Nonlinear Microwave Circuit Design

F. Giannini

University of Roma, Tor Vergata, Italy

and

G. Leuzzi

University of L'Aquila, Italy

John Wiley & Sons, Ltd

Nonlinear Microwave Circuit Design

Nonlinear Microwave Circuit Design

F. Giannini

University of Roma, Tor Vergata, Italy

and

G. Leuzzi

University of L'Aquila, Italy

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Giannini, Franco, 1944-
Nonlinear microwave circuit design / Franco Giannini and Giorgio Leuzzi.
p. cm.
Includes bibliographical references and index.
ISBN 0-470-84701-8 (cloth: alk. paper)
1. Microwave circuits. 2. Electric circuits, Nonlinear. I. Leuzzi, Giorgio. II. Title.
TK7876.G53 2004
621.381′33 – dc22
2004004941

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-84701-8

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

Nonlinear microwave circuits is a field still open to investigation; however, many basic concepts and design guidelines are already well established. Many researchers and design engineers have contributed in the past decades to the development of a solid knowledge that forms the basis of the current powerful capabilities of microwave engineers.

This book is composed of two main parts. In the first part, some fundamental tools are described: nonlinear circuit analysis, nonlinear measurement, and nonlinear modeling techniques. In the second part, basic structure and design guidelines are described for some basic blocks in microwave systems, that is, power amplifiers, oscillators, frequency multipliers and dividers, and mixers. Stability in nonlinear operating conditions is also addressed.

A short description of fundamental techniques is needed because of the inherent differences between linear and nonlinear systems and because of the greater familiarity of the microwave engineer with the linear tools and concepts. Therefore, an introduction to some general methods and rules proves useful for a better understanding of the basic behaviour of nonlinear circuits. The description of design guidelines, on the other hand, covers some well-established approaches, allowing the microwave engineer to understand the basic methodology required to perform an effective design.

The book mainly focuses on general concepts and methods, rather than on practical techniques and specific applications. To this aim, simple examples are given throughout the book and simplified models and methods are used whenever possible. The expected result is a better comprehension of basic concepts and of general approaches rather than a fast track to immediate design capability. The readers will judge for themselves the success of this approach.

Finally, we acknowledge the help of many colleagues. Dr. Franco Di Paolo has provided invaluable help in generating simulation results and graphs. Prof. Tom Brazil, Prof. Aldo Di Carlo, Prof. Angel Mediavilla, and Prof. Andrea Ferrero, Dr. Giuseppe Ocera and Dr. Carlo Del Vecchio have contributed with relevant material. Prof. Giovanni Ghione and Prof. Fabrizio Bonani have provided important comments and remarks, although responsibility for eventual inaccuracies must be ascribed only to the authors. To all these people goes our warm gratitude.

Authors' wives and families are also acknowledged for patiently tolerating the extra work connected with writing a book.

Franco Giannini
Giorgio Leuzzi

1

Nonlinear Analysis Methods

1.1 INTRODUCTION

In this introduction, some well-known basic concepts are recalled, and a simple example is introduced that will be used in the following paragraphs for the illustration of the different nonlinear analysis methods.

Electrical and electronic circuits are described by means of voltages and currents. The equations that fulfil the topological constraints of the network, and that form the basis for the network analysis, are Kirchhoff's equations. The equations describe the constraints on voltages (mesh equations) or currents (nodal equations), expressing the constraint that the sum of all the voltages in each mesh, or, respectively, that all the currents entering each node, must sum up to zero. The number of equations is one half of the total number of the unknown voltages and currents. The system can be solved when the relation between voltage and current in each element of the network is known (constitutive relations of the elements). In this way, for example, in the case of nodal equations, the currents that appear in the equations are expressed as functions of the voltages that are the actual unknowns of the problem. Let us illustrate this by means of a simple example (Figure 1.1).

$$i_s + i_g + i_C = 0 \qquad \text{Nodal Kirchhoff's equation} \qquad (1.1)$$

$$\left.\begin{array}{l} i_s = i_s(t) \\ i_g = g \cdot v \\ i_C = C \cdot \dfrac{dv}{dt} \end{array}\right. \qquad \text{Constitutive relations of the elements} \qquad (1.2)$$

where $i_s(t)$ is a known, generic function of time. Introducing the constitutive relations into the nodal equation we get

$$C \cdot \frac{dv(t)}{dt} + g \cdot v(t) + i_s(t) = 0 \qquad (1.3)$$

Since in this case all the constitutive relations (eq. (1.2)) of the elements are linear and one of them is differential, the system (eq. (1.3)) turns out to be a linear differential

Nonlinear Microwave Circuit Design F. Giannini and G. Leuzzi
 ISBN: 0-470-84701-8

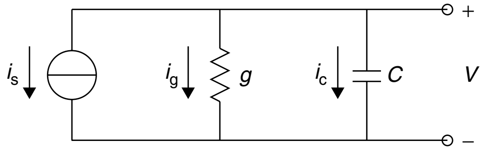

Figure 1.1 A simple example circuit

system in the unknown $v(t)$ (in this case a single equation in one unknown). One of the elements ($i_s(t)$) is a known quantity independent of voltage (known term), and the equation is non-homogeneous. The solution is found by standard solution methods of linear differential equations:

$$v(t) = v(t_0) \cdot e^{-\frac{g}{C}\cdot(t-t_0)} - \int_{t_0}^{t} \frac{e^{-\frac{g}{C}\cdot(t-\tau)}}{C} \cdot i_s(\tau) \cdot d\tau \qquad (1.4)$$

More generally, the solution can be written in the time domain as a convolution integral:

$$v(t) = v(t_0) + \int_{t_0}^{t} h(t-\tau) \cdot i_s(\tau) \cdot d\tau \qquad (1.5)$$

where $h(t)$ is the impulse response of the system.

The linear differential equation system can be transformed in the Fourier or Laplace domain. The well-known formulae converting between the time domain and the transformed Fourier domain, or frequency domain, and vice versa, are the Fourier transform and inverse Fourier transform respectively:

$$V(\omega) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} v(t) \cdot e^{-j\omega t} \cdot dt \qquad (1.6)$$

$$v(t) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} V(\omega) \cdot e^{j\omega t} \cdot d\omega \qquad (1.7)$$

By Fourier transforming eq. (1.3), after simple manipulation (Appendix A.1) we have

$$V(\omega) = H(\omega) \cdot I_s(\omega) \qquad (1.8)$$

where $H(\omega)$ and $I_s(\omega)$ are obtained by Fourier transformation of the time-domain functions $h(t)$ and $i_s(t)$; $H(\omega)$ is the transfer function of the circuit.

We can describe this approach from another point of view: if the current $i_s(t)$ is sinusoidal, and we look for the solution in the permanent regime, we can make use of phasors, that is, complex numbers such that

$$v(t) = \mathrm{Im}[V \cdot e^{j\omega t}] \qquad (1.9)$$

and similarly for the other electrical quantities; the voltage phasor V corresponds to the $V(\omega)$, defined above. Then, by replacing in eq. (1.3) we get

$$j\omega C \cdot V + g \cdot V + I_s = (g + j\omega C) \cdot V + I_s = Y \cdot V + I_s = 0 \qquad (1.10)$$

and the solution is easily found by standard solution methods for linear equations:

$$V = \frac{I_s}{Y} \qquad \frac{1}{Y(\omega)} = H(\omega) = \frac{1}{g + j\omega C} \qquad (1.11)$$

Let us now introduce nonlinearities. Nonlinear circuits are electrical networks that include elements with a nonlinear relation between voltage and current; as an example, let us consider a nonlinear conductance (Figure 1.2) described by

$$i_g(v) = i_{max} \cdot tgh\left(\frac{g \cdot v}{i_{max}}\right) \qquad (1.12)$$

that is, a conductance saturating to a maximum current value i_{max} (Figure 1.3).

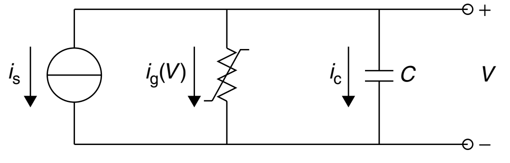

Figure 1.2 The example circuit with a nonlinear conductance

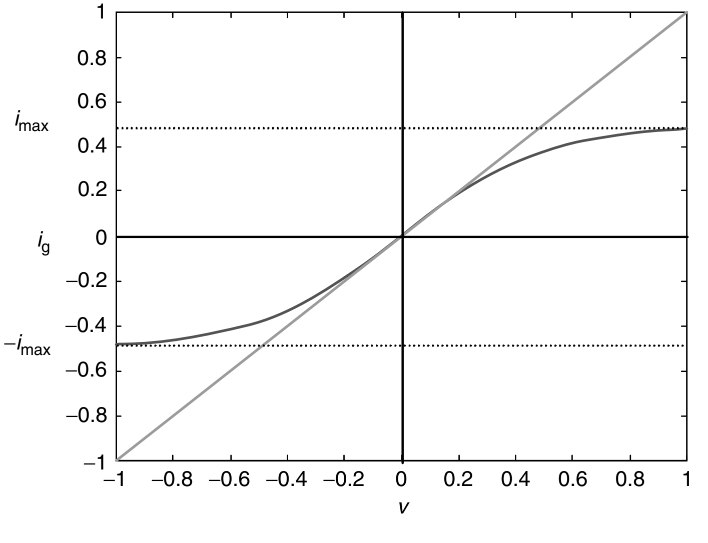

Figure 1.3 The current–voltage characteristic of the nonlinear conductance

When we introduce this relation in Kirchhoff's equation, we have a nonlinear differential equation (in general, a system of nonlinear differential equations)

$$C \cdot \frac{\mathrm{d}v(t)}{\mathrm{d}t} + i_{\max} \cdot tgh\left(\frac{g \cdot v(t)}{i_{\max}}\right) + i_{\mathrm{s}}(t) = 0 \tag{1.13}$$

that has no explicit solution. Moreover, contrary to the linear case, transformation into the Fourier or Laplace domain is not applicable.

Practical solutions to this type of problems fall into two main categories: direct numerical integration in the time domain, and numerical solution through series expansion; they are described in some detail in the following paragraphs.

1.2 TIME-DOMAIN SOLUTION

In this paragraph, the solution of the nonlinear differential Kirchhoff's equations by direct numerical integration in the time domain is described. Advantages and drawbacks are described, together with some improvements to the basic approach.

1.2.1 General Formulation

The time-domain solution of the nonlinear differential equations system that describes the circuit (Kirchhoff's equations) can be performed by means of standard numerical integration methods. These methods require the discretisation of the time variable, and likewise the sampling of the known and unknown time-domain voltages and currents at the discretised time instants.

The time variable, in general a real number in the interval $[t_0, \infty]$, is discretised, that is, considered as a discrete variable:

$$t = t_k \qquad k = 1, 2, \ldots \quad t \in [t_0, \infty] \tag{1.14}$$

All functions of time are evaluated only at this set of values of the time variable. The differential equation becomes a finite-difference equation, and the knowledge of the unknown function $v(t)$ is reduced to the knowledge of a discrete set of values:

$$v_k = v(t_k) \qquad k = 1, 2, \ldots \quad t \in [t_0, \infty] \tag{1.15}$$

Similarly, the known function $i_{\mathrm{s}}(t)$ is computed only at a discrete set of values:

$$i_{\mathrm{s},k} = i_{\mathrm{s}}(t_k) \qquad k = 1, 2, \ldots \quad t \in [t_0, \infty] \tag{1.16}$$

The obvious advantage of this scheme is that the derivative with respect to time becomes a finite-difference incremental ratio:

$$\frac{\mathrm{d}v(t)}{\mathrm{d}t} = \frac{v_k - v_{k-1}}{t_k - t_{k-1}} \tag{1.17}$$

Let us apply the discretisation to our example. Equation (1.13) becomes

$$C \cdot \left(\frac{v_k - v_{k-1}}{t_k - t_{k-1}} \right) + i_{\max} \cdot tgh \left(\frac{g \cdot v_k}{i_{\max}} \right) + i_{s,k} = 0 \qquad k = 1, 2, \ldots \tag{1.18}$$

where we have replaced the derivative with respect to time, defined in the continuous time, with the incremental ratio, defined in the discrete time. In this formulation, the discrete derivative is computed between the current point k, where also the rest of the equation is evaluated, and the previous point $k-1$. There is, however, another possibility:

$$C \cdot \left(\frac{v_k - v_{k-1}}{t_k - t_{k-1}} \right) + i_{\max} \cdot tgh \left(\frac{g \cdot v_{k-1}}{i_{\max}} \right) + i_{s,k-1} = 0 \qquad k = 1, 2, \ldots \tag{1.19}$$

In the second case, the rest of the equation, including the nonlinear function of the voltage, is evaluated in the previous point $k-1$. In both cases, if an initial value is known for the problem, that is, if the value $v_0 = v(t_0)$ is known, then the problem can be solved iteratively, time instant after time instant, starting from the initial time instant $t_0$at $k = 0$. In the case of our example, the initial value is the voltage at which the capacitance is initially charged.

The two cases of eq. (1.18) and eq. (1.19) differ in complexity and accuracy. In the case of eq. (1.19), the unknown voltage v_k at the current point k appears only in the finite-difference incremental ratio; the equation can be therefore easily inverted, yielding

$$v_k = v_{k-1} - \frac{(t_k - t_{k-1})}{C} \cdot i_{\max} \cdot tgh \left(\frac{g \cdot v_{k-1}}{i_{\max}} \right) + i_{s,k-1} \qquad k = 1, 2, \ldots \tag{1.20}$$

This approach allows the explicit calculation of the unknown voltage v_k at the current point k, once the solution at the previous point $k-1$ is known. The obvious advantage of this approach is that the calculation of the unknown voltage requires only the evaluation of an expression at each of the sampling instants t_k. A major disadvantage of this solution, usually termed as ‘explicit’, is that the stability of the solution cannot be guaranteed. In general, the solution found by any discretised approach is always an approximation; that is, there will always be a difference between the actual value of the exact (unknown) solution $v(t)$ at each time instant t_k and the values found by this method

$$v(t_k) \neq v_k \qquad v(t_k) - v_k = \Delta v_k \qquad k = 1, 2, \ldots \tag{1.21}$$

because of the inherently approximated nature of the discretisation with respect to the originally continuous system. The error Δv_k due to an explicit formulation, however, can increase without limits when we proceed in time, even if we reduce the discretisation step $t_k - t_{k-1}$, and the solution values can even diverge to infinity. Even if the values do not diverge, the error can be large and difficult to reduce or control; in fact, it is not guaranteed that the error goes to zero even if the time discretisation becomes arbitrarily dense and the time step arbitrarily small. In fact, for simple circuits the explicit solution is usually adequate, but it is prone to failure for strongly nonlinear circuits. This explicit formulation is also called ‘forward Euler’ integration algorithm in numerical analysis [1, 2].

In the case of the formulation of eq. (1.18), the unknown voltage v_k appears not only in the finite-difference incremental ratio but also in the rest of the equation, and in particular within the nonlinear function. At each time instant, the unknown voltage v_k must be found as a solution of the nonlinear implicit equation:

$$C \cdot \left(\frac{v_k - v_{k-1}}{t_k - t_{k-1}} \right) + i_{\max} \cdot tgh \left(\frac{g \cdot v_k}{i_{\max}} \right) + i_{s,k} = F(v_k) = 0 \qquad k = 1, 2, \ldots \tag{1.22}$$

This equation in general must be solved numerically, at each time instant t_k. Any zero-searching numerical algorithm can be applied, as for instance the fixed-point or Newton–Raphson algorithms. A numerical search requires an initial guess for the unknown voltage at the time instant t_k and hopefully converges toward the exact solution in a short number of steps; the better the initial guess, the shorter the number of steps required for a given accuracy. As an example, the explicit solution can be a suitable initial guess. The iterative algorithm is stopped when the current guess is estimated to be reasonably close to the exact solution. This approach is also called 'backward Euler' integration scheme in numerical analysis [1, 2].

An obvious disadvantage of this approach w.r.t. the explicit one is the much higher computational burden, and the risk of non-convergence of the iterative zero-searching algorithm. However, in this case the error Δv_k can be made arbitrarily small by reducing the time discretisation step $t_k - t_{k-1}$, at least in principle. Numerical round-off errors due to finite number representation in the computer is however always present.

The discretisation of the t_k can be uniform, that is, with a constant step Δt, so that

$$t_{k+1} = t_k + \Delta t \qquad t_k = t_0 + k \cdot \Delta t \qquad k = 1, 2, \ldots \tag{1.23}$$

This approach is not the most efficient. A variable time step is usually adopted with smaller time steps where the solution varies rapidly in time and larger time steps where the solution is smoother. The time step is usually adjusted dynamically as the solution proceeds; in particular, a short time step makes the solution of the nonlinear eq. (1.22) easier. A simple procedure when the solution of eq. (1.22) becomes too slow or does not converge at all consists of stopping the zero-searching algorithm, reducing the time step and restarting the algorithm.

There is an intuitive relation between time step and accuracy of the solution. For a band-limited signal in permanent regime, an obvious criterion for time discretisation is given by Nykvist's sampling theorem. If the time step is larger than the sampling time required by Nykvist's theorem, the bandwidth of the solution will be smaller than that of the actual solution and some information will be lost. The picture is not so simple for complex signals, but the principle still holds: the finer the time step, the more accurate the solution. Since higher frequency components are sometimes negligible for practical applications, a compromise between accuracy and computational burden is usually chosen. In practical algorithms, more elaborate schemes are implemented, including modified nodal analysis, advanced integration schemes, sophisticated adaptive time-step schemes and robust zero-searching algorithms [3–7].

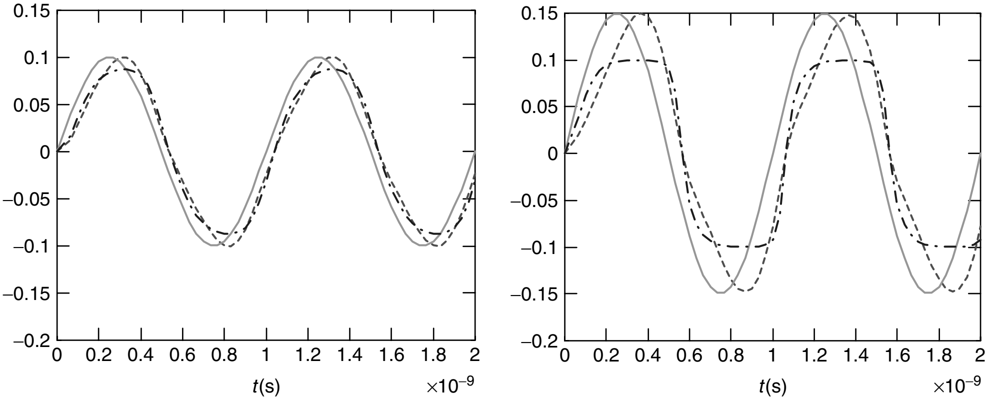

Figure 1.4 Currents and voltages in the example circuit for two different amplitudes of a sinusoidal input current

With the view to illustrate, the time-domain solution of our example circuit is given for a sinusoidal input current, for the following values of the circuit elements (Figure 1.4):

$$g = 10 \text{ mS} \qquad C = 500 \text{ fF} \qquad f = 1 \text{ GHz} \tag{1.24}$$

A simple implicit integration scheme is used, with a uniform time step of $\Delta t =$ 33.3 ps (30 discretisation points per period). The plots show the input current i_s (———), the voltage v (- - - -) and the current in the nonlinear resistor i_g (-·-·-·), for an input current of $i_{s,max} = 100$ mA (a) and for a larger input of $i_{s,max} = 150$ mA (b).

As an additional example, the response of the same circuit to a 1 mA input current step is shown in Figure 1.5, where a uniform time step of $\Delta t = 10$ ps is used.

Time-domain direct numerical integration is very general. No limitation on the type or stiffness of the nonlinearity is imposed. Transient as well as steady state behaviour are computed, making it very suitable, for instance, for oscillator analysis, where the determination of the onset of the oscillations is required. Instabilities are also well predicted, provided that the time step is sufficiently fine. Also, digital circuits are easily analysed.

1.2.2 Steady State Analysis

Direct numerical integration is not very efficient when the steady state regime is sought, especially when large time constant are present in a circuit, like those introduced by the bias circuitry. In this case, a large number of microwave periods must be analysed before the reactances in the bias circuitry are charged, starting from an arbitrary initial state. Since the time step must be chosen small enough in order that the microwave voltages and currents are sufficiently well sampled, a large number of time steps must be computed before the steady state is reached. The same is true when the spectrum of the signal includes components both at very low and at very high frequencies, as in the case

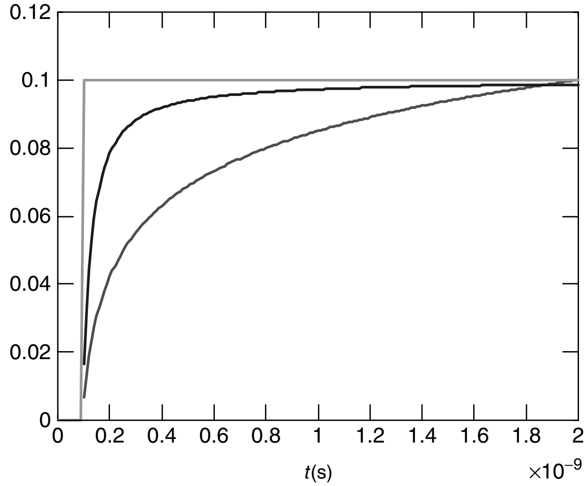

Figure 1.5 Currents and voltages in the example circuit for a step input current

of two sinusoids with very close frequencies, or of a narrowband modulated carrier. The time step must be small enough to accurately sample the high-frequency carrier, but the overall repetition time, that is, the period of the envelope, is comparatively very long.

The case when a long time must be waited for the steady state to be reached can be coped with by a special arrangement of the time-domain integration, called 'shooting method' [8–11]. It is interesting especially for non-autonomous circuits, when an external periodic input signal forces the circuit to a periodic behaviour; in fact, in autonomous circuits like oscillators, the analysis of the transient is also interesting, for the check of the correct onset of the oscillation and for the detection of spurious oscillations and instabilities. In the shooting method, the period of the steady state solution must be known in advance: this is usually not a problem, since it is the period of the input signal. The time-domain integration is carried over only for one period starting from a first guess of the initial state, that is, the state at the beginning of a period in steady state conditions, and then the state at the end of the period is checked. In the case of our example, the voltage at the initial time t_0 is guessed as

$$v_0 = v(t_0) \tag{1.25}$$

and the voltage at the end of the period T is computed after integration over one period:

$$v(T) = v(t_K) = v_K \qquad (k = 1, 2, \ldots, K) \tag{1.26}$$

This final voltage is a numerical function of the initial voltage:

$$v(T) = f(v_0) \tag{1.27}$$

If the initial voltage is the actual voltage at which the capacitor is charged at the beginning of a period $t = t_0$ in the periodic steady state regime, that is, if it is the solution to our periodic problem, the final voltage after a period must be identical to it:

$$v(T) = f(v_0) = v_0 \tag{1.28}$$

In case this is not true, the correct value of the initial voltage is searched by adjustment of the initial guess v_0 until the final value $v(T)$ comes out to be equal to it. This can be done automatically by a zero-searching algorithm, where the unknown is the initial voltage v_0, and the function to be made equal to zero is

$$F(v_0) = v(T) - v_0 = f(v_0) - v_0 \tag{1.29}$$

Each computation of the function $F(v_0)$ consists of the time-domain numerical integration over one period T from the initial value of the voltage v_0 to the final value $v(T) = v_K$. The zero-searching algorithm will require several iterations, that is, several integrations over a period; if the number of iterations required by the zero-searching algorithm to converge to the solution is smaller than the number of periods before the attainment of the steady state by standard integration from an initial voltage, then the shooting algorithm is a convenient alternative.

1.2.3 Convolution Methods

The time-domain numerical integration method has in fact two major drawbacks: on the one hand the number of equations grows with the dimension of the circuit, even when the largest part of it is linear. On the other hand, all the circuit elements must have a time-domain constitutive relation in order for the equations to be written in time domain. It is well known that in many practical cases the linear part of nonlinear microwave circuits is large and that it is best described in the frequency domain; as an example, consider the matching and bias networks of a microwave amplifier. In particular, distributed elements are very difficult to represent in the time domain. A solution to these problems is represented by the 'convolution method' [12–17]. By this approach, a linear subcircuit is modelled by means of frequency-domain data, either measured or simulated; then, the frequency-domain representation is transformed into time-domain impulse response, to be used for convolution in the time domain with the rest of the circuit. In fact, this mixed time-frequency domain approach is somehow dual to the harmonic balance method, to be described in a later paragraph. In order to better understand the approach, a general scheme of time-frequency domain transformations for periodic and aperiodic functions is shown in Appendix A.2.

The basic scheme of the convolution approach is based on the application of eq. (1.5), with the relevant impulse response, to the linear subcircuit. Let us illustrate this principle with our test circuit, where a shunt admittance has been added (Figure 1.6).

Equation (1.13) becomes

$$C \cdot \frac{\mathrm{d}v(t)}{\mathrm{d}t} + i_{\max} \cdot tgh\left(\frac{g \cdot v(t)}{i_{\max}}\right) + i_{\mathrm{y}}(t) + i_{\mathrm{s}}(t) = 0 \tag{1.30}$$

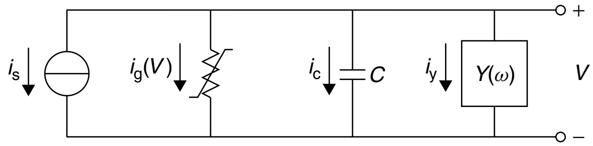

Figure 1.6 The example nonlinear circuit with an added shunt network

where the current through the shunt admittance is defined in the frequency domain:

$$I_{\mathrm{y}}(\omega) = Y(\omega) \cdot V(\omega) \tag{1.31}$$

The time-domain current through the shunt admittance is expressed by means of the convolution integral (1.5) as

$$i_{\mathrm{y}}(t) = i_{\mathrm{y}}(t_0) + \int_{t_0}^{t} y(t-\tau) \cdot v(\tau) \cdot \mathrm{d}\tau \tag{1.32}$$

where

$$y(t) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} Y(\omega) \cdot \mathrm{e}^{j\omega t} \cdot \mathrm{d}\omega \tag{1.33}$$

The integral in eq. (1.32) is computed numerically; if the impulse response $y(t)$ is limited in time, this becomes

$$i_{\mathrm{y}}(t_k) = i_{\mathrm{y},k} = \sum_{m=0}^{M} y_m \cdot v_{k-m} \tag{1.34}$$

Discretisation of eq. (1.30) then yields

$$C \cdot \left(\frac{v_k - v_{k-1}}{t_k - t_{k-1}}\right) + i_{\max} \cdot tgh\left(\frac{g \cdot v_k}{i_{\max}}\right) + \sum_{m=0}^{M} y_m \cdot v_{k-m} + i_{\mathrm{s},k} = F(v_k) = 0 \tag{1.35}$$

where the unknown v_k appears also in the convolution summation with a linear term. This is a modified form of eq. (1.18) and must be solved numerically with the same procedure.

A first remark on this approach is that the algorithm becomes heavier: on the one hand, the convolution with past values of the electrical variables must be recomputed at each time step k, increasing computing time; on the other hand, the values of the electrical variables must be stored for as many time instants as corresponding to the duration of the impulse response, increasing data storage requirements.

An additional difficulty is related to the time step. A time-domain solution may use an adaptive time step for better efficiency of the algorithm; however, the time step of the discrete convolution in eq. (1.34) is a fixed number. This means that at the time instants where the convolution must be computed, the quantities to be used in the convolution

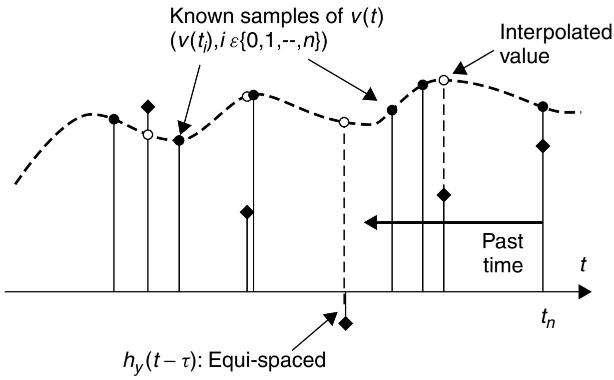

Figure 1.7 Sampling time instants and convolution time instants

are not available. An interpolating algorithm must be used to allow for the convolution to be computed, introducing an additional computational overhead and additional error (Figure 1.7).

The assumption of an impulse response limited in time requires some comments. An impulse response of infinite duration corresponds to an infinite bandwidth of the frequency-domain admittance. The latter however is usually known only within a limited frequency band, both in the case of experimental data and in the case of numerical modelling. A truncated frequency-domain admittance produces a non-causal impulse response when the inverse Fourier transform (eq. (1.33)) is applied (Figure 1.8).

As an alternative, the frequency-domain admittance can be 'windowed' by means, for example, of a low-pass filter, forcing the admittance to (almost) zero just before the limiting frequency f_m; however, this usually produces a severe distortion in phase, so that the accuracy will be unacceptably affected.

An alternative approach is to consider the impulse response as a discrete function of time, with finite duration in time. From the scheme in Appendix A.2, the corresponding spectrum is periodic in the frequency domain. Therefore, the admittance must be extended periodically in the frequency domain (Figure 1.9).

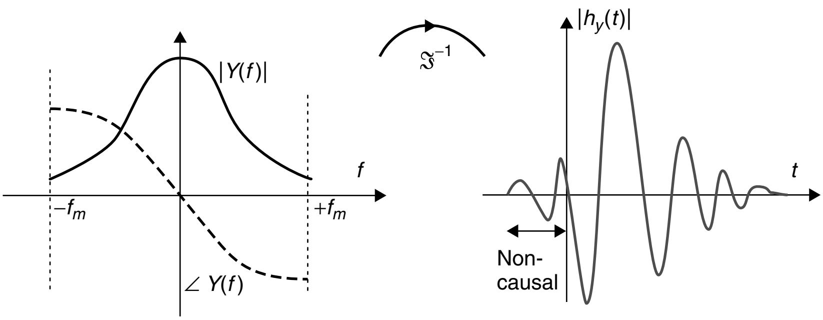

Figure 1.8 Non-causal impulse response generated by artificially band-limited frequency data

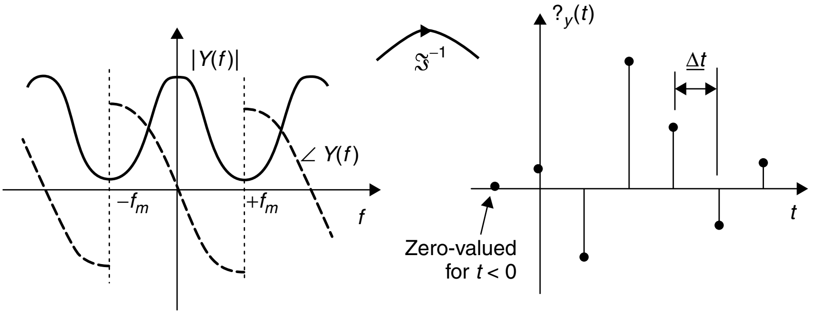

Figure 1.9 Periodical extension of frequency-domain data

In order to satisfy causality, however, the periodic extension must satisfy the Hilbert transform:

$$\hat{Y}(\omega) = \hat{G}(\omega) + j\hat{B}(\omega) \tag{1.36a}$$

$$\begin{aligned} \hat{G}(\omega) &= \hat{G}(0) + \frac{1}{2\pi} \cdot \int_{-\pi}^{+\pi} \hat{B}(\alpha) \cdot \cot\left(\frac{\omega - \alpha}{2}\right) \cdot \mathrm{d}\alpha \\ \hat{B}(\omega) &= -\frac{1}{2\pi} \cdot \int_{-\pi}^{+\pi} \hat{G}(\alpha) \cdot \cot\left(\frac{\omega - \alpha}{2}\right) \cdot \mathrm{d}\alpha \end{aligned} \tag{1.36b}$$

This can be done by suitable procedures [18]. Care must be taken that the frequency-domain data be available in a band wide enough to make the extension error negligible. This is true if the spectrum of the voltages and currents in the circuit are narrower than the frequency 'window'. In practice, the frequency data must extend to frequencies where the signal spectrum has a negligible amplitude (Figure 1.10).

Several microwave or general CAD programmes are now commercially available implementing this scheme, allowing easy inclusion of passive networks described in the frequency domain; as an example, ultra-wide-band systems using short pulses often require the evaluation of pulse propagation through the transmit antenna/channel/receive antenna path, typically described in the frequency domain.

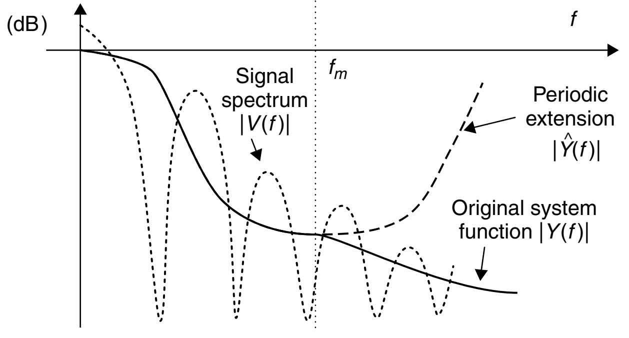

Figure 1.10 Approximation in the periodical extension of frequency-domain data

1.3 SOLUTION THROUGH SERIES EXPANSION

An alternative to direct discretisation of a difficult equation is the assumption of some hypotheses on the solution, in this case, on the unknown function $v(t)$. A typical hypothesis is that the solution can be expressed as an infinite sum of simple terms, and that the terms are chosen in such a suitable way that the first ones already include most of the information on the function. The series is therefore truncated after the first few terms. When replaced in the original equation, the solution in the form of a series allows the splitting of the original equation into infinite simpler equations (one per term of the series). Only a few of the simpler equations are solved however, corresponding to the first terms of the series.

In the following sections, two types of series expansions will be described: the Volterra and the Fourier series expansions, which are the only ones currently used.

1.3.1 Volterra Series

In this paragraph, the solution of the nonlinear differential Kirchhoff's equations by means of the Volterra series is described. Advantages and drawbacks are illustrated, together with some examples.

It has been shown above that the solution of our example circuit in the linear case is (eq. (1.4))

$$v(t) = v(t_0) \cdot \mathrm{e}^{-\frac{g}{C}\cdot(t-t_0)} - \int_{t_0}^{t} \frac{\mathrm{e}^{-\frac{g}{C}\cdot(t-\tau)}}{C} \cdot i_{\mathrm{s}}(\tau) \cdot \mathrm{d}\tau \tag{1.37}$$

that can be put in the general form known as the convolution integral (eq. (1.5)):

$$y(t) = y(t_0) + \int_{t_0}^{t} h(t-\tau) \cdot x(\tau) \cdot \mathrm{d}\tau \tag{1.38}$$

Equation (1.38) can be interpreted in the following way: the output signal of a linear system is the infinite sum (integral) of all contributions due to the input signal at all the time instants in the past, weighted by a function called impulse response, representing the effect of the transfer through the system. In fact, the transfer function represents the 'memory' of the system, and normally becomes smaller as the time elapsed from the time instant of the input contribution to the current time instant becomes larger. If the system is instantaneous (e.g. a resistance), the impulse response is a delta function $k \cdot \delta(t)$, and the integral becomes a simple product:

$$y(t) = k \cdot x(t) \tag{1.39}$$

In this case, the output signal at any given time responds only to the input at that time and has no memory of past values of the input itself.

As we have seen above, given the linearity of the system, its response can be transformed in the Laplace or Fourier domain:

$$y(t) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} Y(\omega) \cdot \mathrm{e}^{j\omega t} \cdot \mathrm{d}\omega = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} H(\omega) \cdot X(\omega) \cdot \mathrm{e}^{j\omega t} \cdot \mathrm{d}\omega \quad (1.40)$$

where

$$X(\omega) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} x(t) \cdot \mathrm{e}^{-j\omega t} \cdot \mathrm{d}t \qquad H(\omega) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} h(t) \cdot \mathrm{e}^{-j\omega t} \cdot \mathrm{d}t \quad (1.41)$$

Equation (1.40) can be interpreted in the following way: the output signal of a linear system is the infinite sum (integral) of all spectral contributions of the input signal, weighted by a function of frequency called transfer function that represents the effect of the transfer through the system. We note explicitly that the spectrum occupancy of the output signal is the same of the spectrum of the input signal, or smaller if the transfer function suppresses a part of it, as for example in a filter. If the system is instantaneous, the transfer function is a constant k and does not alter the harmonic content of the input signal:

$$Y(\omega) = k \cdot X(\omega) \quad (1.42)$$

An extension of this type of formulation to nonlinear circuits has been proposed by the mathematician Vito Volterra early in last century [19–29], in the form

$$\begin{aligned} y(t) = & \int_{-\infty}^{t} h_1(t - \tau_1) \cdot x(\tau_1) \cdot \mathrm{d}\tau_1 \\ & + \int_{-\infty}^{t} \int_{-\infty}^{t} h_2(t - \tau_1, t - \tau_2) \cdot x(\tau_1) \cdot x(\tau_2) \cdot \mathrm{d}\tau_2 \cdot \mathrm{d}\tau_1 + \cdots \end{aligned} \quad (1.43)$$

where the first term is the linear one (first-order term) and the following ones are higher-order terms that take into account the effect on nonlinearities. The hypothesis in this case of series expansion is that the nonlinearities are weak and that only a few higher-order terms will be sufficient to describe their effect. The generalised transfer functions of nth order $h_n(t_1, \ldots, t_n)$ are called nuclei of nth order. In order to compute the nuclei analytically, it is also required that the nonlinearity be expressed as a power series:

$$i_{\mathrm{g}}(v) = g_0 + g_1 \cdot v + g_2 \cdot v^2 + g_3 \cdot v^3 + \cdots \quad (1.44)$$

a requirement that will be justified below. It is clear that any nonlinearity can be expanded in power series, but only within a limited voltage and current range.

The Volterra series can be interpreted in the following way: the output signal of a nonlinear system is composed by an infinite number of terms of increasing order; each term is the infinite sum (integral) of all contributions due to the input signal multiplied by itself n times, where n is the order of the term, in any possible combination of time instants in the past, weighted by a function called nucleus of nth order, representing the effect of the transfer through the system for that order. The nuclei represent also, in this case, the 'memory' of the system, and they represent the way in which the system responds to the presence of an input signal at different time instants in the past; since the system is nonlinear, its response to the input signal applied at a certain time instant is not independent of the value of the input signal at a different time instant. All the combinations must therefore be taken into account through multiple integration. The nuclei become normally smaller as the time elapsed from the time instants of the input contributions and the current time instant becomes larger. If the system is instantaneous, the nuclei are delta functions, and the *n*th order integral becomes the *n*th power of the input:

$$y(t) = k_1 \cdot x(t) + k_2 \cdot x^2(t) + \cdots \tag{1.45}$$

A generalisation of the Fourier transform can be defined for the nonlinear case: if we define

$$\begin{aligned} H_n(\omega_1, \ldots, \omega_n) &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \\ &\quad \times \mathrm{e}^{-j(\omega_1\tau_1 + \cdots \omega_n\tau_n)} \cdot \mathrm{d}\tau_n, \cdots, \mathrm{d}\tau_1 \end{aligned} \tag{1.46a}$$

$$\begin{aligned} h_n(\tau_1, \ldots, \tau_n) &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(\omega_1, \ldots, \omega_n) \\ &\quad \times \mathrm{e}^{j(\omega_1\tau_1 + \cdots \omega_n\tau_n)} \cdot \mathrm{d}\omega_n, \ldots, \mathrm{d}\omega_1 \end{aligned} \tag{1.46b}$$

the Volterra series becomes (Appendix A.3)

$$\begin{aligned} Y(\omega) = \cdots + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(\omega_1, \ldots, \omega_n) \cdot X(\omega_1) \cdot \ldots \cdot X(\omega_n) \\ \times \delta(\omega - \omega_1 - \cdots - \omega_n) \cdot \mathrm{d}\omega_n, \ldots, \mathrm{d}\omega_1 + \cdots \end{aligned} \tag{1.47}$$

Equation (1.47) can be interpreted in the following way: the output signal of a nonlinear system is the sum of an (infinite) number of terms of given orders; each term is the infinite sum (integral) of all spectral contributions of the input signal multiplied by itself *n* times, where *n* is the order of the term, in any possible combination of frequencies, weighted by a function of frequency called frequency-domain nucleus of *n*th order, which represents the effect of the transfer through the system for that order. The frequency of each spectral contribution to the output signal is the algebraic sum of the frequencies of the contributing terms of the input signal; in other words, the spectrum of the output signal will not be zero at a given frequency if there is a combination of the input frequency *n* times that equals this frequency. We note explicitly that the spectrum occupancy of the output signal is now broader than that of the spectrum of the input signal.

Let us clarify these concepts by illustrating the special case of periodic signals. If the input signal is a periodic function, its spectrum is discrete and the integrals become summations; in the case of an ideal, complex single tone

$$x(t) = A \cdot \mathrm{e}^{j\omega_0 t} \qquad X(\omega) = A \cdot \delta(\omega - \omega_0) \tag{1.48}$$

the output signal is given by

$$y(t) = A \cdot H_1(\omega_0) \cdot \mathrm{e}^{j\omega_0 t} + A^2 \cdot H_2(\omega_0, \omega_0) \cdot \mathrm{e}^{j2\omega_0 t} + \cdots \tag{1.49a}$$

$$Y(\omega) = A \cdot H_1(\omega_0) \cdot \delta(\omega - \omega_0) + A^2 \cdot H_2(\omega_0, \omega_0) \cdot \delta(\omega - 2\omega_0) + \cdots \tag{1.49b}$$

The second-order term generates a signal component at second-harmonic frequency, and so on for higher-order terms. In the case of a real single tone, that is, a couple of ideal single tones at opposite frequencies,

$$x(t) = A \cdot \frac{(e^{j\omega_0 t} + e^{-j\omega_0 t})}{2} = A \cdot \cos(\omega_0 t) \tag{1.50a}$$

$$X(\omega) = \frac{A}{2} \cdot \delta(\omega - \omega_0) + \frac{A}{2} \cdot \delta(\omega + \omega_0) \tag{1.50b}$$

The output signal is given by

$$\begin{aligned}
y(t) &= y_1(t) + y_2(t) + y_3(t) + \cdots \qquad (1.51)\\
y_1(t) &= A \cdot \overline{H_1(\omega_0)} \cdot \cos(\omega_0 t)\\
y_2(t) &= \frac{A^2}{2} \cdot \overline{H_2(\omega_0, -\omega_0)} + \frac{A^2}{2} \cdot \overline{H_2(\omega_0, \omega_0)} \cdot \cos(2\omega_0 t)\\
y_3(t) &= \frac{3A^3}{4} \cdot \overline{H_3(\omega_0, \omega_0, -\omega_0)} \cdot \cos(\omega_0 t) + \frac{A^3}{4} \cdot \overline{H_3(\omega_0, \omega_0, \omega_0)} \cdot \cos(3\omega_0 t)\\
Y(\omega) &= Y_1(\omega) + Y_2(\omega) + Y_3(\omega) + \cdots\\
Y_1(\omega) &= \frac{A}{2} \cdot H_1(\omega_0) \cdot \delta(\omega - \omega_0) + \frac{A}{2} \cdot H_1(-\omega_0) \cdot \delta(\omega + \omega_0)\\
Y_2(\omega) &= \frac{A^2}{4} \cdot H_2(-\omega_0, -\omega_0) \cdot \delta(\omega + 2\omega_0) + \frac{A^2}{4} \cdot H_2(\omega_0, \omega_0) \cdot \delta(\omega - 2\omega_0)\\
&\quad + \frac{A^2}{4} \cdot H_2(-\omega_0, \omega_0) \cdot \delta(\omega) + \frac{A^2}{4} \cdot H_2(\omega_0, -\omega_0) \cdot \delta(\omega)\\
Y_3(\omega) &= \frac{A^3}{8} \cdot H_3(-\omega_0, -\omega_0, -\omega_0) \cdot \delta(\omega + 3\omega_0) + \frac{A^3}{8} \cdot H_2(\omega_0, \omega_0, \omega_0) \cdot \delta(\omega - 3\omega_0)\\
&\quad + \frac{A^3}{8} \cdot H_2(-\omega_0, -\omega_0, \omega_0) \cdot \delta(\omega + \omega_0) + \frac{A^3}{8} \cdot H_2(-\omega_0, \omega_0, -\omega_0) \cdot \delta(\omega + \omega_0)\\
&\quad + \frac{A^3}{8} \cdot H_2(\omega_0, -\omega_0, -\omega_0) \cdot \delta(\omega + \omega_0) + \frac{A^3}{8} \cdot H_2(-\omega_0, \omega_0, \omega_0) \cdot \delta(\omega - \omega_0)\\
&\quad + \frac{A^3}{8} \cdot H_2(\omega_0, -\omega_0, \omega_0) \cdot \delta(\omega - \omega_0) + \frac{A^3}{8} \cdot H_2(\omega_0, \omega_0, -\omega_0) \cdot \delta(\omega - \omega_0)
\end{aligned}$$

The first-order terms generate the linear output signal at input frequency; the second-order terms generate a zero-frequency signal (rectified signal) and a double-frequency signal (second harmonic); the third-order terms generate a signal at input frequency (compression or expansion) and at triple frequency (third harmonic); and so on. The higher-order terms are the nonlinear contribution to the distortion of the signal and are proportional to the nth power of the input where n is the order of the term. A graphical representation of the spectra is depicted in Figure 1.11.

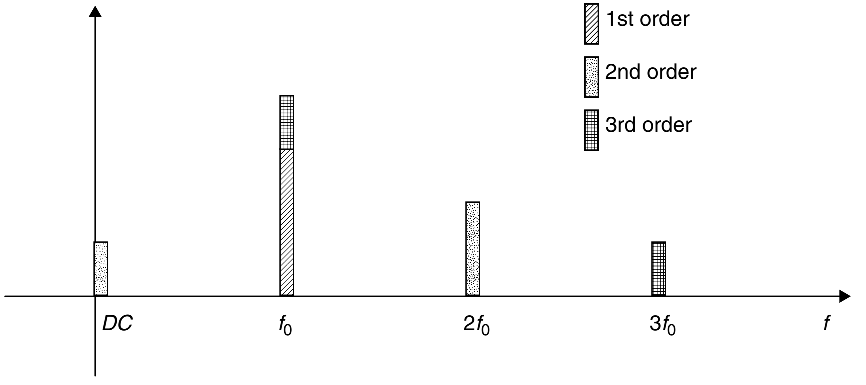

Figure 1.11 Contributions of the terms of the Volterra series to the spectrum of a single-tone signal

Let us now consider a two-tone input signal, in the form

$$x(t) = A_1 \cdot \cos(\omega_1 t) + A_2 \cdot \cos(\omega_2 t) \tag{1.52}$$

The output signal is given by

$$\begin{aligned}
y(t) &= y_1(t) + y_2(t) + y_3(t) + \cdots \qquad (1.53)\\
y_1(t) &= A_1 \cdot \overline{H_1(\omega_1)} \cdot \cos(\omega_1 t) + A_2 \cdot \overline{H_1(\omega_2)} \cdot \cos(\omega_2 t)\\
y_2(t) &= \frac{A_1^2}{2} \cdot \overline{H_2(\omega_1, -\omega_1)} + \frac{A_2^2}{2} \cdot \overline{H_2(\omega_2, -\omega_2)}\\
&+ \frac{A_1^2}{2} \cdot \overline{H_2(\omega_1, \omega_1)} \cdot \cos(2\omega_1 t) + \frac{A_1 A_2}{2} \cdot \overline{H_2(\omega_1, \omega_2)} \cdot \cos((\omega_1 + \omega_2)t)\\
&+ \frac{A_2^2}{2} \cdot \overline{H_2(\omega_2, \omega_2)} \cdot \cos(2\omega_2 t)\\
y_3(t) &= \frac{3A_1^3}{4} \cdot \overline{H_3(\omega_1, \omega_1, -\omega_1)} \cdot \cos(\omega_1 t) + \frac{3A_2^3}{4} \cdot \overline{H_3(\omega_2, \omega_2, -\omega_2)} \cdot \cos(\omega_2 t)\\
&+ \frac{3A_1 A_2^2}{4} \cdot \overline{H_3(\omega_1, \omega_2, -\omega_2)} \cdot \cos(\omega_1 t) + \frac{3A_1^2 A_2}{4} \cdot \overline{H_3(\omega_2, \omega_1, -\omega_1)} \cdot \cos(\omega_2 t)\\
&+ \frac{3A_1^2 A_2}{4} \cdot \overline{H_3(\omega_1, \omega_1, -\omega_2)} \cdot \cos((2\omega_1 - \omega_2)t)\\
&+ \frac{3A_1 A_2^2}{4} \cdot \overline{H_3(\omega_2, \omega_2, -\omega_1)} \cdot \cos((2\omega_2 - \omega_1)t)\\
&+ \frac{3A_1^3}{4} \cdot \overline{H_3(\omega_1, \omega_1, \omega_1)} \cdot \cos(3\omega_1 t) + \frac{3A_1^2 A_2}{4} \cdot \overline{H_3(\omega_1, \omega_1, \omega_2)} \cdot \cos((2\omega_1 + \omega_2)t)\\
&+ \frac{3A_1 A_2^2}{4} \cdot \overline{H_3(\omega_1, \omega_2, \omega_2)} \cdot \cos((2\omega_2 - \omega_1)t) + \frac{A_2^3}{4} \cdot \overline{H_3(\omega_2, \omega_2, \omega_2)} \cdot \cos(3\omega_2 t)
\end{aligned}$$

The first-order terms generate the linear output signals at input frequencies. The second-order terms generate three components: a zero-frequency signal that is the rectification of both input signals; a difference-frequency signal and three second-harmonic or mixed-harmonic signals. The third-order terms generate four components: two compression components at input frequencies; two desensitivisation components again at input frequencies, due to the interaction of the two input signals, that add to compression; two intermodulation signals at $2\omega_1 - \omega_2$ and at $2\omega_2 - \omega_1$ and four third-harmonic or mixed-harmonic signals. The higher-order terms are proportional to suitable combinations of powers of the input signals. A graphical representation of the spectrum is depicted in Figure 1.12.

From the formulae above, it is clear that the output signal is easily computed when the nuclei are known. In fact, the nuclei are computed by a recursive method if the nonlinearity is expressed as a power series [23, 29]; in the case of our example (eq. (1.44))

$$i_g(v) = g_0 + g_1 \cdot v + g_2 \cdot v^2 + g_3 \cdot v^3 + \cdots \tag{1.54}$$

An input 'probing' signal in the form of an ideal tone of unit amplitude (eq. (1.48)) is first used:

$$i_s(t) = e^{j\omega_1 t} \tag{1.55}$$

The output can formally be written as (see eq. (1.49))

$$v(t) = H_1(\omega_1) \cdot e^{j\omega_1 t} + H_2(\omega_1, \omega_1) \cdot e^{j2\omega_1 t} + \cdots \tag{1.56}$$

where the nuclei are still unknown. Kirchhoff's equation (eq. (1.3)) with the nonlinearity in power-series form (eq. (1.54), limited to second order for brevity) is

$$C \cdot \frac{dv(t)}{dt} + g_1 \cdot v(t) + g_2 \cdot v^2(t) + \cdots + i_s(t) = 0 \tag{1.57}$$

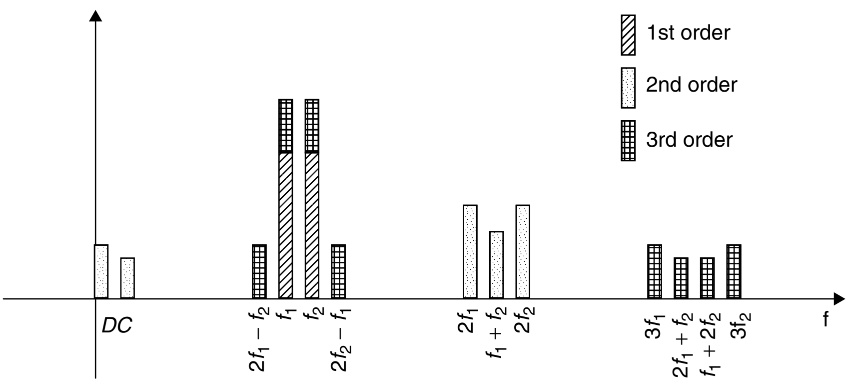

Figure 1.12 Contributions of the terms of the Volterra series to the spectrum of a two-tone signal