) M_1(\omega_3)}{(j\omega_1)(j\omega_2)(j\omega_3)}. \end{aligned} \quad (23)$$

Expanding the second term on the RHS and noting $N = 1$ in the third term leads to the results shown in (24), shown at the bottom of the next page. It can be seen M_3 depends on M_1 . M_5 , were we to write it out, would depend on both M_3 and M_1 ; M_7 on M_5 , M_3 , and M_1 , and so on. The rapid increase of complexity in writing analytic formulas for M_n as n increases suggests numerical computer evaluation is the best approach; the recursive nature of M_n means such an evaluation is not difficult to program. For weak-enough nonlinearity, an assumption we shall make, the terms above M_3 contribute negligibly to the final result.

Why go to the trouble of calculating the Volterra transfer functions? Suppose the input is $x(t) = \sum_{i=0}^2 V_i \cos(\omega_i t)$ where ω_0 is the desired tone and $\omega_1 = \omega_0 - \Delta\omega$ and $\omega_2 = \omega_0 - 2\Delta\omega$ are interferers. Then, from (14), the following output terms will appear at frequency $\exp(j\omega_0 t)$:

$$\begin{aligned}
 \text{Linear gain:} & \quad \frac{V_0}{2} M_1(\omega_0) \\
 \text{Compression:} & \quad \frac{V_0^3}{16} M_3(\omega_0, \omega_0, -\omega_0) \\
 \text{Desensitization:} & \quad \frac{V_0 V_1^2}{8} M_3(\omega_0, \omega_1, -\omega_1), \\
 & \quad \frac{V_0 V_2^2}{8} M_3(\omega_0, \omega_2, -\omega_2) \\
 \text{Intermodulation:} & \quad \frac{V_1^2 V_2}{16} M_3(\omega_1, \omega_1, -\omega_2). \quad (25)
 \end{aligned}$$

The Volterra transfer functions are useful because they enable us to write analytic expressions for the linear gain and different types of distortion at the output. Their true power is illustrated in the two examples below.

Suppose we wish to know how changing the Q of the filter affects the distortion. With ω_0 and A_0 arbitrarily fixed at 100 MHz and 20 dB, Fig. 3(a) plots M_1 against frequency for three different values of Q . Thanks to (25), it is now trivial to plot Fig. 3(b), which shows ω_1 desensitization and (ω_1, ω_2) intermodulation versus Q . From (18), Q is increased by raising R_1 while lowering g_{mi} simultaneously to keep A_0 constant. We have assumed a -70 dBm desired tone at ω_0 and two -50 dBm interferers at $\omega_1 = 0.98\omega_0$ and $\omega_2 = 0.96\omega_0$ and $\text{IP}_3 = 0$ dBm for each TA. In this paper, units of dBm are specified assuming a $50\text{-}\Omega$ resistance.

The exact shape of the curves is rather complicated to explain, and perhaps somewhat counterintuitive. However, it helps to note that for fixed $\omega_1 \neq \omega_0$, M_1 in (22) is constant for low Q but starts falling at 6 dB/oct as Q gets large enough. The denominator of M_3 in (24) can be shown to be inversely proportional to Q ; meanwhile, the numerator has the product of three M_1 terms in it. For low Q , therefore, M_3 rises because the M_1 in its numerator are constant and its denominator falls with Q . Then, as Q increases, the M_1 in its numerator start to fall, which makes M_3 decrease. Finally, M_1 can be seen to flatten out for very large Q . Essentially, we may think of the interfering tones as moving “out of band” as Q increases since the filter gain at the frequency of the interfering tones drops with Q if A_0 is held constant.

The shape of the intermodulation curve suggests that for a given set of conditions, there is an optimum Q at which to operate (which, in this case, is $Q \approx 1000$). However, other

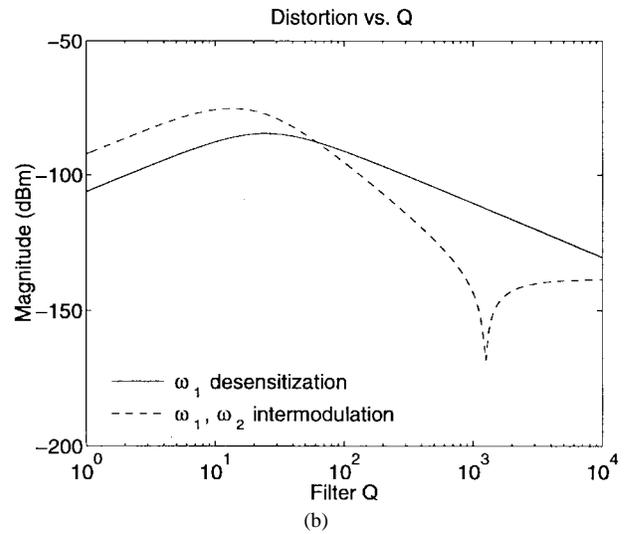
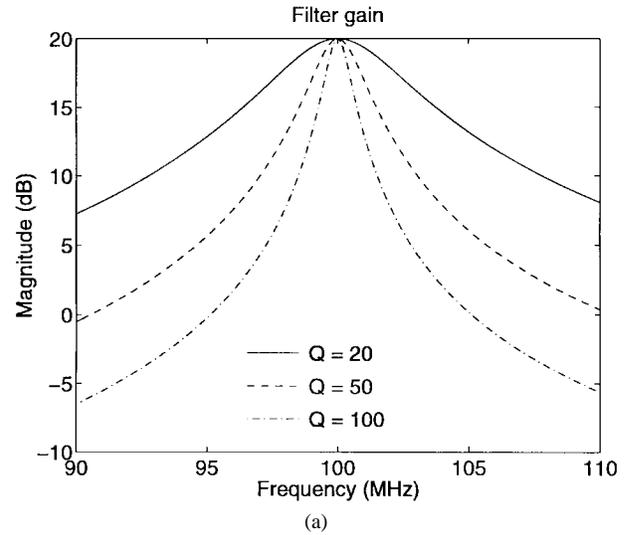


Fig. 3. (a) Effect of Q on M_1 and (b) distortion as a function of Q .

constraints (such as the finite bandwidth of the actual radio input, or limited accuracy in a Q -tuning circuit) may make this point unreachable.

Lastly, we can easily show how interferer proximity affects distortion at the desired frequency. Define the *channel separation* as $1 - \Delta\omega/\omega_0$ where $\Delta\omega$ is the frequency separation between the desired signal and the closest interferer. Fig. 4 plots distortion against channel separation for the same conditions as in the previous example and a filter Q of 100. Not surprisingly, moving the interferers closer worsens distortion; using Volterra transfer functions enables us to quantify the worsening.

$$M_3(\omega_1, \omega_2, \omega_3) = \frac{6\epsilon_i - 6M_1(\omega_1)M_1(\omega_2)M_1(\omega_3) \left[\frac{\epsilon_2 g_{m1}}{C_2 \sum_3 j\omega_i} + \frac{\epsilon_1 g_{m2}^3}{C_2^3 \prod_3 j\omega_i} \right]}{C_1 \sum_3 j\omega_i + \frac{1}{R_1} + \frac{g_{m1}g_{m2}}{C_2 \sum_3 j\omega_i}} \quad (24)$$

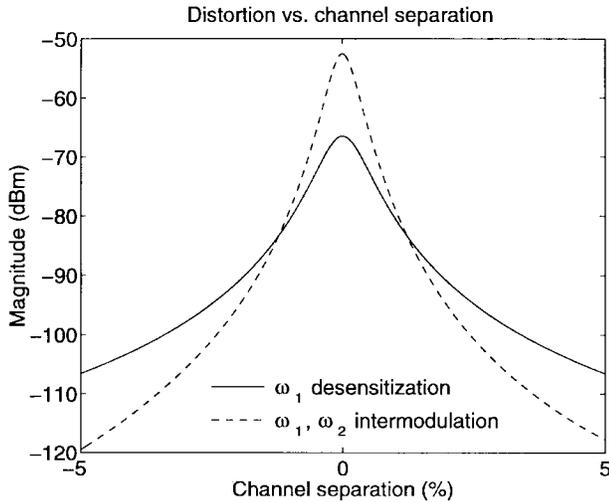


Fig. 4. Distortion as a function of interferer proximity to desired signal.

As noted, Volterra series' main advantage is that they give approximate *analytical* design formulae. Harmonic balance techniques [18], by contrast, are numerical. However, their principal virtues are that they work even when nonlinearities *are not* weak, and that the number of harmonics can be prescribed.

IV. LOWERING DISTORTION USING A BANK OF FILTERS

The curves of Fig. 3(b) suggest that it is desirable to notch out the interferers as soon as possible so they cannot excite M_3 . A typical biquad notch filter does not do this, since the notch is only for one amplifier and the others are still driven by interferers. Cascade and ladder structures will exhibit similar problems, at least up to the stage that includes the notch for a given interferer.

The filter bank in Fig. 5 [14], though, has the property we want: It consists of a number of infinite- Q resonators, each tuned to a different frequency, with feedback. The linear transfer function from $X(s)$ to $E(s)$ has notches at the resonator frequencies because $E(s)$ is followed by an infinite-gain block at those frequencies. Straight linear analysis of the filter bank shows that each resonator has an output

$$Y_i(s) = \frac{M_{i1}(s)}{1 + \Phi(s) \sum M_{i1}(s)} X(s). \quad (26)$$

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

$$E(s) = \frac{1}{1 + \Phi(s) \sum M_{i1}(s)} X(s) \quad (27)$$

and an M_{i1} term goes to infinity at each frequency. Hence, the relevant third-order distortion term should be nulled.

To go about verifying this, the Volterra transfer functions for the filter bank are derived in Appendix C. Let us compare distortion in the filter bank and two competing biquads whose linear transfer functions are depicted in Fig. 6. The low- Q biquad has similar pass band performance to the filter bank while the high- Q biquad has similar stop band performance.

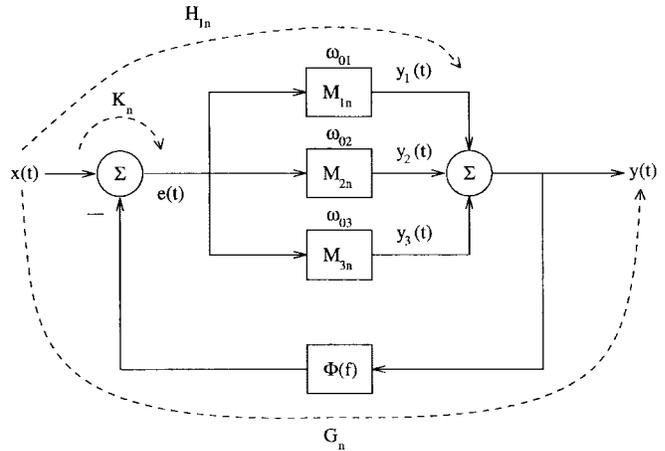


Fig. 5. A filter bank with desirable distortion properties.

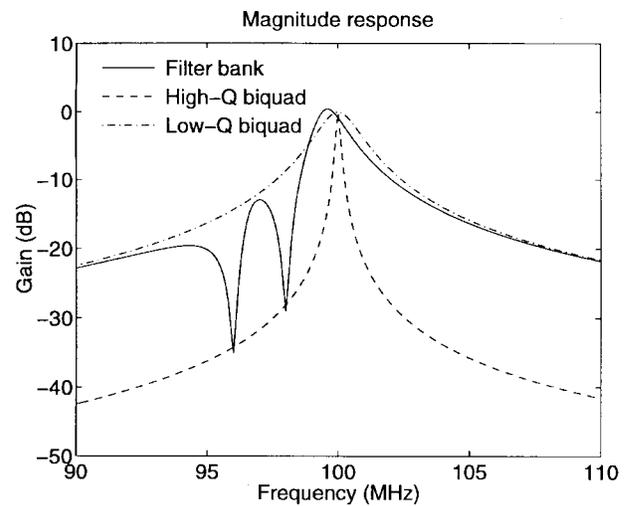


Fig. 6. Transfer functions for competing architectures.

TABLE II
FILTER BANK DISTORTION IMPROVEMENT OVER SINGLE BIQUADS

Third-order term	Filter bank win over low- Q biquad (dB)	Filter bank win over high- Q biquad (dB)
Compression	99.7	119.7
ω_1 desensitization	74.7	56.9
ω_2 desensitization	9.1	-3.8
Intermodulation	25.8	17.3

The notches in the filter bank have finite depth because in a real circuit $Q \neq \infty$; this places a practical limit on the distortion improvement we can expect. The interferers in Fig. 6 are assumed to fall in the notches of the filter bank, at $\omega_1 = 0.98\omega_0$ and $\omega_2 = 0.96\omega_0$, and the linear gain of all architectures is $A_0 = 0$ dB at ω_0 .

Table II shows the improvement in distortion offered by the filter bank. These numbers are independent of the assumed value of IIP₃ in the TA's. The filter bank is at least as good as either biquad, and most importantly, it performs better in the crucial area of intermodulation distortion where, as stated earlier, the presence of large adjacent-channel interferers can severely tax the ability of a radio receiver to demodulate a weak desired signal.

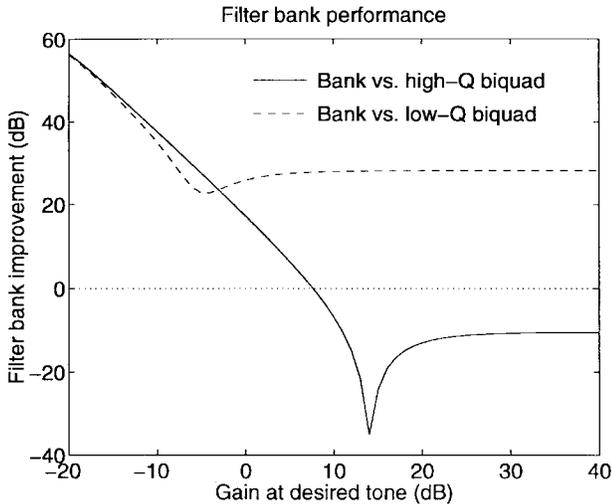
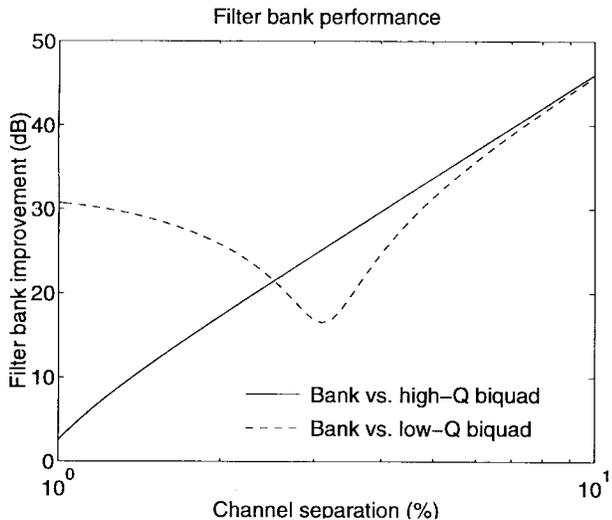

 Fig. 7. Intermodulation distortion improvement as a function of A_0 .


Fig. 8. Intermodulation distortion improvement as a function of interferer proximity.

Two important parameters that affect the performance of the filter bank are the gain at the desired signal frequency, A_0 , and the proximity of interferers to the desired signal. Fig. 7 plots just the filter bank intermodulation distortion improvement as a function of A_0 ; essentially, the x -axis in Fig. 7 measures the vertical shift of the curves in Fig. 6. We see that more vertical shift means less distortion improvement. Lastly, Fig. 8 plots intermodulation distortion performance against channel separation as defined in Section III-C. As the interferers move closer to the desired tone, the filter bank shows a smaller improvement over the single biquads. The low- Q biquad curve is nonmonotonic because of a complicated interaction between the phases of the complex terms that make up H_{13} in (71).

In order for the filter bank to work well, the notches must be deep; deep notches require ever higher stable Q as they move closer together. Because of this, a practical limit on the channel separation is on the order of 0.5%. In terms of applications of the filter bank, consider GSM, that has a channel spacing of 400 kHz. The filter bank would be impractical at the RF of 900 MHz because interferers are only 0.05% away from the desired

signal; at a second IF of 70 MHz, however, the filter bank becomes more reasonable because the interferers become 0.6% away. Similarly, in DECT, which has a channel spacing of 1.7 MHz, the interferers are 0.1% away at the RF of 1.9 GHz, but 0.9% away at an IF of 200 MHz. Thus, filter bank seems suited more to IF rather than RF filtering. Although this paper assumes signals that are single tones, the filter bank should work with modulated signals as long as they are relatively narrow band.

V. MEASURING DISTORTION FROM SIMULATIONS

Usually, one would think to use a simulator that performs harmonic balance to extract distortion terms, both because working directly in the frequency domain is faster than working in the time domain, and because it will work even in circuits with strong nonlinearity. The technique illustrated in this section still merits attention because it can be adapted to work on real circuits in the laboratory. We derive formulae and illustrate their application to SPICE transient simulation here, and carry this over to real life in Section VI.

A. Formula Derivation

Until now, we have supposed the system equations are known, and that therefore the M_n can be derived and the distortion components distinguished as in (25). We shall now show that it is possible to distinguish overlapping distortion components from simulation and measurement, that is, even when the system equations are unknown or are too complex to apply analytical Volterra series analysis.

Consider the case where there is a single tone at ω_0 of amplitude V_{01} into a weakly nonlinear system. From (13), the output tone at ω_0 is

$$e^{j\omega_0 t} \left[\frac{V_{01}}{2} M_1(\omega_0) + \frac{V_{01}^3}{16} M_3(\omega_0, \omega_0, -\omega_0) \right] + e^{-j\omega_0 t} \left[\frac{V_{01}}{2} M_1(-\omega_0) + \frac{V_{01}^3}{16} M_3(-\omega_0, -\omega_0, \omega_0) \right] \quad (28)$$

assuming higher order terms are negligible. Define $M_1(\omega_0) = a + jb$ and $M_3(\omega_0, \omega_0, -\omega_0) = c + jd$ where (a, b, c, d) are the unknowns we wish to find. From (10) it follows that the terms inside the second set of square brackets in (28) are the complex conjugates of the terms inside the first set. We may simplify (28) as

$$e^{j\omega_0 t} \left[\left(\frac{aV_{01}}{2} + \frac{cV_{01}^3}{16} \right) + j \left(\frac{bV_{01}}{2} + \frac{dV_{01}^3}{16} \right) \right] + e^{-j\omega_0 t} \left[\left(\frac{aV_{01}}{2} + \frac{cV_{01}^3}{16} \right) - j \left(\frac{bV_{01}}{2} + \frac{dV_{01}^3}{16} \right) \right] = 2 \left(\frac{aV_{01}}{2} + \frac{cV_{01}^3}{16} \right) \cos(\omega_0 t) - 2 \left(\frac{bV_{01}}{2} + \frac{dV_{01}^3}{16} \right) \sin(\omega_0 t) = \left(aV_{01} + \frac{cV_{01}^3}{8} \right) \angle 0^\circ - \left(bV_{01} + \frac{dV_{01}^3}{8} \right) \angle -90^\circ = \left(aV_{01} + \frac{cV_{01}^3}{8} \right) + j \left(bV_{01} + \frac{dV_{01}^3}{8} \right) \quad (29)$$

where we have suppressed the time variable t by thinking in terms of phasors. Define the last line of (29) as

$$y_{R1} + y_{I1} = \left(aV_{01} + \frac{cV_{01}^3}{8} \right) + j \left(bV_{01} + \frac{dV_{01}^3}{8} \right). \quad (30)$$

If the input tone at ω_0 has amplitude V_{02} instead, then the output tone at ω_0 will be

$$y_{R2} + y_{I2} = \left(aV_{02} + \frac{cV_{02}^3}{8} \right) + j \left(bV_{02} + \frac{dV_{02}^3}{8} \right) \quad (31)$$

(30) and (31) are linear equations. Separating real and imaginary parts and solving yields

$$\begin{aligned} M_1(\omega_0) &= a + jb \\ &= \frac{V_{01}^3 y_{R2} - V_{02}^3 y_{R1}}{V_{01}^3 V_{02} - V_{01} V_{02}^3} + j \frac{V_{01}^3 y_{I2} - V_{02}^3 y_{I1}}{V_{01}^3 V_{02} - V_{01} V_{02}^3} \end{aligned} \quad (32)$$

$$\begin{aligned} M_3(\omega_0, \omega_0, -\omega_0) &= c + jd \\ &= 8 \frac{V_{01} y_{R2} - V_{02} y_{R1}}{V_{01} V_{02}^3 - V_{01}^3 V_{02}} \\ &\quad + j 8 \frac{V_{01} y_{I2} - V_{02} y_{I1}}{V_{01} V_{02}^3 - V_{01}^3 V_{02}}. \end{aligned} \quad (33)$$

This suggests the following method to extract the linear gain and compression terms at ω_0 from a circuit simulation: simulate the circuit twice, once with input amplitude V_{01} and once with V_{02} , and take the fast Fourier transform (FFT) of each transient output to determine (y_{R1}, y_{I1}) and (y_{R2}, y_{I2}) . Then, substitute into (32) and (33) to find $M_1(\omega_0)$ and $M_3(\omega_0, \omega_0, -\omega_0)$. Care must be taken to ensure:

- input amplitudes are not so small that the compression component is buried in the FFT noise floor;
- input amplitudes are not so large that M_5 contributes significantly to the output at ω_0 ;
- simulation is long enough so that transients die out and the circuit is in steady state.

A similar procedure can be used to extract a desensitization term. To find $M_3(\omega_0, \omega_1, -\omega_1) = e + jf$, apply a two-tone input (amplitude V_{01} at frequency ω_0 and amplitude V_{11} at frequency ω_1) and measure the output at ω_0 , which from (14) and (25) is

$$\begin{aligned} e^{j\omega_0 t} &\left[\frac{V_{01}}{2} M_1(\omega_0) + \frac{V_{01}^3}{16} M_3(\omega_0, \omega_0, -\omega_0) \right. \\ &\quad \left. + \frac{V_{01} V_{11}^2}{8} M_3(\omega_0, \omega_1, -\omega_1) \right] \\ &+ e^{-j\omega_0 t} \left[\frac{V_{01}}{2} M_1(-\omega_0) + \frac{V_{01}^3}{16} M_3(-\omega_0, -\omega_0, \omega_0) \right. \\ &\quad \left. + \frac{V_{01} V_{11}^2}{8} M_3(-\omega_0, -\omega_1, \omega_1) \right]. \end{aligned} \quad (34)$$

Applying a time-suppression simplification as we did to (29) means we can write the tone at ω_0 as

$$\begin{aligned} y_{R1} + jy_{I1} &= \left(aV_{01} + c \frac{V_{01}^3}{8} + e \frac{V_{01} V_{11}^2}{4} \right) \\ &\quad + j \left(bV_{01} + d \frac{V_{01}^3}{8} + e \frac{V_{01} V_{11}^2}{4} \right). \end{aligned} \quad (35)$$

Changing the amplitude of the tone at ω_1 to V_{12} and remeasuring the output tone at ω_0 gives

$$\begin{aligned} y_{R2} + jy_{I2} &= \left(aV_{01} + c \frac{V_{01}^3}{8} + e \frac{V_{01} V_{12}^2}{4} \right) \\ &\quad + j \left(bV_{01} + d \frac{V_{01}^3}{8} + e \frac{V_{01} V_{12}^2}{4} \right). \end{aligned} \quad (36)$$

Solving (35) and (36) gives us

$$\begin{aligned} M_3(\omega_0, \omega_1, -\omega_1) &= e + jf \\ &= \frac{4}{V_{01}} \frac{y_{R2} - y_{R1}}{V_{12}^2 - V_{11}^2} \\ &\quad + j \frac{4}{V_{01}} \frac{y_{I2} - y_{I1}}{V_{12}^2 - V_{11}^2}. \end{aligned} \quad (37)$$

Thus, a desensitization term can be measured with another pair of simulations, this time altering the amplitude of the *interferer*. An intermodulation term $M_3(\omega_1, \omega_1, -\omega_2) = g + jh$ can be measured with a single simulation: apply two input tones at ω_1 , amplitude V_1 , and ω_2 , amplitude V_2 , and measure the output at ω_0 from (25)

$$\begin{aligned} y_R + jy_I &= e^{j\omega_0 t} \left[\frac{V_1^2 V_2}{16} M_3(\omega_1, \omega_1, -\omega_2) \right] \\ &\quad + e^{-j\omega_0 t} \left[\frac{V_1^2 V_2}{16} M_3(-\omega_1, -\omega_1, \omega_2) \right] \\ &= g \frac{V_1^2 V_2}{8} + jh \frac{V_1^2 V_2}{8}. \end{aligned} \quad (38)$$

Solving for (g, h) gives

$$M_3(\omega_1, \omega_1, -\omega_2) = g + jh = 8 \frac{y_R}{V_1^2 V_2} + j 8 \frac{y_I}{V_1^2 V_2}. \quad (39)$$

In summary, given three frequencies $(\omega_0, \omega_1, \omega_2)$, it takes a total of seven simulations to extract numerical values for the five terms listed in (25). The requisite formulas are (32), (33), (37), and (39).

B. Example Applications

To test the results from the previous subsection, the filter in Fig. 2 and the filter bank in Fig. 5 were implemented mathematically in a C program that performed Runge–Kutta numerical integration [15]. The two single-biquads and filter bank from Section IV were compared, and the values in Table II were verified as correct.

A more realistic and challenging test was performed using SPICE on the circuit of Fig. 9. It is based on [16] and is an integrated filter with positive feedback to enhance the low Q of the on-chip inductors. The two tank circuits have slightly different resonant frequencies; the overall filter center frequency can be altered by tuning the frequency control to adjust the relative current in each tank. As in Section IV, two single-filters were compared with a bank of three filters in the feedback configuration of Fig. 5. Fig. 10 shows the SPICE ac analysis of the three architectures; once again, the single filters have similar pass-band and stop-band performances to the bank of filters. This time, however, $A_0 = 40$ dB and the

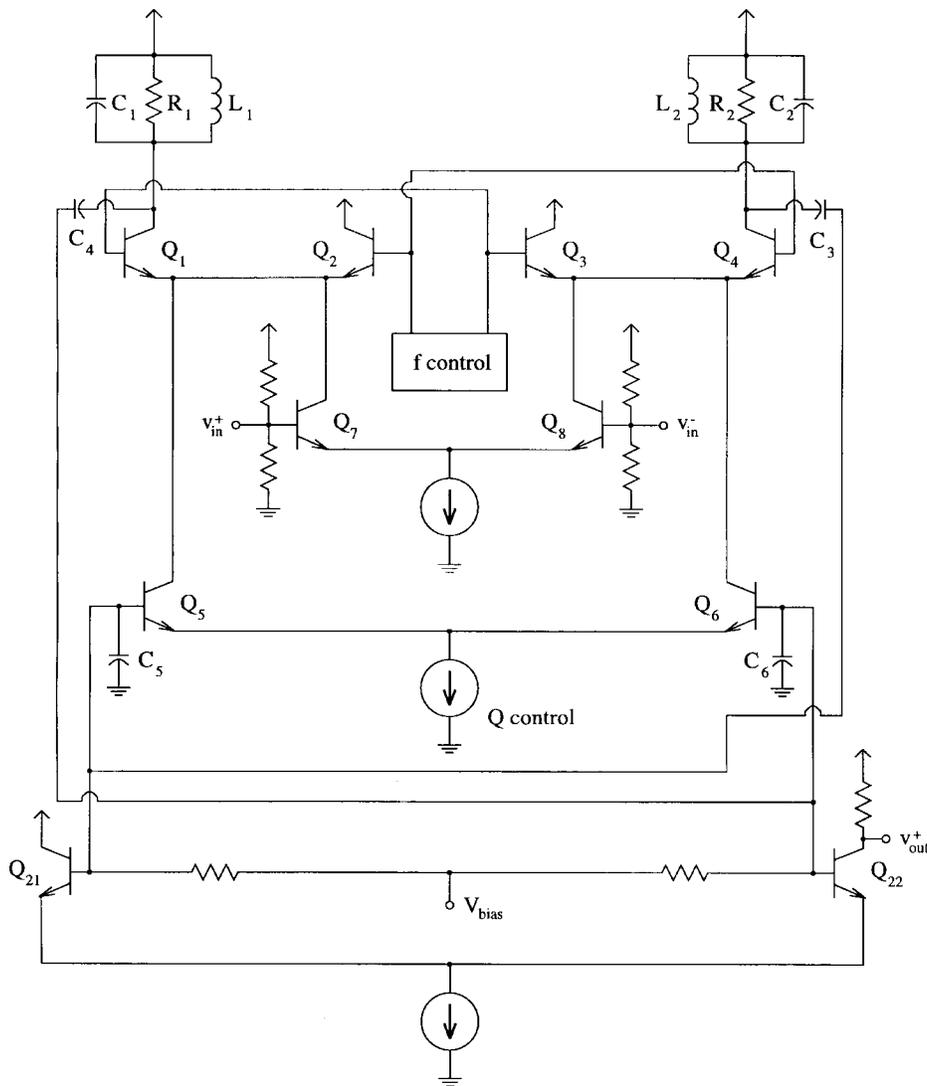


Fig. 9. Realistic circuit for testing extraction method.

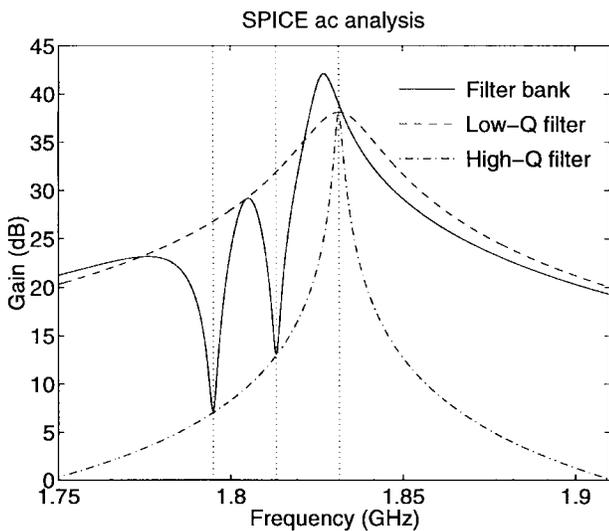


Fig. 10. SPICE ac analyses of the three realistic filter configurations.

interferers are only 1% away from the desired signal at 1.83 GHz.

To apply the extraction method, the following procedure is found to be effective.

- 1) Use $V_{02} = 2 \times V_{01}$ for the two input amplitudes.
- 2) Measure M_1 and M_3 . From (30), calculate the relative sizes of the linear and compression terms at the larger input amplitude as $y_L = |M_1|V_{02}$ and $y_C = |M_3|V_{02}^3/8$.
- 3) It is found that $0.10 < y_C/y_L < 0.20$ means (V_{01}, V_{02}) are suitable input levels. That is, the third-order term should be between 10 and 20% as large as the linear term to ensure M_3 is large enough but M_5 is negligible.

Furthermore, several pairs of simulations with levels close to the appropriate ones should be undertaken to verify that the extraction formulae are giving consistent values for the distortion terms. Taking account of all this, the extraction results are summarized in Table III. The intermodulation improvements are seen to be less than those in Table II, but Fig. 7 suggests that this is due to the inherently high value of A_0 (40 dB) in this filter. Still, the results are useful because intermodulation distortion is improved somewhat by the filter bank and the extraction method was applied successfully to a circuit in SPICE.

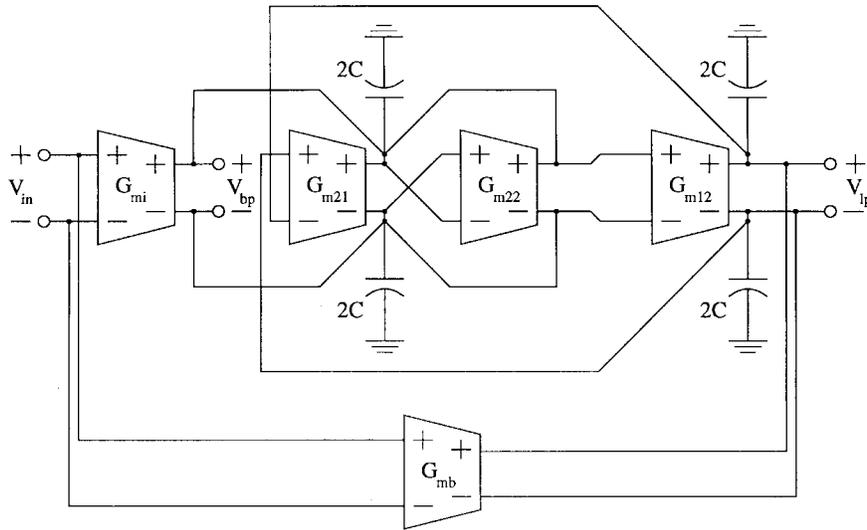


Fig. 11. Filter circuit for laboratory tests.

TABLE III
REALISTIC FILTER BANK DISTORTION IMPROVEMENT

Third-order term	Filter bank win over low-Q filter (dB)	Filter bank win over high-Q filter (dB)
Compression	-5.6	47.4
ω_1 desensitization	15.0	34.0
ω_2 desensitization	16.5	32.9
Intermodulation	4.6	5.7

VI. EXPERIMENTAL RESULTS

By and large, the literature on measuring Volterra transfer functions does not deal with terms that overlap in frequency [17]–[21]. Some papers [9], [22] use tones that are close together but not overlapping so that their products may still be distinguished on a spectrum analyzer. Indeed, such an approach would work in theory here—instead of [30] that has separate real and imaginary parts, on a spectrum analyzer we would be able to obtain an equation containing magnitude information only such as

$$\sqrt{y_{R1}^2 + y_{I1}^2} = \sqrt{\left(aV_{01} + \frac{cV_{01}^3}{8}\right)^2 + \left(bV_{01} + \frac{dV_{01}^3}{8}\right)^2}. \quad (40)$$

We could write four equations in the four unknowns (a, b, c, d) and solve them. But the equations are *nonlinear* rather than linear like (30) and (31) and therefore much more difficult to solve. Consequently, it is easier to use a *network analyzer* to obtain both magnitude *and* phase information so the extraction method of Section V can be applied directly.

The filter used in the lab is shown in Fig. 11. This G_m - C filter [23] was manufactured in a $0.8 \mu\text{m}$ BiCMOS process and has TA's tunable over about a decade. Both low-pass and band-pass differential outputs are available; the linear band-pass transfer function is

$$V_{BP}(s) = \frac{\frac{G_{mi}}{C} s - \frac{G_{m21}G_{mb}}{C^2}}{s^2 + \frac{G_{m22}}{C} s + \frac{G_{m12}G_{m21}}{C^2}}. \quad (41)$$

The center frequency ω_0 can thus be tuned by varying G_{m12} and G_{m21} , filter Q with G_{m22} , and center frequency gain A_0 with G_{mi} . The bipolar and MOS transistors in the TA's cause the i - v characteristics of the TA's to deviate from linear as the power supply rails are approached, whence the major source of the weak nonlinearity.

The packaged chip was mounted on a circuit board and connected to complementary power supplies. The effects of nonlinearity can be seen in Fig. 12, where the network analyzer sweeps the frequency close to band center with a fixed amplitude tone. As the amplitude is increased, the gain at $\omega_0 = 99 \text{ MHz}$ decreases, illustrating gain compression; at lower frequencies, the gain increases, illustrating gain expansion. Next, the technique of Section V was applied. The network analyzer was made to emit a tone fixed at frequency ω , and a program controlled its amplitude to calculate $M_1(\omega)$ and $M_3(\omega, \omega, -\omega)$ using (32) and (33). ω was then swept over frequencies near ω_0 . Several problems had to be overcome for the measurement to work.

- 1) Measurement noise meant (y_R, y_I) values changed from one measurement to the next; averaging helped alleviate this.
- 2) The filter ω_0 , A_0 , and Q tended to drift slowly over time, so automating measurements with a program helped to speed them.
- 3) Component insertion loss had to be known so that the input signal amplitude at the filter input was known.
- 4) Suitable signal levels had to be used so that M_3 was not lost in noise, but M_5 was not too large. As suggested in Section V-B, input levels that made the third-order term 15% of the linear term were deemed suitable.

The results are plotted in Fig. 13. M_1 is the linear gain, so the agreement between M_1 in Fig. 13 and the solid line in Fig. 12 is expected. The phase of M_1 is not centered at 0° because of phase shifts caused by the connecting cables. The M_3 magnitude graph plots $20 \log_{10} |M_3|$ versus ω ; we find that (24) predicts *both* that M_3 should peak where M_1 does *and* the ratio of the slopes of the magnitude graphs

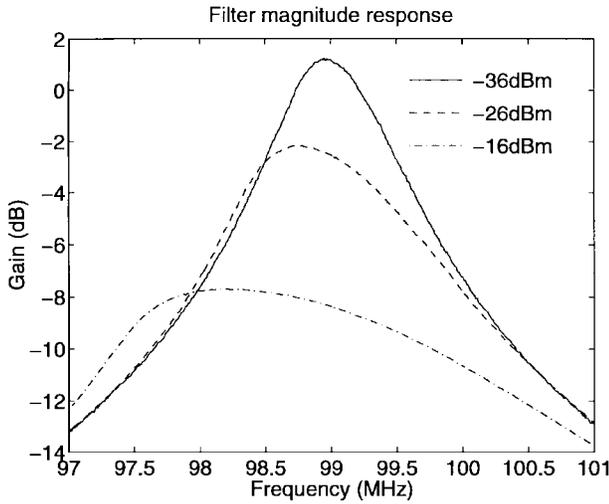


Fig. 12. Compression in measured linear gain characteristics.

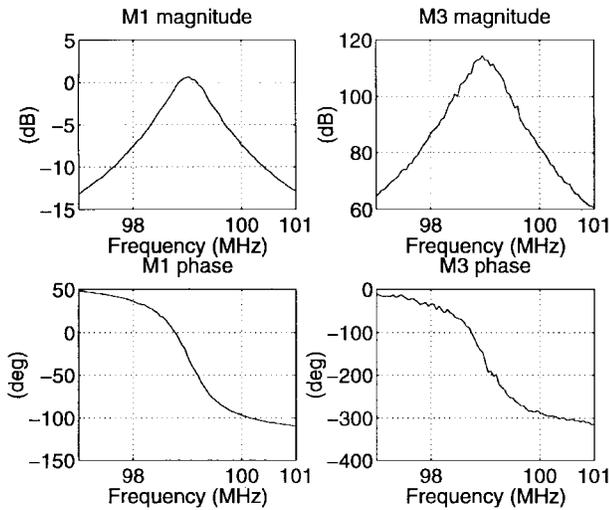
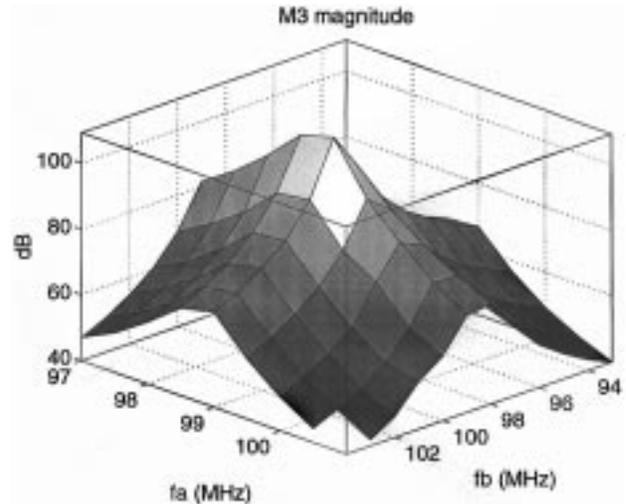


Fig. 13. Measured linear and compression terms.

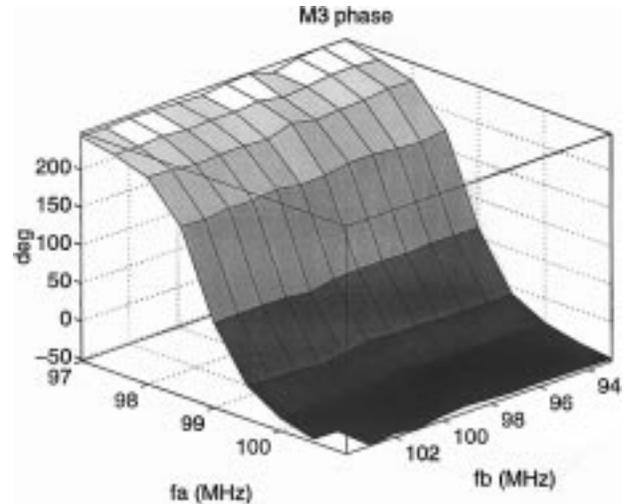
should be $M_3 : M_1 = 4 : 1$, which is what the graphs show. In other words, the nonlinearities in the filter of Fig. 11 are well-modeled as weak cubics.

Moreover, Fig. 13 offers further insight into Fig. 12. From (28), when the M_1 and M_3 vectors are in the *same* direction, they will add constructively to produce a larger vector, i.e., gain *expansion* occurs. When M_1 and M_3 are in *opposite* directions, they result in gain *compression*. In Fig. 13 we have that $|\angle M_3 - \angle M_1| < 90^\circ$ for $\omega < 98.5$ MHz, and gain expansion can be seen in Fig. 12 for $\omega < 98.5$ MHz. For $\omega > 98.5$ MHz we see $|\angle M_3 - \angle M_1| > 90^\circ$ and hence gain compression in Fig. 12.

As a final test of the results, desensitization measurements were performed by adding a signal generator to the test setup to provide an interfering tone. A second program swept both frequencies, chose suitable input amplitudes, and used (37) to calculate the desensitization term $M_3(\omega_a, \omega_b, -\omega_b)$ at frequencies near ω_0 . The results are shown in Fig. 14. These surfaces, too, are consistent with those for the simple band pass biquad with weak cubic nonlinearities.



(a)



(b)

Fig. 14. (a) Measured magnitude and (b) phase of desensitization term.

The test setup was not easily modified to measure a bank of filters like Fig. 5 to verify the results from Section IV. Nonetheless, the measured results presented here are useful because they confirm the formulae in Section V, and to the authors' knowledge, this is the first time Volterra series *phase* measurements have been published in this TRANSACTIONS.

VII. CONCLUSIONS

In this paper we have talked about both the *characterization* of distortion in weakly nonlinear filters, using Volterra series, and the *reduction* of distortion, using the filter bank, and concluded that the filter bank is best used in IF stages. We have also shown that understanding Volterra series eases the extraction of distortion tones, both in theory and in practice.

APPENDIX A

DERIVATION OF VOLTERRA TRANSFER FUNCTIONS

When the system equations are known, the *harmonic input method* may be used to find the Volterra transfer functions

G_n and the *direct expansion method* may be used to find the Volterra kernels g_n [13]. The former is more useful here, and it is described below. The description here is very similar to the one in [13], but the authors feel its inclusion is necessary to aid in the understanding of other work presented here. If the amount of detail here seems insufficient, more may be found by referring to that paper.

It can be proven that in order to find $G_n(\omega_1, \dots, \omega_n)$, the input should be set to a sum of n sinusoids incommensurate in frequency

$$x(t) = \sum_{i=1}^n \exp(j\omega_i t), \quad \omega_i \text{ incommensurate.} \quad (42)$$

This is substituted in the system equation, and then $G_n(\omega_1, \dots, \omega_n)$ is the coefficient of $\exp[j(\omega_1 + \dots + \omega_n)t]$ in the resulting expression for $y(t)$. The proof involves straightforward manipulation of (9). Often, the system equation cannot be explicitly solved for $y(t)$, instead having powers of the output such as $[y(t)]^l$, l a positive integer [see, for example, (20)]. In such cases, it can be shown with a more involved proof that the Volterra transfer function is $G_n^{(l)}(\omega_1, \dots, \omega_n)$, where [13]

$$\begin{aligned} G_n^{(l)}(\omega_1, \dots, \omega_n) = & l! \sum_{(v;l,n)} \sum_N' G_{v_1}(\omega_1, \dots, \omega_{v_1}) \\ & \times G_{v_2}(\omega_{v_1+1}, \dots, \omega_{v_1+v_2}) \times \dots \\ & \times G_{v_l}(\omega_\mu, \dots, \omega_n). \end{aligned} \quad (43)$$

An explanation of the notation in (43) is required.

The $(v;l,n)$ under the first summation sign means "the sum over all partitions of n into l parts," which means the sum over all sets $\{v_i\}_{i=1}^l$ such that

$$v_1 + \dots + v_l = n, \quad 1 \leq v_1 \leq v_2 \leq \dots \leq v_l. \quad (44)$$

If $l > n$, then $G_n^{(l)} = 0$. The second summation \sum_N' extends over the *nonidentical products* of the G_{v_i} . $G_{v_{i1}}$ and $G_{v_{i2}}$ are identical if $v_{i1} = v_{i2}$ and their arguments are simply permutations of one another, such as $G_2(\omega_1, \omega_2)$ and $G_2(\omega_2, \omega_1)$. A simple combinatorial argument gives the number of nonidentical products N as

$$N = \frac{n!}{v_1! \dots v_n! r_1! \dots r_k!} \quad (45)$$

where r_1 is the number of equal v_i in the first run of inequalities in $v_1 \leq v_2 \leq \dots \leq v_l$ from (44), r_2 is the number in the second run, and so on. $r_j = 1$ if a v_i is not equal to any others. The μ inside the G_{v_i} term in (43) is defined as

$$\mu = v_1 + \dots + v_{l-1} + 1 = n - v_l + 1. \quad (46)$$

This is more easily understood with some examples. What is $G_2^{(2)}(\omega_1, \omega_2)$? Here, $n = 2$ and $l = 2$, and the only partition of n into l parts is $(v_1, v_2) = (1, 1)$. Thus, $r_1 = 2$ in (45) and

$$N = \frac{2!}{v_1! v_2! r_1!} = \frac{2!}{1!1!2!} = 1. \quad (47)$$

Finally,

$$\begin{aligned} G_2^{(2)}(\omega_1, \omega_2) &= 2! \sum_1' G_1(\omega_1) G_1(\omega_2) \\ &= 2G_1(\omega_1) G_1(\omega_2). \end{aligned} \quad (48)$$

What about $G_3^{(2)}(\omega_1, \omega_2, \omega_3)$? Now, $n = 3$ and $l = 2$, and there is still only one partition of n into l parts: $(v_1, v_2) = (1, 2)$. So $r_1 = 1$, $r_2 = 1$, and from (45)

$$N = \frac{3!}{v_1! v_2! r_1! r_2!} = \frac{3!}{1!2!1!1!} = 3. \quad (49)$$

Therefore

$$\begin{aligned} G_3^{(2)}(\omega_1, \omega_2, \omega_3) &= 2! \sum_3' G_1(\omega_1) G_2(\omega_2, \omega_3) \\ &= 2[G_1(\omega_1) G_2(\omega_2, \omega_3) \\ &\quad + G_1(\omega_2) G_2(\omega_1, \omega_3) \\ &\quad + G_1(\omega_3) G_2(\omega_1, \omega_2)] \\ &= 2\{[1][2, 3] + [2][1, 3] + [3][1, 2]\} \end{aligned} \quad (50)$$

where the last line is an abbreviation for the middle line. Lastly, what is $G_4^{(2)}(\omega_1, \omega_2, \omega_3, \omega_4)$? $n = 4$ and $l = 2$, and there are now two partitions: $(v_{11}, v_{12}) = (1, 3)$ and $(v_{21}, v_{22}) = (2, 2)$. For each partition,

$$N_1 = \frac{4!}{v_{11}! v_{12}! r_1! r_2!} = \frac{4!}{1!3!1!1!} = 4 \quad (51)$$

$$N_2 = \frac{4!}{v_{21}! v_{22}! r_1!} = \frac{4!}{2!2!2!} = 3 \quad (52)$$

giving an overall value

$$\begin{aligned} G_4^{(2)}(\omega_1, \omega_2, \omega_3, \omega_4) &= 2! \sum_{N_1}' G_1(\omega_1) G_3(\omega_2, \omega_3, \omega_4) \\ &\quad + 2! \sum_{N_2}' G_2(\omega_1, \omega_2) G_2(\omega_3, \omega_4) \\ &= 2\{[1][2, 3, 4] + [2][1, 3, 4] \\ &\quad + [3][1, 2, 4] + [4][1, 2, 3] \\ &\quad + [1, 2][3, 4] + [1, 3][2, 4] \\ &\quad + [1, 4][2, 3]\}. \end{aligned} \quad (53)$$

APPENDIX B

VOLTERRA TRANSFER FUNCTIONS FOR A BANDPASS BIQUAD

Equation (20) is not explicitly solved for y , but using the formulas in Appendix A, the Volterra transfer functions M_n can be derived nonetheless. Using the harmonic input method, we *assume* $y(t)$ is made up of terms like

$$M_n(\omega_1, \dots, \omega_n) \exp[j(\omega_1 + \dots + \omega_n)t] \quad (54)$$

and proceed a term at a time through (20), following the example of [13] for terms such as $\int y dt$.

Starting on the left-hand side of (20), if y is made up of terms like (54), then y/R_1 will be made up of the same terms divided by R_1 . Thus, the coefficient of $\exp[j(\omega_1 + \dots + \omega_n)t]$ will be M_n/R_1 . For $C_1 dy/dt$, (54) suggests the coefficient will be

$$C_1[j(\omega_1 + \dots + \omega_n)]M_n(\omega_1, \dots, \omega_n). \quad (55)$$

Similarly, the coefficient for $(g_{m1}g_{m2})/C_2 \times \int y dt$ will be

$$\frac{g_{m1}g_{m2}}{C_2} \frac{1}{[j(\omega_1 + \dots + \omega_n)]} M_n(\omega_1, \dots, \omega_n). \quad (56)$$

Thus, the three left-hand side terms taken together have a coefficient

$$\left\{ C_1 \sum_{i=1}^n j\omega_i + \frac{1}{R_1} + \frac{g_{m1}g_{m2}}{C_2 \sum_{i=1}^n j\omega_i} \right\} M_n(\omega_1, \dots, \omega_n). \quad (57)$$

The overall formula for M_n can be found by dividing the coefficients for the right-hand side of (20), that we will now derive, by the quantity in (57) inside the braces.

On the right-hand side, starting with the two terms containing x , we set $x(t)$ to the sum of n exponentials and extract the coefficient of $\exp[j(\omega_1 + \dots + \omega_n)t]$. Clearly, for $n = 1$ the coefficient is g_{mi} , for $n = 3$ the coefficient is $6\epsilon_i$, and for all other n the coefficient is zero. For the term containing $\int y^3 dt$, we know that the coefficient for y^3 is $M_n^{(3)}$ from Appendix A; the integral sign produces a result similar to (56), leading to an overall coefficient of

$$\frac{g_{m1}\epsilon_2}{C_2 \sum_{i=1}^n j\omega_i} M_n^{(3)}(\omega_1, \dots, \omega_n) \quad (58)$$

for that term. In the term containing $[\int y dt]^3$, the integration happens *before* the cube. This means the $1/\sum j\omega_i$ terms must be taken for each individual M_{v_i} in the final expression, leading to a coefficient as given in (59), shown at the bottom of the page, or, in an abbreviated (yet hopefully clear) notation

$$\frac{\epsilon_1 g_{m2}^3}{C_2^3} 3! \sum_{(v;3,n)} \sum_N' \frac{M_{v_1} M_{v_2} M_{v_3}}{\sum_{v_1} j\omega_i \sum_{v_2} j\omega_i \sum_{v_3} j\omega_i}. \quad (60)$$

Using a similar argument, the coefficient for the term involving $[\int y^3 dt]^3$ can be seen to be

$$\frac{\epsilon_1 \epsilon_2^3}{C_2^3} \sum_{(v;3,n)} \sum_N' \frac{3! M_{v_1}^{(3)} M_{v_2}^{(3)} M_{v_3}^{(3)}}{\sum_{v_1} j\omega_i \sum_{v_2} j\omega_i \sum_{v_3} j\omega_i}. \quad (61)$$

The coefficients for the remaining two terms are slightly more difficult. One of the terms is of the form

$$\left[\int y dt \right] \left[\int y dt \right] \left[\int y^3 dt \right]$$

which suggests a coefficient involving M_{v_1} , M_{v_2} , and $M_{v_3}^{(3)}$ where $v_1 + v_2 + v_3 = n$. After careful examination, the

coefficient can be seen to be

$$\frac{3g_{m2}^2 \epsilon_1 \epsilon_2}{C_2^3} \sum_{(v;3,n)} \left(\sum_{N_{12}}' \frac{2! M_{v_1} M_{v_2}}{\sum_{v_1} j\omega_i \sum_{v_2} j\omega_i} \right) \left(\frac{M_{v_3}^{(3)}}{\sum_{v_3} j\omega_i} \right). \quad (62)$$

The $2!$ inside the first brackets is for the permutation of v_1 and v_2 , and

$$N_{12} = \frac{n!}{v_1! v_2! r!} \quad \text{where } r = \begin{cases} 1, & v_1 \neq v_2 \\ 2, & v_1 = v_2 \end{cases}. \quad (63)$$

Finally, by similar reasoning, the coefficient for the term with $[\int y dt][\int y^3 dt]^2$ is

$$\frac{3g_{m2}\epsilon_1\epsilon_2^2}{C_2^3} \sum_{(v;3,n)} \left(\sum_{N_1}' \frac{M_{v_1}}{\sum_{v_1} j\omega_i} \right) \times \left(\sum_{N_{23}}' \frac{2! M_{v_2}^{(3)} M_{v_3}^{(3)}}{\sum_{v_2} j\omega_i \sum_{v_3} j\omega_i} \right) \quad (64)$$

where

$$N_1 = \frac{n!}{v_1!(n-v_1)!}$$

$$N_{23} = \frac{(n-v_1)!}{v_2! v_3! r!} \quad \text{where } r = \begin{cases} 1, & v_2 \neq v_3 \\ 2, & v_2 = v_3 \end{cases}. \quad (65)$$

The results for general n are shown in Table I. Although the coefficients in the last three rows are nonzero only for $n \geq 5$, the authors have not seen coefficients for terms such as $[\int y dt]^2[\int y^3 dt]$ published before, and hence feel they are worth including.

APPENDIX C

VOLTERRA TRANSFER FUNCTIONS FOR THE FILTER BANK

Let us find the Volterra transfer functions for the filter bank of Fig. 5. We shall use the naming conventions shown there: The resonators have transfer functions M_{1n} , M_{2n} , and M_{3n} ; the transfer function from input to output is G_n ; the transfer function from input to "error" is K_n ; and the transfer functions from input to each resonator's output are H_{1n} , H_{2n} , and H_{3n} . We assume the feedback $\Phi(\omega)$ is linear, but possibly frequency dependent.

Finding the Volterra transfer functions for Fig. 5 is tantamount to performing the same calculation on Fig. 15 because (it is easily verified) the three parallel resonators can be reduced to one resonator with

$$M_n = M_{1n} + M_{2n} + M_{3n}. \quad (66)$$

$$\frac{\epsilon_1 g_{m2}^3}{C_2^3} 3! \sum_{(v;3,n)} \sum_N' \frac{M_{v_1}(\omega_1, \dots, \omega_{v_1}) M_{v_2}(\omega_{v_1+1}, \dots, \omega_{v_1+v_2}) M_{v_3}(\omega_\mu, \dots, \omega_n)}{\sum_{i=1}^{v_1} j\omega_i \sum_{i=v_1+1}^{v_1+v_2} j\omega_i \sum_{i=\mu}^n j\omega_i} \quad (59)$$

$$H_{13}(\omega_1) = K_1(\omega_1)K_1(\omega_2)K_1(\omega_3)M_{13}(\omega_1, \omega_2, \omega_3) + K_3(\omega_1, \omega_2, \omega_3)M_{13}(\omega_1, \omega_2, \omega_3)$$

where $K_1(\omega) = \frac{1}{1 + \Phi(\omega) \left[\sum_3 M_{i1}(\omega) \right]}$

and $G_3(\omega_1, \omega_2, \omega_3) = \frac{\prod_3 K_1(\omega_i) \left[\sum_3 M_{i3}(\omega_1, \omega_2, \omega_3) \right]}{1 + \Phi(\omega_1 + \omega_2 + \omega_3) \left[\sum_3 M_{i1}(\omega_1 + \omega_2 + \omega_3) \right]}$ (71)

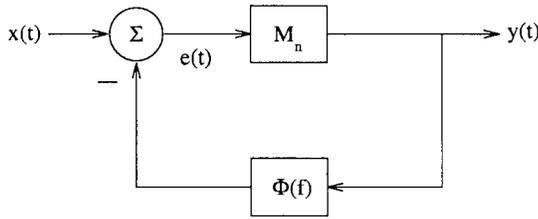


Fig. 15. Simplified filter bank structure.

The input–output Volterra transfer function G_n for Fig. 15 is derived in [13]:

$$G_n(\omega_1, \dots, \omega_n) = [1 + \Phi(\omega_1 + \dots + \omega_n)M_1(\omega_1 + \dots + \omega_n)]^{-1} \times \sum_{l=2}^n \sum_{(v;l,n)} \sum_{N'} K_{v_1}(\omega_1, \dots, \omega_{v_1}) \times K_{v_2}(\omega_{v_1}, \dots, \omega_{v_1+v_2}) \times \dots \times K_{v_l}(\omega_\mu, \dots, \omega_n) \times M_l \left(\sum_{i=1}^{v_1} \omega_i, \sum_{i=v_1}^{v_1+v_2} \omega_i, \dots, \sum_{i=\mu}^n \omega_i \right) \quad (67)$$

with $\mu = n - \omega_{v_l} - 1$ as in Appendix A and

$$K_n(\omega_1, \dots, \omega_n) = -G_n(\omega_1, \dots, \omega_n)\Phi(\omega_1 + \dots + \omega_n) + \begin{cases} 1, & n = 1 \\ 0, & n > 1 \end{cases} \quad (68)$$

To derive H_{1n} , we note that it is simply a *series connection* of K_n and M_{1n} . Extrapolating from the derivation in [13], we can write the result

$$H_{1n}(\omega_1, \dots, \omega_n) = \sum_{l=1}^n \sum_{(v;l,n)} \sum_{N'} K_{v_1}(\omega_1, \dots, \omega_{v_1}) \times K_{v_2}(\omega_{v_1}, \dots, \omega_{v_1+v_2}) \times \dots \times K_{v_l}(\omega_\mu, \dots, \omega_n) \times M_{1l} \left(\sum_{i=1}^{v_1} \omega_i, \sum_{i=v_1}^{v_1+v_2} \omega_i, \dots, \sum_{i=\mu}^n \omega_i \right). \quad (69)$$

Expanding (69) for $n = 1$ and $n = 3$ gives, respectively,

$$H_{11}(\omega_1) = K_1(\omega_1)M_{11}(\omega_1) = \frac{M_{11}(\omega_1)}{1 + \Phi(\omega_1) \left[\sum_3 M_{i1}(\omega_1) \right]} \quad (70)$$

[the same result (26) as the linear analysis] and, using (66)–(68), gives the result shown in (71), at the top of the page.

ACKNOWLEDGMENT

The authors are grateful to A. Shoval for the use of his test board in measuring the laboratory results. Thanks are also due to the anonymous reviewers, whose comments helped correct errors in and improve the clarity of the paper.

REFERENCES

- [1] J. A. Cherry, "Distortion analysis of weakly nonlinear filters using Volterra series," M.Eng. thesis, Carleton Univ., Ottawa, Ont., Canada, 1994.
- [2] J. A. Cherry and W. M. Snelgrove, "Analog filter banks with low intermodulation distortion," in *Int. Symp. on Circuits and Systems*, 1995, pp. 1195–1198.
- [3] R. A. Duncan, K. W. Martin, and A. S. Sedra, "A Q -enhanced active-RLC bandpass filter," in *Int. Symp. Circuits and Systems*, 1993, pp. 1416–1419.
- [4] S. Pilipos and Y. Tsvividis, "Design of active RLC integrated filters with applications in the GHz range," in *Int. Symp. Circuits and Systems*, 1994, pp. 645–649.
- [5] N. M. Nguyen and R. G. Meyer, "Si IC-compatible inductors and LC passive filters," *IEEE J. Solid-State Circuits*, vol. 25, pp. 1028–1031, Aug. 1990.
- [6] B. S. Song, "A 10.7MHz switched-capacitor bandpass filter," *IEEE J. Solid-State Circuits*, vol. 24, pp. 320–334, Apr. 1989.
- [7] W. M. Snelgrove and A. Shoval, "A balanced 0.9 μ m CMOS transconductance-C filter tunable over the VHF range," *IEEE J. Solid-State Circuits*, vol. 27, pp. 314–323, Mar. 1992.
- [8] R. E. Fisher, "A subscriber set for the equipment test," *Bell Syst. Tech. J.*, vol. 58, p. 133, Jan. 1979.
- [9] J. J. Bussgang, L. Ehrman, and J. W. Graham, "Analysis of nonlinear systems with multiple inputs," *Proc. IEEE*, vol. 62, pp. 1088–1119, Aug. 1974.
- [10] H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*. New York: Wiley, 1980.
- [11] V. Volterra, *Theory of Functionals and of Inegro and Integro-Differential Equations*. New York: Dover, 1959 (reprint of 1930 edition).
- [12] N. Wiener, *Nonlinear Problems in Random Theory*. Cambridge, MA: MIT Press, 1958.
- [13] E. Bedrosian and S. O. Rice, "The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian inputs," *Proc. IEEE*, vol. 59, pp. 1688–1707, Dec. 1971.

- [14] M. Padmanabhan and K. Martin, "Resonator-based filter-banks for frequency-domain applications," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 1145–1159, Oct. 1991.
- [15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge U.K.: Cambridge Univ. Press, 1992.
- [16] N. M. Nguyen and R. G. Meyer, "A 1.8-GHz monolithic *LC* voltage-controlled oscillator," *IEEE J. Solid-State Circuits*, vol. 27, pp. 444–450, Mar. 1992.
- [17] Y. L. Kuo, "Distortion analysis of bipolar transistor circuits," *IEEE Trans. Circuit Theory*, vol. 20, pp. 709–716, Nov. 1973.
- [18] K. Gopal, M. S. Nakhla, K. Singhal, and J. Vlach, "Distortion analysis of transistor networks," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 99–106, Feb. 1978.
- [19] R. G. Meyer and M. L. Stephens, "Distortion in variable capacitor diodes," *IEEE J. Solid-State Circuits*, vol. 10, pp. 47–54, Feb. 1975.
- [20] S. Narayanan, "Application of Volterra series to intermodulation distortion analysis of transistor feedback amplifier," *IEEE Trans. Circuit Theory*, vol. 17, pp. 518–527, Nov. 1970.
- [21] S. Boyd and L. O. Chua, "Measuring Volterra kernels," *IEEE Trans. Circuits Syst.*, vol. 30, pp. 571–577, Mar. 1983.
- [22] R. G. Meyer, M. J. Shensha, and P. Eschenbach, "Cross modulation and intermodulation in amplifiers at high frequencies," *IEEE J. Solid-State Circuits*, vol. 7, pp. 16–23, Feb. 1972.
- [23] A. Shoval, D. A. Johns, and W. M. Snelgrove, "A wide-range tunable BiCMOS transconductor," *Microelectron. J.*, vol. 24, pp. 555–564, Aug. 1993.



James A. Cherry (S'97) received the B.A.Sc. degree in computer engineering from the University of Waterloo, Ont., Canada, in 1992, graduating at the top of his class, and the M.Eng. degree in electronics from Carleton University, Ottawa, Ont., in 1994. He is currently completing the Ph.D. degree at Carleton University, in the area of high-speed continuous-time sigma delta modulators.

His research interests lie in the rapid simulation of such circuits, and in the rapid identification and determination of the effect on performance of circuit

nonidealities.



W. Martin Snelgrove (S'75–M'81) received the B.A.Sc. degree in chemical engineering and the M.A.Sc. and Ph.D. degrees in electrical engineering, from the University of Toronto, Ont., Canada, in 1975, 1977, and 1982, respectively.

In 1982, he was with INAOE, Mexico, as a Visiting Researcher in CAD. He then taught at the University of Toronto until 1992, when he moved to Carleton University, Ottawa, Ont., as a Professor and holder of the OSCI/NSERC Industrial Research Chair in High Speed Integrated Circuits. He spent sabbatical leaves in 1989 and 1990 as a Resident Visitor at AT&T Bell Laboratories in Reading, PA, working in CMOS analog design, and took research leaves in 1991 and 1992 to work on a VLSI circuits text, from which he taught at the University of Oulu, Finland. His work focuses on architectures and circuits for high-performance integrated circuits for signal processing applications, including RF signal conditioning, high-speed data conversion, real-time DSP, and CAD for signal processing.

Dr. Snelgrove was the winner of the 1986 Circuits and Systems Society Guillemin–Cauer Award for a 1986 paper coauthored with A. S. Sedra. He serves as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING.