3. BASICS OF VOLTERRA ANALYSIS

Volterra analysis is a way to extend small-signal analysis and the use of integral transforms to nonlinear dynamical systems. Volterra functionals (i.e. functional series consisting of sum of i’th order convolution integrals) were first developed by Volterra 1930, applied to electrical systems by Wiener 1942 and systematically used to analysis of distortion since 1970’s, starting from Bell Labs. Currently Volterra analysis is implemented at least in Berkeley SPICE (.DISTO analysis), some RF simulators, and symbolic analysis tool ISAAAC.

Volterra digital filters have been used for years e.g. to linearise the response of high-power loudspeakers. IEEE Online search engine finds ca. 650 hits with search word ‘Volterra’. References to basic litterature can be found from Wambacq and Maas.

NOTATIONS

As remembered, the time response $y(t)$ of a linear system can be calculated as a convolution integral of the impulse response $h(t)$ and input signal $x(t)$

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau$$

which can be either Fourier or Laplace transformed into frequency domain, where output and input spectrums are related by

$$Y(j\omega) = H(j\omega) \cdot X(j\omega).$$

where $Y(s)$ and $X(s)$ are output and input spectrums and $H(s)$ is the integral transform of $h(t)$.

In a nonlinear system, to model the effect of signal dependent impulse response, the convolution integral must be extended to series of 1 to N-dimensional convolutions.

$$y(t) = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_n h_k(u_1, \ldots, u_n) \prod_{r=1}^{n} x(t-u_r)$$

$$= \int_{-\infty}^{\infty} h_1(u_1)x(t-u_1)du_1$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(u_1, u_2)x(t-u_1)x(t-u_2)du_1du_2$$

$$+ \ldots$$
here

\[ h_n(t_1, t_2, \ldots, t_n) \]

are n-dimensional impulse responses or kernels. For causality,

\[ h_n(t_1, t_2, \ldots, t_n) = 0 \text{ when any } t_i < 0. \]

In many analysis it is required that the kernels are symmetric, i.e. (for a 2nd order kernel)

\[ h_2(t_1, t_2) = h_2(t_2, t_1) \]

An asymmetric kernel \( h_2^+ (t_1, t_2) \) can be symmetrised by

\[ h_2(t_1, t_2) = 0.5 \left( h_2^+(t_1, t_2) + h_2^+(t_2, t_1) \right) \]

If each convolution integral is marked as \( H_i(x(t)) \), a Volterra series can be drawn as system of parallel transfer functions of different order. Here, \( H_1 \) is the normal linear transfer function, and \( H_2 \) is a second order term and so forth.

**INTEGRAL TRANSFORMS**

The basic integral transforms, Fourier and Laplace transforms can be extended to Volterra series representation. In the following, main interest is in Fourier transform and discrete line spectrums, and following rules are useful:

\[ H_2(-j\omega_1, -j\omega_2) = H_2^* (j\omega_1, j\omega_2) \]

(negative frequencies mean complex conjugate) and if the impulse response \( h() \) is symmetric, then also

\[ H_2(j\omega_1, j\omega_2) = H_2(j\omega_2, j\omega_1) \]

Due to nonlinear nature, output spectrum is not merely a product of transfer function and input spectrum \( X(f) \), but convolution integral is needed to calculate higher order spectrums.

\begin{align*}
  H_2(f_1, f_2) &= H_1(f_1)X(f) + \int_{-\infty}^{f_1} df_1 H_2(f_1, f-f_1)X(f)X(f-f_1) \\
  &+ \int_{-\infty}^{f_2} df_2 \int_{-\infty}^{f_2} df_3 H_3(f_1, f_2, f-f_2, f-f_3)X(f_1)X(f_2)X(f_2-f_3)X(f-f_2) \\
  &+ \ldots
\end{align*}
THE MEANING OF $H_N(S_1,..SN)$

In an $n$th order transfer function

$$H_n(s_1, ..., s_n)$$

is a general transfer function that describes the response to any of the $n$th order output frequencies. It has $n$ frequency arguments $s_1, s_2, .., s_n$, because there are $2^n$ possible output frequencies that are of form

$$s = \pm s_1 \pm s_2 \pm ... \pm s_n$$

In a 2-tone test, only frequencies $w_h$ and $w_l$ are used, but any of their permutation may appear in the transfer function. Thus, e.g. in a 3rd order transfer function the possible positive output frequencies are

<table>
<thead>
<tr>
<th>Freq.</th>
<th>$H_3(x)$</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_l+wl-wh$</td>
<td>$H_3(w_l,w_l,-wh)$</td>
<td>IM3L</td>
</tr>
<tr>
<td>$w_l+wl-wl$</td>
<td>$H_3(w_l,w_l,-wl)$</td>
<td>FUND</td>
</tr>
<tr>
<td>$wh+wh-wh$</td>
<td>$H_3(wh,wh,-wh)$</td>
<td>FUND</td>
</tr>
<tr>
<td>$-wl+wh+wh$</td>
<td>$H_3(-wl,wh,wh)$</td>
<td>IM3H</td>
</tr>
<tr>
<td>$wl+wl+wl$</td>
<td>$H_3(w_l,w_l,w_l)$</td>
<td>HD3</td>
</tr>
<tr>
<td>$wl+wl+wh$</td>
<td>$H_3(w_l,w_l,wh)$</td>
<td></td>
</tr>
<tr>
<td>$wl+wh+wh$</td>
<td>$H_3(w_l,wh,wh)$</td>
<td></td>
</tr>
<tr>
<td>$wh+wh+wh$</td>
<td>$H_3(wh,wh,wh)$</td>
<td>HD3</td>
</tr>
</tbody>
</table>

RESPONSE TO SINUSOIDS

Suppose a second order system with a sinusoidal excitation

$$x(t) = \frac{A}{2} \cdot (e^{j\omega t} + e^{-j\omega t}) = x_a(t) + x_b(t)$$

Now, when $x(t) = \exp(j\omega t)$, the double convolution of the impulse response is the same as the 2-D Fourier transform $H_2(j\omega_1, j\omega_2)$ of the impulse response, and the time domain output $y(t)$ is

$$y(t) = H_2[x_a(t)] + H_2[x_a(t)]$$
$$+ H_2[x_a(t), x_b(t)] + H_2[x_a(t), x_b(t)]$$
$$= \frac{A^2}{4} \cdot H_2(j\omega, j\omega) \cdot e^{j2\omega t}$$
$$+ \frac{A^2}{4} \cdot H_2(-j\omega, -j\omega) \cdot e^{-j2\omega t}$$
$$+ 2\frac{A^2}{4} \cdot H_2(j\omega, -j\omega) \cdot 1$$

(C) 1999- Timo Rahkonen, University of Oulu, Oulu, Finland
RESPONSE TO A 2-TONE TEST

A sum of two sinusoids is

\[ x(t) = \left( \frac{A_1}{2} e^{j\omega_1 t} + \frac{A_1^*}{2} e^{-j\omega_1 t} \right) + \left( \frac{A_2}{2} e^{j\omega_2 t} + \frac{A_2^*}{2} e^{-j\omega_2 t} \right) = x_a(t) + x_b(t) + x_c(t) + x_d(t) \]

In a second order nonlinearity, this produces the following outputs

\[ y(t) = 2 \cdot \left( \frac{|A_1|^2}{4} \cdot H_2(j\omega_1, -j\omega_1) + \frac{|A_2|^2}{4} \cdot H_2(j\omega_2, -j\omega_2) \right) \]
\[ + \frac{A_1^2}{4} \cdot H_2(j\omega_1, j\omega_1) \cdot e^{j2\omega_1 t} + \frac{A_2^2}{4} \cdot H_2(-j\omega_1, -j\omega_1) \cdot e^{-j2\omega_1 t} \]
\[ + \frac{A_1^2}{4} \cdot H_2(j\omega_2, j\omega_2) \cdot e^{j2\omega_2 t} + \frac{A_2^2}{4} \cdot H_2(-j\omega_2, -j\omega_2) \cdot e^{-j2\omega_2 t} \]
\[ + \frac{A_1 A_2}{2} \cdot H_2(j\omega_1, j\omega_2) \cdot e^{j(\omega_1 + \omega_2) t} + \frac{A_1 A_2^*}{2} \cdot H_2(-j\omega_1, -j\omega_2) \cdot e^{-j(\omega_1 + \omega_2) t} \]
\[ + \frac{A_1 A_2^*}{2} \cdot H_2(j\omega_1, -j\omega_2) \cdot e^{j(\omega_1 - \omega_2) t} + \frac{A_1 A_2}{2} \cdot H_2(-j\omega_1, j\omega_2) \cdot e^{-j(\omega_1 - \omega_2) t} \]

Note that amplitudes contain also phase information, and for negative frequencies a complex conjugate has been used. The relative amplitudes of the terms depends on how many terms at that frequency is produced, when the input signal is multiplied by itself. For reference, all frequencies produced in a 2nd order system are listed below.

| Table 1: Output frequencies in a 2nd order nonlinearity |
|---------------------------------|--------------|
| +jω1  | -jω1  | +jω2  | -jω2 |
| j2ω1 | 0 | +j(ω1+ω2) | +j(ω1-ω2) |
| -jω1 | 0 | +j(ω2-w1) | -j(ω1+w2) |
| +j(ω1+w2) | +j(ω2-w1) | j2ω2 | 0 |
| +j(ω1-ω2) | -j(ω1+w2) | 0 | -j2ω2 |

(C) 1999- Timo Rahkonen, University of Oulu, Oulu, Finland
As shown before, Volterra kernels are general Nth-order transfer functions valid at all frequencies. In a 2-tone test the relative amplitudes of different terms depend on how many times these frequencies appear in the convolution of the original spectrum. The different outputs are shown below.

Table 1:

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>0.5</td>
</tr>
<tr>
<td>2ω1−ω1</td>
<td>A1^*A2 H2(j</td>
</tr>
<tr>
<td>2ω1−ω2</td>
<td>0.75A1^2 A2^*H2(j</td>
</tr>
<tr>
<td>ω1</td>
<td>A1 H1(j</td>
</tr>
<tr>
<td>3ω1</td>
<td>0.25A1^3H2(j</td>
</tr>
<tr>
<td>2ω1+ω2</td>
<td>0.75A1^2 A2 H2(j</td>
</tr>
<tr>
<td>2ω2</td>
<td>0.5A2^2 H2(j</td>
</tr>
<tr>
<td>3ω2</td>
<td>0.25A2^3H2(j</td>
</tr>
</tbody>
</table>

(Wambacq, p. 76)

An important simplification is a sort of block diagram presentation, where frequency dependent signal paths are multiplied so that the multiplication does not affect the time constants of the filters.

It can be shown that from x(t) to output y(t)

\[ h_{tot}(\tau_1, \tau_2) = \int_{-\infty}^{\infty} h_1(\sigma)h_2(\tau_1-\sigma)h_3(\tau_2-\sigma)d\sigma \]

which results in frequency domain into second order transfer function of

\[ H_2(j\omega_1, j\omega_2) = H_1(j\omega_1) \cdot H_2(j\omega_2) \cdot H_3(j\omega_1 + j\omega_2) \]

Thus, Hc() is evaluated at the frequency components produced by the multiplication.
Example 1. Amp with input and output filters
(see Wambacq p.71-)

First order response

\[ H_1(j\omega) = \frac{K_1}{(1 + j\omega \tau_1) \cdot (1 + j\omega \tau_2)} \]

2nd order response

2nd order response is shown left: input filter is evaluated at different frequencies, responses are multiplied and further multiplied by 2nd order gain and output filter response at the output frequency.

Example: input freqs are \( w_1 = 5 \) and \( w_2 = 12 \) rad/s. then responses at following frequencies are

- 2w1: \( H_2(5,5) \)
- w2-w1: \( H_2(12,-5) \)
- w1+w2: \( H_2(5,12) \)
- 2w2: \( H_2(12,12) \)

3rd order response

\[ H_3(j\omega_1, j\omega_2, j\omega_3) = \frac{K_3}{(1 + j\omega_1 \tau_1) \cdot (1 + j\omega_2 \tau_1) \cdot (1 + j\omega_3 \tau_1) \cdot (1 + j(\omega_1 + \omega_2 + \omega_3) \tau_2)} \]
Example 2. Block level

In this simple 2nd order example, the nonlinearity is on the output side and affects the gain of the circuit. As will be seen later, the frequency response of this circuit is

Here

\[ u(t) = \frac{i}{g + j\omega C} = H_1(j\omega) \cdot i \]

and

\[ H_2(\omega_1, \omega_2) = K_2 \cdot H_1(j\omega_1) \cdot H_1(j\omega_2) \]

\[ = \frac{(1/C)^2}{(j\omega_1 + a) \cdot (j\omega_2 + a)} \]

where \( K_2 = 1 \) and \( a = g/C \). After some algebra the 2-dimensional impulse response can be calculated to be

\[ h_2(t_1, t_2) = \exp(-a(t_1 + t_2)) \]

Example 3.

In this example, a nonlinear conductance of type

\[ i = g \cdot (v + K_2 \cdot v^2) \]

causes the system to behave according to nonlinear differential equation

\[ i(t) = C \cdot \frac{dv}{dt} + g \cdot v(t) + K_2 g \cdot v(t)^2 \]

In this case, the nonlinearity affects also the system time constants, and compared to example 1, more terms appear. The frequency response is

\[ H_2(\omega_1, \omega_2) = \frac{K_2 g}{(j\omega_1 + \omega_o) \cdot (j\omega_2 + \omega_o) \cdot (j(\omega_1 + \omega_2) + \omega_o)} \]

where \( \omega_o = g/C \). The 2-dimensional impulse response gets complicated and cannot be shown in closed form: it consists of integral functions of the exponential function \( \text{Ei}(t,n) = \text{Int}(\exp(t)/t^n) \).
Example 3 Cont...

The previous system is described as a signal flow graph left. This shows a very common way of analyzing Volterra kernels by evaluating the terms by order, i.e. linear term first, then the 2nd order and so forth. Thus, denoting \( v = v_1 + v_2 + v_3 + \ldots \) and supposing input \( i(t) \) has linear term only,

\[
 i(t) = C \frac{dv_1}{dt} + g \cdot v_1(t) \\
 0 = C \frac{dv_2}{dt} + g \cdot v_2(t) + K_{2g} \cdot v_1^2(t) \\
 0 = C \frac{dv_3}{dt} + g \cdot v_3(t) + (K_{3g} \cdot v_1(t) \cdot v_2(t)) \\
\]

where

\[
 v_1 = \int_{-\infty}^{\infty} h_1(u_1) i_{in}(t - u_1) du_1 \\
 v_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(u_1, u_2) i_{in}(t - u_1) i_{in}(t - u_2) du_1 du_2 \\
\]

CASCADe OF TWO STAGES

In a cascade of two stages, nonlinearities are caused in both stages. Moreover, IM3 terms are produced by 3rd order nonlinearities in both blocks, and also by up or downmixed second order terms. Thus,

\[
 Q_1(s_1) = H_1(s_1) F_1(s_1) \\
 Q_2(s_1, s_2) = \frac{H_2(s_1, s_2)}{H_1(s_1)} F_1(s_1 + s_2) + \\
 \frac{H_1(s_1) H_1(s_2) F_2(s_1, s_2)}{H_1(s_1)} \\
 Q_3(s_1, s_2, s_3) = \frac{H_3(s_1, s_2, s_3)}{H_1(s_1)} F_1(s_1 + s_2 + s_3) + \\
 \frac{H_1(s_1) H_1(s_2) H_1(s_3) F_3(s_1, s_2, s_3)}{H_1(s_1)} + \\
 \frac{H_2(s_1, s_2, s_3) F_2(s_1, s_2 + s_3)}{H_1(s_1)} + \\
 \frac{H_3(s_1, s_2, s_3) F_3(s_1, s_2 + s_3)}{H_1(s_1)} + \\
 \frac{H_1(s_2, s_3) F_1(s_1 + s_2 + s_3)}{H_1(s_1)} + \\
 \frac{H_1(s_3) F_1(s_1 + s_2 + s_3)}{H_1(s_1)}
\]
CASCADE EQUATIONS IN SYSTEM SIMULATIONS

IM3 level in receiver systems is easy to estimate using e.g. spreadsheet calculators (Microsoft Excel). In most system calculations, following simplifications are used:

- IM3 due to second order terms is not modelled. This does not cause errors, if interstage filters remove envelope and 2nd harmonic terms.
- Usually, IM3 contributions from different stages is supposed to add coherently. This gives a worst case estimate, and its justification is discussed in Maas95
- The effect of interstage filtering has amplitude response only, showing IIP3 improvement due to passive filtering

Moreover, one should note that that input-output pair of an amplifier is also a cascade. Thus, the impedances seen at envelope and 2nd harmonic affect the total IM3 level.

ANALYSIS OF INVERSE FUNCTIONS

Suppose that we have function $y = f(x)$ and its inverse function $x = g(y) = f^{-1}(y)$ (e.g. $i = f(v)$ and $v = f^{-1}(i)$). Volterra series is based on nth order derivatives, so we will use differentiating rules for inverse functions:

$$ (f^{-1})'(a) = \frac{1}{f'(f^{-1}(a))} $$
$$ (f^{-1})''(a) = \frac{-f''(f^{-1}(a))}{f'(f^{-1}(a))^3} $$
$$ (f^{-1})'''(a) = \frac{-f'''(f^{-1}(a)) + 3f''(f^{-1}(a))}{f'(f^{-1}(a))^4} $$

As a result of this, Volterra expansions for inverse functions below (here ac conductance and ac resistance) can be calculated as shown left:

$$ v = r \cdot i + K_{2r} \cdot i^2 + K_{3r} \cdot i^3 $$
$$ i = g \cdot u + K_{2g} \cdot u^2 + K_{3g} \cdot u^3 $$
ANALYSIS OF FEEDBACK

In a feedback system shown left, the linear response is

\[ Q_1(s_1) = H_1(s_1) \cdot R(s_1) \]

where \( R \) is the so-called gain-reduction function

\[ R(s_1) = \frac{1}{1 + H_1(s_1) \cdot F_1(s_1)} \]

Supposing both the amplifier \( H \) and the feedback \( F \) are nonlinear, total response can be derived to be as shown on the next slide.
SPECIAL CASES: LINEAR FEEDBACK

If the feedback portion is linear, the kernels reduce to the ones shown left. Marking loop gain as

\[ T(s) = H_1(s) \cdot F_1(s) \]

and supposing that \( T(s) \gg 1 \), the transfer functions can be estimated as

\[ Q_1(s_1) = \frac{H_1(s_1)}{T(s_1)} \]

\[ Q_2(s_1, s_2) = \frac{H_2(s_1, s_2)}{T^2(s_2)} \]

\[ Q_3(s_1, s_2, s_3) = \frac{1}{T^3(s_2)} \times \left[ H_3(s_1, s_2, s_3) - \frac{2H_2(s_1, s_2)F_1(s_1 + s_2)R_1(s_1 + s_2)R_2(s_1 + s_2)H_2(s_1 + s_2)}{H_1(s_1 + s_2)} \right] \]

In wideband case the loop gain \( T \) can be considered constant.

Then,

\[ Q_1(s_1) = \frac{H_1(s_1)}{T} \]

\[ Q_2(s_1, s_2) = \frac{H_2(s_1, s_2)}{T^2} \]

\[ Q_3(s_1, s_2, s_3) = \frac{1}{T^3} \times \left[ H_3(s_1, s_2, s_3) - \frac{2H_2(s_1, s_2)H_2(s_3, s_1 + s_2)}{H_1(s_1 + s_2)} \right] \]

Now the distortion in the linear feedback case can be compared to what would be achieved in open loop, where for the same output amplitude, the amplifier would be driven with attenuated amplitude \( V_a/T \) compared to feedback amplitude. In these cases, linear and 2nd and 3rd harmonic amplitudes D2 and D3 are

<table>
<thead>
<tr>
<th></th>
<th>Closed-loop</th>
<th>Open-loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin</td>
<td>( H_1/T \times V_a )</td>
<td>( H_1/T \times V_a )</td>
</tr>
<tr>
<td>D2</td>
<td>( H_2/T^2 \times V_a^2 )</td>
<td>( H_2/T^2 \times V_a^2 )</td>
</tr>
<tr>
<td>D3</td>
<td>( H_3/T^3 \times V_a^3 \times \left(1 - 2H_2/H_1H_3\right) )</td>
<td>( H_3/T^3 \times V_a^3 )</td>
</tr>
</tbody>
</table>

Thus, in general the feedback causes \( 1/T \) reduction in distortion compared to open-loop configuration with same output amplitude. However, also 2nd order nonlinearity is mixed to 3rd order distortion, and it may either increase or decrease the total distortion.

For example, in a purely exponential BJT this upconverted 2nd order term would be dominant (\( Q_3 = gm(K_3' - 2(K_2')^2)/T^4 = gm(250-800)/T^4 \)). Thus, the feedback is seriously changing the shape of the distortion, in this case from expanding to compressing.
Here, emitter degeneration is used in a BJT to change the type of distortion. Originally, IM3 is dominated by expanding 3rd order nonlinearity. When the strength of the feedback is increased, 2nd order nonlinearity at emitter will mix to a compressing 3rd order distortion.

\[ Ut = 25e^{-3}; \]
\[ gm0 = 1e^{-2}; \]
\[ beta = 50; \]
\[ ft = 5e9; \]
\[ tf = 1/(2\pi ft); \]
\[ gpi0 = gm0/beta; \]
\[ cpi0 = tf*gm0; \]
\[ Kp = \begin{bmatrix} 1 & 1/2/Ut & 1/6/Ut^2/Ut \\ gm - Ymu & go + Ymu & 0 & -gm & ... \\ 0 & 0 & 1 & 0 & ... \\ -gm & 0 & 0 & Ye + gm \end{bmatrix}; \]

Left above, fundamental amplitudes are shown as a sum of linear and expansion/compression term for two different values of RE. Below, fundamental amplitude and phase are shown as functions of RE.

This clearly shows that while the feedback resistor is increased the expanding behaviour of open-loop BJT turns to compressing behaviour due to the feedback.
VOLTERRA ANALYSIS OF ACTIVE CIRCUITS: TEST SIGNAL METHOD

General principle
A simple method to evaluate Volterra kernels of real circuits is described next. All nonlinear conductances and capacitances produce distortion currents. Thus, they are modelled by the component itself with parallel distortion current sources. The value of the distortion current depends on lower order products.

There are two methods for calculating the responses:
- General method, based on Volterra transfer functions
- Direct method, based on calculating node voltages of different order.

1. GENERAL METHOD

In the general method, general transfer functions (kernels) \( H_1(s_1) \) ... \( H_n(s_1, ..., s_n) \) are calculated in symbolic form, and response at different frequencies is obtained by evaluating the functions at these frequencies and by scaling with proper amplitudes and coefficients, as shown before.

In the calculations, kernels \( H_n() \) are directly related to \( n \)th order voltages and are obtained by solving network equations, where \( H_n() \) represents voltage, \( Y \) network admittance matrix at frequency \( s_1+ ... \), \( I \) linear input current and \( I_{NLn} \) nonlinear currents of the devices:

\[
Y(s_1) \cdot H_1(s_1) = I_1
\]

\[
I_{NL2} = f(H_1)
\]

\[
Y(s_1 + s_2) \cdot H_2(s_1, s_2) = I_{NL2}
\]

\[
I_{NL3} = f(H_1, H_2)
\]

\[
Y(s_1 + s_2 + s_3) \cdot H_3(s_1, s_2, s_3) = I_{NL3}
\]

\[
H_n(s_1, ...) = Y(s_1 + ...)^{-1} \cdot I_{NLn}
\]
2. DIRECT METHOD

In the direct method, entire kernels are not solved, but responses at different frequencies are solved directly. First, calculate 1st order voltages. Using them, 2nd order voltages can be calculated, and so forth. This is shown in more detail on the next page and also implemented in the Matlab program nli_main().

Notation used in 2-tone test:

\[ v_{i,m,n} \]

node \( m \) \( \omega_1 \) \( n \) \( \omega_2 \)

Examples:

- 2nd harmonic \( 2\omega_1 \): \( v_{i,2,0} \)
- IM3 \( 2\omega_1-\omega_2 \): \( v_{i,2,-1} \)
- Fundamental \( \omega_2 \): \( v_{i,0,1} \)

VOLterra Expansion for a 2-Tone Test Signal

- Evaluate the fundamental (1st order) node voltages \( v_1 \) using linear ac analysis.
- Evaluate 2nd order distortion currents \( i_{NL2} \) using fundamental voltage amplitudes and \( K_2 \) terms. These will appear on five different sum and difference frequencies.
- Based on these distortion currents, calculate distortion voltages \( v_2 \) in each node using linear analysis.
- Using 1st and 2nd order voltages \( v_1 \) and \( v_2 \) and \( K_2 \) and \( K_3 \) nonlinearities, calculate third order distortion currents \( i_{NL3} \) in nonlinear components. These will appear on 8 different frequencies. It is important to note that also 2nd order nonlinearity can create 3rd order distortion, if the exciting voltage contains 2nd order terms.
- Again at the frequencies of \( i_{NL3} \) terms, perform ac analysis to find 3rd order node voltages \( v_3 \).
EXAMPLE SOFTWARE

Soon, a couple of examples are presented. There are a couple of ways to get Volterra kernels:

• ISAAC is a symbolic circuit analysis tool for KU Leuven, Belgium. It gives directly transfer functions containing nonlinear terms.
• Maple is symbolic mathematics program that (as well as Mathematica) can be used to form symbolic Volterra transfer functions.
• Matlab can be used to to plot vector plots of dominant distortion terms, when numerical values are known.

In the latter two cases, the following procedure is used:

• the circuit is described as MNA (modified nodal analysis) matrix and a source vector
• besides MNA, nonlinear components are described as a list containing component value and nonlin coeffs, and controlling and output node pairs.

1-DIMENSIONAL GM

A nonlinear conductance is described by

\[ i = g_m \cdot (v + K_2 \cdot v^2 + K_3 \cdot v^3) \]

It will will be excited not only by the linear voltages, but also by the 2nd order voltages caused by its own distortion. Thus, it can be considered as a V-I-V cascade, and the values of the nonlinear current sources-parallel to the conductance are:

<table>
<thead>
<tr>
<th>order</th>
<th>( i_{NL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>( K_2 g \cdot H1(s1) \cdot H1(s2) )</td>
</tr>
<tr>
<td>3</td>
<td>( K_3 g \cdot H1(s1) \cdot H1(s2) \cdot H1(s3) + \frac{2}{3} K_2 g \cdot (H1(s1) \cdot H2(s2,s3) + H1(s2) \cdot H2(s1,s3) + H1(s3) \cdot H2(s1,s2)) )</td>
</tr>
</tbody>
</table>

(C) 1999- Timo Rahkonen, University of Oulu, Oulu, Finland
Previous expressions can be applied for calculating response at any frequency, when scaled with proper amplitudes and coefficients. When calculating directly with nth order node voltages, different expressions are needed for different frequencies. The most usual ones for 1-dimensional gm are shown left.

Here, $V_{i,m,n}$ refers to controlling voltage $V_i = V_{k-V1}$ at frequency $m\omega_1+n\omega_2$.

### Table 1: $gm$ iNL for direct calculation

<table>
<thead>
<tr>
<th>Freq. $\omega$</th>
<th>iNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2\omega_1$ $\pm \omega_2$</td>
<td>$0.75K_3g(V_{i,1,0})^2V_{i,0,\pm1}$ + $K_2gV_{i,1,0}V_{i,1,\pm1}$ + $K_2gV_{i,0,\pm1}V_{i,2,0}$</td>
</tr>
<tr>
<td>$3\omega_1$</td>
<td>$0.25K_3g(V_{i,1,0})^3$ + $K_2gV_{i,1,0}V_{i,2,0}$</td>
</tr>
</tbody>
</table>

### Table 1: $gm$ iNL for general analysis

<table>
<thead>
<tr>
<th>order</th>
<th>iNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$K_2gH1(s1)H1(s2)$</td>
</tr>
<tr>
<td>3</td>
<td>$K_3gH1(s1)H1(s2)H1(s3)$ + $2/3 K_2g (H1(s1)H2(s2,s3)$ + $H1(s2)H2(s1,s3)$ + $H1(s3)H2(s1,s2)$ )</td>
</tr>
</tbody>
</table>

| Table 1: $gm$ iNL for general analysis |

**General**

Starting from general kernels

**Direct**

Response at IM3 = 0.75*H3(w1,w1,-w2)

\[
\begin{align*}
&K_3g(V_{i,1,0})^3V_{i,0,-1} + (2/3)K_2gV_{i,1,0}(V_{i,1,-1}/1) + (2/3)K_2gV_{i,1,0}(V_{i,1,-1}/2) + (2/3)K_2gV_{i,0,-1}(V_{i,2,0}/0.5) \\
&= 0.75K_3g(V_{i,1,0})^3V_{i,0,-1} + K_2gV_{i,1,0}V_{i,1,-1} + K_2gV_{i,0,-1}V_{i,2,0}
\end{align*}
\]
**EXAMPLE OF NONLINEARITY (E.G. GΠ)**

Suppose a conductance G having 2nd order nonlinearity only. When it is driven by a voltage, 2nd order distortion current is generated.

If G is driven from low impedance (left), the distortion current \( i_{NL2} \) is short circuited and does not cause 2nd order voltage signal over G. However, if driving impedance Zs is high (right), \( i_{NL2} \) develops 2nd order distortion voltage across G and it, mixed with the linear voltage components, creates also 3rd order distortion current \( i_{NL3} \), even though the device itself has only 2nd order nonlinearity.

In BJT this appears so that when driven by current (Zs >>), a 3rd order nonlinear voltage develops in the base due to \( g_\pi \) and cancels the 3rd order nonlinearity of gm. When driven by voltage, the base voltage remains sinusoidal and gm nonlinearity appears in the output. Thus, Zs may linearize the response.

---

**2-DIMENSIONAL GM**

A nonlinear 2-dimensional conductance is controlled by two voltages k and l (e.g. \( V_{BE} \), \( V_{BC} \)). This will be replaced by its small signal value and a parallel nonlinear current source, the value of which is calculated by supposing independent nonlinearities in the input and output and the following crossterms:
$i = g_m \cdot (v_k + K_{2gm} \cdot v_k^2 + K_{3gm} \cdot v_k^3) + g_o \cdot (v_l + K_{2go} \cdot v_l^2 + K_{3go} \cdot v_l^3) + K_{2k\&l} \cdot v_k \cdot v_l + K_{32k\&l} \cdot v_k^2 \cdot v_l + K_{3k\&2l} \cdot v_k \cdot v_l^2$

Table 1: Cross terms in a 2-dimensional conductance

<table>
<thead>
<tr>
<th>Order</th>
<th>$i_{NL}$ (crossterms only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$0.5 (K_{2g1&amp;g2} H_{1k}(s1) H_{1l}(s2) + K_{2g1&amp;g2} H_{1k}(s2) H_{1l}(s1))$</td>
</tr>
<tr>
<td>3</td>
<td>$1/3 K_{2g1&amp;g2} (H_{1k}(s1) H_{2g}(s2,s3) + H_{1k}(s2) H_{2g}(s1,s3) + H_{1k}(s3) H_{2g}(s1,s2) + H_{3g}(s1,s2) H_{1l}(s3) + H_{2g}(s1,s3) H_{1l}(s2) + H_{2g}(s2,s3) H_{1l}(s1))$</td>
</tr>
<tr>
<td></td>
<td>$+ 1/3 K_{3g1&amp;2g2} (H_{1k}(s1) H_{1k}(s2) H_{1l}(s3) + H_{1k}(s1) H_{1k}(s3) H_{1l}(s2) + H_{1k}(s2) H_{1k}(s3) H_{1l}(s1))$</td>
</tr>
<tr>
<td></td>
<td>$+ 1/3 K_{3g1&amp;2g2} (H_{1k}(s1) H_{1l}(s2) H_{1l}(s3) + H_{1k}(s1) H_{1l}(s1) H_{1l}(s2))$</td>
</tr>
</tbody>
</table>

**Meaning of the Crossterms**

The meaning of the various terms in 2-dimensional gm is the following:

- $g_m$, $K_{2gm}$, $K_{3gm}$ describe input behaviour, i.e. what is the input-output characteristics at a fixed output voltage.
- $g_o$, $K_{2go}$, $K_{3go}$ describe output characteristics at a fixed input voltage.
- Crossterms $K_{2k\&l}$, $K_{32k\&l}$, $K_{3k\&2l}$ model the fact that ID-VDS curve has different shape at different VGS voltages. They are essential in modelling e.g. Early effect.
1-DIMENSIONAL CAP

Derivation of capacitance model is started from non-linear charge equation \( Q(V) \) which is differentiated to get current. This results in the following distortion currents:

\[
Q(V_o + v) = Q(V_o) + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{\partial V^k} Q(V) \bigg|_{V_o} \cdot v^k
\]

\( C, K_2, K_3, \ldots \)

\[
i_C = j\omega C \cdot (v_k + K_2C' \cdot v_k^2 + K_3C' \cdot v_k^3)
\]

\[
C = \frac{\partial}{\partial V} Q(V)
\]

\[
K_2C = \frac{1}{2!} \frac{\partial^2}{\partial V^2} Q(V) = \frac{1}{2} \frac{\partial}{\partial V} C(V)
\]

\[
K_3C = \frac{1}{3!} \frac{\partial^3}{\partial V^3} Q(V) = \frac{1}{3!} \frac{\partial^2}{\partial V^2} C(V)
\]

---

### Table 1: CapiNL for general \( H_n() \)

<table>
<thead>
<tr>
<th>order</th>
<th>( i_{NL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>((s1+s2) K_{2c} H1(s1) H1(s2))</td>
</tr>
<tr>
<td>3</td>
<td>((s1 + s2 + s3) x (\ldots + 2/3 K_{2c} (H1(s1) H2(s2,s3) + H1(s2) H2(s1,s3) + H1(s3) H2(s1,s2)) )</td>
</tr>
</tbody>
</table>

---

### Table 1: Cap iNL for direct calculation

<table>
<thead>
<tr>
<th>Freq.</th>
<th>( i_{NL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_1 \pm \omega_2 )</td>
<td>( j(\omega_1 \pm \omega_2) K_{2c} V_{i,1,0} V_{i,0,\pm 1} )</td>
</tr>
<tr>
<td>( 2\omega_1 )</td>
<td>( j(2\omega_1) 0.5 K_{2c} (V_{i,1,0})^2 )</td>
</tr>
<tr>
<td>( 2\omega_1 \pm \omega_2 )</td>
<td>( j(2\omega_1 \pm \omega_2) \cdot \left( 0.75 K_{3c} (V_{i,1,0})^2 V_{i,0,\pm 1} \right) )</td>
</tr>
<tr>
<td>( 3\omega_1 )</td>
<td>( j3\omega_1 \cdot \left( 0.25 K_{3c} (V_{i,1,0})^3 \right) )</td>
</tr>
</tbody>
</table>

---

(C) 1999- Timo Rahkonen, University of Oulu, Oulu, Finland
MEASURING VOLTERRA KERNELS

Numerical values for Volterra kernels of complete blocks or subcircuits can be measured using vector analyzers and AM/AM and AM/PM curves. However, different contributions can only be measured using device level measurements.

IM3 ASYMMETRY

As shown before, IM3 term consists partially of 2nd order terms. In a nonlinear gm using a 2-tone test, IM3 terms are as shown left. Supposing equal amplitudes at s1 and s2, terms caused by K_{3g} match, but asymmetry may be caused by

- unequal response at fundamental tones s1 and s2. Usually, this is not the case as flat passband is desired
- the fact that envelope s2-s1 term appears in opposite phase (complex conjugate) in the lower and higher IM3 terms
- the response of 2nd harmonics (that are already quite far away from each other) may differ.

In more complex systems, partially compensating distortion mechanisms may appear (e.g. in a current mirror), but the accuracy of the cancellation tends to deteriorate with increasing frequency.
EXTENSIONS: TIME-VARYING VOLterra ANALYSIS

As such, Volterra analysis is suited for small signal, time-invariant systems. Thus, it is not well suited for analysing circuits with large periodic excitations like MOS samplers and mixers, although some mixers consisting of linear multipliers have been analysed using Volterra analysis (Wambacq et al., IEEE TCAS-II, March 99).

Volterra analysis can be extended to periodically time-varying systems, and e.g. a MOS sampler has been analysed in Yu, Sen & Leung: Time Varying Volterra Series and Its Application to the Distortion Analysis of a Sampling Mixer.