DESIGN OF SINUSOIDAL OSCILLATOR WITH CONTROLLABLE AMPLITUDE USING DESCRIBING FUNCTION METHOD

PONGRAPEE KAEWSAIHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN CONTROL ENGINEERING

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2008

KMITL-2008-EN-M-080-019

หัวข้อวิทยานิพนธ์	การออกแบบไซนูซอยด์ออสซิลเลเตอร์ที่สามารถควบคุมขนาดสัญญาณได้โดยใช้วิธีเดสไครบิงฟังก์ชัน
นักศึกษา	นายพงศ์ระพี แก้วไทรฮะ
รหัสนักศึกษา	48060511
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมระบบควบคุม
พ.ศ.	2551
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.ปรเมษฐ์ ประนยานันทน์

บทคัดย่อ

วิทยานิพนธ์นี้นำเสนอวิธีการออกแบบออสซิลเลเตอร์ชนิดไม่เป็นเชิงเส้นเพื่อให้ได้เอาท์พุทเป็นสัญญาณรูปซายน์ซึ่งมีขนาดและความถี่ในการสั่นเป็นไปตามที่ต้องการโดยใช้วิธีเดส-ไครบิงฟังก์ชันในการทำนายการเกิดลิมิตไซเคิลและคำนวณหาขนาดของสัญญาณ วงจรที่ใช้เป็นลักษณะของระบบควบคุมป้อนกลับซึ่งประกอบด้วยตัวกรองสัญญาณชนิดแถบความถี่ผ่านถูกป้อนกลับด้วยอุปกรณ์ไม่เป็นเชิงเส้น จากการทดลองออกแบบวงจรไซนูซอยด์ออสซิลเลเตอร์ชนิดไม่เป็นเชิงเส้นโดยใช้โอทีเอเป็นอุปกรณ์ไม่เป็นเชิงเส้น ผลการทดลองแสดงให้เห็นว่าวงจรออสซิลเลเตอร์ที่ออกแบบจะมีการสั่นที่มีขนาดคงที่เสมอโดยไม่ขึ้นกับสภาวะเริ่มต้นของวงจร

Thesis Title	Design of Sinusoidal Oscillator with Controllable Amplitude Using Describing Function Method
Student	Mr. Pongrapee Kaewsaiha
Student ID.	48060511
Degree	Master of Engineering
Program	Control Engineering
Year	2008
Thesis Advisor	Assist.Prof.Dr. Poramate Pranayanuntana

ABSTRACT

This thesis presents a design of sinusoidal nonlinear oscillators which operate at the desired output amplitude and frequency of oscillation. The describing function method is used for predicting the existence of limit cycles and for determining the amplitude of oscillation. An equivalent circuit of the oscillator used is a feedback configuration of a linear system with band-pass filter characteristic fed-back via a nonlinear element. The design of a sinusoidal nonlinear oscillator using an operational transconductance amplifier (OTA) as a nonlinear element has been done. The experimental results show that the oscillator circuits can keep the amplitude of oscillation constant and do not depend on the circuit's initial conditions.

ACKNOWLEDGEMENT

My most sincere thanks go to my thesis advisor Asst. Prof. Dr. Poramate Pranayanuntana for his valuable helps, suggestion and encouragement through the course of this thesis. Without his generous support, this thesis would not come true.

I wish to express my gratitude to Assoc.Prof. Vanchai Riewruja and his student, P'Chu. They help me a lot with an OTA IC and its characteristic test.

I sincerely thank Heng, Bank, Ton and all my friends who give me supports, helps and friendship that I will never forget. They always remind me when the deadline is coming. Also thank P'Amorntep, a store manager for the help in using lab instruments.

I wish to thank a group of 4th-year project students in my working room for the valuable help in scanning and typing.

At last, I would like to thank my parents for their support, understanding, loves and encouragement even though the process in finishing this thesis is too long.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (Cont.)

Figure **Page**

CHAPTER 1
Introduction

1.1 Background

A nonlinear oscillator is a nonlinear system that can display oscillation of fixed amplitude and fixed period without external excitation. The oscillations of this kind are called limit cycles, or self-excited oscillations. An advantage of nonlinear oscillator over the linear one is due to a magnitude stabilization phenomenon that keeps the amplitude of oscillation constant and not depending on the circuits initial condition.

The describing function method is a first-order version of the method of harmonic balance used to find periodic solutions for nonlinear systems by fitting a truncated Fourier series. This method applies to a system shown in Figure 1.1 with the linear block having low-pass or band-pass filtering characteristics and $\psi(\cdot)$ as a time-invariant nonlinear element. The solution is found using graphical constructions involving the Nyquist locus. This study employs a numerical integration ability of mathematical softwares, for example, MAPLE or MATLAB to obtain the describing function $D(A)$ used in oscillator circuit design.

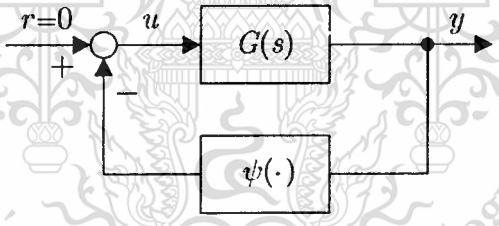


Figure 1.1 Feedback connection

Also we give a systematic approach to synthesizing an OTA-based nonlinear oscillator given specifications on amplitude and period (or frequency) of oscillation. An equivalent feedback configuration of an oscillator circuit with a linear dynamic system $G(s)$ as a feedforward element and a memoryless nonlinearity $\psi(\cdot)$ as a feedback element as shown in Figure 1.1, is used in the design.

Traditionally, the solution is found using graphical constructions involving the Nyquist locus. This tradition has not died out with the universal availability of computers; rather, it has been strengthened because of the vastly more informative nature of graphic displays over lists of numbers.

1.2 Objective

This thesis describes a systematic approach to designing the nonlinear oscillator to achieve the desired specifications on amplitude and frequency of oscillation. The approach make use of the understanding of the describing function of nonlinear elements. Experimental and simulation results are included.

1.3 Scope of the Study

We study here the oscillator system that contains a linear block which has a band-pass filter characteristic fed back with a single nonlinear element. The describing function method which is used for predicting the existence of limit cycles and, more generally, for analyzing the magnitude stabilization phenomena, is used in synthesizing the nonlinear oscillator.

1.4 Thesis Outline

The oscillator fundamentals in Chapter 2 show a basic form of an oscillator and the amplitude stabilizing problems. Phase plane analysis [1], a graphical method used for analyzing the system is described in Chapter 3. The describing function analysis [1] used for describing characteristic of nonlinear elements is mentioned in Chapter 4. Chapter 5 shows the design of a sinusoidal nonlinear oscillator. Chapter 6 shows the simulation results by MATLAB SIMULINK and the experimental result by using an OTA (LM13600) as a nonlinear element. Conclusions of this thesis are discussed in Chapter 7.

CHAPTER 2

Oscillator Fundamentals

2.1 Introduction

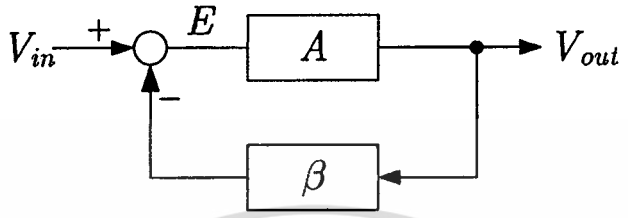


Figure 2.1 Canonical form of a feedback system with positive or negative feedback

Figure 2.1 shows a block diagram for the system in which V_{in} is the input voltage, V_{out} is the output voltage of the amplifier block (gain A), and β is called a feedback factor that fed the signal back to the summing junction. E represents the error term which is the difference of the input voltage and the feedback signal.

The corresponding classic expression for a feedback system is derived as

$$\frac{V_{out}}{V_{in}} = \frac{A}{1 + A\beta} \tag{2.1}$$

Oscillators do not require any external input signal, instead, they use some fraction of the output signal created by the feedback network as an input signal.

Oscillation occurs when the feedback system is unable to find a stable steady-state because its transfer function can not be satisfied. The system goes unstable when the denominator in (2.1) becomes zero, i.e., when $1 + A\beta = 0$, or $A\beta = -1$. The key for designing the oscillator is ensuring that $A\beta = -1$. This is called the *Barkhausen criterion*[2]. Satisfying this criterion requires that the magnitude of the loop gain is unity with a corresponding phase shift of 180° as indicated by the minus sign. An equivalent complex expression is $A\beta = 1\angle -180°$ for the negative feedback system. For the positive feedback system, the expression is $A\beta = 1\angle 0°$ and the sign of the $A\beta$ term is negative in (2.1).

As the phase shift approaches 180° and $|A\beta| \rightarrow 1$, the output voltage of the (now unstable) system tends to infinity but, of course, is limited to finite values by an energy-limited power supply when the output voltage approaches either power rail or the active devices in the amplifiers change gain. This causes the value of A to change and forces $A\beta$ away from the singularity; thus the trajectory towards an infinite voltage slows and eventually halts. At this stage, one of three things can occur:

1. Nonlinearity in saturation or cut-off causes the system to become stable and lock up at the current power rail.

2. The initial change causes the system to saturate (or cut-off) and stay that way for a long time before it becomes linear and heads for the opposite power rail.

3. The system stays linear and reverses direction, heading for the opposite power rail.

The second alternative produces highly distorted oscillations (usually quasi-square waves), the resulting oscillators being called relaxation oscillators. The third produces a sine-wave oscillator.

2.2 Basic Wien-Bridge Oscillator

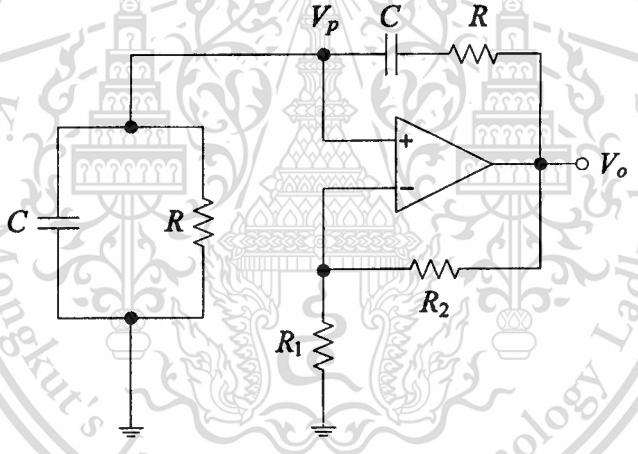


Figure 2.2 Wien-bridge oscillator circuit

A Wien bridge oscillator [2, 3] is a type of electronic oscillator that generates sine waves without having any input source. It can generate a large range of output frequencies. The bridge comprises four resistors and two capacitors. The circuit is based on a network originally developed by Max Wien in 1891. The circuit of Figure 2.2 uses both negative feedback, via R_2 and R_1, and positive feedback, via the series and parallel RC networks. Circuit behavior is strongly affected by whether positive or negative feedback prevails. The components of the RC networks need not to be equal-valued. However, making them so simplifies analysis as well as inventory.

The circuit can be viewed as a noninverting amplifier that amplifies V_p by the amount

$$A = \frac{V_o}{V_p} = 1 + \frac{R_2}{R_1} \tag{2.2}$$

assuming an ideal op-amp for simplicity. In turn, V_p is supplied by the op-amp itself via the two RC networks as $V_p = [Z_p/(Z_p + Z_s)]V_o$, where $Z_p = R \parallel (1/j\omega C)$. Expanding,we get

$$B(j\omega) = \frac{V_p}{V_o} = \frac{1}{3 + j(\omega/\omega_0) - \omega_0/\omega} \tag{2.3}$$

where $\omega_0 = 1/RC$. The overall gain experienced by a signal in going around the loop is $G(j\omega) = AB$, or

$$G(j\omega) = \frac{1 + R_2/R_1}{3 + j(\omega/\omega_0) - \omega_0/\omega} \tag{2.4}$$

This is a band-pass function since it approaches zero at both high and low frequencies. Its peak value occurs at $\omega = \omega_0$, that is

$$G(j\omega) = \frac{1 + R_2/R_1}{3} \tag{2.5}$$

The fact that $G(j\omega)$ is real indicates that a signal of frequency ω_0 will experience a net phase shift of zero in going around the loop. Depending on the magnitude of $G(j\omega)$, we have three distinct possibilities.

1. $G(j\omega_0) < 1$, that is, $A < 3$ V/V. Any disturbance of frequency f_0 arising at the input of the op-amp is first amplified by $A < 3$ V/V, and then by $B(j\omega_0) = \frac{1}{3}$ V/V, for a net gain of less than unity. Any disturbance decrease each times it goes around the loop until it eventually decays to zero. We can state that negative feedback (via R_2 and R_1) prevails over positive feedback (via Z_S and Z_p), resulting in a stable system. Consequently, the system's poles lie in the left half of the complex plane.

2. $G(j\omega_0) > 1$, that is, $A > 3$ V/V. Now, positive feedback prevails over negative feedback, indicating that a disturbance of frequency ω_0 will be amplified regeneratively, causing the circuit to break out into oscillations of growing magnitude. The circuit is now ***unstable***, and its poles lie in the right half plane. As we know, the oscillations build up until the saturation limits of the op-amp, the power supply voltages, are reached. Therefore, V_o will appear as a clipped sine wave when observed with the oscilloscope or visualized via PSPICE.

3. $G(j\omega_0) = 1$, or $A = 3$ V/V exactly. A condition referred to as *neutral stability* because positive and negative feedback are now applied in equal amounts. Any disturbance of frequency ω_0 is first amplified by 3 V/V and then by 1/3 V/V, indicating that once start, it will be sustained indefinitely. As we know, this corresponds to a pole pair right on the $j\omega$ axis. The conditions $\angle G(j\omega_0) = 0°$ and $|G(j\omega)| = 1$ are together referred to as the *Barkhausen criterion* for oscillation at $\omega = \omega_0$. The band-pass nature of $G(j\omega)$ allows for oscillation to occur only at $\omega = \omega_0$. Any attempt to oscillate at other frequency is naturally discouraged because $\angle G(j\omega) \neq 0°$ and $|G(j\omega)| < 1$ there. By (2.5), neutral stability is

only achieved with

$$\frac{R_2}{R_1} = 2 \tag{2.6}$$

It is apparent that when this condition is met, the components around the op-amp form a *balanced bridge* at $\omega = \omega_0$.

In a real-life circuit, component drift makes it different to keep the bridge exactly balanced. Moreover, provisions must be made so that (i) oscillation starts spontaneously at power turn-on, and (ii) its amplitude is kept below the op-amp saturation limits to avoid excessive distortion. These objectives are met by making the ratio R_2/R_1 *amplitude-dependent* such that at low signal levels it is slightly greater than 2 to ensure oscillation start-up, and that at high signal levels, it is slightly less than 2 to limit amplitude. Then, once the oscillation has started, it will grow and automatically stabilized at some intermediate level where $R_2/R_1 = 2$ exactly.

Amplitude stabilization takes on many forms, all of which use nonlinear elements to either decrease R_2 or increase R_1 with signal amplitude.

2.3 Automatic Amplitude Control

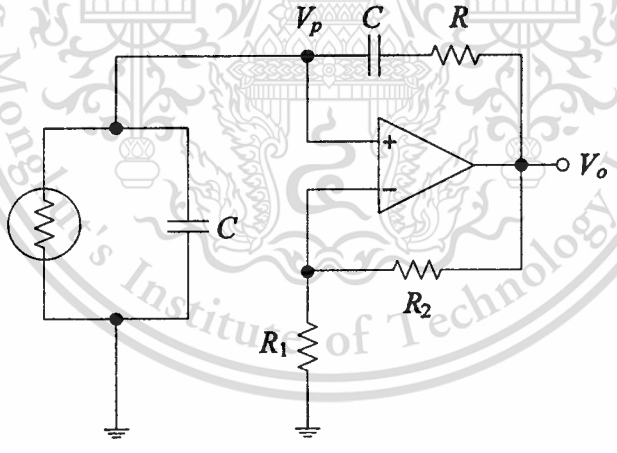


Figure 2.3 Hewlett's low-distortion oscillator

The developed circuit is derived from William Hewlett's 1939 Stanford University master's degree thesis [2]. Hewlett, along with David Packard co-founded Hewlett-Packard. Their first product was the HP 200A, a precision sine wave oscillator based on the Wien-bridge. The 200A is a classic instrument known for its low distortion. Hewlett used an incandescent bulb in the oscillator feedback path to limit the gain, see Figure 2.3. The resistance of light bulbs and similar heating elements increases as their temperature increases. If the gain is inversely proportional to the oscillation amplitude, the oscillator gain stage reaches a steady state and operates at very low distortion.

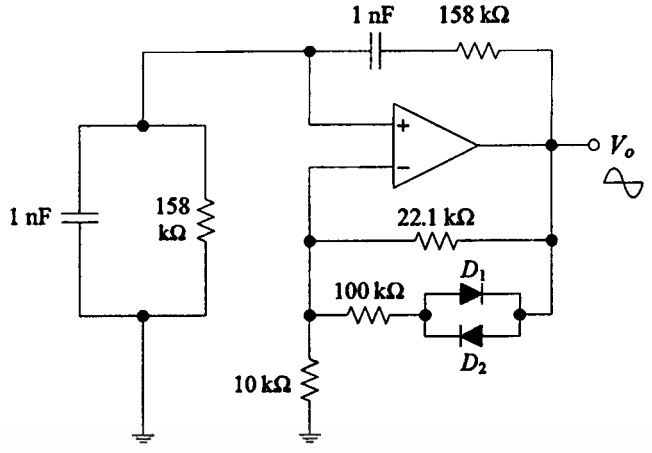


Figure 2.4 Practical Wien-bridge oscillator

Modern Wien bridge oscillators have used field effect transistors or photocells for amplitude stabilization in place of light bulbs [3]. The circuit of Figure 2.4 uses a simple diode-resistor network to control the effective value of R_2. At low signal levels, the diodes are off so the 100-kΩ resistance has no effect. We thrust have $R_2/R_1 = 22.1/10.0 = 2.21$, or $G(j\omega_0) = (1 + 2.21)/3 = 1.07 > 1$, indicating oscillation buildup. As the oscillation grows, the diodes are gradually brought into conduction on alternate half-cycles. In the limit of heavy diode conduction, R_2 would effectively change to $(22.1 \parallel 100) = 18.1k\Omega$, giving $G(j\omega_0) = 0.937 < 1$. However, before this limiting condition is reached, amplitude will automatically stabilize at some intermediate level of diode conduction where $R_2/R_1 = 2$ exactly, or $G(j\omega_0) = 1$.

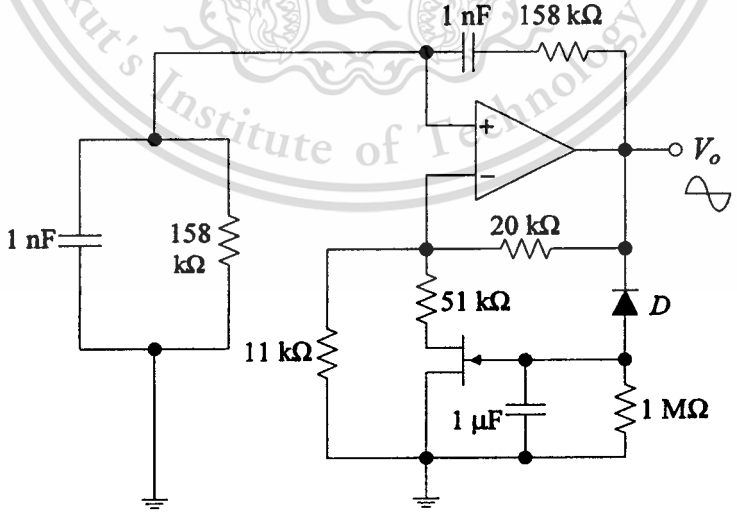


Figure 2.5 Wien-bridge oscillator using JFET

A disadvantage of the circuit in Figure 2.4 is that V_{om} is quite sensitive to variations in the diode-forward voltage drops. The circuit of Figure 2.5 overcomes the drawback by using an n-JFET effectively shorts the 51-kΩ resistance

to ground to give $R_2/R_1 \cong 2.21 > 2$, so oscillation starts to build up. The diode and the 1-μF capacitance from a negative peak detector whose voltage becomes progressively more negative as the oscillation grows. This gradually reduces the conductivity of the JFET until, in the limit of complete cut-off we would have $R_2/R_1 = 20.0/11.0 = 1.82 < 2$, However, amplitude stabilizes automatically at some intermediate level where $R_2/R_1 = 2$ exactly. Denoting the corresponding gate-source voltage as $V_{GS(crit)}$, and the output peak amplitude as V_{om}, we have $-V_{om} = V_{GS(crit)} - V_{D(on)}$. For instant, with $V_{GS(crit)} = -4.3V$ we get $V_{om} \cong 4.3 + 0.7 = 5V$.

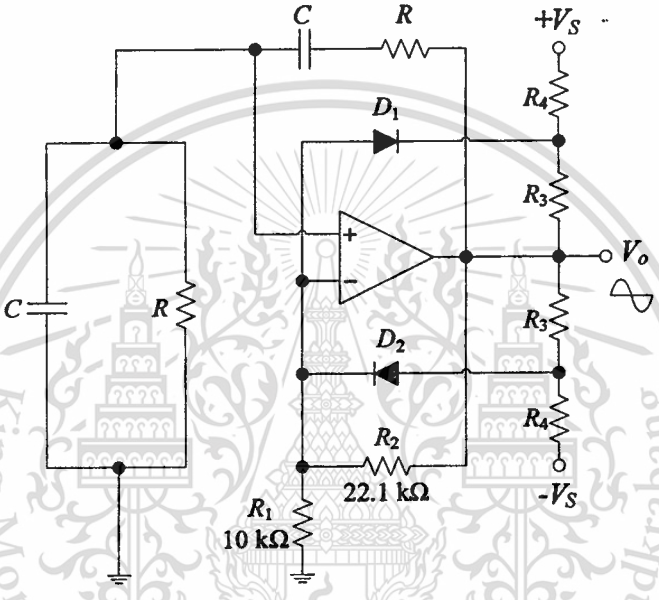


Figure 2.6 Wien-bridge oscillator using a voltage limiter

Figure 2.6 shows yet another popular amplitude stabilization scheme using a diode limiter for easier programming of amplitude. As usual, for low output levels the diodes are biased in cutoff, yielding $R_2/R_1 = 2.21 > 2$. The oscillation grows until the diodes become conductive on alternate output peaks. These peaks are likewise symmetric, or $\pm V_{om}$. To estimate V_{om}, consider the instant when D_2 starts to conduct. Assuming the current through D_2 is still negligible, and denoting the voltage at the anode of D_2 as V_2, we use KCL to write $(V_{om} - V_2)/R_3 \cong [V_2 - (-V_S)]/R_4$, where $V_2 = V_n + V_{D2(on)} \cong V_{om}/3 + V_{D2(on)}$. Eliminating V_2 and solving gives $V_{om} \cong 3[(1 + R_4/R_3)V_{D2(on)} + V_S]/(2R_4/R_3 - 1)$. For example, with $R_3 = 3k\Omega$, $R_4 = 20k\Omega$, $V_S = 15V$, and $V_{D(on)} = 0.7V$, we get $V_{om} \cong 5V$.

Another way to stabilize the output amplitude is by using a nonlinear element (in the feedback path of Figure 1.1). An analysis of the amplitude stabilization in LC sinusoidal oscillators were done by Bayard and Ayachi [4] in 2002. They considered two nonlinear elements, OTA and CFOA, and realized that OTA- and CFOA-based oscillators of this type are nonlinear and could not be

simply analyzed using linear approximation, since lots of nonlinear information and characteristics would be lost. They instead used third-order Taylor polynomial approximation in analysis, and state the difference in oscillation existence between OTA-based and CFOA-based oscillators.

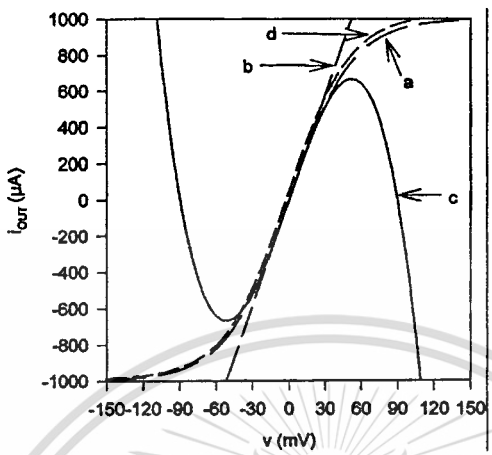


Figure 2.7 Variation of the output current i_{out} against the command voltage v for an OTA: (a) Without simplification; (b) First-order approximation; (c) Third-order approximation; (d) SPICE simulation.

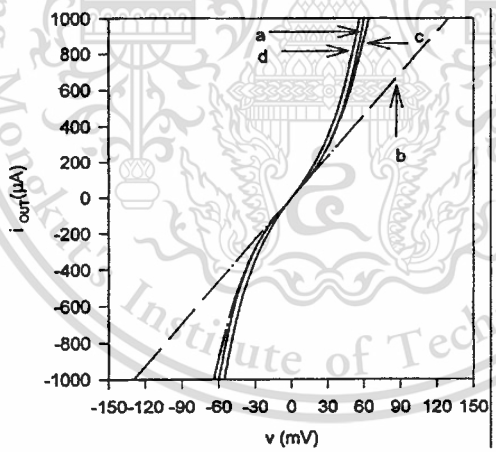


Figure 2.8 Variation of the output current i_{out} against the command voltage v for a CFOA: (a) Without simplification; (b) First-order approximation; (c) Third-order approximation; (d) SPICE simulation.

We present here the describing function method that gives an insight into analyzing and designing a nonlinear oscillator circuit. Limit cycle analysis can describe the difference between OTA and CFOA characteristic (OTA has only one limit cycle which is stable but CFOA has two limit cycles, the one with smaller amplitude is unstable while the one with the larger amplitude is stable) [5] and will be used in our oscillator design.

CHAPTER 3

Phase Plane Analysis

3.1 Introduction

Phase plane analysis [1] is a graphical method for studying second-order systems, which was introduced well before the turn of the previous century by mathematicians such as Henri Poincare. The basic idea of the method is to generate, in the state space of a second-order dynamic system (a two-dimensional plane called the phase plane), motion trajectories corresponding to various initial conditions, and then to examine the qualitative features of the trajectories. In such a way, information concerning stability and other motion patterns of the system can be obtained. In this chapter, The basic tools of phase plane analysis is used to acquire intuitive insights on the effects of nonlinearities on control systems.

Phase plane analysis has a number of useful properties. First, as a graphical method, it allows us to visualize what goes on in a nonlinear system starting from various initial conditions, without having to solve the nonlinear equations analytically. Second, it is not restricted to small or smooth nonlinearities, but applies equally well to strong nonlinearities and to hard nonlinearities. Finally, some practical control systems can indeed be adequately approximated as second-order systems, and the phase plane method can be used easily for their analysis. Conversely, the fundamental disadvantage of this method is that it is restricted to second-order (or first-order) systems, because the graphical study of higher-order systems is computationally and geometrically complex.

3.2 Concept of Phase Plane Analysis

3.2.1 Phase Portraits

The phase plane method is concerned with the graphical study of second-order systems described by

$$\dot{x}_1 = f_1(x_1, x_2) \tag{3.1a}$$

$$\dot{x}_2 = f_2(x_1, x_2) \tag{3.1b}$$

where x_1 and x_2 are the states of the system, and f_1 and f_2 are nonlinear functions of the states. Geometrically, the state space of this system is a plane having x_1 and x_2 as coordinates. This plane is called the *phase plane*.

Given a set of initial conditions $x(0) = x_0$, Equation (3.1) defines a solution $x(t)$. With time t varied from zero to infinity, the solution $x(t)$ can be represented geometrically as a curve in the phase plane. Such a curve is called a phase plane *trajectory*. A family of phase plane trajectories corresponding to various initial conditions is called a *phase portrait* of a system.

To illustrate the concept of phase portrait, consider the following simple system.

Example 3.1. Phase portrait of a mass-spring system

The governing equation of the mass-spring system in Figure 3.1a is the linear second-order differential equation

$$\ddot{x} + x = 0 \tag{3.2}$$

Assume that the mass is initially at rest, at length x_0 . Then the solution of the equation is

$$\begin{aligned} x(t) &= x_0 \cos t \\ \dot{x}(t) &= -x_0 \sin t \end{aligned}$$

Eliminate time t from the above equations to obtain the equation of the trajectory

$$x^2 + \dot{x}^2 = x_0^2$$

This represents a circle in the phase plane. Corresponding to different initial conditions, circles of different radius can be obtained. Plotting these circles on the phase plane, a phase portrait for the mass-spring system is shown in Figure 3.1b.

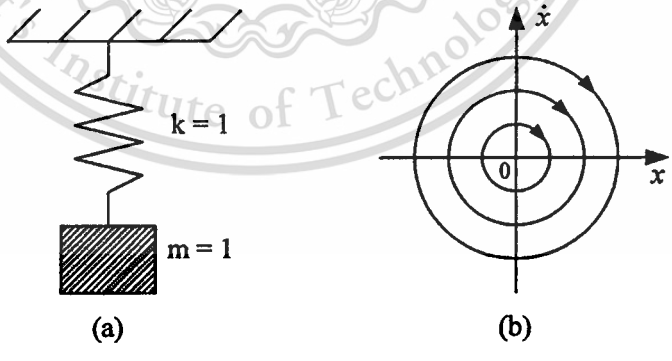


Figure 3.1 A mass-spring system and its phase portrait

The power of the phase portrait lies in the fact that once the phase portrait of a system is obtained, the nature of the system response corresponding to various initial conditions is directly displayed on the phase plane. In the above example, the system trajectories neither converge to the origin nor diverge to infinity. They simply circle around the origin, indicating the marginal nature of the system's stability.

A major class of second-order systems can be described by differential equations of the form

$$\ddot{x} + f(x, \dot{x}) = 0 \tag{3.3}$$

In state space form, this dynamics can be represented as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -f(x_1, x_2) \end{aligned}$$

with $x_1 = x$ and $x_2 = \dot{x}$. Most second-order systems in practice, such as mass-spring-damper systems in mechanics, or resistor-coil-capacitor systems in electrical engineering, can be represented in or transformed into this form. For these systems, the states are x and its derivative $\dot{x}$. Traditionally, the phase plane method is developed for the dynamics (3.3), and the phase plane is defined as the plane having x and $\dot{x}$ coordinates. But it's easy to extend the method to more general dynamics of the form (3.1), with the (x_1, x_2) plane as the phase plane.

3.2.2 Singular Points

An important concept in phase plane analysis is that of a singular point. A singular point is an equilibrium point in the phase plane. Since an equilibrium point is defined as a point where the system states can stay forever, its location can be obtained from solving $\dot{x} = 0$ for the state x, *i.e.*, from solving

$$f_1(x_1, x_2) = 0 \qquad f_2(x_1, x_2) = 0 \tag{3.4}$$

The singular point of system (3.2) is simply the origin $(0, 0)$ of the phase plane

For a linear system, there is usually only one singular point (although in some cases there can be a continuous infinity of singular points, as seen by the system $\ddot{x} + \dot{x} = 0$, for which all points on the real axis are singular points). However, a nonlinear system often has more than one isolated singular point.

Example 3.2. A nonlinear second-order system

Consider the system

$$\ddot{x} + 0.5\dot{x} + 2x + x^2 = 0$$

whose phase portrait is plotted in Figure 3.2. The system has two singular points, one at $(0, 0)$ and the other at $(-2, 0)$. The motion patterns of the system trajectories in the vicinity of the two singular point have different natures. The trajectories move towards the point $x = 0$ while moving away from the point $x = -2$.

One may wonder why an equilibrium point of a second-order system is called a *singular point*. To answer this, examine the slope of the phase trajectories. From (3.1), the slope of the phase trajectory passing through a point (x_1, x_2)

is determined by

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)} \tag{3.5}$$

With the functions f_1 and f_2 assumed to be single valued, there is usually a definite value for this slope at any given point in phase plane. This implies that the phase trajectories will not intersect. At singular points, however, the value of the slope is $0/0$, *i.e.*, the slope is indeterminate. Many trajectories may intersect at such points, as seen from Figure 3.2. This indeterminacy of the slope accounts for the adjective "singular".

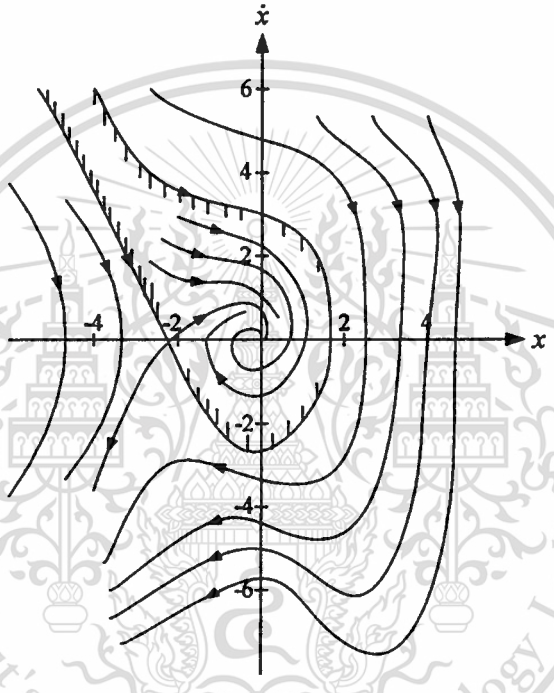


Figure 3.2 Phase portrait of a nonlinear system

Singular points are very important features in the phase plane. Examination of the singular points can reveal a great deal of information about the properties of a system. In fact, the stability of linear systems is uniquely characterized by the nature of their singular points. For nonlinear systems, besides singular points, there may be more complex features, such as limit cycles.

Note that, although the phase plane method is developed primarily for second-order systems, it can also be applied to the analysis of first-order systems of the form

$$\dot{x} + f(x) = 0$$

The idea is still to plot $\dot{x}$ with respect to x in the phase plane. The difference now is that the phase portrait is composed of a single trajectory.

Example 3.3. A first-order system

Consider the system

$$\dot{x} = -x + x^3$$

A phase trajectory of this system is the curve shown in Figure 3.3. The arrows in the figure denote the direction of the motion, and whether they point toward the left or the right at a particular point is determined by the sign of $\dot{x}$ at that point. There are three singular points for this system, defined by $-x + x^3 = 0$, *i.e.*, $x = 0, -1$, and 1. It is seen from the phase portrait of this system that the equilibrium point $x = 0$ is stable, while the other two are unstable.

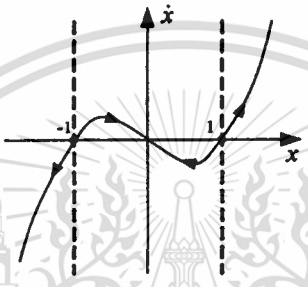


Figure 3.3 Phase trajectory of a first-order system

3.2.3 Symmetry in Phase Plane Portraits

In some cases, the phase portrait is symmetric with respect to the x-axis, or the y-axis, or both. The exploitation of such symmetry properties can simplify the generation and study of the phase portrait. Usually, symmetry in the phase portrait can be established directly from the original differential equations (3.1).

The slope of trajectories in the phase plane is of the form

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)}$$

Let the slope be denoted by f_0, *i.e.*,$f_0(x_1, x_2) = f_2/f_1$. Then, the following symmetry situations may occur:

Symmetry about the x_1 axis: In order for the phase portrait to be symmetric about the x-axis, the slopes at points (x_1, x_2) and $(x_1, -x_2)$ should be equal in absolute values, but of opposite signs

$$f_0(x_1, x_2) = f_0(x_1, -x_2)$$

This implies that the function f_0 should be even in x_2. The mass-spring system in Example 3.1 satisfies this condition, and its phase portrait is seen to be symmetric about the x_1-axis.

Symmetry about the y-axis: Similarly,

$$f_0(x_1, x_2) = f_0(-x_1, x_2)$$

implies symmetry with respect to the x_2 axis. The mass-spring system also satisfies this condition.

Symmetry about the origin: In order for the phase portrait to be symmetric about the origin, one must have

$$f_0(x_1, x_2) = -f_0(-x_1, -x_2)$$

3.3 Constructing Phase Portraits

Today, phase portraits are routinely computer-generated. In fact, it is largely the advent of the computer in the early 1960's, and the associated ease of quickly generating phase portraits, which motivated many advances in the study of complex nonlinear dynamic behaviors such as chaos. However, of course (as *e.g.*, in the case of root locus for linear systems), it is still practically useful to learn how to roughly sketch phase portraits or quickly verify the plausibility of computer outputs.

There are a number of methods for constructing phase plane trajectories for linear or nonlinear systems, such as the so-called analytical method, the method of isoclines, the delta method, Lienard's method, and Pell's method. Two of them are discussed in this section, namely, the analytical method and the method of isoclines. These methods are chosen primarily because of their relative simplicity. The analytical method involves the analytical solution of the differential equations describing the systems. It is useful for some special nonlinear systems, particularly piecewise linear systems, whose phase portraits can be constructed by piecing together the phase portraits of the related linear systems. The method of isoclines is a graphical method which can conveniently be applied to construct phase portraits for systems which cannot be solved analytically, which represent by far the most common case.

3.3.1 Analytical Method

There are two techniques for generating phase plane portraits analytically. Both techniques lead to a functional relation between the two phase variables x_1 and x_2 in the form

$$g(x_1, x_2, c) = 0 \tag{3.6}$$

where the constant c represents the effects of initial conditions (and, possibly, of external input signals). Plotting this relation in the phase plane for different initial conditions yields a phase portrait.

The first technique involves solving equations (3.1) for x_1 and x_2 as functions of time t, *i.e.*,

$$x_1(t) = g_1(t) \qquad x_2(t) = g_2(t)$$

and then eliminating time t from these equations, leading to a functional relation in the form of (3.6). This technique was already illustrated in Example 3.1.

The second technique, on the other hand, involves directly eliminating the time variable, by noting that

$$\frac{dx_2}{dx_1} = \frac{f_2(x_1, x_2)}{f_1(x_1, x_2)}$$

and then solving this equation for a functional relation between x_1 and x_2 . This technique is used to solve the mass-spring equation in the following example.

Example 3.4. Mass-spring system

By noting that $\ddot{x} = (d\dot{x}/dx)(dx/dt)$, Equation (3.3) can be rewritten as

$$\dot{x}\frac{d\dot{x}}{dx} + x = 0$$

Integration of this equation yields

$$\dot{x}^2 + x^2 = x_0^2$$

One sees that the second technique is more straightforward in generating the equations for the phase plane trajectories.

Most nonlinear systems cannot be easily solved by either of the above two techniques. However, for piece-wise linear systems, an important class of nonlinear systems, this method can be conveniently used, as the following example shows.

Example 3.5. A satellite control system

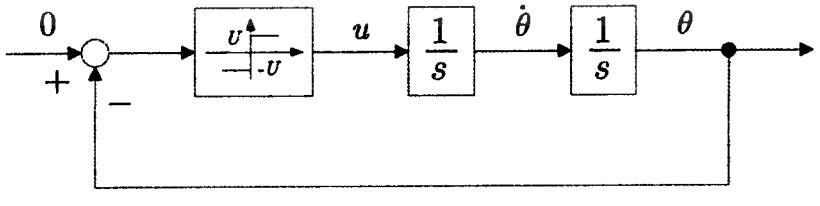


Figure 3.4 On-off control system

A one-dimensional model of a satellite in free space is depicted in Figure 3.4. It simply consists of a rotational inertia controlled by a pair of thrusters, which can provide either a positive constant torque U (positive firing) or a negative torque $-U$ (negative firing). The purpose of the control system is to maintain the satellite

attitude at the zero angle by firing the thrusters. The mathematical model of the system is

$$\ddot{\theta} = u$$

where u is the torque provided by the thrusters and θ is the satellite angle.

First examine the response of the system in the presence of positive and negative firing, and then the response when the thrusters fire according to the control law

$$u(t) = \begin{cases} -U & if \quad \theta > 0 \\ U & if \quad \theta < 0 \end{cases} \tag{3.7}$$

which means that the thrusters push in the counterclockwise direction if θ is positive, and vice versa.

When the thrusters fire with positive torque U, the dynamics of the system is

$$\ddot{\theta} = U$$

which implies that $\dot{\theta}d\dot{\theta} = Ud\theta$. Therefore, the phase trajectories are a family of parabolas defined by

$$\dot{\theta}^2 = U\theta + c_1$$

where c_1 is a constant. The corresponding phase portrait of the system is shown in Figure 3.5b.

When the thrusters fire with torque $-U$, the equation of phase trajectories is similarly found to be

$$\dot{\theta}^2 = -Ux + c_1$$

whose phase portrait is shown in Figure 3.5c.

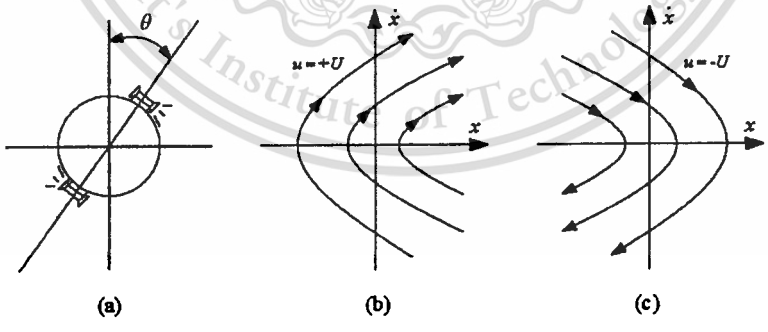


Figure 3.5 Satellite control using on-off trusters

Now, consider the phase plane trajectory when the control law (3.7) is used. It can be obtained simply by connecting the trajectories on the left half of the phase plane in Figure 3.5b. with those on the right half of the phase plane in Figure 3.5c. The vertical axis represents a switching line, because the control input and thus the phase trajectories are switched on that line. The resulting

phase portrait. shown in Figure 3.6, is a collection of closed curves which represent periodic motions. One concludes from this phase portrait that the system is marginally stable, similarly to the mass-spring system in Example 3.1.

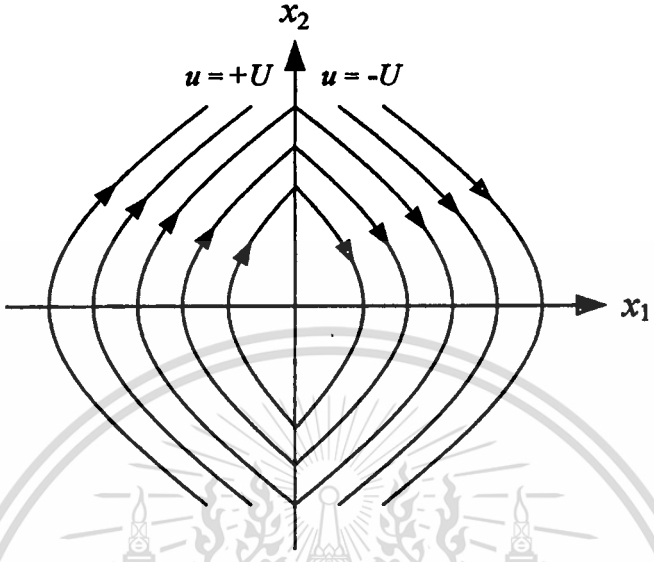


Figure 3.6 Complete phase plane portrait of the on-off controller

The Method of Isoclines

Consider the dynamics in (3.1). At a point (x_1, x_2) in the phase plane, the slope of the tangent to the trajectory can be determined by (3.5). An isocline is defined to be the locus of the points with a given tangent slope. An isocline with slope α is thus defined to be

$$\frac{dx_2}{dx_1} = \frac{f_1(x_1, x_2)}{f_2(x_1, x_2)} = \alpha$$

This is to say that points on the curve

$$f_2(x_1, x_2) = \alpha f_1(x_1, x_2)$$

all have the same tangent slope α.

In the method of isoclines, the phase portrait of a system is generated in two steps. In the first step, a field of directions of tangents to the trajectories is obtained. In the second step, phase plane trajectories are formed from the field of directions.

The isocline method on the mass-spring system is explained in (3.2). The slope of the trajectories is easily seen to be

$$\frac{dx_2}{dx_1} = -\frac{x_1}{x_2}$$

Therefore, the isocline equation for a slope α is

$$x_1 + \alpha x_2 = 0$$

i.e., a straight line. Along the line, it's possible to draw a lot of short line segments with slope α. By taking α to be different values, a set of isoclines can be drawn, and a field of directions of tangents to trajectories are generated, as shown in Figure 3.7. To obtain trajectories from the field of directions, assume that the the tangent slopes are locally constant. Therefore, a trajectory starting from any point in the plane can be found by connecting a sequence of line segments.

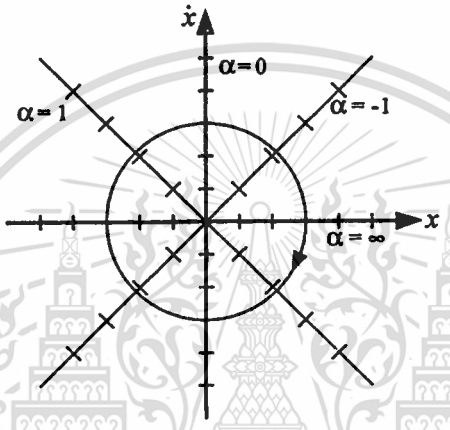


Figure 3.7 Isoclines for the mass-spring system

The method of isoclines is used to study the Van der Pol equation, a non-linear equation in the following example.

Example 3.6. The Van der Pol equation

For the Van der Pol equation

$$\ddot{x} + \mu(x^2 - 1)\dot{x} + x = 0$$

an isocline of slope α is defined by

$$\frac{d\dot{x}}{dx} = \frac{-\mu(x^2 - 1)\dot{x} - x}{\dot{x}} = \alpha$$

Therefore, the points on the curve

$$\dot{x} = \frac{x}{(\mu - \mu x^2) - \alpha}$$

all have the same slope α.

By using $\mu = 0.2$ and taking α of different values, different isoclines can be obtained, as plotted in Figure 3.8. Short line segments are drawn on the isoclines to generate a field of tangent directions. The phase portraits can then

be obtained, as shown in the plot. It is interesting to note that there exists a closed curve in the portrait, and the trajectories starting from both outside and inside converge to this curve. This closed curve corresponds to a limit cycle, as reflected by the motion patterns in Figure 3.8.

Note that the same scales should be used for the x_1 axis and x_2 axis of the phase plane, so that, the derivative dx_2/dx_l equals the geometric slope of the trajectories. Also note that since in the second step of phase portrait construction, it's essential to assume that the slope of the phase plane trajectories is locally constant, more isoclines should be plotted in regions where the slope varies quickly, to improve accuracy.

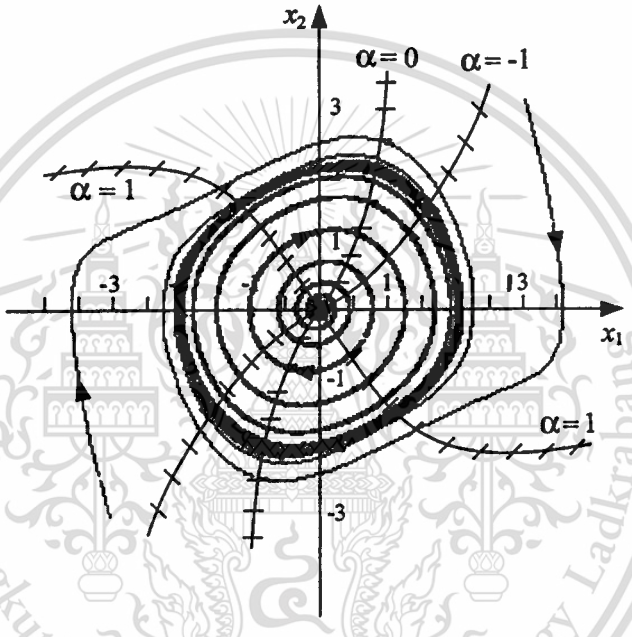


Figure 3.8 Phase portrait of the Van der Pol equation

3.3.2 Determining Time from Phase Portraits

Note that time t does not explicitly appear in the phase plane having x_1 and x_2 as coordinates. However, in some cases, time information might be interested. For example, one might want to know the time history of the system states starting from a specific initial point. Another relevant situation is when one wants to know how long it takes for the system to move from a point to another point in a phase plane trajectory. There are two techniques for computing time history from phase portraits. Both techniques involve a step-by-step procedure for recovering time.

Obtaining time from $\Delta t \approx \Delta x/\dot{x}$

In a short time Δt, the change of x is approximately

$$\Delta x \approx \dot{x}\Delta t \tag{3.8}$$

where $\dot{x}$ is the velocity corresponding to the increment Δx. Note that for a Δx of finite magnitude, the average value of velocity during a time increment should be used to improve accuracy. From (3.8), the length of time corresponding to the increment Δx is

$$\Delta t \approx \frac{\Delta x}{\dot{x}}$$

The above reasoning implies that, in order to obtain the time corresponding to the motion from one point to another point along a trajectory, one should divide the corresponding part of the trajectory into a number of small segments (not necessarily equally spaced), find the time associated with each segment, and then add up the results. To obtain the time history of states corresponding to a certain initial condition, one simply computes the time t for each point on the phase trajectory, and then plots x with respect to t and $\dot{x}$ with respect to t.

Obtaining time from $t = \int(1/\dot{x})dx$

Since $\dot{x} = dx/dt$, or $dt = dx/\dot{x}$. Therefore,

$$t - t_0 = \int_x^{x_0} (1/\dot{x})dx$$

where x corresponds to time t and x_0 corresponds to time t_0. This equation implies that, if we plot a phase plane portrait with new coordinates x and $(1/\dot{x})$, then the area under the resulting curve is the corresponding time interval.

3.4 Phase Plane Analysis of Linear Systems

The phase plane analysis of linear systems not only useful in visually observe the motion patterns of linear systems, but also help in developing of nonlinear system analysis because the same motion patterns can be observed in the local behavior of nonlinear systems.

The general form of a linear second-order system can be written as

$$\dot{x}_1 = ax_1 + bx_2 \tag{3.9a}$$

$$\dot{x}_2 = cx_1 + dx_2 \tag{3.9b}$$

which can be transformed into a scalar second-order differential equation

$$b\dot{x}_2 = bcx_1 + d(\dot{x}_1 - ax_1)$$

Consequently, differentiation of (3.9a) and then substitution of (3.9b) leads to

$$\ddot{x}_1 = (a+d)\dot{x}_1 + (cb - ad)x_1$$

Therefore, simply consider the second-order linear system described by

$$\ddot{x} + a\dot{x} + bx = 0 \tag{3.10}$$

To obtain the phase portrait of this linear system, first solve for the time history

$$x(t) = k_1 e^{\lambda_1 t} + k_2 e^{\lambda_2 t} \qquad for \lambda_1 \neq \lambda_2 \tag{3.11a}$$

$$x(t) = k_1 e^{\lambda_1 t} + k_2 e^{\lambda_1 t} \qquad for \lambda_1 = \lambda_2 \tag{3.11b}$$

where the constants λ_1 and λ_2 are the solutions of the characteristic equation

$$\lambda_1 = (-a + \sqrt{a^2 - 4b})/2, \qquad \lambda_2 = (-a - \sqrt{a^2 - 4b})/2$$

For linear systems described by (3.10), there is only one singular point (assuming $b \neq 0$), namely the origin. However, the trajectories in the vicinity of this singularity point can display quite different characteristics, depending on the values of a and b. The following cases can occur

1. λ_1 and λ_2 are both real and have the same sign (positive or negative)
2. λ_1 and λ_2 are both real and have opposite signs
3. λ_1 and λ_2 are complex conjugate with non-zero parts
4. λ_1 and λ_2 are complex conjugate with real parts equal to zero

CASE 1 : STABLE OR UNSTABLE MODE

The first case corresponds to a *node*. If the signs of the eigenvalues are negative, both $x(t)$ and $\dot{x}(t)$ converge to zero exponentially, according to (3.4), and the singularity point is called a stable node. There is no oscillation in this convergence, and therefore, the phase trajectories have the features illustrated in Figure 3.9a. If both eigenvalues are positive, then both $x(t)$ and $\dot{x}(t)$ diverge from zero exponentially, as shown in Figure 3.9b, which is called an unstable node.

CASE 2 : SADDLE POINT

The second case ($\lambda_1 < 0$ and $\lambda_2 > 0$) corresponds to a *saddle point*. The phase portrait of the system has the interesting shape shown in Figure 3.9c. Because of the unstable pole λ_2, almost all of the system trajectories diverge to infinity. In Figure 3.9c, the straight line in the second and four quadrant that passes through the origin is the only exception. This line corresponds to initial conditions which make k_2 (i.e., the unstable component) equal to zero. The other straight line in the plot corresponds to initial conditions which make k_1 equal to zero.

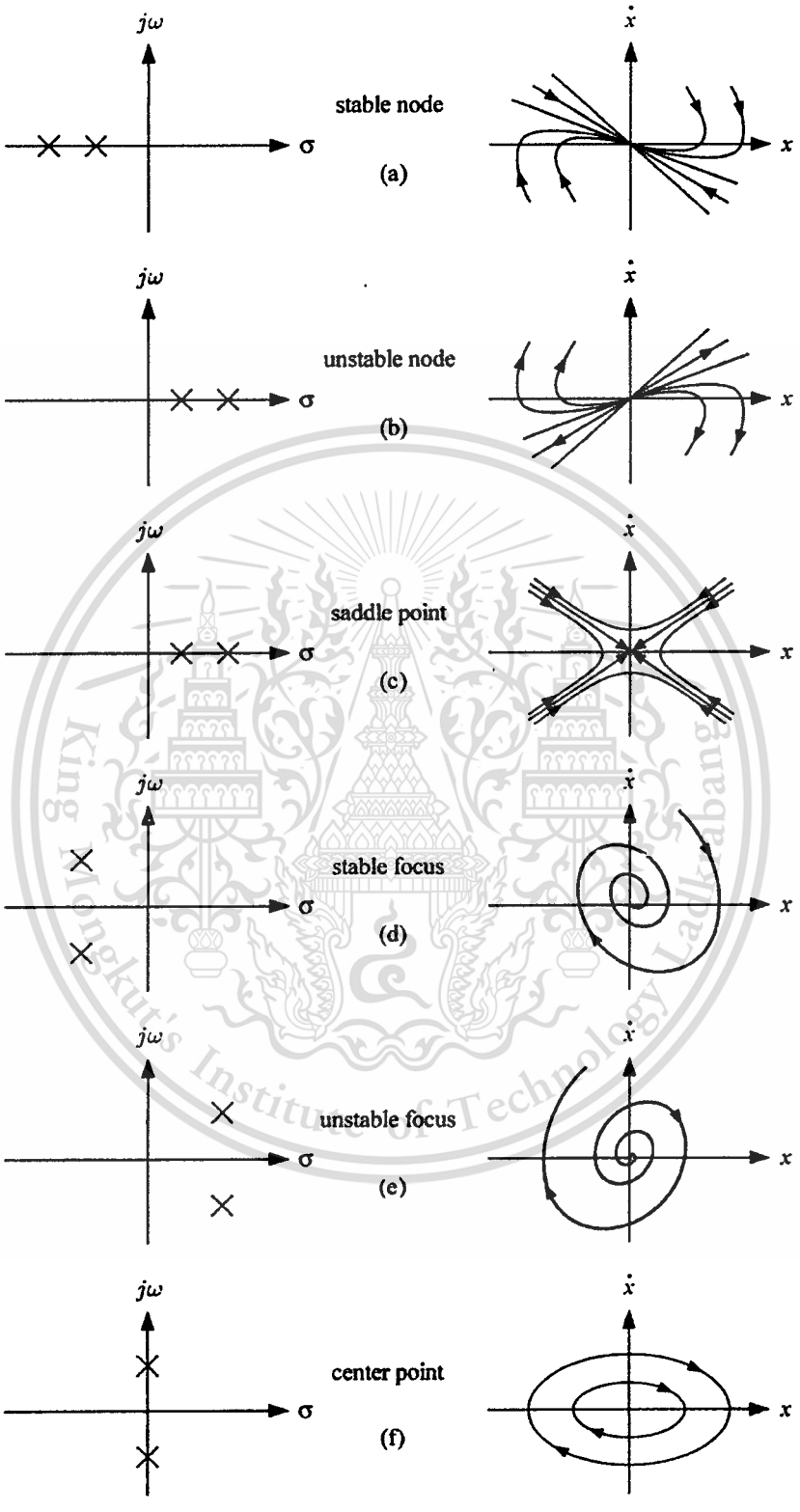


Figure 3.9 Phase-portrait of linear systems

CASE 3 : STABLE OR UNSTABLE FOCUS

The third case corresponds to a *focus*. If the real part of the eigenvalue is negative, then $x(t)$ and $\dot{x}(t)$ both converge to zero. The system trajectories in the vicinity of a stable focus are depicted in Figure 3.9d. Note that the trajectories encircle the origin one or more times before converging to it, unlike the situation for a stable node. If the real part of the eigenvalues is positive, then $x(t)$ and $\dot{x}(t)$ both diverge to infinity and the singularity point is called an unstable focus. The trajectories corresponding to an unstable focus are plotted in Figure 3.9e.

CASE 4 : CENTER POINT

The last case corresponds to a center point, as shown in Figure 3.9f, because all trajectories are ellipses and the singularity point is the center of these ellipses. The phase portrait of the undamped mass-spring system belongs to this category.

Note that the stability characteristics of linear systems are uniquely determined by the nature of their singularity points. However, this is not true for nonlinear systems.

3.5 Phase Plane Analysis of Nonlinear Systems

Phase plane analysis of nonlinear systems is related to that of linear systems, because the local behavior of a nonlinear system can be approximated by a linear one. Moreover, nonlinear systems can display much more complicated patterns in the phase plane, such as multiple equilibrium points and limit cycles.

3.5.1 Local Behavior of Nonlinear Systems

In the phase portrait of Figure 3.2, one notes that, in contrast to linear systems, there are two singular points, $(0, 0)$ and $(-2, 0)$. However, the features of the phase trajectories in the neighborhood of the two singular points look very much like those of linear systems. This similarity to a linear system in the local region of each singular point can be formalized by linearizing the nonlinear system, which is shown below.

If the singular point of interest is not at the origin, by defining the difference between the original state and the singular point as a new set of state variables, one can always shift the singular point to the origin. therefore, without loss of generality, just simply consider Equation (3.1). Using Taylor expansion, Equation (3.1a) and (3.1b) can be written as

$$\begin{aligned} \dot{x}_1 &= ax_1 + bx_2 + g_1(x_1, x_2) \\ \dot{x}_2 &= cx_1 + dx_2 + g_2(x_1, x_2) \end{aligned}$$

where g_1 and g_2 contain higher order terms.

In the neighborhood of the origin, the higher order terms can be neglected, and therefore, the nonlinear system trajectories essentially satisfy the linearized equation

$$\begin{aligned} \dot{x}_1 &= ax_1 + bx_2 \\ \dot{x}_2 &= cx_1 + dx_2 \end{aligned}$$

As a result, the local behavior of the nonlinear system can be approximated by the patterns shown in Figure 3.9

3.5.2 Limit Cycles

In the phase portrait of the nonlinear Van der Pol equation, shown in Figure 3.8, one observes that the system has an unstable node at the origin. Furthermore, there is a closed curve in the phase portrait. Trajectories inside the curve and those outside the curve all tend to this curve, while a motion started on this curve will stay on it forever, circling periodically around the origin. This curve is an instance of the so-called "limit cycle" phenomenon. Limit cycles are unique features of nonlinear systems.

In the phase plane, a limit cycle is defined as an isolated closed curve. The trajectory has to be both closed, indicating the periodic nature of the motion, and isolated, indicating the limiting nature of the cycle (with nearby trajectories converging or diverging from it). Thus, while there are many closed curves in the phase portraits of the mass-spring-damper system in Example 3.1 or the satellite system in Example 3.5, these are not limit cycles because they are not isolated.

Depending on the motion patterns of the trajectories in the neighborhood of the limit cycle, one can distinguish three kinds of limit cycles.

1. All trajectories in the neighborhood of the limit cycle converges to it as $t \to \infty$ (Figure 3.10a). In this case, the limit cycle is said to be *stable.*

2. All trajectories in the neighborhood of the limit cycle diverges from it as $t \to \infty$ (Figure 3.10b). In this case, the limit cycle is said to be *unstable.*

3. Some of the trajectories in the neighborhood converges to it, while the others diverge from it as $t \to \infty$ (Figure 3.10c). In this case, the limit cycle is is called *semi-stable.*

As seen from the phase portrait of Figure 3.8, the limit cycle of the Van der Pol equation is clearly stable. Consider some additional examples of stable, unstable, and semi-stable limit cycles.

Example 3.7. Stable, unstable, and semi-stable limit cycles

Consider the following nonlinear systems

$$(a)\ \dot{x}_1 = x_2 - x_1(x_1^2 + x_2^2 - 1) \qquad \dot{x}_2 = -x_1 - x_2(x_1^2 + x_2^2 - 1) \tag{3.12}$$

$$(b)\ \dot{x}_1 = x_2 + x_1(x_1^2 + x_2^2 - 1) \qquad \dot{x}_2 = -x_1 + x_2(x_1^2 + x_2^2 - 1) \tag{3.13}$$

$$(c)\ \dot{x}_1 = x_2 - x_1(x_1^2 + x_2^2 - 1)^2 \qquad \dot{x}_2 = -x_1 - x_2(x_1^2 + x_2^2 - 1)^2 \tag{3.14}$$

First, study system (a) by introducing polar coordinates

$$r = (x_1^2 + x_2^2)^{1/2} \qquad \theta = \tan^{-1}(x_2/x_1)$$

the dynamic equations (3.12) are transformed as

$$\frac{dr}{dt} = -r(r^2 - 1)$$
$$\frac{d\theta}{dt} = -1$$

When the state starts on the unit circle, the above equation shows that $\dot{r}(t)$=0. Therefore, the state will circle around the origin with a period $1/2\pi$. When $r < 1$, then $\dot{r} > 0$. This implies that the state tends to the circle frc:n inside. When $r > 1$, then $\dot{r} < 0$. This implies that the state tends toward the unit circle from outside. Therefore, the unit circle is a stable limit cycle. This can also be concluded by examining the analytical solution of (3.12)

$$r(t) = \frac{1}{(1 + c_0 e^{-2t})^{1/2}} \qquad \theta(t) = \theta_0 - t$$

where

$$c_0 = \frac{1}{r_0^2} - 1$$

Similarly. one can find that the system (b) has an unstable limit cycle and system (c) has a semi-stable limit cycle.

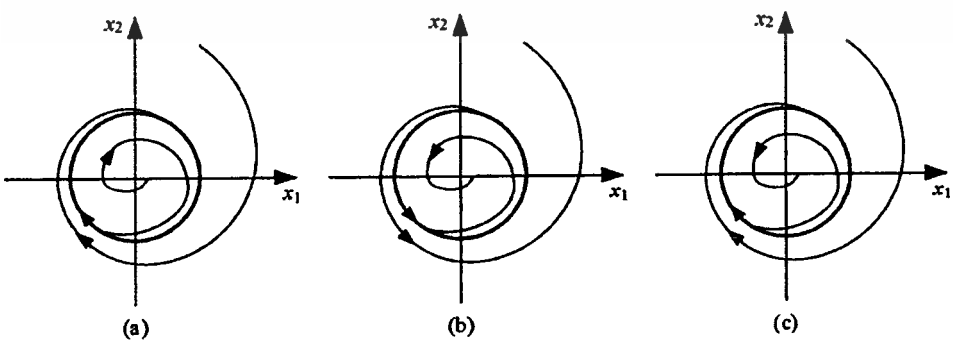


Figure 3.10 Stable, unstable, and semi-stable limit cycles

3.5.3 Detection of Limit Cycles

It is of great importance for control engineers to predict the existence of limit cycles in control systems. Phase plane analysis, Lyapunov analysis, and the describing function method can all be used to predict the existence of limit cycles, although they differ considerably in terms of application ranges. This section state a few simple classical results for predicting the existence of limit cycles in second-order systems of the form (3.1).

The first theorem to be presented indicates a relationship between the existence of a limit cycle and the number of singular points it encloses. In the statement of the theorem, N represents the number of nodes, centers, and foci enclosed by a limit cycle, and S represents the number of enclosed saddle points.

Theorem 3.5.1. (Poincare)[1] If a limit cycle exists in the second-order system, then $N = S + 1$ where N represents the number of nodes, centers, and foci enclosed by a limit cycle, and S for the number of enclosed saddle points.

Theorem 3.5.2. (Poincare-Bendixson)[1] If a trajectory of the second-order system remains in a finite region ω, then one of the following is true:

- The trajectory goes to an equilibrium point.
- The trajectory tends to an asymptotically stable limit cycle.
- The trajectory is itself a limit cycle.

Theorem 3.5.3. (Bendixson)[1] For the nonlinear system, no limit cycle can exist in a region ω of the phase plane in which $\partial f_1/\partial x_1 + \partial f_2/\partial x_2$ appears and does not change it's sign.

CHAPTER 4

Describing Function Analysis

4.1 Introduction

The frequency response method is a powerful tool for the analysis and design of linear control systems. It is based on describing a linear system by a complex-valued function, the frequency response, instead of a differential equation. The power of the method comes from a number of source. First, the analysis of a system's behavior, such as stability and sinusoidal response, is reduced to the relatively simple examination of a set of *algebraic* equations in the frequency domain. Second, *graphical* representations can be used to facilitate analysis and design. Third, *physical* insights can be used, because the frequency response functions have clear physical meanings. Finally, the method's complexity only increases mildly with system order. Frequency domain analysis, however, cannot be directly applied to nonlinear systems because frequency response functions cannot be defined for nonlinear systems.

For some nonlinear systems and certain conditions, an extended version of the frequency response method, called the *describing function method* [1], can be used to approximately analyze and predict nonlinear behavior. Even though it is only an approximation method, the desirable properties it inherits from the frequency response method of linear control, and the shortage of other systematic tools for nonlinear system analysis. The main use of describing function method is for the prediction of limit cycles in nonlinear systems, although the method has a number of other practical applications.

4.2 Describing Function Fundamentals

This section provides some basic concepts associated with the describing function analysis start with a simple example.

4.2.1 Example of Describing Function Analysis

Consider the Van der Pol equation

$$\ddot{x} + \mu(x^2 - 1)\dot{x} + x = 0 \tag{4.1}$$

Determine whether there exists a limit cycle in this system and, if so, calculate the amplitude and frequency of the limit cycle (pretending that the phase portrait of the Van der Pol equation in chapter 2 is unseen). To this effect, first assume the existence of a limit cycle with undetermined amplitude and frequency, and then determine whether the system equation can indeed sustain such a solution.

This is quite similar to the assumed-variable method in differential equation theory substitute it into the differential equation, and then attempt to determine the coefficients in the solution.

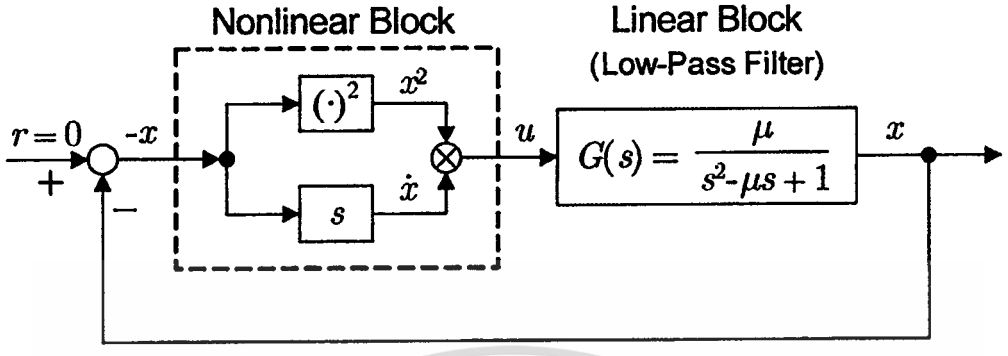


Figure 4.1 Feedback interpretation of the Van der Pol oscillator

Consider the system dynamics in a block diagram form, as shown in Figure 4.1 It is seen that the feedback system in (4.1) contains a linear block and a nonlinear block, where the linear block is a low-pass filter.

Assume that there is a limit cycle in the system and the oscillation signal x is in the form of

$$x(t) = A\sin(\omega t)$$

with A being the limit cycle amplitude and ω being the frequency. Thus,

$$\dot{x}(t) = A\omega\cos(\omega t)$$

Therefore, the output of the nonlinear block is

$$\begin{aligned} u &= -x^2\dot{x} = -A^2\sin^2(\omega t)A\omega\cos(\omega t) \\ &= -\frac{A^3\omega}{2}(1-\cos(2\omega t))\cos(\omega t) = -\frac{A^3\omega}{4}(\cos(\omega t)-\cos(3\omega t)) \end{aligned}$$

It is seen that u contains a third harmonic term. Since the linear block is a low-pass filter, it's reasonable to assume that this third harmonic term is sufficiently attenuated by the linear block and its effect is not present in the signal flow after the linear block. This means that u can be approximated by

$$u \approx -\frac{A^3}{4}\omega\cos\omega t = \frac{A^2}{4}\frac{d}{dt}[-A\sin(\omega t)]$$

so that the nonlinear block in Figure 4.1 can be approximated by the equivalent "quasi-linear" block in Figure 4.2. The "transfer function" of the quasi-linear block depends on the signal amplitude A, unlike a linear system transfer function (which is independent of the input magnitude).

In the frequency domain, this corresponds to

$$u = D(A,\omega)(-x) \tag{4.2}$$

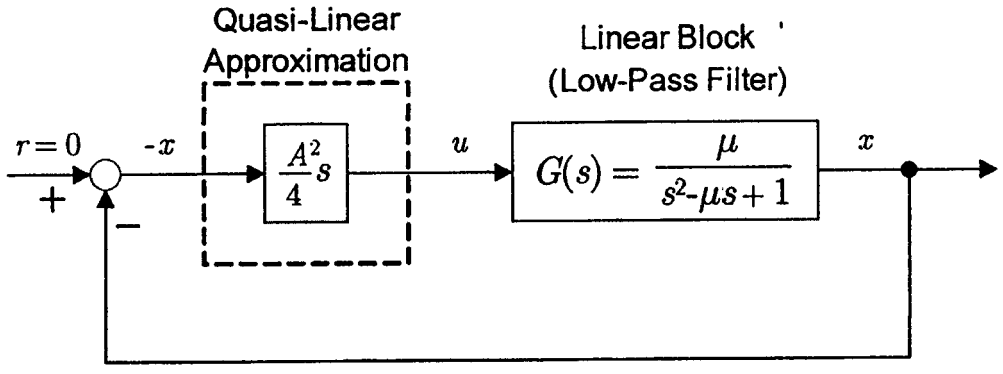


Figure 4.2 Quasi-linear approximation of the Van der Pol oscillator

where

$$D(A,\omega) = \frac{A^2}{4}(j\omega)$$

That is, the nonlinear block can be approximated by the frequency response function $D(A,\omega)$. Since the system is assumed to contain a sinusoidal oscillation,

$$x = A\sin(\omega t) = G(j\omega)u = G(j\omega)D(A,\omega)(-x)$$

where $G(j\omega)$ is the linear component transfer function. This implies that

$$1 + \frac{A^2(j\omega)}{4}\frac{\mu}{(j\omega)^2 - \mu(j\omega) + 1} = 0$$

Solving this equation, then

$$A = 2 \qquad \omega = 1$$

Note that in terms of the Laplace variable s, the closed-loop characteristic equation of this system is

$$1 + \frac{A^2 s}{4}\frac{\mu}{s^2 - \mu s + 1} = 0 \tag{4.3}$$

whose eigenvalues are

$$\lambda_{1,2} = -\frac{1}{8}\mu(A^2 - 4) \pm \sqrt{\frac{1}{64}\mu^2(A^2 - 4)^2 - 1} \tag{4.4}$$

Corresponding to $A = 2$, then the eigenvalues $\lambda_{1,2} = \pm j$. This indicates the existence of a limit cycle of amplitude 2 and frequency 1. It is interesting to note neither the amplitude nor the frequency obtained above depends on the parameter μ in Equation 4.1.

In the phase plane, the above approximation analysis suggests that the limit cycle is a circle of radius 2. To verify the plausibility of this result, the real limit cycles corresponding to the different values of μ are plotted. It is seen that the above approximation is reasonable for small value of μ, but the inaccuracy

grows as μ increases. This is understandable because as μ grows the nonlinearity becomes more significant and the quasi-linear approximation becomes less accurate.

The stability of the limit cycle can also be studied using the above analysis. Assume that the limit cycle amplitude A is increased to a value larger than 2. Then, (4.4) indicates that the closed-loop poles now have a negative real part. This indicates that the system becomes exponentially stable and thus the signal magnitude will decrease. Similar conclusions are obtained assuming that the limit cycle amplitude A is decreased to a value less than 2. Thus, the limit cycle is stable with an amplitude of 2.

Note that, in the above approximate analysis, the critical step is to replace the nonlinear block by the quasi-linear block which has the frequency response function $(A^2/4)(j\omega)$. Afterwards, the amplitude and frequency of the limit cycle can be determined from $1 + G(j\omega)D(A,\omega) = 0$. The function $D(A,\omega)$ is called the describing function of the nonlinear element. The above approximate analysis can be extended to predict limit cycles in other nonlinear systems which can be represented into the block diagram similar to Figure 4.1.

4.2.2 Applications Domain

Before moving on to the formal treatment of the describing function method, there is a briefly discuss what kind of nonlinear systems is applies to, and what kind of information it can provide about nonlinear system behavior.

The Systems

Any systems which can be transformed into the configuration in Figure 4.3 can be studied using describing functions. There are at least two important classes of systems in this category.

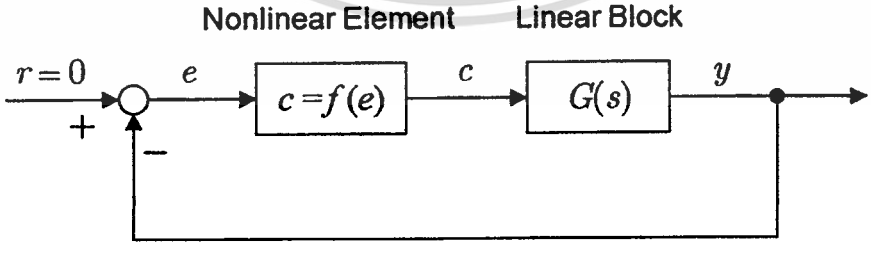


Figure 4.3 A nonlinear system

The first important class consists of "almost" linear systems. By "almost" linear systems, refer to systems which contain one hard nonlinearity in the control loop but are otherwise linear. Such systems arise when a control system is designed using linear control but its implementation involves hard nonlinearities,

such as motor saturation, actuator or sensor dead-zones, Colomb friction, or hysteresis in the plant. An example is shown in Figure 4.4, which involves hard nonlinearities in the actuator.

Example 4.1. A system containing only one nonlinearity.

Consider the control system shown in Figure 4.4. The plant is linear and the controller is also linear. However, the actuator involves a hard nonlinearity. This system can be rearranged into the form of Figure 4.3 by regarding $G_pG_1G_2$ as the linear component G, and the actuator nonlinearity as the nonlinear element N.

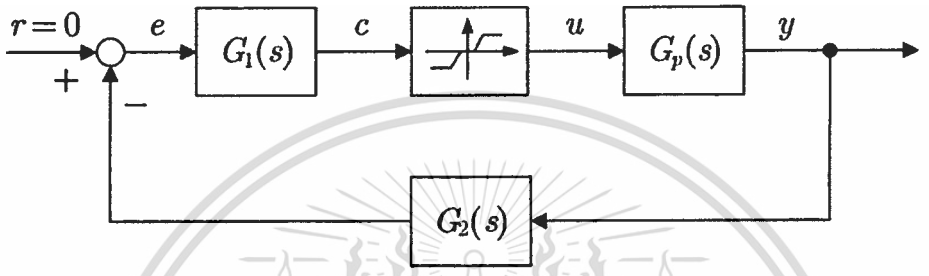


Figure 4.4 A control system with hard nonlinearity

"Almost" linear systems involving sensor or plant nonlinearities can be similarly rearranged into the form of Figure 4.3.

The second class of systems consists of genuinely nonlinear systems whose dynamic equations can actually be rearranged into the form of Figure 4.3. For example, the nonlinear equation

$$m\ddot{x} + c\dot{x} + kx + k_1x^3 = 0$$

can be rewritten as

$$m\ddot{x} + c\dot{x} + kx = -k_1x^3$$

which can be represented by the block diagram shown in Figure 4.5.

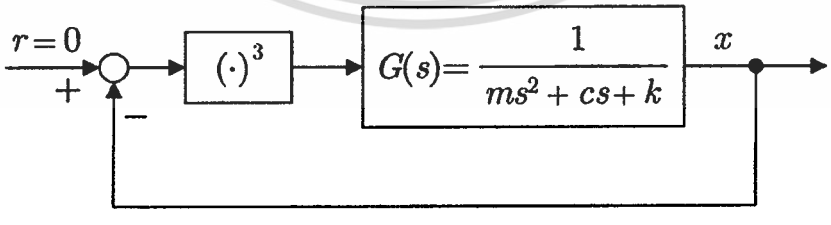


Figure 4.5 Transforming a nonlinear system into the standard form

Applications of Describing Functions

For systems such as the one in Figure 4.4, limit cycles can often occur due to the nonlinearity. However, linear control cannot predict such problems. Describing functions, on the other hand, can be conveniently used to discover the existence of limit cycles and determine their stability, regardless of whether the nonlinearity is "hard" or "soft". the applicability to limit cycle analysis is due to the fact that the form of the signals in a limit-cycling system is usually approximately sinusoidal. This can be conveniently explained on the system in Figure 4.5. Assume that there is a limit cycle in the system, the system signals must all be periodic according to definition. Since the input to the linear element in Figure 4.5, as a periodic signal, can be expanded as the sum of many harmonics, and the linear element, because of its low-pass property, filters out higher frequency signals, the output $x(t)$ must be composed of mostly the lowest harmonics. Therefore, it is appropriate to assume the signals in the whole system are basically sinusoidal in form. Remark that the describing function method can also be used for some other purposes, such as predicting the closed-loop response of the systems representable by Figure 4.3 in the presence of external sinusoidal excitation $r(t) = r_0 \sin(\omega t)$.

Prediction of limit cycles is very important, because limit cycles can occur in any kind of physical nonlinear system. Sometimes, a limit cycle can be desirable. This is the case of limit cycles in the electronic oscillators used in laboratories. Another example is the so-called dither technique which can be used to minimize the negative effects of Coulomb friction in mechanical systems. In most control systems, however, limit cycles are undesirable. This may be due to a number of reasons:

1. Limit cycling, as a way of instability, tends to cause poor control accuracy.
2. The constant oscillation associated with the limit cycles can cause increasing wear or even mechanical failure of the control system hardware.
3. Limit cycling may also cause other undesirable effects, such as passenger discomfort in an aircraft under autopilot.

In general, although a precise knowledge of the waveform of a limit cycle is usually not mandatory, the knowledge of the limit cycle's existence, as well as that of its approximate amplitude and frequency, is critical. The describing function method can be used for this purpose. It can also guide the design of compensators so as to avoid limit cycles.

4.2.3 Basic Assumptions

Consider a nonlinear system in the general form of Figure 4.3. In order to develop the basic version of the describing function method, the system has to satisfy the following four condition

1. There is only a single nonlinear component.
2. The nonlinear component is time-invariant.
3. Corresponding to a sinusoidal input $e = \sin(\omega t)$, only the fundamental component $c_1(t)$ in the output $c(t)$ has to be considered.
4. The nonlinearity is odd

The first assumption implies that if there are two or more nonlinear components in a system,one either has to lump them together as a single nonlinearity (as can be done with two nonlinearities in parallel), or retain only the primary nonlinearity and neglect the others.

The second assumption implies that only autonomous nonlinear systems are considered. It is satisfied by many nonlinearities in practice, such as saturation in amplifiers, backlash in gears, Coulomb friction between surfaces, and hysteresis in relays. The reason for this assumption is that the Nyquist criterion, on which the describing function method is largely based, applies only to linear time-invariant systems.

The third assumption is the fundamental assumption of the describing function method. It represents an approximation, because the output of a nonlinear element corresponding to a sinusoidal input usually contains higher harmonics besides the fundamental. This assumption implies that the higher-frequency harmonics can all be neglected in the analysis, as compared with the fundamental component. For this assumption to be valid, it is important for the linear element following the nonlinearity to have low-pass properties, *i.e.,*

$$|G(j\omega)| >> |G(jn\omega)| \qquad for \;\; n = 2, 3, \ldots \tag{4.5}$$

This implies that higher harmonics in the output will be filtered out significantly. Thus, the third assumption is often referred to as the filtering hypothesis.

The forth assumption means that the plot of the nonlinearity relation $f(e)$ between the input and output of the nonlinear element is symmetric about the origin. This assumption is introduced for simplicity, *i.e.,* so that the static term in the Fourier expansion of the output can be neglected. Note that the common nonlinearities discussed before all satisfy this assumption.

The relaxation of the above assumptions has been widely studied in literature, leading to describing function approaches for general situations, such as multiple nonlinearities, time-varying nonlinearities, or multiple-sinusoids. However, these methods based on relaxed conditions are usually much more complicated than the basic version, which corresponds to the above assumptions.

4.2.4 Basic Definitions

The describing function is a representation of a nonlinear component that looks like an application of the frequency response method. First, consider a sinusoidal input to the nonlinear element, of amplitude A and frequency ω, *i.e.*, $e(t) = A\sin(\omega t)$, as shown in Figure 4.6. The output of the nonlinear component $c(t)$ is often a periodic, though generally non-sinusoidal, function. Note that this is always the case if the nonlinearity $f(e)$ is single-valued, because the output if $f[A\sin(\omega(t + 2\pi/\omega))] = f[A\sin(\omega t)]$. Using Fourier series, this periodic function can be expanded as

$$c(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n\cos(n\omega t) + b_n\sin(n\omega t)] \tag{4.6}$$

where the Fourier coefficients a_i and b_i are generally functions of A and ω, determined by

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} c(t)d(\omega t) \tag{4.7a}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} c(t)\cos(n\omega t)d(\omega t) \tag{4.7b}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} c(t)\sin(n\omega t)d(\omega t) \tag{4.7c}$$

Due to the forth assumption above, one has $a_0 = 0$. Furthermore, the third assumption implies that we only need to consider the fundamental component $c_1(t)$, namely

$$c(t) \approx c_1(t) = a_1\cos(\omega_0 t) + b_1\sin(\omega_0 t) = M\sin(\omega_0 t + \phi), \tag{4.8}$$

where

$$M(A,\omega_0) = \sqrt{a_1^2 + b_1^2} \quad and \quad \phi(A,\omega_0) = \arctan(a_1/b_1)$$

Expression above indicates that the fundamental component corresponding to a sinusoidal input is a sinusoidal at the same frequency. In complex representation, this sinusoid can be written as

$$c_1 = Me^{j(\omega_0 t+\phi)} = (b_1 + ja_1)e^{j\omega_0 t}$$

Similarly to the concept of frequency response function, which is the frequency-domain ratio of the sinusoidal input and the sinusoidal output of a system, the *describing function* is defined as a complex ratio of the fundamental component of the nonlinear element by the input sinusoidal,

$$D(A,\omega_0) = \frac{Me^{j(\omega_0 t+\phi)}}{Ae^{j\omega_0 t}} = \frac{M}{A}e^{j\phi} = \frac{1}{A}(b_1 + ja_1) \tag{4.9}$$

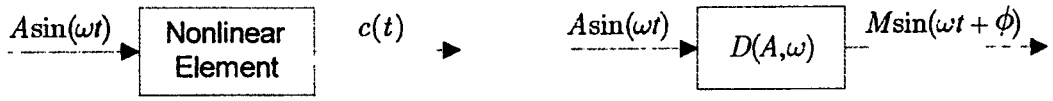


Figure 4.6 A nonlinear element and its describing function representation

With a describing function representing the nonlinear component, the nonlinear element, in the presence of sinusoidal input, can be treated as if it were a linear element with a transfer function $D(A,\omega)$, as shown in Figure 4.6. The concept of the describing function can thus be regarded as an extension of the notion of frequency response. For a linear dynamic system with transfer function $H(j\omega)$, the describing function is independent of the input gain, as can be easily shown. However, the describing function of a nonlinear element differs from the frequency response function of a linear element in that it depends on the input amplitude A.

Generally, the describing function depends on the frequency an amplitude of the input signal. There are, however, a number of special cases. When the nonlinearity is single-valued, the describing function $D(A,\omega)$ is real and independent of the input frequency ω. The realness of D is due to the fact that $a_1 = 0$, which is true because $f[A\sin(\omega t)]\cos(\omega t)$, the integrand in the expression (4.7b) for a_1, is an odd function of ωt, and the domain of integration is the symmetric interval $[-\pi,\pi]$. The frequency-independent nature is due to the fact that the integration of the single-valued function $f[A\sin(\omega t)]\sin(\omega t)$ in expression (4.7c) is done for the variable ωt, which implies that ω does not explicitly appear in the integration.

Although we have implicitly assumed the nonlinear element to be a scalar nonlinear function, the definition of the describing function equally applies to the case when the nonlinear element contains dynamic (*i.e.*, it should be described by differential equations instead of a function). The describing function method indeed applies to such systems, but the evaluation of the describing functions for such nonlinear elements is much more complicated and requires experimental determination.

4.2.5 Computing Describing Function

A number of methods are available to determine the describing functions of nonlinear elements in control systems, based on definition (4.9). There are three such methods: analytical calculation, experimental determination, and numerical integration. Convenience and cost in each particular application determine which method should be used. One thing to remember is that precision is not critical in evaluating describing functions of nonlinear elements, because the describing function method is itself an approximate method.

Analytical Calculation

When the nonlinear characteristic $c = f(e)$ (where e is the input and c is the output) of the nonlinear element are describe by an explicit function and the integration in (4.7) can be easily carried out, then analytical evaluation of the describing function based on (4.7) is desirable. The explicit function $f(e)$ of the nonlinear element may be an idealized representation of simple nonlinearities such as saturation and dead-zone, or it may be the curve-fit of an input-output relationship for the element.

Numerical Integration

For nonlinearities whose input-output relationship $c = f(e)$ is given by graphs or tables, it is convenient to use numerical integration to evaluate the describing function. The idea is to approximate integrals in (4.7) by discrete sums over small intervals. Various numerical integration schemes can be applied for this purpose. It is obviously important that the numerical integration be easily implementable by computer programs. The result is a plot representing the describing function, which can be used to predict limit cycles.

Experimental Evaluation

The experimental method is particularly suitable for complex nonlinearities and dynamic nonlinearities. When a system nonlinearities can be isolated and excited with sinusoidal input of known amplitude and frequency, experimental determination of the describing function can be obtained by using a harmonic analyzer on the output of the nonlinear element. This is quite similar to the experimental determination of transfer functions for linear elements. The difference is that not only the frequencies, but also the amplitudes of the input sinusoidal should be varied. The results of the experiments are a set of curves on complex planes representing the describing function $D(A, \omega)$, instead of analytical expressions. Specializes instruments are available which automatically compute the describing functions of nonlinear elements based on the measurement of nonlinear element response to harmonic excitation.

The next example illustrates how to evaluate describing functions by using the analytical technique on a simple nonlinearity.

Example 4.2. Describing function of a hardening spring

The characteristics of a hardening spring are given by

$$y = x + \frac{x^3}{2}$$

with x being the input and y being the output. Given an input $x(t) = A\sin(\omega t)$, the output

$$y(t) = A\sin(\omega t) + \frac{1}{2}A^3\sin^3(\omega t)$$

can be expanded as a fourier series, with the fundamental being

$$y_1(t) = a_1\cos(\omega t) + b_1\sin(\omega t)$$

Because $y(t)$ is an odd function, one has $a_1 = 0$, according to (4.7). The coefficient b_1 is

$$\begin{aligned} b_1 &= \frac{1}{\pi}\int_{-\pi}^{\pi}\left[A\sin(\omega t) + \frac{1}{2}A^3\sin^3(\omega t)\right]\sin(\omega t)d(\omega t) \\ &= A + \frac{3}{8}A^3 \end{aligned}$$

Therefore, the fundamental is

$$y_1 = \left(A + \frac{3}{8}A^3\right)\sin(\omega t)$$

and the describing function of this nonlinear component is

$$D(A,\omega) = D(A) = 1 + \frac{3}{8}A^2$$

Note that due to the odd nature of this nonlinearity, the describing function is real, being a function only of the amplitude of the sinusoidal input.

4.3 Common Nonlinearities in Control Systems

This section takes a closer look at the nonlinearities found in control systems. Consider the typical system block shown in Figure 4.7. It is composed of four parts: a plant to be controlled, sensors for measurement, actuators for control action. and a control law, usually implemented on a computer. Nonlinearities may occur in any part of the system, and thus make it a nonlinear control system.

Continuous and Discontinuous Nonlinearities

Nonlinearities can be classified as *continuous* and *discontinuous*. Because discontinuous nonlinearities cannot be locally approximated by linear functions. They are also called "hard" nonlinearities. Hard nonlinearities are commonly found in control systems, both in small range operation and large range operation. Whether a system in small range operation should be regarded as nonlinear or linear depends on the magnitude of the hard nonlinearities and on the extent of their effects on the system performance.

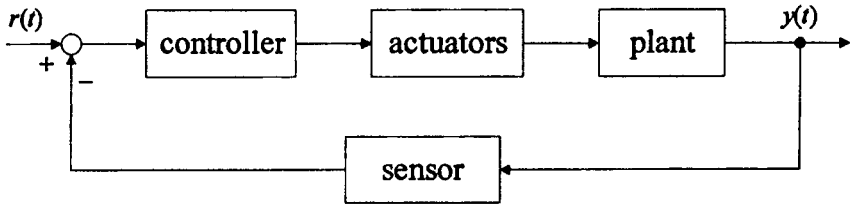


Figure 4.7 Block diagram of a control system

Saturation

Saturation is probably the most commonly encountered nonlinearity in control engineering. The saturation nonlinearity is usually caused by limits on component size, properties of materials, and available power. It is often associated with amplifiers and actuators, both of which are important components of control systems. In transistor amplifiers and magnetic amplifiers, the output varies linearly with the input only for small amplitude inputs. When the input amplitude gets out of the linear range of the amplifier, the output changes very little and stays close to its maximum value, giving rise to the saturation phenomenon. The characteristics of saturation nonlinearity is plotted in Figure 4.8, where the solid line is the real nonlinearity and the dotted line is an idealized saturation nonlinearity.

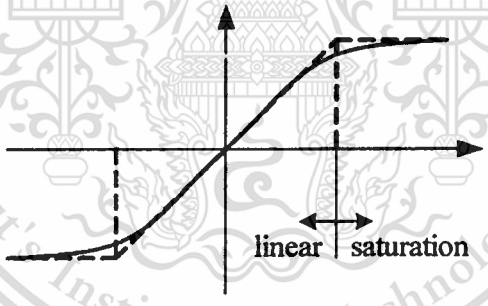


Figure 4.8 Saturation nonlinearity

Most actuators display saturation characteristics. For example, the output torque of a two-phase servo motor cannot increase infinitely and tends to saturate, due to the properties of the magnetic material. Similarly, valve-controlled hydraulic servo motors are saturated by the maximum flow rate.

Saturation can have complicated effects on control system performance. Roughly speaking, the occurence of saturation amounts to reducing the gain of the component (*e.g.*, the amplifier) as the input signals are increased. As a result, if a system is unstable in its linear range, its divergent behavior may be suppressed into a self-sustained oscillation, due to the inhibition created by the saturating component on the system signals. On the other hand, in a linearly stable system, saturation tends to slow down the response of the system, because it reduces the effective gain.

On-off nonlinearity

An extreme case of saturation is the on-off or relay nonlinearity. It occurs when the linearity range is shrunken to zero and the slope in the linearity range becomes vertical. Important examples of on-off nonlinearities include output torques of gas jets for spacecraft control and, of course, electrical relays. On-off nonlinearities have effects similar to those of saturation nonlinearities.

Dead-zone

In many physical devices, the output is zero until the magnitude of the input exceeds a certain value. Consider for instance a D.C. motor. In linear system analysis, assume that any voltage applied to the armature windings will cause the armature to rotate, if the field current is maintained constant. In reality, due to the static friction at the motor shaft, rotation will occur only if the torque provided by the motor is sufficiently large. This corresponds to a so-called dead-zone for small voltage signals. Similar dead-zone phenomena occur in valve-controlled pneumatic actuators and in hydraulic components.

Generally speaking, a dead-zone nonlinearity may occur in various components of control systems, including sensors, amplifiers, and actuators. Dead-zones can have a number of possible effects on control systems. Their most common effect is to decrease static output accuracy. They may also lead to limit cycles or system instability. In some cases, however, they may actually stabilize a system or suppress self-oscillations. For example, if a dead-zone is incorporated into an ideal relay, it may lead to the avoidance of the oscillation at the contact point of the relay, thus eliminating sparks and reducing wear at the contact point.

Backlash and hysteresis

Another kind of common nonlinearity in control components, particularly in mechanical components, is backlash. It is usually caused by the small gaps which exist in transmission mechanisms. In gear trains, there always exist small gaps between a pair of mating gears (Figure 4.9), due to the unavoidable errors in manufacturing and assembly. As a result, when the driving gear rotates a smaller angle than the gap b. the driven gear does not move at all, which corresponds to the dead-zone (OA segment in Figure 4.9); after contact has been established between the two gears. the driven gear follows the rotation of the driving gear in a linear fashion (AB segment). When the driving gear rotates in the reverse direction by a distance of $2b$, the driven gear again does not move, corresponding to the BC segment in Figure 4.9. After the contact between the two gears is re-established, the driven gear follows the rotation of the driving gear in the reverse direction (CD segment). Therefore, if the driving gear is in periodic motion, the driven gear will move in the fashion rep'resented by the closed path $EBCD$.

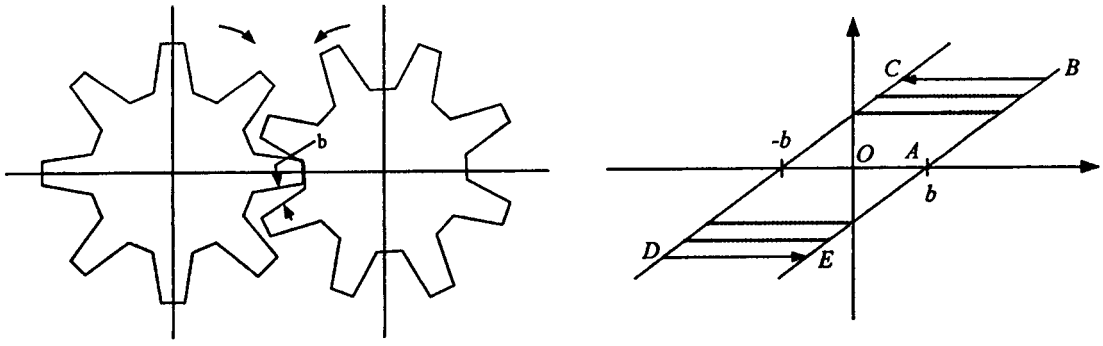


Figure 4.9 A backlash nonlinearity

A critical feature of backlash is its multi-valued nature. Corresponding to each input, two output values are possible. Which one of the two occur depends on the history of the input. Remark that a similar multi-valued nonlinearity is hysteresis. which is frequently observed in relay components.

Multi-valued nonlinearities like backlash and hysteresis usually lead to energy storage in the system. Energy storage is a frequent cause of instability and self-sustained oscillation.

4.4 Describing Functions of Common Nonlinearities

In this section, there is a computation of the describing functions for a few common nonlinearities. This will not only allow us to familiarize ourselves with the frequency domain properties of these common nonlinearities, but also will provide further examples of how to derive describing functions for nonlinear elements.

4.4.1 Saturation

The input-output relationship for a saturation nonlinearity is plotted in Figure 4.10, with a and k denoting the range and slope of the linearity. Since this nonlinearity is single-valued, we expect the describing function to be a real function of the input amplitude.

Consider the input $e(t) = A\sin(\omega t)$. If $A \leq a$, then the input remains in the linear range, and therefore, the output is $y(t) = kA\sin(\omega t)$. Hence, the describing function is simply a constant k.

$$c(t) = \begin{cases} kA\sin(\omega)t & 0 \leq \omega t \leq \omega t_1 \\ ka & \omega t_1 < \omega t \leq \pi/2 \end{cases}$$

where $\omega t_1 = \sin^{-1}(a/A)$. The odd nature of $c(t)$ implies that $a_1 = 0$ and the

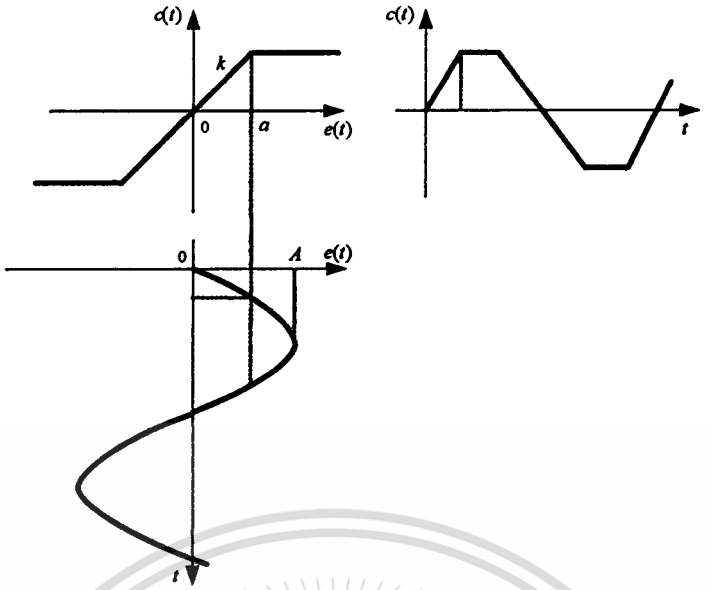


Figure 4.10 Saturation nonlinearity and the corresponding input-output relationship

symmetry over the four quarters of a period implies that

$$\begin{aligned} b_1 &= \frac{4}{\pi}\int_0^{\pi/2} c(t)\sin(\omega t)d(\omega t) \\ &= \frac{4}{\pi}\int_0^{\omega t_1} kA\sin^2(\omega t)d(\omega t) + \frac{4}{\pi}\int_{\omega t_1}^{\pi/2} ka\sin(\omega t)d(\omega t) \\ &= \frac{2kA}{\pi}\left[\omega t_1 + \frac{a}{A}\sqrt{1-\frac{a^2}{A^2}}\right] \end{aligned} \tag{4.10}$$

Therefore, the describing function is

$$D(A) = \frac{b_1}{A} = \frac{2k}{\pi}\left[\sin^{-1}\frac{a}{A} + \frac{a}{A}\sqrt{1-\frac{a^2}{A^2}}\right] \tag{4.11}$$

The normalized describing function $D(A)/k$ is plotted in Figure 4.11 as a function of A/a. The curve indicates that saturation can be described as an amplitude-dependent gain, which equals k if the input amplitude A is smaller than the linearity range, and decreases as the input amplitude increases. The saturation nonlinearity is inhibitory to input signals larger than a, but it causes no phase shift because $D(A)$ is real.

As a special case, one can obtain the describing function for the relay-type (on-off) nonlinearity shown in Figure 4.12. This case corresponds to shrinking the linearity range in the saturation function to zero, *i.e.*, $a \to 0$, $k \to \infty$, but $ka = M$. Though hi can be obtained from (4.10) by taking the limit, it is more

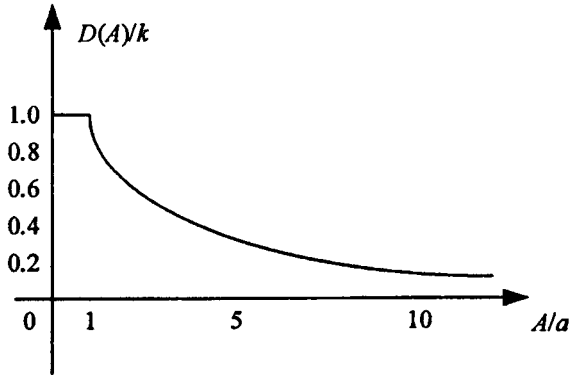


Figure 4.11 Describing function of the saturation of the saturation nonlinearity

easily obtained directly as

$$b_1 = \frac{4}{\pi}\int_0^{\pi/2} M\sin(\omega t)d(\omega t) = \frac{4}{\pi}M$$

Therefore, the describing function of the relay nonlinearity is

$$D(A) = \frac{4M}{\pi A} \tag{4.12}$$

The normalized describing function (D/M) is plotted in Figure 4.12 as a function of input amplitude. The flat segment seen in Figure 4.11 is missing in this plot, due to the completely nonlinear nature of the relay. However, we can also observe the inhibitory nature of the relay, *i.e.*, the describing function decreases as the input amplitude increases.

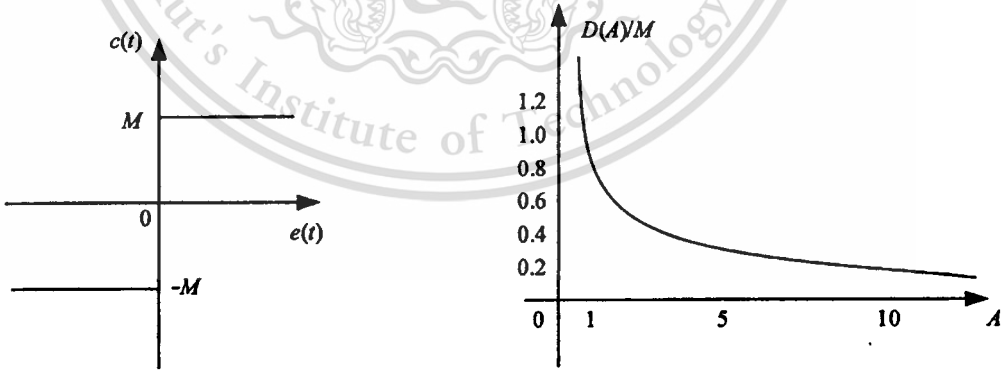


Figure 4.12 Relay nonlinearity and its describing function

4.4.2 Dead-Zone

The dead-zone characteristics are shown in Figure 4.13, with the dead-zone width being 2δ and its slope k. The response corresponding to a sinusoidal input $e(t) = A\sin(\omega t)$ into a dead-zone of width 2δ and slope k, with $A \geq \delta$, is plotted in Figure 5.13. Since the characteristics is an odd function, $a_1 = 0$. The response is also seen to be symmetric over the four quarters of a period. In one quarter of a period, *i.e.*, when $0 \leq \omega t \leq \pi/2$, one has

$$c(t) = \begin{cases} 0 & 0 \leq \omega t \leq \omega t_1 \\ k(A\sin(\omega t) - \delta) & \omega t_1 < \omega t \leq \pi/2 \end{cases} \tag{4.13}$$

where $\omega t_1 = \sin^{-1}(\delta/A)$. The coefficient b_l can be computed as follows

$$\begin{aligned} b_1 &= \frac{4}{\pi}\int_0^{\pi/2} c(t)\sin(\omega t)d(\omega t) = \frac{4}{\pi}\int_{\omega t_1}^{\pi/2} k(A\sin(\omega t) - \delta)\sin(\omega t)d(\omega t) \\ &= \frac{2kA}{\pi}\left[\frac{\pi}{2} - \sin^{-1}\frac{\delta}{A} - \frac{\delta}{A}\sqrt{1 - \frac{\delta^2}{A^2}}\right] \end{aligned} \tag{4.14}$$

Therefore, the describing function is

$$D(A) = \frac{2k}{\pi}\left[\frac{\pi}{2} - \sin^{-1}\frac{\delta}{A} - \frac{\delta}{A}\sqrt{1 - \frac{\delta^2}{A^2}}\right] \tag{4.15}$$

This describing function $D(A)$ is a real function and, therefore, there is no phase shift. It is seen that D/k is zero when $A/\delta < 1$, and increases up to 1 with A/δ. This increase indicates that the effect of the dead-zone gradually diminishes as the amplitude of the input signal is increased, consistent with intuition.

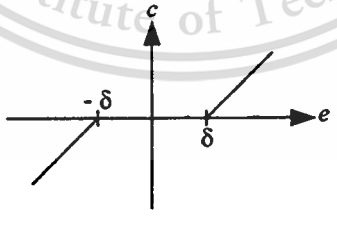


Figure 4.13 Characteristics of a dead-zone nonlinearity

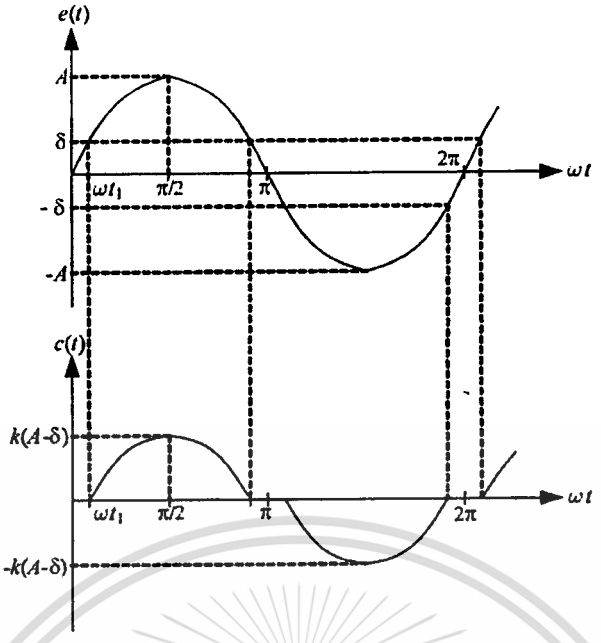


Figure 4.14 Input and general form of the output for a dead-zone nonlinearity

4.4.3 Backlash

A backlash nonlinearity is plotted in Figure 4.15, with slope k and width 2δ. When the input is $e(t) = A\sin(\omega t)$, $A \geq \delta$, the output $c(t)$ of the nonlinearity is as shown in the figure. In one cycle, the function c(t) can be represented as

$$\begin{aligned}
c(t) &= (A-\delta)k && \tfrac{\pi}{2} < \omega t \leq \pi - \omega t_1 \\
c(t) &= (A\sin(\omega t)+\delta)k && \pi - \omega t_1 < \omega t \leq \tfrac{3\pi}{2} \\
c(t) &= -(A-\delta)k && \tfrac{3\pi}{2} < \omega t \leq 2\pi - \omega t_1 \\
c(t) &= (A\sin(\omega t)-\delta)k && 2\pi - \omega t_1 \leq \tfrac{5\pi}{2}
\end{aligned}$$

Unlike the previous nonlinearities, the function $c(t)$ here is neither odd nor even. Therefore, a_1 and b_1 are both nonzero. Using (5.7b) and (5.7c), it possible to state that

$$a_1 = \frac{4k\delta}{\pi}\left(\frac{\delta}{A} - 1\right) \tag{4.16}$$

$$b_1 = \frac{Ak}{\pi}\left[\frac{\pi}{2} - \sin^{-1}\left(\frac{2\delta}{A} - 1\right) - \left(\frac{2\delta}{A} - 1\right)\sqrt{1 - \left(\frac{2\delta}{A} - 1\right)^2}\right] \tag{4.17}$$

Therefore, the describing function of the backlash is givenby

$$|D(A)| = \frac{1}{A}\sqrt{a_1^2 + b_1^2} \tag{4.18}$$

$$\angle D(A) = \tan^{-1}(a_1/b_1) \tag{4.19}$$

Note that a phase shift (up to 90) is introduced for larger input signals. This phase lag may create stability problems in feedback control systems.

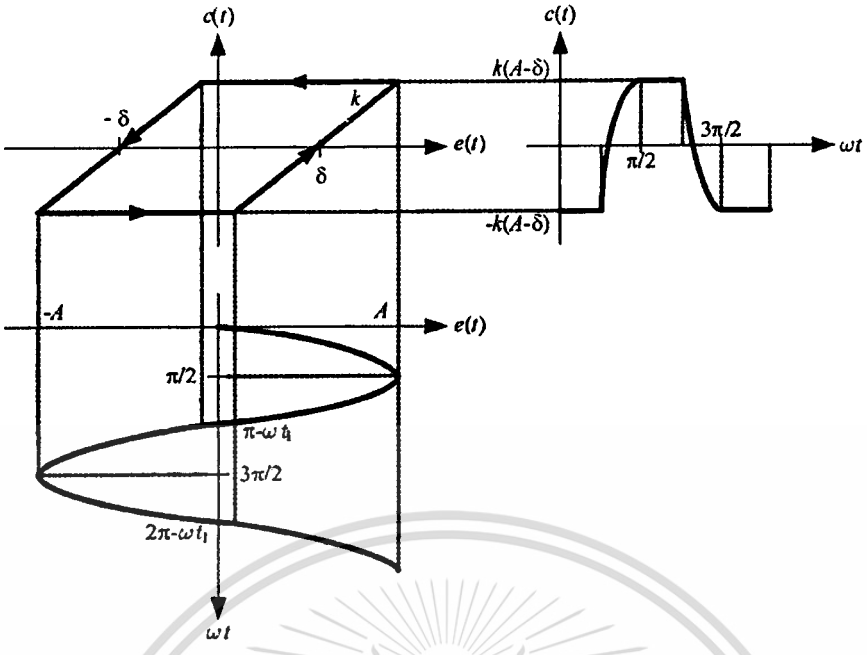


Figure 4.15 Input and general form of the output waveform for a backlash nonlinearity

4.5 Describing Function Analysis of Nonlinear Systems

The most important application of the describing function method in nonlinear system analysis is to predict the existence of limit cycles and determine their, stability. This application is discussed in this section, following a short review of the Nyquist criterion in classical control. Other applications of the describing function method, such as the determination of the closed-loop system response in the presence of external excitation, will be discussed in later sections.

4.5.1 The Nyquist Criterion and Its Extension

Since the describing function method is a generalization of Nyquist analysis, a short review and extension of the Nyquist criterion is helpful to a clear understanding of the describing function method.

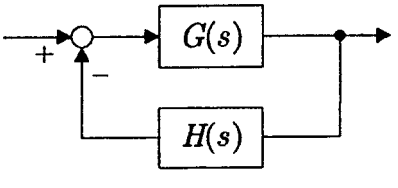


Figure 4.16 Closed-loop linear system

Consider the linear system of Figure 4.16. The characteristic equation of this system is

$$\delta(s) = 1 + G(s)H(s) = 0$$

Note that $\delta(s)$, often called the *loop transfer function*, is a rational function of s, with its zeros being the poles of the closed-loop system, and its poles being the poles of the open-loop transfer function $G(s)H(s)$. Rewrite the characteristic equation as

$$G(s)H(s) = -1$$

Based on this equation, the famous Nyquist criterion can be derived straightforwardly from the Cauchy theorem in complex analysis. The criterion can be summarized in the following procedure:

1. Draw, in the s plane, a so-called Nyquist path enclosing the right-half plane.
2. Map this path into another complex plane through $G(s)H(s)$.
3. Determine N, the number of clockwise encirclements of the plot of $G(s)H(s)$ around the point $(-1, 0)$.
4. Compute Z, the number of zeros of the loop transfer function $\delta(s)$ in the right-half s plane, by $Z = N + P$, where P is the number of unstable poles of $\delta(s)$.

Then the value of Z is the number of unstable poles of the closed-loop system. Figure 4.17 shows the example of Nyquist contour in s-plane and the $G(s)H(s)$ locus in GH-plane where $G(s)H(s) = 6/[(s+1)(s+2)]$

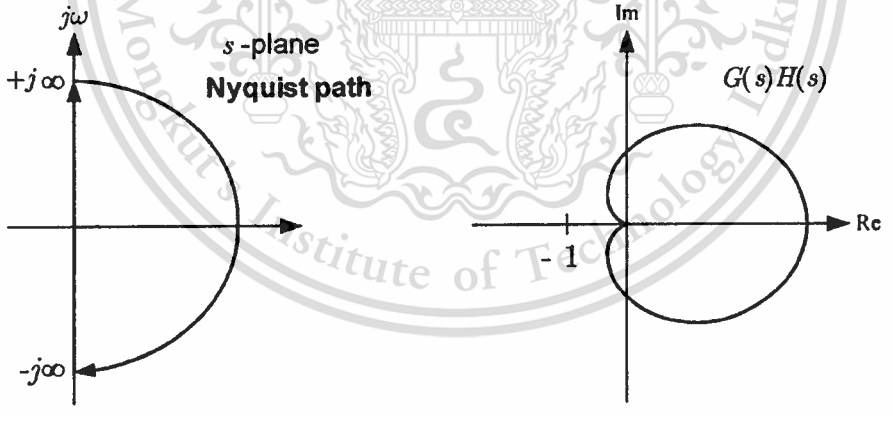


Figure 4.17 The Nyquist criterion

A simple formal extension of the Nyquist criterion can be made to the case when a constant gain K (possibly a complex number) is included in the forward path in Figure 4.18. This modification will be useful in interpreting the stability analysis of limit cycles using the describing function method. The loop transfer function becomes

$$\delta(s) = 1 + KG(s)H(s)$$

with the corresponding characteristic equation

$$G(s)H(s) = -1/K$$

The same arguments as used in the derivation of Nyquist criterion suggest the same procedure for determining unstable closed-loop poles, with the minor difference that now Z represents the number of clockwise encirclements of the $G(s)H(s)$ plot around the point $-1/K$. Figure 4.18 shows the corresponding extended Nyquist plot.

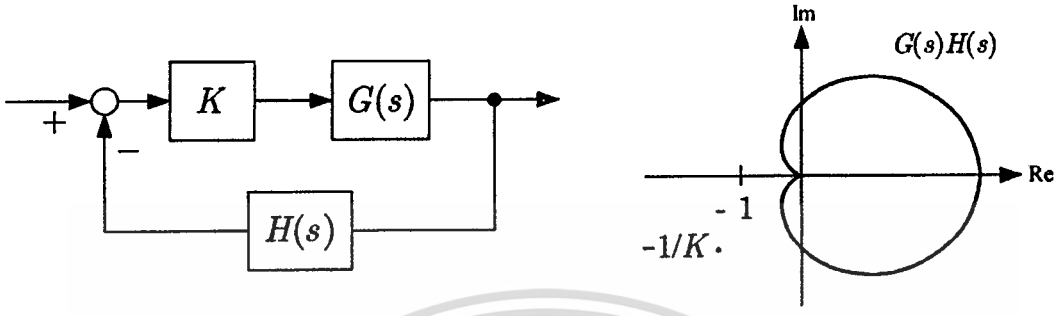


Figure 4.18 Extension of the Nyquist criterion

4.5.2 Existence of Limit Cycles

Assume that there exists a self-sustained oscillation of amplitude A and frequency ω in the system of Figure 4.19. Then the variables in the loop must satisfy the following relations

$$\begin{aligned} e &= -y \\ c &= D(A,\omega)e \\ y &= G(j\omega)c \end{aligned}$$

Therefore, $y = G(j\omega)D(A,\omega)(-y)$. Since $y \neq 0$, this implies

$$G(j\omega)D(A,\omega) + 1 = 0 \tag{4.20}$$

which can be written as

$$G(j\omega) = -\frac{1}{D(A,\omega)} \tag{4.21}$$

Therefore, the amplitude A and frequency ω of the limit cycles in the system must satisfy (4.21). If the above equation has no solutions, then the nonlinear system has no limit cycles.

Expression (4.21) represents two nonlinear equations (the real part and imaginary part) in the two variables A and ω. There are usually a finite number of

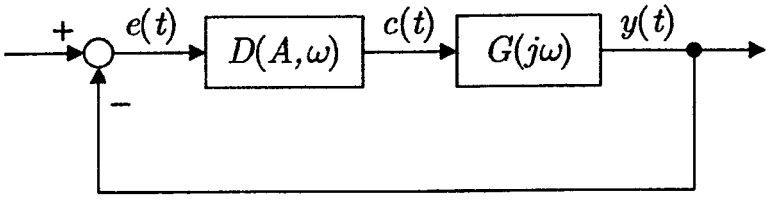


Figure 4.19 A nonlinear system

solutions. It is generally very difficult to solve these equations analytical methods, particularly for high-order systems. Therefore, a graphical approach is usually taken. The idea is to plot both sides of (4.21) in the complex plane and find the intersection points of the two curves.

Frequency-Independent Describing Function

First, consider the simpler case when the describing function D being a function of the gain A only, *i.e.*,$D(A,\omega) = D(A)$. This includes all single-valued nonlinearities and important double-valued nonlinearities such as backlash and. hysteresis. The equality becomes

$$G(j\omega) = -\frac{1}{D(A)} \tag{4.22}$$

Plot both the frequency response function $G(j\omega)$ (varying ω) and the negative inverse describing function $-1/D(A)$ (varying A) in the complex plane, as in Figure 4.20. If the two curves intersect, then there exist limit cycles, and the values of A and co corresponding to the intersection point are the solutions of Equation (4.22). If the curves intersect n times, then the system has n possible limit cycles. Which one is actually reached depends on the initial conditions. In Figure 4.20, the two curves intersect at one point P. This indicates that there is one limit cycle in the system. The amplitude of the limit cycle is A_p, the value of A corresponding to the point P on the $-1/D(A)$ curve. The frequency of the limit cycle is ω_p, the value of co corresponding to the point P on the $G(j\omega)$ curve.

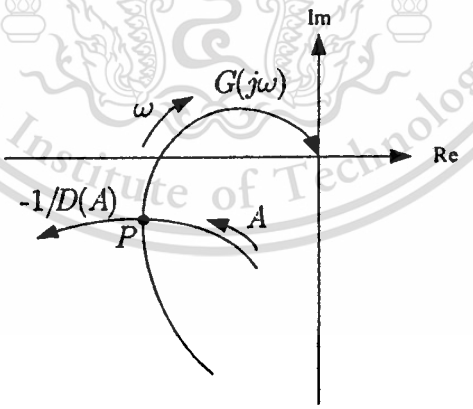


Figure 4.20 Detection of limit cycles

Note that for single-valued nonlinearities, D is real and therefore the plot of $-1/D$ always lies on the real axis. It is also useful to point out that the above procedure only gives a prediction of the existence of limit cycles. The validity and accuracy of this prediction should be confirmed by computer simulations.

Frequency-Dependent Describing Function

For the general case, where the describing function depends on both input amplitude and frequency ($D = D(A,\omega)$), the method can be applied, but with more complexity. Now the right-hand side of (4.21), $-1/D(A,\omega)$, corresponds to a family of curves on the complex plane with A as the running parameter and ω fixed for each curve, as shown on the left of Figure 4.21. There are generally an infinite number of intersection points between the $G(j\omega)$ curve and the $-1/D(A,\omega)$ curves. Only the intersection points with matched ω indicate limit cycles.

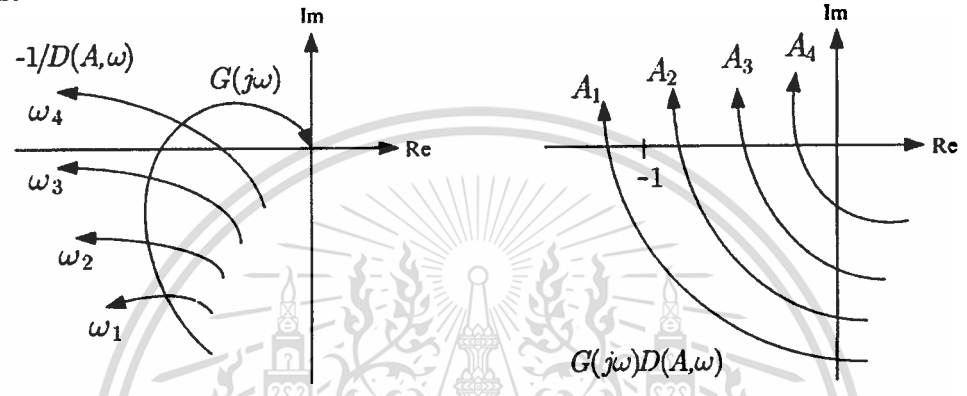


Figure 4.21 Limit cycle detection for frequency-dependent describing functions

To avoid the complexity of matching frequencies at inter ection points, it may be advantageous to consider the graphical solution of (4.21) directly, based on the plots of $G(j\omega)D(A,\omega)$. With A fixed .find ω varying from 0 to ∞, we obtain a curve representing $G(j\omega)D(A,\omega)$. Different values of A correspond to a family of curves, as shown on the right of Figure 4.21. A curve passing through the point $(-1,0)$ in the complex plane indicates the existence of a limit cycle, with the value of A for the curve being the amplitude of the limit cycle, and the value of ω at the point $(-1,0)$ being the frequency of the limit cycle. This technique is much more straightforward than the previous one, but it requires repetitive computation of the $G(j\omega)$ in generating the family of curves, which may be handled easily by computer.

The following example shows the limit cycle prediction for a system involving a hard nonlinearity.

Example 4.3.

Consider the system of Figure 4.22, which is typical of the dynamics of electronic oscillators used in laboratories.

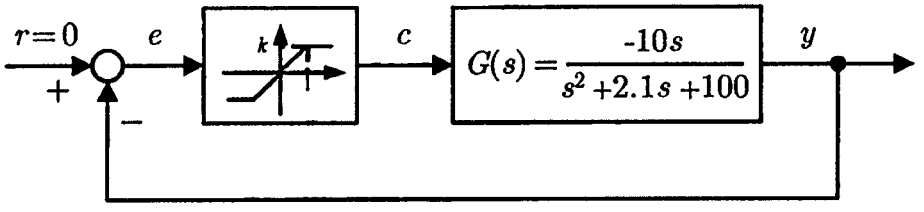


Figure 4.22 Oscillator dynamics

The locus of $G(j\omega)$ is plotted in Figure 4.23. Since

$$G(j10) = \frac{-j100}{-100 + j21 + 100} = -4.76$$

it intersects the real axis at the point $(-4.76, 0)$, and the frequency at the intersection point is $\omega = 10$.

The describing function of this saturation element is, according to (4.11),

$$D(A) = \frac{2k}{\pi}\left[\sin^{-1}\frac{1}{A} + \frac{1}{A}\sqrt{1 - \frac{1}{A^2}}\right]$$

The locus of $-1/D(A)$ is a line along the negative real axis. the starting point of this line is $(-1/k, 0)$ because $D(A) < k$, $\forall A > 0$. Therefore, the locus of $-1/D(A)$ will intersect with the locus of $G(j\omega)$ if $1/k < 4.76$, *i.e.*, $k > 0.21$. To be specific, set $k = 0.25$ and find the amplitude of the limit cycle. Since at the intersection point

$$-\frac{1}{D(A)} = G(j10) = -4.76$$

one has $D(A)/k = 0.84$. From Figure 4.11, the corresponding value of A is roughly 1.36, since $a = 1$. Therefore, the describing function analysis predicts a limit cycle with amplitude 1.36 and frequency 10. *i.e.*, $e(t) = 1.36\sin(10t)$.

One might be interested in knowing why such a limit cycle occurs for this system. To gain some intuitive understanding, assume that the system starts at a

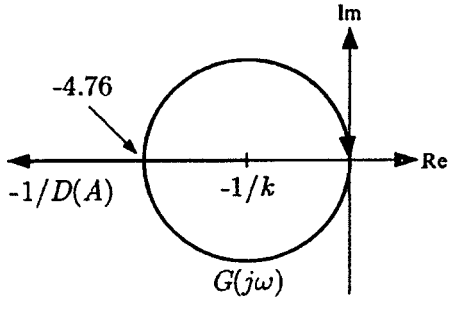


Figure 4.23 Graph of describing function

small initial state. Since the signals are below the saturation level, the closed-loop system dynamics are

$$\ddot{y} + (2.1 - 10k)\dot{y} + 100y = 0$$

If $k > 0.21$, the system has negative damping, *i.e.*, increases its energy. The small signal operation is thus unstable and the signals will diverge. However, the signals cannot diverge infinitely because of the saturation. Can the signal $c(t)$ stay at the saturation level? No, because the output of this linear element corresponding to a constant input decays to zero, as seen by the final-value theorem, or by noticing the differentiator s in the numerator of $G(s)$. Therefore, the system signals can stay neither at small values, because of instability, nor at saturation values, because of the above reasoning. Thus, it oscillates and a limit cycle results. The above describing function analysis confirms this intuitive argument.

4.5.3 Stability of Limit Cycles

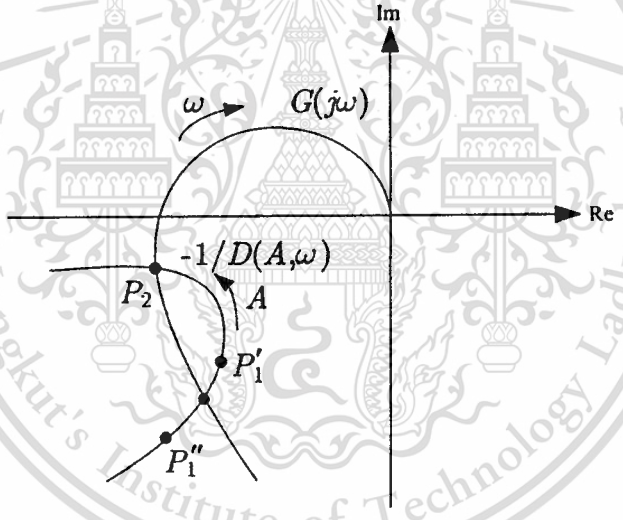


Figure 4.24 Limit cycle stability

Consider the plots of frequency response and inverse describing function in Figure 4.24. There are two intersection points, predicting that the system has two limit cycles. Note that the value of A corresponding to point P_1 is smaller than the value of A corresponding to P_2. For simplicity of the discussion, let the linear transfer function $G(j\omega)$ has no unstable poles.

Assume that the system initially operates at point P_1, with the limit cycle amplitude being A_1 and its frequency being ω_1. Due to a slight disturbance, the amplitude of the input to the nonlinear element is slightly increased, and the system operating point is moved from P_1 to P_1'. Since the new point P_1' is encircled by the curve of $G(j\omega)$, according to the extended Nyquist criterion, the system at this operating point is unstable, and the amplitudes of the system signals will increase. Therefore, the operating point will continue to move along the curve

$-1/D(A)$ toward the other limit cycle point P_2. On the other hand, if the system is disturbed so that the amplitude A is decreased, with the operating point moved to the point P_1'', then A will continue to decrease and the operating point moving away from P_1 in the other direction. This is because P_1'' is not encircled by the curve $G(j\omega)$ and thus the extended Nyquist plot confirms the stability of the system. This indicates that a slight disturbance can destroy the oscillation at point P_1 and, therefore, that this limit cycle is unstable. A similar analysis for the limit cycle at point P_2 indicates that the limit cycle is stable.

Summarizing the above discussion and the result in the previous subsection, there is a criterion for existence and stability of limit cycles:

Limit Cycle Criterion[1]

Each intersection point of the curve $G(j\omega)$ and the curve $-1/D(A)$ corresponds to a limit cycle. If points near the intersection and along the increasing-A side of the curve $-1/D(A)$ are not encircled by the curve $G(j\omega)$, then the corresponding limit cycle is stable. Otherwise, the limit cycle is unstable.

4.5.4 Reliability of Describing Function Analysis

Empirical evidence over the last three decades, and later theoretical justification, indicate that the describing function method can efficiently solve a large number of practical control problems involving limit cycles. However, due to the approximate nature of the technique, it is not surprising that the analysis results are sometimes not very accurate. Three kinds of inaccuracies are possible:

1. The amplitude and frequency of the predicted limit cycle are not accurate.
2. A predicted limit cycle does not actually exist.
3. An existing limit cycle is not predicted.

The first kind of inaccuracy is quite common. Generally, the predicted amplitude and frequency of a limit cycle always deviate somewhat from the true values. How much the predicted values differ from the true values depends on how well the nonlinear system satisfies the assumptions of the describing function method. In order to obtain accurate values of the predicted limit cycles, simulation of the nonlinear system is necessary.

The occurence of the other two kinds of inaccuracy is less frequent but has more serious consequences. Usually, their occurence can be detected by examining the linear element frequency response and the relative positions of the G plot and $-1/D(A)$ plot.

Violation of filtering hypothesis

The validity of the describing function method relies on the filtering hypothesis defined by (4.5). For some linear elements, this hypothesis is not satisfied and errors may result in the describing function analysis. Indeed, a number of failed cases of describing function analysis occur in systems whose linear element has resonant peaks in its frequency response $G(j\omega)$.

Graphical Conditions

If the $G(j\omega)$ locus is tangent or almost tangent to the $-1/D$ locus, then the conclusions from a describing function analysis can be erroneous. This is because effects of neglected higher harmonics or system model uncertainty may cause the change of the intersection situations, particularly when filtering in the linear element is weak. As a result, the second and third types of errors listed above may occur. A classic case of this problem involves a second-order servo with backlash studied by Nychols. While describing function analysis predicts' two limit cycles (a stable one at high frequency and an un table one at low frequency), it can be shown that the low-frequency unstable limit cycle does not exist.

Conversely, if the $-1/D$ locus intersects the G locus almost perpendicularly, then the results of the describing function are usually good.

CHAPTER 5

Design of Sinusoidal Nonlinear Oscillator

5.1 Approach and Methods

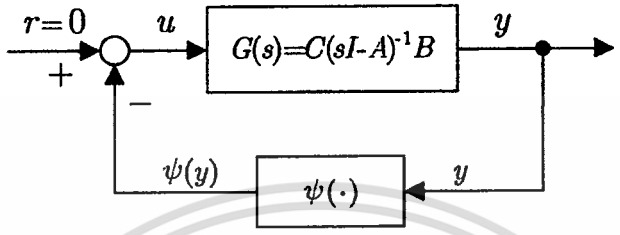


Figure 5.1 Feedback connection

We consider the system shown in Figure 5.1, with the linear system $G(s)$ having low-pass or band-pass filtering characteristics and $\psi(\cdot)$ as a time-invariant nonlinear element. The design needs to meet some basic requirements. In order to develop the basic version of the describing function method, the system has to satisfy the previous assumption.

5.1.1 Periodic Function

Consider a sinusoidal input to the nonlinear element, of amplitude a and frequency ω_0. The output of the nonlinear component $w(t) = \psi(y(t))$ in Figure 5.1 is always periodic: Corresponding to a sinusoidal input $y = a\sin\omega t$, the output becomes $\psi(a\sin(\omega_0(t + 2\pi/\omega_0))) = \psi(a\sin(\omega_0 t))$. Using Fourier series, this periodic function can be expanded as

$$w(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)\right]. \tag{5.1}$$

Since the set $\{\frac{1}{\sqrt{2}}, \cos(\omega_0 t), \sin(\omega_0 t), \cos(2\omega_0 t), \sin(2\omega_0 t), \ldots\}$ is a complete orthonormal basis with respect to the inner product $< f(t), g(t) >:= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)g(t)d(\omega_0 t)$. The coefficients can be found as $a_0 =< f(t), 1 >$, $a_n =< f(t), \cos(n\omega_0 t) >$, $b_n =< f(t), \sin(n\omega_0 t) >$ or

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} w(t)d(\omega_0 t) \tag{5.2}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} w(t)\cos(n\omega_0 t)d(\omega_0 t) \tag{5.3}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} w(t)\sin(n\omega_0 t)d(\omega_0 t) \tag{5.4}$$

Furthermore, since $G(s)$ used has band-pass or low-pass characteristics, this implies that only the fundamental component $w_1(t)$ is needed to be considered, namely

$$\begin{aligned} w(t) \approx w_1(t) &= a_1 \cos(\omega_0 t) + b_1 \sin(\omega_0 t) \\ &= M \sin(\omega_0 t + \phi), \end{aligned} \tag{5.5}$$

where $M(a, \omega_0) = \sqrt{a_1^2 + b_1^2}$ and $\phi(a, \omega_0) = \tan^{-1}(a_1/b_1)$, which indicate that the fundamental component corresponding to a sinusoidal input is a sinusoidal at the same frequency. In complex representation, this sinusoidal can be written as $w_1 = Me^{j(\omega_0 t + \phi)} = (b_1 + ja_1)e^{j\omega_0 t}$.

5.1.2 Using of Describing Function Method

The describing function of the nonlinear element is defined to be the complex ratio of the fundamental component of the output of the nonlinear element by the input sinusoid, i.e.,

$$D(a, \omega_0) = \frac{w_1}{ae^{j\omega_0 t}} = \frac{Me^{j(\omega_0 t + \phi)}}{ae^{j\omega_0 t}} = \frac{(b_1 + ja_1)e^{j\omega_0 t}}{ae^{j\omega_0 t}} = \frac{1}{a}(b_1 + ja_1).$$

Since $w(t)$ is an odd function, by assumption 4, (5.3) becomes

$$a_1 = \frac{1}{\pi} \int_{-\pi}^{\pi} w(t) \cos(\omega_0 t) d(\omega_0 t) = 0$$

Then

$$D(a, \omega_0) = \frac{b_1}{a} = \frac{1}{a\pi} \int_{-\pi}^{\pi} w_1(t) \sin(\omega_0 t) d(\omega_0 t)$$

For the case of single-valued nonlinearities, the describing function D is a function of the gain a only, $D(a, \omega_0) = D(a)$.

$$D(a) = \frac{1}{a\pi} \int_{-\pi}^{\pi} \psi(a \sin\theta) \sin\theta d(\theta); \qquad \theta = \omega_0 t$$

5.1.3 OTA Nonlinear Behavior

The simplified equivalent circuit diagram is given in Figure 5.2. According to [4] and [5], i_{OUT} of the OTA equivalent circuit in Figure 5.2 is

$$i_{OUT} = I \frac{\exp\left(\frac{qv}{kT}\right) - 1}{\exp\left(\frac{qv}{kT}\right) + 1} = I \tanh\left(\frac{v}{2V_T}\right) \tag{5.6}$$

where $v = v^+ - v^-$, and $V_T = \frac{kT}{q}$.

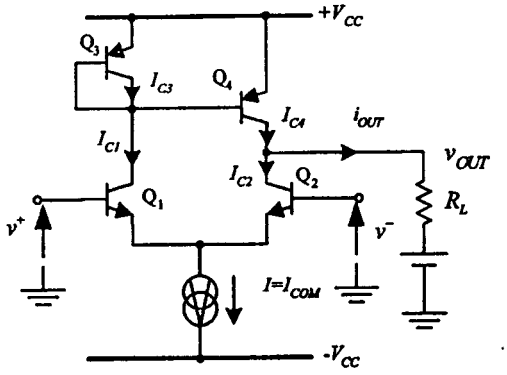


Figure 5.2 OTA equivalent circuit

5.2 Calculation of System Parameters

For the linear system in Figure 5.1 having band-pass filter characteristics of the form

$$G(s) = \frac{-\omega_0 s}{s^2 + 2\alpha s + \omega_0^2} \tag{5.7}$$

where ω_0 is the resonant frequency in rad/sec and 2α is the bandwidth in rad/sec of the band-pass filter. If a self-excited oscillation exists the frequency of oscillation will be $\omega = \omega_0$. Therefore the condition of oscillation (condition for existence of limit cycle) becomes

$$\frac{1}{Q} = \frac{2\alpha}{\omega_0} = \frac{2}{a\pi}\int_0^{\pi} \psi(a\sin\theta)\sin\theta\, d\theta \tag{5.8}$$

where Q is the quality factor of the band-pass filter. The far right-hand side of (5.8) is the describing function of the function $\psi(\cdot)$, and for a fixed $\psi(\cdot)$ it depends only on the magnitude of oscillation a, as follows:

$$D(a) := \frac{2}{a\pi}\int_0^{\pi} \psi(a\sin\theta)\sin\theta\, d\theta. \tag{5.9}$$

For known values of α and ω_0, (5.8) is an equation in only one unknown variable, a, and with the aid of modern computers, one can solve (5.8) numerically for a or plot $D(a)$ against a. In the next section, we will give an example showing how to design an OTA-based sinusoidal nonlinear oscillator of specified amplitude and frequency of oscillation.

CHAPTER 6

Simulation and Experimental Results

6.1 Introduction

This chapter describes a systematic approach to designing a sinusoidal nonlinear oscillator which uses an operational transconductance amplifier (OTA) as a nonlinear element to achieve the desired specifications on amplitude and frequency of oscillation. The describing function method is used for predicting the existence of limit cycles and, more generally, for analyzing the magnitude stabilization phenomena. The results are simulated using MAPLE and MATLAB SIMULINK.

6.2 System Design

As an illustration of the design procedure, let us consider the following example.

Example 6.1.

Given the specifications on the frequency of oscillation $\omega_0 = 120$ rad/s and on the amplitude of oscillation $a = 0.3$ volts, then follows the following steps in designing an OTA-based sinusoidal nonlinear oscillator of the feedback configuration in Figure 5.1.

First, consider the OTA nonlinear behavior in (5.6). With $I = 1mA$, we have

$$i_{OUT} = 0.001 \tanh(20v) \tag{6.1}$$

since $V_T \approx 25$ mV. Then we can realize $\psi(y) = \frac{1}{50}\tanh(20y)$ using an additional amplifier of gain 20 V/A at the output of the OTA.

From (5.9), plot $D(a)$ with respect to a, for $\psi(y) = \frac{1}{50}\tanh(20y)$, as shown below in Figure 6.1.

With $\omega_0 = 120$ rad/sec and $a = 0.3$ V, then $Q = \frac{1}{D(0.3)} = \frac{1}{0.08388617} = 11.92091526$, where $D(0.3) = 0.08388617$ was evaluated numerically from (5.9) by MAPLE program, so was the bandwidth $2\alpha = \omega_0/Q = 10.06634116$. Therefore the transfer function of our band-pass filter in (5.7) becomes

$$G(s) = \frac{-120s}{s^2 + 10.06634116s + 14400} \tag{6.2}$$

The condition of oscillation in (5.8) can be written as

$$Q = \frac{\omega_0}{2\alpha} = -G(j\omega_0) = \frac{1}{D(a)} \tag{6.3}$$

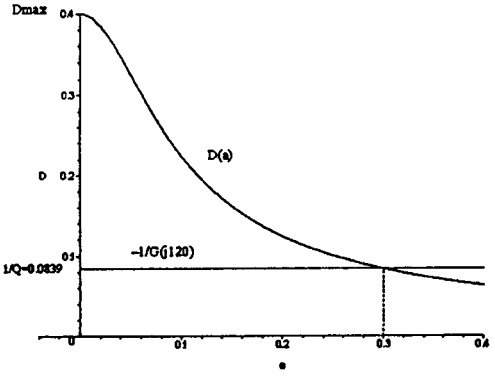


Figure 6.1 $D(a)$ for $\psi(y) = \frac{1}{50}\tanh(20y)$.

and in order for the limit cycle to exist it is required that

$$0 < \frac{2\alpha}{\omega_o} = \frac{1}{Q} < D_{max} = D(0), \tag{6.4}$$

and in the case here, (6.4) is satisfied with Q high enough for $G(s)$ to be very selective.

Note: Condition (6.4) confirms an existence of a limit cycle which can also be seen as in the graph of $D(a)$ in Figure 6.1. If the desired value of amplitude of oscillation a yields a too low value of Q in (6.3) or if D_{max} is not greater than $1/Q$ in (6.4) then either an OTA circuit have to be redesigned or an amplifier may be required at the input or/and output of $\psi(\cdot)$.

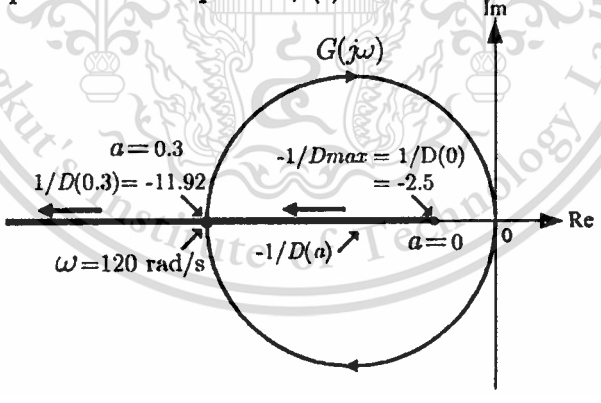


Figure 6.2 Plots of $G(j\omega)$ and $-1/D(a)$

The plot of $G(j\omega)$ and $\frac{1}{D(a)}$ of the obtained sinusoidal nonlinear oscillator on complex plane is shown in Figure 6.2. Both graphs intersect at $\omega = 120$ rad/sec and $a = 0.3$ volts, at this point $G(j120) \approx -11.9209$. The corresponding limit cycle is stable since points near the intersection and along the increasing-a side of the curve $-1/D(a)$ are not encircled by the curve $G(j\omega)$. Simulation using MATLAB SIMULINK was done as follows:

6.3 MATLAB SIMULINK Simulation

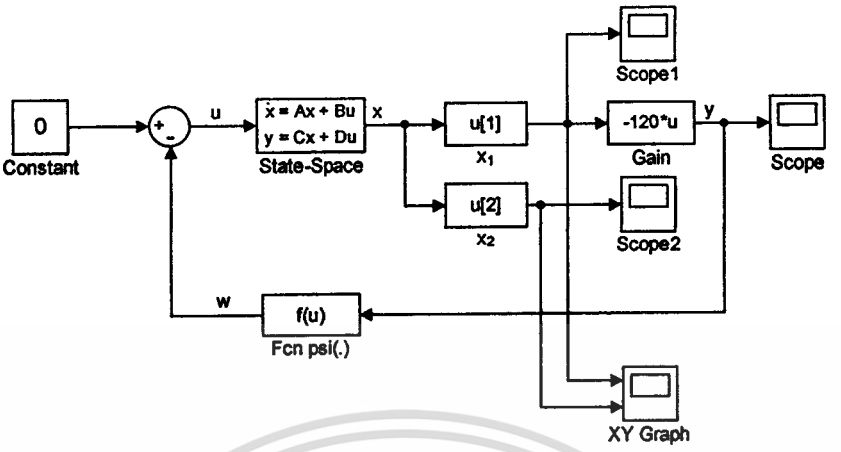


Figure 6.3 MATLAB SIMULINK block diagram

State and output equation of the transfer function $G(s)$ in (6.2) are as follows:

$$\dot{x} = \begin{bmatrix} -10.07 & 14400 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u \tag{6.5a}$$

$$y = \begin{bmatrix} -120 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{6.5b}$$

which can be used to create a MATLAB SIMULINK block diagram as shown in Figure 6.3. The phase plane plot of $x(t)$ is shown in Figure 6.4.

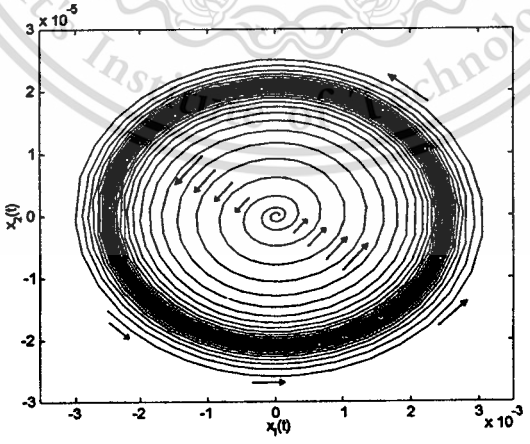


Figure 6.4 Phase plane plot of $x = [x_1, x_2]^T$.

The limit cycle in Figure 6.4 shows that $x_1(t)$ has amplitude of 0.0025 volts, therefore after multiplied by the gain of -120, the amplitude of $y(t)$ becomes 0.3 volts as expected.

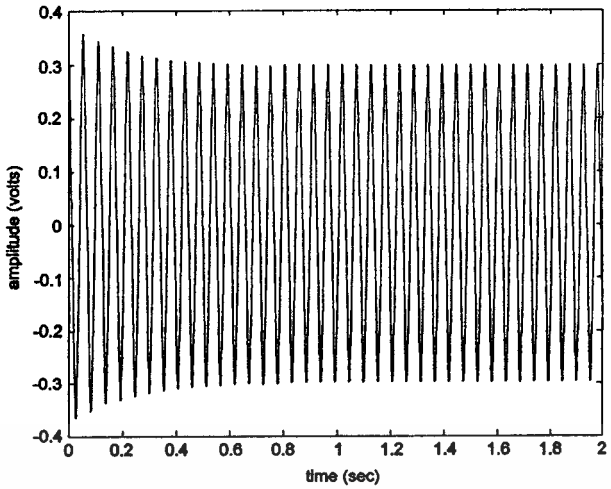


Figure 6.5 Plot of $y(t)$ with $y(0) = 0.25$.

The graphs of the output $y(t)$ for two different initial conditions are also plotted against time t. Figure 6.5 shows the graph of $y(t)$ with initial condition $y(0) = 0.25$, and Figure 6.6 shows the graph of $y(t)$ with initial condition $y(0) = -0.005$. Both graphs show the same frequency of oscillation $\omega_0 = 120$ rad/sec (about 23 cycles in 1.2 seconds) and also show the magnitude stabilization phenomenon of output amplitude of oscillation $a = 0.3$ volts. This feature make a sinusoidal nonlinear oscillator superior to the linear ones.

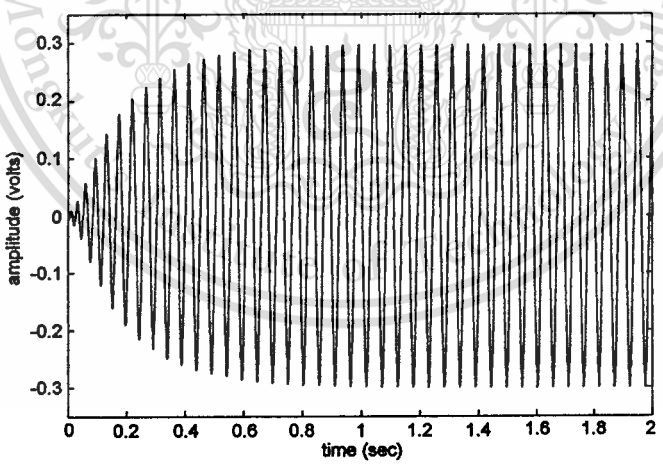


Figure 6.6 Plot of $y(t)$ with $y(0) = -0.005$.

6.4 Experimental Results

6.4.1 OTA Characteristic Test

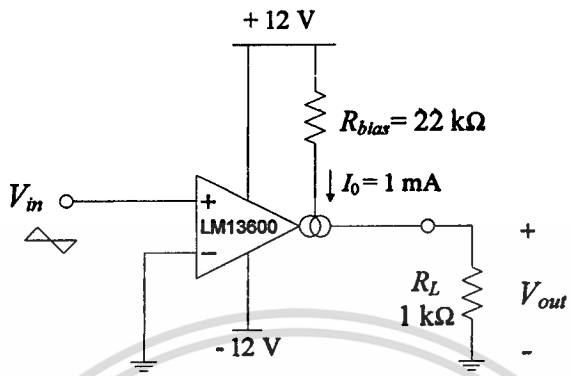


Figure 6.7 OTA test circuit

The OTA IC LM13600 is used as a nonlinear element for a nonlinear oscillator system in Figure 5.1. First an OTA characteristic, as in (5.6), is tested by applying a triangular signal[1] into a test circuit in Figure 6.7.

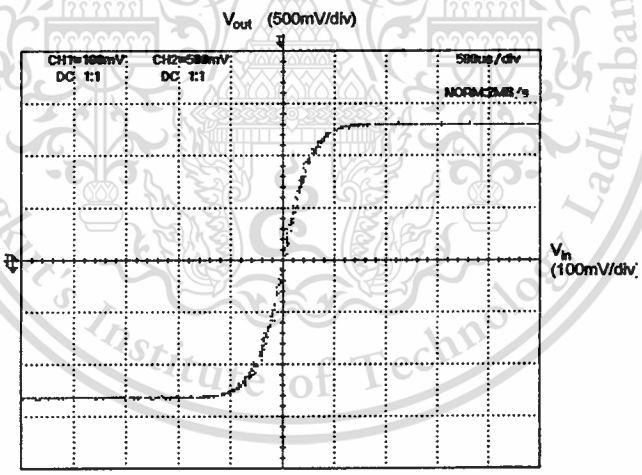


Figure 6.8 OTA characteristic

The input/output relationship is shown as an xy-graph in Figure 6.8 where x-axis and y-axis represents V_{in} and V_{out} respectively. Sample data from a graph of Figure 6.8 is shown as following.

[1]According to a data sheet, input voltage is ± 5 V maximum and input bias current I_0 is 2 mA maximum.

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} V_{in} \\ V_{out} \end{bmatrix} = \begin{bmatrix} -0.5 & -0.4 & -0.3 & -0.2 & -0.1 & -0.08 & -0.06 \\ -1.3 & -1.3 & -1.3 & -1.3 & -1.25 & -1.18 & -1.08 \\ -0.04 & -0.02 & 0 & 0.02 & 0.04 & 0.06 & 0.08 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ -0.85 & -0.55 & 0 & 0.55 & 0.85 & 1.08 & 1.18 & 1.25 & 1.3 & 1.3 & 1.3 & 1.3 \end{bmatrix}$$

Make data fitting by using a least squares method.

$$\min_x \sum_{i=1}^{m} (y_i - f(t_i, x))^2 \tag{6.6}$$

where

$$\begin{aligned} f(t, a, b) &= a \tanh(bt); \qquad a = 1.3 \\ f(t, b) &= 1.3 \tanh(bt) \end{aligned} \tag{6.7}$$

With the data from metrix above, summation in (6.6) can be expanded as

$$\begin{aligned} S &:= (y_1 - f(t_1, b))^2 + (y_2 - f(t_2, b))^2 + \ldots + (y_{19} - f(t_{19}, b))^2 \\ &= (-1.3 - f(-0.5b))^2 + (-1.3 - f(-0.4b))^2 + \ldots + (1.3 - f(0.5b))^2 \end{aligned} \tag{6.8}$$

then differentiate (6.8) to find a minimum point (solution by MAPLE).

$$\begin{aligned} \frac{dS(b)}{db} &= 0 \quad \Rightarrow \quad b = 20.21582696 \\ \frac{d^2S(b)}{db^2} &= 0.008308718086 \end{aligned}$$

Note that the second derivative test confirms that $b = 20.21582696$ is the minimum point.

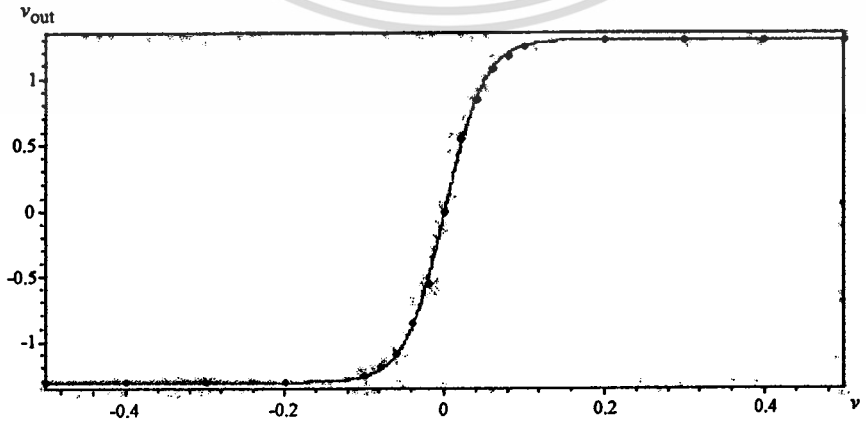


Figure 6.9 Data fitting using a least squares method

By fitting the data to a graph of tanh function[2], and from the derivation above, we obtain a practical OTA characteristic as

$$v_{out} = 1.3\tanh(20v) \tag{6.9}$$

According to $R_L = 1k\Omega$, it cancelled the bias current term 1 mA in (5.8). Note that the load resistance R_L is used to convert the output current from OTA into a voltage that will be used to feedback the system.

6.4.2 High-Q Low-Bandwidth Op-Amp Band-Pass Filter

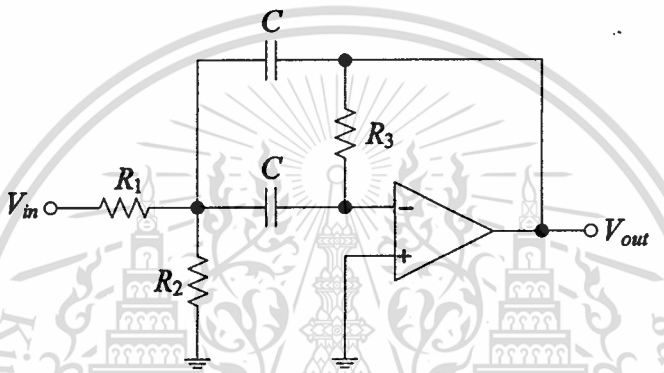


Figure 6.10 Band-pass filter circuit

The best choice for $G(s)$ block in Figure 5.1 is an op-amp based high-Q low-bandwidth band-pass filter, as shown in Figure 6.10, which has the following characteristics

$$\frac{V_{out}}{V_{in}} = \frac{-(H_0\beta)s}{s^2 + \beta s + \omega_0^2} \tag{6.10}$$

where

$$\omega_0 = 2\pi f_0 = \frac{1}{C\sqrt{\frac{R_1R_2R_3}{R_1+R_2}}} \tag{6.11}$$

$$H_0 = \frac{R_3}{2R_1} \tag{6.12}$$

$$\beta = 2\alpha = \frac{2}{CR_3} \tag{6.13}$$

[2]Plot a graph of tanh function using MAPLE with the desired parameters and compare with the experimental result (graph captured from an oscilloscope)

As the requirement in (6.2) where $\omega_0 = 120$ rad/s, $\beta = 2\alpha = 10.07$, using[3] $C = 4.7\mu F$ we get the following results.

$$\begin{aligned} H_0 &= \frac{120}{10.07} = 11.92 \\ R_3 &= \frac{2}{C\beta} = \frac{2}{4.7 \times 10.07 \times 10^{-6}} = 42.2k\Omega \\ R_1 &= \frac{R_3}{2H_0} = \frac{42.2 \times 10^3}{2 \times 11.92} = 1.77k\Omega \end{aligned}$$

For R_2,

$$\begin{aligned} \frac{R_1 R_2 R_3}{R_1 + R_2} &= \frac{1}{C^2\omega_0^2} \\ \frac{(1.77 \times 10^3) \times (42.2 \times 10^3) \times R_2}{(1.77 \times 10^3) + R_2} &= \frac{1}{(4.7 \times 10^{-6})^2 \times 120^2} \\ 23.76R_2 &= 1770 + R_2 \\ 22.76R_2 &= 1770 \\ R_2 &= 77.7\Omega \end{aligned}$$

Band-Pass Filter Test

Applying a sinusoidal signal into the circuit of Figure 6.10 with input amplitude of $1V$ and vary input frequency between 0 and 100 Hz, the frequency response center frequency is

$$f_0 = \frac{\omega_0}{2\pi} = \frac{120}{2 \times 3.14} \approx 19Hz$$

and from (6.2) at $\omega_0 = 120$ rad/s, a filter gain becomes

$$\begin{aligned} G(j120) &= \frac{120(j120)}{(j120)^2 + 10.07(j120) + 120^2} \\ &= \frac{14400j}{1208.4j} \\ &\approx 12 \end{aligned}$$

[3]Capacitance can be varied but the best choice is one that make all resistances not too high or too low.

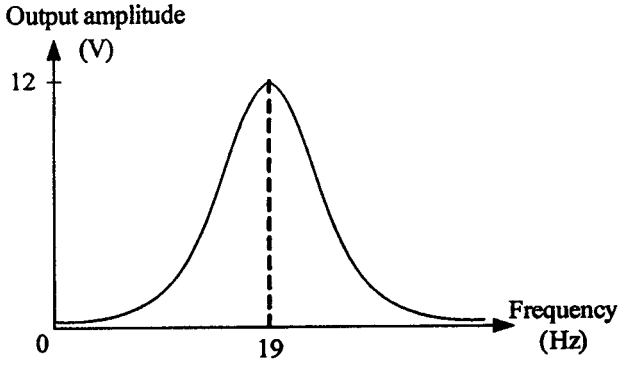


Figure 6.11 Band-pass filter characteristic

6.4.3 Proposed Oscillator

The oscillator circuit is the feedback system of the band-pass filter block $G(s)$ (Figure 6.10) and OTA network (Figure 6.8). A buffer is inserted between both block in the feed-forward path to prevent a loading effect and an inverting amplifier is used to create a negative feedback. A complete circuit is shown in Figure 6.12

Since the desired nonlinearity for this design is $\psi(y) = \frac{1}{50}\tanh(20y)$, we use $R_L = 15.38\Omega$ instead of $R_L = 1k\Omega$ as in the test circuit previously.

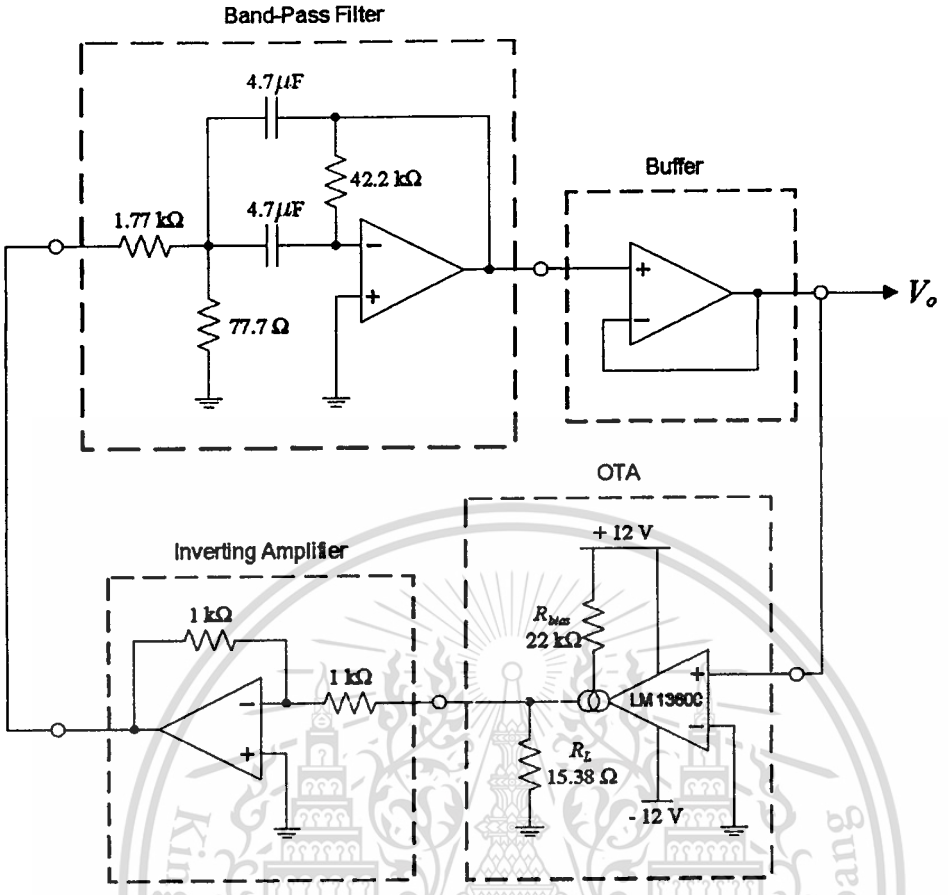


Figure 6.12 Oscillator circuit

6.4.4 Output Signal

The oscillation starts when applying a supply voltage to the circuit of Figure 6.12, without any external excitaion. The oscillator generates a sinusoidal output, the oscillation is recorded and has the wave form as shown in Figure 6.13.

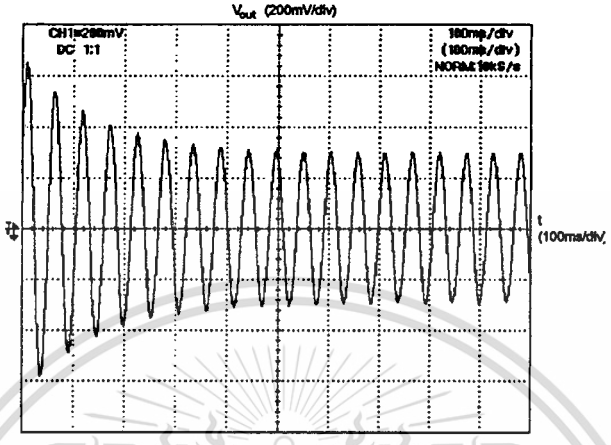


Figure 6.13 Output signal from oscillator

The output amplitude is 0.3 V as expected and the frequency of oscillation read from the waveform is $f_0 = 18.52$ Hz ($\omega_0 \approx 116$ rad/s). The error occurred is due to the actual values of resistance used are different from the designed values. This problem can be solved using adjustable resistance in a band-pass filter block to adjust the resistance ratio in (6.11).

Consider the characteristic equation (6.5). To observe the system's limit cycle, plot the xy-graph of $x_1(t) = -\frac{1}{120}y(t)$ against $x_2(t) = \int x_1(t)dt$. Note that the state x_2 can be simply measured by applying an integrator to the output y. The result is shown in Figure 6.14.

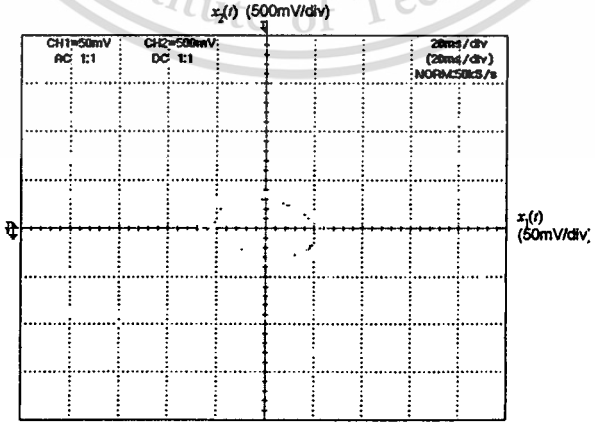


Figure 6.14 Detection of the oscillator's limit cycle

Using the same method with new desired parameters ($a = 0.3$ V, $f_0 = 100$ Hz) and ($a = 0.5$ V, $\omega_0 = 100$ Hz), results are shown in Figure 6.15 and 6.16.

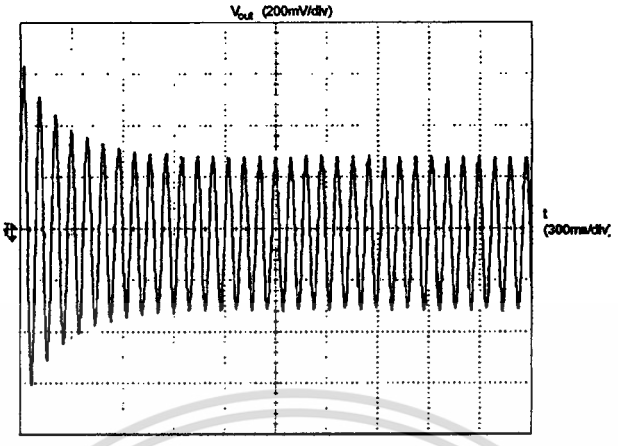


Figure 6.15 Output signal at $a = 0.3$ V, $f_0 = 100$ Hz

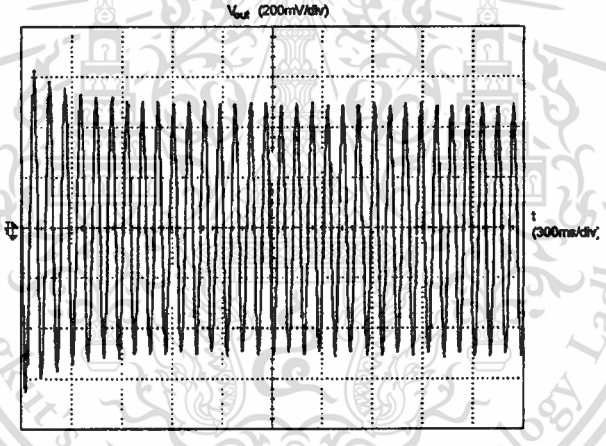


Figure 6.16 Output signal at $a = 0.5$ V, $f_0 = 100$ Hz

CHAPTER 7
Conclusions

Experimental results from the previous chapter show that nonlinear oscillators designed by this method oscillate at the desired amplitude and frequency of oscillation. The result from a real circuit is similar to a simulation result from MATLAB and satisfy the design specifications on amplitude and frequency of oscillation. Once the supply voltage is applied, the oscillator generates a sinusoidal signal by itself without any external excitation. The oscillation amplitude is stable and independent of initial conditions.

The describing function works well on this kind of oscillator as the system's nonlinearity satisfy the assumptions and it can describe a magnitude stabilization phenomena. A graphical method is also easier to understand and solve.

Further Improvements and Studies

The benefits of this design method is a clear description in nonlinear behaviors (using a describing function) and its magnitude stabilization phenomena. Another applications can be something that display or work at a fixed amplitude, *i.e.*, temperature, voltage, pressure, etc. The oscillator itself can also be applied as a signal function generator.

REFERENCES

[1] Slotine J.J.E., Li W. **Applied Nonlinear Control.** New Jersey : Prentice-Hall, Inc. 1991.

[2] Mancini R., Palmer R. **"Sine-Wave Oscillator."** [Online]. Available : http://www.calvin.edu/~pribeiro/courses/engr332/Handouts/oscillators.pdf. 2001.

[3] Franco S. **Design with Operational Amplifiers and Analog Integrated Circuits.** 3rd Edition. New York : McGraw-Hill. 2002.

[4] Bayard J., Ayachi M. "OTA- or CFOA-Based LC Sinusoidal Oscillators Analysis of the Magnitude Stabilization Phenomenon." **IEEE Transactions on Circuits and Systems.,** vol. 49, no. 8, Aug. 2002. pp. 1231-1236

[5] Pranayanuntana P., Anuntahirunrat K., Fongsamut C. and Kaewsaiha P. "The Describing Function Method and the Analysis of the Magnitude Stabilization Phenomenon in a Nonlinear OSC." **The International Symposium on Communications and Information Technologies 2005 (ISCIT 2005), Beijing, China,** October, 2005. pp. 420-423

[6] Khalil H.K. **Nonlinear Systems.** New Jersey : Prentice-Hall, Inc. 2002.

[7] Sedra A.S.,. Smith K.C. **Microelectronic Circuits.** New York : Oxford University Press, Inc. 1998.

[8] Williams J. **Analog Circuit Design : Art, Science and Personalities.** Boston : Newnes, Inc. 1998.

[9] Gratz, A. **"Operational Transconductance Amplifiers."** [Online]. Available : http://Stromeko.Synth.net/diy/OTA.pdf. 2006.

[10] National Semiconductor. **"LM13600 Dual Operational Transconductance Amplifiers with Linearizing Diodes and Buffers."** [Online]. Available : http://www.datasheetcatalog.com. 1995.

Appendix A

Operational Transconductance Amplifiers

A.1 Preface

The OTA is popular for implementing voltage controlled oscillators (VCO) and filters (VCF) for analog music synthesizers, because it can act as a two-quadrant multiplier as well see later. For this application the control input has to have a wide dynamic range of at least 60 dB, while the OTA should behave sensibly when overdriven from the signal input (in particular, it should not lock up or phase reverse). Viewed from a slightly different angle an OTA can be used to implement an electrically tunable resistor that is referenced to ground, with extra circuitry floating resistors are possible as well.

The primary application for an OTA is however to drive low-impedance sinks such as coaxial cable with low distortion at high bandwith. Hence, "improved" OTA such as the MAX436 or OPA660 have optimized these characteristics, but made it either impossible (MAX436) or considerably harder (OPA660) to use them as two-quadrant multipliers. Four quadrant multipliers on the other hand are hideously expensive, so that "obsolete" OTA like the CA3080 are still in widespread use.

A.2 Principle of Operation

An OTA is a voltage controlled current source, more specifically the term "operational" comes from the fact that it takes the difference of two voltages as the input for the current conversion. The ideal transfer characteristic is therefore

$$I_{out} = g_m(V_+ - V_-) \tag{A.1}$$

or, by taking the pre-computed difference as the input,

$$I_{out} = g_m V_{in} \tag{A.2}$$

with the ideally constant transconductance g_m as the proportionality factor between the two. In reality the transconductance [1] is also a function of the input differential voltage and dependent on temperature, as we will later see.

To summarize, an ideal OTA has two voltage inputs with infinite impedance (i.e. there is no input current). The common mode input range is also infinite,

[1] The term "transconductance" comes about because the ratio of the output current over the input voltage, g_m, has the unit of a conductance if looked at "across the amplifier". The proportional factor of output vs input for an amplifier with current input and voltage output has the unit of a resistance and such an amplifier is called a transresistance amplifier.

while the differential signal between these two inputs is used to control an ideal current source (i.e. the output current does not depend on the output voltage) that functions as an output. The proportionality factor between output current and input differential voltage is called transconductance.

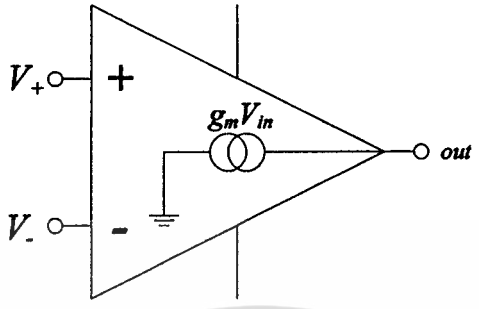


Figure A.1 An ideal OTA

Any real OTA will thus have circuitry to process the input voltages with low input current over a wide common mode input range, to produce an internal representation of the input differential voltage and to provide a current to the output that is relatively independent of the output voltage. Since an OTA can be used without feedback, the maximum output current and with it the transconductance can often be adjusted.

A.2.1 The Bipolar OTA

The most simple bipolar OTA consists of a differential pair to convert the input voltage difference to two currents I_+ and I_-. These two currents are then mirrored to the output so that their difference becomes the output of the OTA, while the rest of the OTA is made up of bias circuitry. The truly great feature of this "long-tailed differential pair" as it is often called is that the tail current, which is a necessary part of the biasing, can be used to control the transconductance.

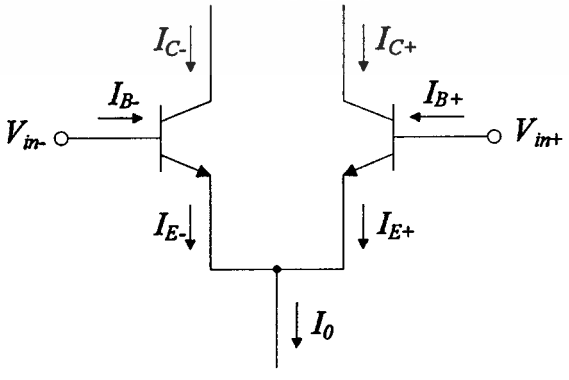


Figure A.2 Bipolar differential pair (with npn transistors, biasing not shown)

The Bipolar Differential Pair

The collector current of an npn transistor is (with some simplifying assumptions) related to its base-emitter voltage V_{BE} by

$$I_C = I_S \exp \frac{V_{BE}}{V_T}, \tag{A.3}$$

with the temperature voltage (k is the Boltzmann constant and q the elementary charge)

$$V_T = \frac{kT}{q} \tag{A.4}$$

The collector current can also be expressed as a multiple of the base current, viewing the transistor as a current amplifier with a gain of β,

$$I_C = \beta I_B \tag{A.5}$$

which makes the emitter current[2]

$$I_E = -(I_C + I_B) = -(\beta + 1) I_B \tag{A.6}$$

The tail current I_0 of the differential pair is composed of the emitter currents of the individual transistors.

$$I_0 = I_{E+} + I_{E-} \tag{A.7}$$

$$I_0 = \frac{\beta_+ + 1}{\beta_+} I_+ + \frac{\beta_- + 1}{\beta_-} I_- \tag{A.8}$$

and finally with $\beta_+ \gg 1$ and $\beta_- \gg 1$, this simplifies to

$$I_0 = I_+ + I_- \tag{A.9}$$

This simply means that as long as β is sufficiently high, its exact value is not at all important. Note however, that the β of a bipolar transistor is dependent on the collector current and therefore exact matching of β_+ and $\beta-$ can only occur at zero differential input voltage. Furthermore at low tail currents the error made in the simplification from (A.8) to (A.9) becomes quite noticeable as β drops off from its maximum value. Nevertheless for now well stick to the simplified equations and proceed to combine (A.3) and (A.9) to

$$I_0 = I_{S+} \exp \frac{V_{BE+}}{V_T} + I_{S-} \exp \frac{V_{BE-}}{V_T} \tag{A.10}$$

When the transistors are matched and at the same temperature this results in

$$I_0 = I_S \left(\exp \frac{V_{BE+}}{V_{T+}} + \exp \frac{V_{BE-}}{V_{T-}} \right) \tag{A.11}$$

[2]Traditionally all currents for a single transistor are directed towards the crystal, hence the minus sign. The positive counting current direction in a circuit is often different for various reasons.

which can be solved for I_S to

$$I_S = \frac{I_0}{\exp\frac{V_{BE+}}{V_T} + \exp\frac{V_{BE-}}{V_T}} \tag{A.12}$$

The output current of the OTA is the difference of the two collector currents in the pair

$$I_{out} = I_+ - I_- \tag{A.13}$$

and using (A.3) and (A.12) this gives the rather unwieldy expression

$$I_0 = I_S \left(\frac{\exp\frac{V_{BE+}}{V_T}}{\exp\frac{V_{BE+}}{V_T} + \exp\frac{V_{BE-}}{V_T}} - \frac{\exp\frac{V_{BE-}}{V_T}}{\exp\frac{V_{BE+}}{V_T} + \exp\frac{V_{BE-}}{V_T}} \right) \tag{A.14}$$

which can be simplified to

$$I_{out} = I_0 \left(\frac{1}{1 + \exp\frac{V_{BE-} - V_{BE+}}{V_T}} - \frac{1}{1 + \exp\frac{V_{BE+} - V_{BE-}}{V_T}} \right) \tag{A.15}$$

and further with $V_{in} = V_{BE+} - V_{BE-}$ to

$$I_{out} = I_0 \left(\frac{1}{1 + \exp\frac{-V_{in}}{V_T}} - \frac{1}{1 + \exp\frac{-V_{in}}{V_T}} \right) \tag{A.16}$$

Notice that the dependence on I_S is gone, due to the matching of both transistors and keeping them at the same temperature, but were still not having an explicit and compact dependence on the input voltage. This is exactly what well develop next and we start by extending to the common denominator

$$I_{out} = I_0 \left[\frac{(1 + \exp V_{in} V_T) - (1 + \exp -V_{in} V_T)}{(1 + \exp V_{in} V_T)(1 + \exp -V_{in} V_T)} \right] \tag{A.17}$$

which reduces to

$$I_{out} = I_0 \frac{\exp\frac{V_{in}}{V_T} - \exp\frac{-V_{in}}{V_T}}{2 + \exp\frac{V_{in}}{V_T} + \exp\frac{-V_{in}}{V_T}} \tag{A.18}$$

or

$$I_{out} = I_0 \frac{2\sinh\frac{V_{in}}{V_T}}{2 + 2\cosh\frac{V_{in}}{V_T}} \tag{A.19}$$

which correspond to[3]

$$I_{out} = I_0 \tanh\frac{V_{in}}{2V_T} \tag{A.20}$$

This puts us into a much better position to find out what g_m really is. The differential definition of the transconductance is:

$$g_m = \frac{dI_{out}}{dV_{in}} \tag{A.21}$$

[3]Another proofs of this equation are derived in the next appendix.

and with (A.20),

$$g_m = \frac{I_0}{2V_T} \operatorname{sech}^2 \frac{N_{in}}{2V_T} \tag{A.22}$$

Thus it shows that the transconductance is anything but constant, depending both on temperature and input voltage as has been stated earlier. The second term is a bell shaped curve that equals 1 at zero input, falling off rapidly at both sides to asymptotically approach zero. The practical input range depends on how much error[4] one is willing to tolerate, but seldom exceeds 20 mV. In fact, using (A.22), to keep the linearity error below one percent (or -40 dB below the signal) the input range is limited to $\pm 0.2V_T$ or 5 mV at room temperature. The maximum input range is approximately $\pm 5V_T$, 125 mV at room temperature or equivalently 28 dB of overdrive beyond the linear input range. Beyond this more than 99% of the tail current flows through just one of the two transistors and no changes in the output can be effected. The limiting action is comparably smooth, so overdriving an OTA from the input can be musically quite useful. The temperature voltage in the argument of that term conspires to make the bell shape wider at higher temperature, which means that the linear input range of the OTA is smaller at low temperature as the g_m drops off more rapidly from its maximum value. Often well find just the first part of the above expression as the transconductance, accompanied by some mumbling about small input voltages. The transconductance is however strictly proportional to the tail current, which provides the function of a two-quadrant multiplier. This is typically used to set and modulate the transconductance, which is useful for instance for building VCO and VCF in analog synthesizers. Making the tail current proportional to absolute temperature (which can be done using a $\triangle V_{BE}$-Arrangement) gets rid of the the temperature dependence in the first part of the expression. Of course this just makes the transconductance for zero input a constant and thus does not compensate the temperature dependence for any useful circuit.

Input Diode Linearization

Making a better OTA involves flattening the transconductance characteristic to achieve a wider input range and of course removing the temperature dependence. Flattening the transconductance curve generally reduces the peak transconductance for any given circuit, however. Both objectives can be achieved by connecting a "differential pair" of diodes to the inputs, fed by another current source. In short, the diodes in connection with a resistive input network will provide a compression of the input voltages to the differential transistor pair which expands them into a current, while through their matching to the input transistors

[4]Of concern would typically be the absolute error in the instantaneous output current for CV processing (after I-V-conversion) and total harmonic distortion (THD) for audio signal processing.

the temperature dependence of the inputs is also canceled.[5]

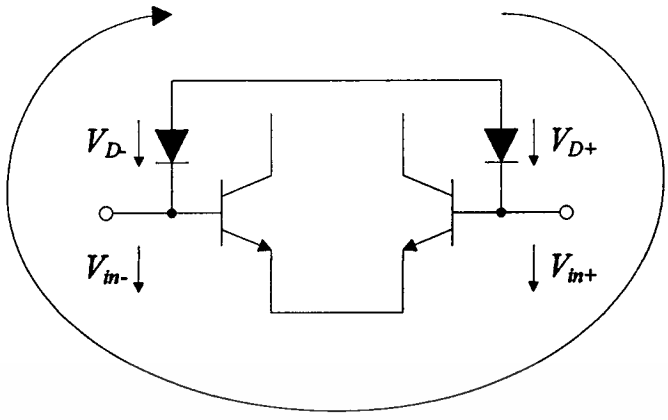


Figure A.3 Principle of input diode linearization

Look at the loop made up of the linearizing diodes[6] and the base-emitter diodes of the differential pair. Now for the voltages in that loop[7]

$$V_{D+} + V_{I+} = V_{D-} + V_{I-} \tag{A.23}$$

holds with some reordering and expressing it in terms of the currents this becomes

$$V_T \ln \frac{I_+}{I_{S+}} - V_T \ln \frac{I_-}{I_{S-}} = V_T \ln \frac{I_{D-}}{I_{S,D-}} - V_T \ln \frac{I_{D+}}{I_{S,D+}} \tag{A.24}$$

When all elements are matched, the saturation currents are identical and with some further simplification we get

$$\ln \frac{I_+}{I_-} = \ln \frac{I_{D-}}{I_{D+}} \tag{A.25}$$

which means that the current ratios must also be equal:

$$\frac{I_+}{I_-} = \frac{I_{D-}}{I_{D+}} \tag{A.26}$$

With (A.9), (A.13) and

$$I_D = I_{D+} + I_{D-} \tag{A.27}$$

$$I_{in} = I_{D-} - I_{D+} \tag{A.28}$$

[5]This is an example of a translinear circuit, whose principle is that the input-output relations are linear even though potentially none of their internal nodes bear any linear relationship with the inputs or outputs.

[6]In an integrated circuit these diodes generally will be transistors with the base and collector shorted. Diode connected transistors have a diode characteristic that is close to ideal over a wider current range and provide better matching than simple diodes.

[7]The linearizing diodes can also be put in parallel to the base-emitter diodes (like it is done in the CA3280). The operating principle is not changed by that modification all equations from (A.25) on are indeed identical, but the biasing requirements are different.

(which again assumes $\beta \gg 1$) we can rewrite the currents

$$I_+ = \frac{1}{2}(I_0 + I_{out}) \quad (A.29)$$

$$I_- = \frac{1}{2}(I_0 - I_{out}) \quad (A.30)$$

$$I_{D+} = \frac{1}{2}(I_D - I_{in}) \quad (A.31)$$

$$I_{D-} = \frac{1}{2}(I_D + I_{in}) \quad (A.32)$$

and simplify further to

$$\frac{I_0 + I_{out}}{I_0 - I_{out}} = \frac{I_D + I_{in}}{I_D - I_{in}} \quad (A.33)$$

$$(I_0 + I_{out})(I_D - I_{in}) = (I_D + I_{in})(I_0 - I_{out}) \quad (A.34)$$

$$I_0(I_D - I_{in}) + I_{out}(I_D - I_{in}) = I_0(I_D + I_{in}) - I_{out}(I_D + I_{in}) \quad (A.35)$$

$$I_{out}(I_D - I_{in} + I_D + I_{in}) = I_0(I_D + I_{in} - I_D + I_{in}) \quad (A.36)$$

and finally arrive at

$$I_{out} = \frac{I_0}{I_D} I_{in} \qquad where \quad |I_{in}| < I_D \quad (A.37)$$

Looking at the last equation we find of course that we have a current amplifier[8] rather than a transconductance amplifier as the independent variable is now a current instead of a voltage. On the positive side, the temperature dependence of the transconductance is compensated. Of course one can use a resistor in front of each input for the voltage to current conversion, which should be dimensioned so that the maximum input current does not exceed the diode bias current at the maximum input voltage. It can also be observed that the maximum transconductance is achieved for vanishing diode biasing. While it appears at first that the transconductance can be made infinitely large, this is not the case as the input range is also zero at that point. We know of course that for vanishing diode bias current the OTA reverts to its non-linearized form.

When driven by voltage signals, resistors can be used to provide voltage to current conversion (the potential at the bases of the input transistors is almost constant). With equal input resistors the transconductance becomes

$$I_{out} = \frac{I_0}{R_{in} I_D} V_{in} \qquad where \quad |V_{in}| < R_{in} I_D \quad (A.38)$$

which also means that compensating for temperature is not as easy as it looked at first, depending on how you produce the currents for the tail and diodes. Overdriving a linearized OTA at the input more or less just clips the signal. Changes in

[8]The equation just derived may look familiar: it is the very same as for the famous Gilbert cell, where gain is the ratio of inner to outer current.

the input potential that are effected by changes in either I_0 or I_D produce common mode inputs and are thus suppressed at the output as long as the common mode input range is not exceeded. The driving stage should be designed with careful consideration of the comparatively low and non-constant input impedance of a linearized OTA.

A.2.2 The FET OTA

An OTA could obviously also be implemented in CMOS technology by replacing the current mirrors and the input differential pair with their FET equivalents. Assuming ideal current mirrors and current sources again, the only real change is the switch to a FET differential pair.

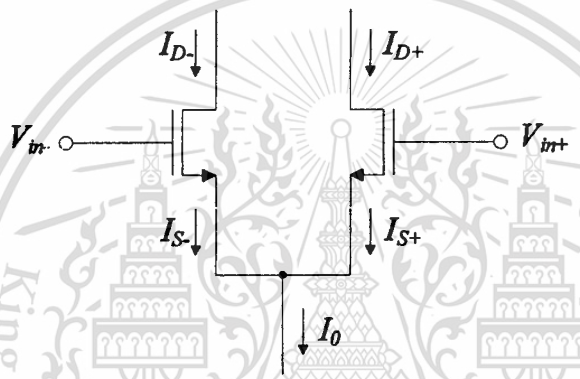


Figure A.4 FET differential pair (with nMOS enhancement FET, biasing not shown)

The FET Differential Pair

We notice that the input resistance is infinite and hence there is no input current. Also, the source and drain currents are equal if we neglect leakage currents. The drain current of the nMOS enhancement FET with a threshold voltage of V_{th} in pinch-off regime is with some simplifying assumptions

$$I_D = I_{Dsat}\left(\frac{V_{GS}}{V_{th}} - 1\right)^2 \qquad \left|V_{GS} \geq V_{th}\right. \tag{A.39}$$

Thus,

$$I_0 = I_+ + I_- = I_{S+} + I_{S-} = I_{D+} + I_{D-} \tag{A.40}$$

$$V_{in} = V_{GS+} - V_{GS-} = V_{in+} - V_{in-} \tag{A.41}$$

and with the transforms

$$i_* = \frac{I_*}{I_{Dsat}} \tag{A.42}$$

$$v_* = \left(\frac{V_*}{V_{th}} - 1\right) \geq 0 \tag{A.43}$$

$$\hat{i}_* = \frac{I_*}{I_0} \leq 1 \tag{A.44}$$

the equations

$$I_+ = I_{D+} = I_{Dsat+}\left(\frac{V_{GS+}}{V_{th}} - 1\right)^2 \tag{A.45}$$

$$I_- = I_{D-} = I_{Dsat-}\left(\frac{V_{GS-}}{V_{th}} - 1\right)^2 \tag{A.46}$$

can be written more simple as

$$i_{D+} = v_{GS+}^2 \tag{A.47}$$

$$i_{D-} = v_{GS-}^2 \tag{A.48}$$

Transformation and substitution into (A.41) yields under the assumption of matched transistors

$$v_{in} + 1 = \sqrt{i_+} - \sqrt{i_-} = \sqrt{i_0}(\sqrt{\hat{i}_+} - \sqrt{\hat{i}_-}) \tag{A.49}$$

Writing out the output current and using the identity $(\sqrt{A} - \sqrt{B})(\sqrt{A} + \sqrt{B}) = |A| - |B|$ together with (A.49) provides

$$\begin{aligned} i_{out} &= I_+ - I_- \\ &= (v_{in} + 1)(\sqrt{i_+} - \sqrt{i_-}) \\ &= (v_{in} + 1)\sqrt{i_0}(\sqrt{\hat{i}_+} - \sqrt{\hat{i}_-}) \end{aligned} \tag{A.50}$$

The maximum input range is therefore $\pm\sqrt{i_0}V_{th}$, the signal is clipped beyond that point as the tail current flows through just one transistor in the differential pair and the other is closed. Recalling that

$$\hat{i}_0 = 1 = \hat{i}_+ + \hat{i}_- \tag{A.51}$$

we can substitute

$$\sin^2 x + \cos^2 x = 1 \tag{A.52}$$

and use trigonometric identities to observe

$$\sqrt{\hat{i}_+} + \sqrt{\hat{i}_-} = \sin x + \cos x = \sqrt{2}\sin(x + \frac{\pi}{4}) \qquad \in [1, \sqrt{2}] \tag{A.53}$$

$$\sqrt{\hat{i}_+} - \sqrt{\hat{i}_-} = \sin x - \cos x = \sqrt{2}\sin(x - \frac{\pi}{4}) \qquad \in [-1, 1] \tag{A.54}$$

Through substitution of (A.54) into (A.49) we solve for

$$x = \arcsin\left(\frac{v_{in}+1}{\sqrt{2i_0}}\right) + \frac{\pi}{4} \tag{A.55}$$

With (A.50), (A.53) and the identity $\sin(x+\frac{\pi}{2}) = \cos x$ we can finally express the output current as a function of input voltage

$$\begin{aligned} i_{out} &= (v_{in}+1)\sqrt{2i_0}\cos\left(\arcsin\frac{v_{in}+1}{\sqrt{2i_0}}\right) && \text{(A.56)} \\ I_{out} &= 2I_0\frac{V_{in}}{V_{th}\sqrt{2i_0}}\cos\left(\arcsin\frac{V_{in}}{V_{th}\sqrt{2i_0}}\right) \\ &= 2I_0 z\cos(\arcsin z) \qquad \left| z = \frac{V_{in}}{V_{th}\sqrt{2i_0}}\right. && \text{(A.57)} \\ &\approx 2I_0 z \qquad \left| z \ll 1\right. \end{aligned}$$

which gives

$$g_m = \frac{dI_{out}}{dV_{in}} = \frac{dI_{out}}{dz}\frac{dz}{dV_{in}} \approx \frac{\sqrt{2I_o I_{Dsat}}}{V_{th}} = \frac{I_{Dsat}}{V_{th}}\sqrt{2i_0} \tag{A.58}$$

This means that the gm of a FET OTA is not proportional to the tail current[9] as for the bipolar OTA, but rather to its square root. As long as one wants exponential control, it is sufficient to double the scale factor. Then each octave of transconductance translates into two octaves of tail current. The square law characteristic of the FET is not nearly as precise as the exponential characteristic of a bipolar transistor, so it is challenging to maintain tracking over many octaves.

For linear control, one could conceivably rig up a circuit with another matched FET to deliver a current proportional to the input voltage (the biasing may be somewhat tricky). Also, the input range of the FET OTA varies considerably with the transconductance, to keep linearity to one percent the input range again has to be in the Millivolt range.

A.3 Applications

The OTA as voltage controlled resistor

If there is a resistor that is referenced to the virtual ground of an operational amplifier, then it is easy to use an OTA to make that resistance voltage controlled. The resistor is replaced by a voltage divider to the real ground so that the divider puts out about 5 mV, which gets connected to the positive input of

[9]The same result is more laboriously arrived at via developing the full expression into a power series, differentiating that and truncating to the linear term. The quadratic term is slightly more than 1.5 times larger than that of the equivalent power series for the $sech^2$ part of the expression for the bipolar OTA.

the OTA. The negative input is connected to ground as well, while the output of the OTA goes into the virtual ground of the operational amplifier. The apparent resistance can then be controlled by adjusting I_0 accordingly.

Another applications can be found in any OTA IC datasheets.

A.4 OTA IC

A.4.1 The CA3080

The CA 3080 is probably the most simple bipolar OTA that you can find. It consists of only the input differential pair and the current mirrors that bias the input transistors and produce the output current. In particular, the mirror for the tail current is a simple Widlar type and emitter degeneration cannot be used as the tail current can vary widely. It is therefore important to keep the differential and current inputs at the same potential, otherwise the transconductance gets modulated by the common mode input voltage. Unfortunately the datasheet does not show the circuit for measuring the CMRR, but it appears that the common mode amplitude was low for the test and the input potentials about the same. The output current mirrors are all Wilson type, the pnp mirrors also use a Darlington pair for the cascode transistor.

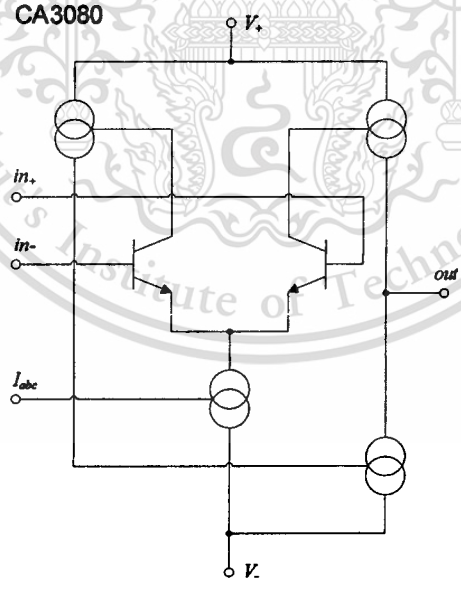


Figure A.5 CA3080

A.4.2 The LM13600/LM13700

The LM13700 improves upon the CA3080 by adding linearization to the OTA inputs. While this improves the linear input range greatly, it lowers input impedance and changes the distortion properties. It uses a Wilson mirror also for the tail current. Since a Wilson mirror needs more voltage headroom, the common mode voltage range is reduced on the negative rail and the potential for the tail current input is increased in comparison with the CA3080, which may become important in certain applications. The LM13600 and the LM13700 differ only in the way the bias current for the buffer (which is not shown here) is produced. The LM13700 uses a constant bias current according to the datasheet, while in the LM13600 the bias is a mirrored copy of the tail current. This can lead to CV feedthrough to the output when the tail current is changed rapidly. However the datasheet for the LM13700 does not show any biasing of the buffer at all, so one can only speculate how it is achieved. What is clear is that there must be some biasing and the only hint one can find of that is some mumbling about "controlled impedance buffers".

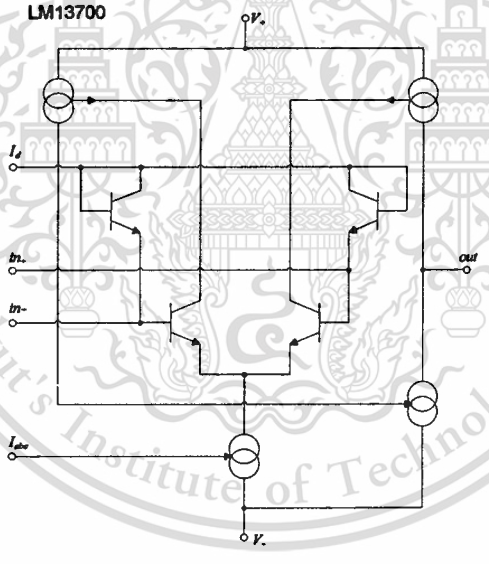


Figure A.6 LM13700

A.4.3 The NE5517

The OTA section of the NE5517 is identical to the LM13700. The buffer bias is almost constant, only somewhat varied with the tail current, presumably to compensate the change in output impedance of the OTA section. The datasheet consequently claims "constant impedance buffers". Since all figures in the datasheet are identical I suspect that the missing bias network in the datasheet of the LM13700 looks the same. This is corroborated by the fact that various distributors use both types as replacements for each other.

A.4.4 The CA3280

The CA3280 also adds linearization, but in a slightly more complicated way then the LM13700 that maximizes the common mode input range when the lineaization diodes are used. It also uses a Wilson mirror for the tail current. The output mirrors are not shown in detail on the datasheet. While its safe to assume theyre Wilson types, it is hard to know exactly if they use Darlington pairs. The relatively wide bandwith leads me to assume that theyre plain pnp transistors like the LM13700, however.

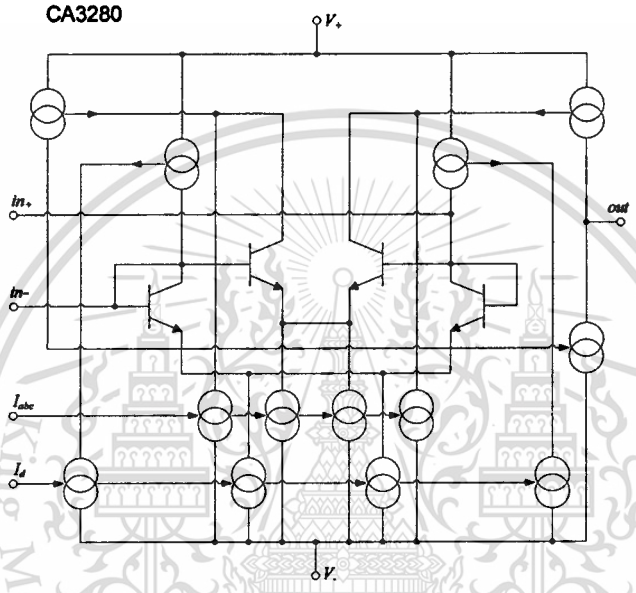


Figure A.7 CA3280

A.4.5 The "Diamond Transistor" OPA660

The "Diamond Transistor" OPA660 has a different tack on the OTA theme. The negative input is a low impedance terminal, in effect becoming both an input and a (differential) output.

A.4.6 The MAX435 / MAX436

The MAX435 is an OTA with differential outputs and a gain setting network, the MAX436 drops the differential output and has a different internal gain factor. The maximum output current is controlled by a set current like the conventional OTA. It is unclear whether these OTA could be used without the gain setting impedance and if the transconductance would then be controllable through the set current.

Appendix B

Proofs of an OTA Characteristic Equation

B.1 First Derivation

This alternative (and a bit shorter) derivation of (A.20) (provided by Ian Fritz) results:

$$\begin{aligned} I_{out} &= I_0 \left(\frac{\exp\frac{V_{in}}{2V_T}}{\exp\frac{V_{in}}{2V_T}} \frac{1}{1+\exp\frac{-V_{in}}{V_T}} - \frac{\exp\frac{-V_{in}}{2V_T}}{\exp\frac{-V_{in}}{2V_T}} \frac{1}{1+\exp\frac{V_{in}}{V_T}} \right) \\ &= I_0 \left(\frac{\exp\frac{V_{in}}{2V_T}}{\exp\frac{V_{in}}{2V_T} + \exp\frac{V_{in}}{2V_T}\exp\frac{-2V_{in}}{2V_T}} - \frac{\exp\frac{-V_{in}}{2V_T}}{\exp\frac{-V_{in}}{2V_T} + \exp\frac{-V_{in}}{2V_T}\exp\frac{2V_{in}}{2V_T}} \right) \\ &= I_0 \frac{\exp\frac{V_{in}}{2V_T} - \exp\frac{-V_{in}}{2V_T}}{\exp\frac{V_{in}}{2V_T} + \exp\frac{-V_{in}}{2V_T}} \\ &= I_0 \tanh\frac{V_{in}}{2V_T} \end{aligned}$$

The different signs of the multiplicands can be motivated by symmetry considerations.

B.2 Second Derivation

The second alternative derivation comes from the lecture notes on analog multiplication by Paul Junor. It starts off with a slightly different reduction of the common denominator, while the introduction of the half-argument can again be motivated by symmetry considerations.

$$I_{out} = I_0 \frac{(1+\exp\frac{V_{in}}{V_T}) - ((1+\exp\frac{-V_{in}}{V_T})}{(1+\exp\frac{V_{in}}{V_T}) + ((1+\exp\frac{-V_{in}}{V_T})}$$

Substituting the identity

$$\exp\frac{x}{2}\exp -x2 = 1$$

gives

$$I_{out} = I_0 \frac{(\exp\frac{-V_{in}}{2V_T}\exp\frac{V_{in}}{2V_T} + \exp 2V_{in}2V_T) - (\exp\frac{-V_{in}}{2V_T}\exp\frac{V_{in}}{2V_T} + \exp -2V_{in}2V_T)}{(\exp\frac{-V_{in}}{2V_T}\exp\frac{V_{in}}{2V_T} + \exp 2V_{in}2V_T) + (\exp\frac{-V_{in}}{2V_T}\exp\frac{V_{in}}{2V_T} + \exp -2V_{in}2V_T)}$$

which enables the following extraction

$$I_{out} = I_0 \frac{\exp\frac{V_{in}}{2V_T}[\exp\frac{-V_{in}}{2V_T} + \exp\frac{V_{in}}{2V_T}] - \exp\frac{-V_{in}}{2V_T}[\exp\frac{V_{in}}{2V_T} + \exp\frac{-V_{in}}{2V_T}]}{\exp\frac{V_{in}}{2V_T}[\exp\frac{V_{in}}{2V_T} + \exp\frac{-V_{in}}{2V_T}] + \exp\frac{-V_{in}}{2V_T}[\exp\frac{-V_{in}}{2V_T} + \exp\frac{V_{in}}{2V_T}]}$$

dropping the terms in brackets gives

$$I_{out} = I_0 \frac{\exp \frac{V_{in}}{2V_T} - \exp \frac{-V_{in}}{2V_T}}{\exp \frac{V_{in}}{2V_T} + \exp \frac{-V_{in}}{2V_T}}$$

which interpreted as hyperbolic function reads

$$\begin{aligned} I_{out} &= I_0 \frac{\sinh \frac{V_{in}}{2V_T}}{\cosh \frac{V_{in}}{2V_T}} \\ &= I_0 \tanh \frac{V_{in}}{2V_T} \end{aligned}$$

B.3 Another Proof

Another proof starting with developing an expression for the individual collector currents via (A.3) and with (A.9) or we can simply take it out from the first part of (A.16):

$$I_+ = \frac{I_0}{1 + \exp \frac{-V_{in}}{V_T}}$$

motivated by the fact that with no signal each branch of the differential pair sees half the tail current we pull this out as the scaling factor

$$I_+ = \frac{I_0}{2} \frac{2}{1 + \exp \frac{-V_{in}}{V_T}}$$

and substitute the 2

$$\begin{aligned} I_+ &= \frac{I_0}{2} \frac{(1 + \exp \frac{-V_{in}}{V_T}) + (1 - \exp \frac{-V_{in}}{V_T})}{1 - \exp \frac{-V_{in}}{V_T}} \\ &= \frac{I_0}{2} \left(1 + \frac{1 - \exp \frac{-V_{in}}{V_T}}{1 + \exp \frac{-V_{in}}{V_T}} \right) \end{aligned}$$

and via one of the addition theorems we find

$$I_+ = \frac{I_0}{2} \left(1 + \tanh \frac{V_{in}}{2V_T} \right)$$

and due to symmetry

$$I_- = \frac{I_0}{2} \left(1 - \tanh \frac{V_{in}}{2V_T} \right)$$

and finally we arrive via (A.13) at

$$\begin{aligned} I_{out} &= \frac{I_0}{2} \left(1 + \tanh \frac{V_{in}}{2V_T} \right) - \frac{I_0}{2} \left(1 - \tanh \frac{V_{in}}{2V_T} \right) \\ &= I_0 \tanh \frac{V_{in}}{2V_T} \end{aligned}$$

Appendix C

LM13600 Datasheet

National Semiconductor

February 1995

LM13600 Dual Operational Transconductance Amplifiers with Linearizing Diodes and Buffers

General Description

The LM13600 series consists of two current controlled transconductance amplifiers each with differential inputs and a push-pull output. The two amplifiers share common supplies but otherwise operate independently. Linearizing diodes are provided at the inputs to reduce distortion and allow higher input levels. The result is a 10 dB signal-to-noise improvement referenced to 0.5 percent THD. Controlled impedance buffers which are especially designed to complement the dynamic range of the amplifiers are provided.

Features

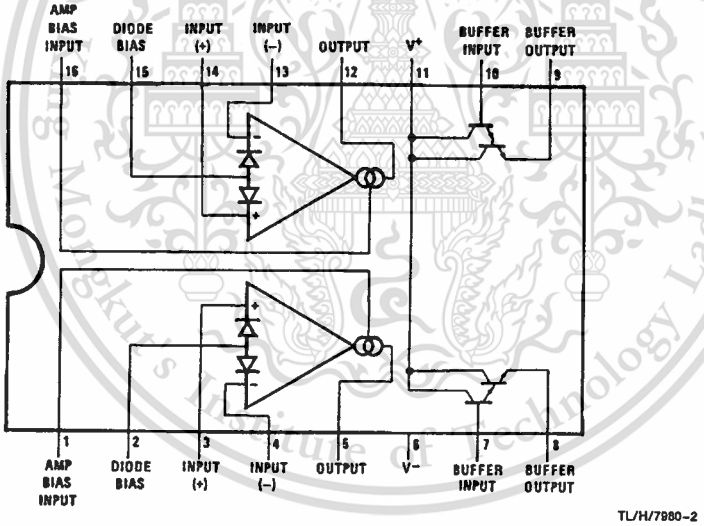
- g_m adjustable over 6 decades
- Excellent g_m linearity
- Excellent matching between amplifiers
- Linearizing diodes
- Controlled impedance buffers
- High output signal-to-noise ratio

Applications

- Current-controlled amplifiers
- Current-controlled impedances
- Current-controlled filters
- Current-controlled oscillators
- Multiplexers
- Timers
- Sample and hold circuits

Connection Diagram

Dual-In-Line and Small Outline Packages



TL/H/7980–2

Top View

Order Number LM13600M, LM13600N or LM13600AN
See NS Package Number M16A or N16A

RRD-B30M115/Printed in U. S. A.

Absolute Maximum Ratings

If Military/Aerospace specified devices are required, please contact the National Semiconductor Sales Office/Distributors for availability and specifications.

Supply Voltage (Note 1)	
LM13600	36 V_{DC} or ±18V
LM13600A	44 V_{DC} or ±22V
Power Dissipation (Note 2) T_A = 25°C	570 mW
Differential Input Voltage	±5V
Diode Bias Current (I_D)	2 mA
Amplifier Bias Current (I_{ABC})	2 mA
Output Short Circuit Duration	Continuous
Buffer Output Current (Note 3)	20 mA
Operating Temperature Range	0°C to +70°C
DC Input Voltage	$+V_S$ to $-V_S$
Storage Temperature Range	−65°C to +150°C
Soldering Information	
Dual-In-Line Package	
Soldering (10 seconds)	260°C
Small Outline Package	
Vapor Phase (60 seconds)	215°C
Infrared (15 seconds)	220°C

See AN-450 "Surface Mounting Methods and Their Effect on Product Reliability" for other methods of soldering surface mount devices.

Electrical Characteristics (Note 4)

Parameter	Conditions	LM13600			LM13600A			Units
		Min	Typ	Max	Min	Typ	Max	
Input Offset Voltage (V_{OS})			0.4	4		0.4	1	mV
	Over Specified Temperature Range						2	mV
	I_{ABC} = 5 μA		0.3	4		0.3	1	mV
V_{OS} Including Diodes	Diode Bias Current (I_D) = 500 μA		0.5	5		0.5	2	mV
Input Offset Change	5 μA ≤ I_{ABC} ≤ 500 μA		0.1	3		0.1	1	mV
Input Offset Current			0.1	0.6		0.1	0.6	μA
Input Bias Current			0.4	5		0.4	5	μA
	Over Specified Temperature Range		1	8		1	7	μA
Forward Transconductance (g_m)		6700	9600	13000	7700	9600	12000	μmho
	Over Specified Temperature Range	5400			4000			μmho
g_m Tracking			0.3			0.3		dB
Peak Output Current	R_L = 0, I_{ABC} = 5 μA		5		3	5	7	μA
	R_L = 0, I_{ABC} = 500 μA	350	500	650	350	500	650	μA
	R_L = 0, Over Specified Temp Range	300			300			μA
Peak Output Voltage								
Positive	R_L = ∞, 5 μA ≤ I_{ABC} ≤ 500 μA	+12	+14.2		+12	+14.2		V
Negative	R_L = ∞, 5 μA ≤ I_{ABC} ≤ 500 μA	−12	−14.4		−12	−14.4		V
Supply Current	I_{ABC} = 500 μA, Both Channels		2.6			2.6		mA
V_{OS} Sensitivity								
Positive	$\Delta V_{OS}/\Delta V+$		20	150		20	150	μV/V
Negative	$\Delta V_{OS}/\Delta V-$		20	150		20	150	μV/V
CMRR		80	110		80	110		dB
Common Mode Range		±12	±13.5		±12	±13.5		V
Crosstalk	Referred to Input (Note 5) 20 Hz < f < 20 kHz		100			100		dB
Differential Input Current	I_{ABC} = 0, Input = ±4V		0.02	100		0.02	10	nA
Leakage Current	I_{ABC} = 0 (Refer to Test Circuit)		0.2	100		0.2	5	nA

Electrical Characteristics (Note 4) (Continued)

Parameter	Conditions	LM13600			LM13600A			Units
		Min	Typ	Max	Min	Typ	Max	
Input Resistance		10	26		10	26		kΩ
Open Loop Bandwidth			2			2		MHz
Slew Rate	Unity Gain Compensated		50			50		V/μs
Buffer Input Current	(Note 5), Except I_{ABC} = 0 μA		0.2	0.4		0.2	0.4	μA
Peak Buffer Output Voltage	(Note 5)	10			10			V

Note 1: For selections to a supply voltage above ±22V, contact factory.

Note 2: For operating at high temperatures, the device must be derated based on a 150°C maximum junction temperature and a thermal resistance of 175°C/W which applies for the device soldered in a printed circuit board, operating in still air.

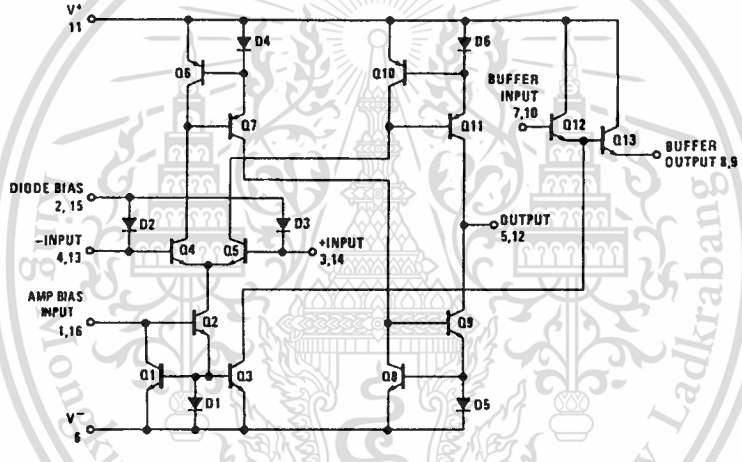
Note 3: Buffer output current should be limited so as to not exceed package dissipation.

Note 4: These specifications apply for V_S = ±15V, T_A = 25°C, amplifier bias current (I_{ABC}) = 500 μA, pins 2 and 15 open unless otherwise specified. The inputs to the buffers are grounded and outputs are open.

Note 5: These specifications apply for V_S = ±15V, I_{ABC} = 500 μA, R_{OUT} = 5 kΩ connected from the buffer output to $-V_S$ and the input of the buffer is connected to the transconductance amplifier output.

Schematic Diagram

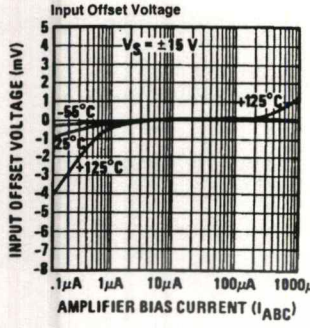
One Operational Transconductance Amplifier



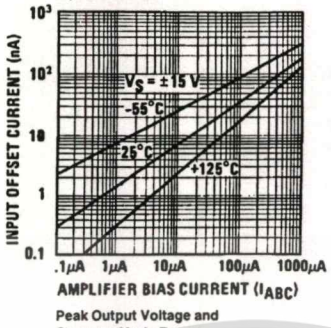
TL/H/7980–1

Typical Performance Characteristics

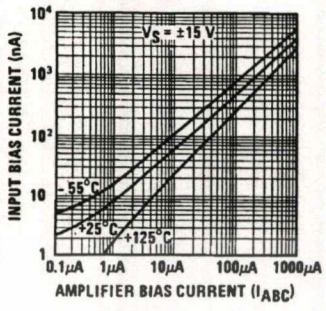
Input Offset Voltage



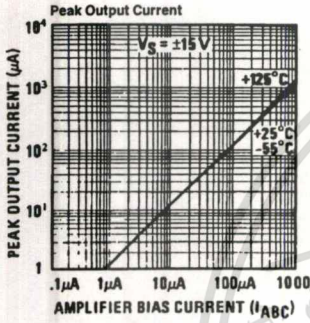
Input Offset Current



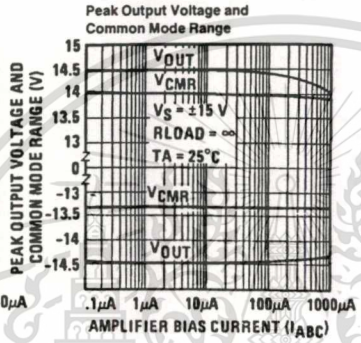
Input Bias Current



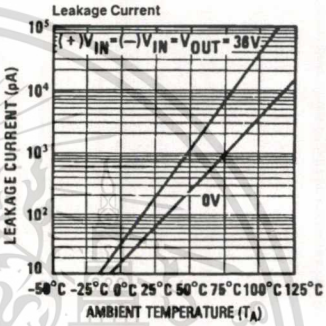
Peak Output Current



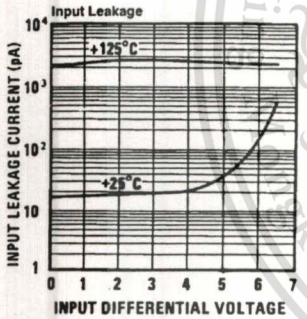
Peak Output Voltage and Common Mode Range



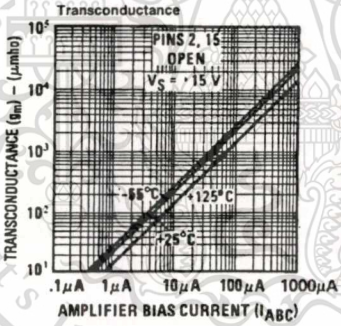
Leakage Current



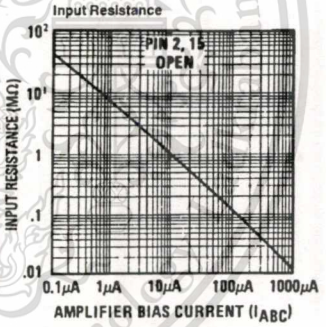
Input Leakage



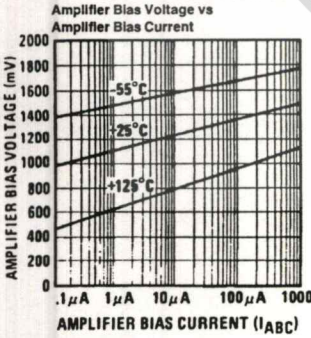
Transconductance



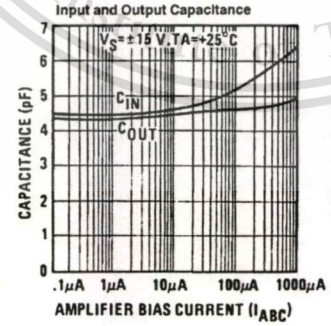
Input Resistance



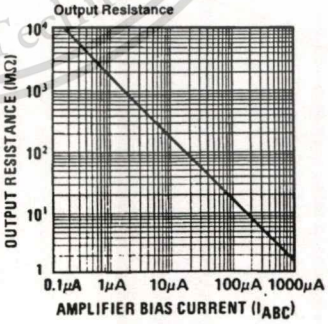
Amplifier Bias Voltage vs Amplifier Bias Current



Input and Output Capacitance

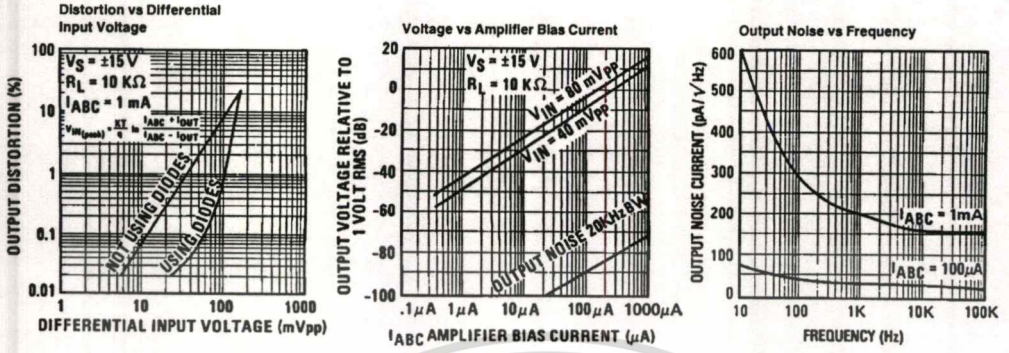


Output Resistance



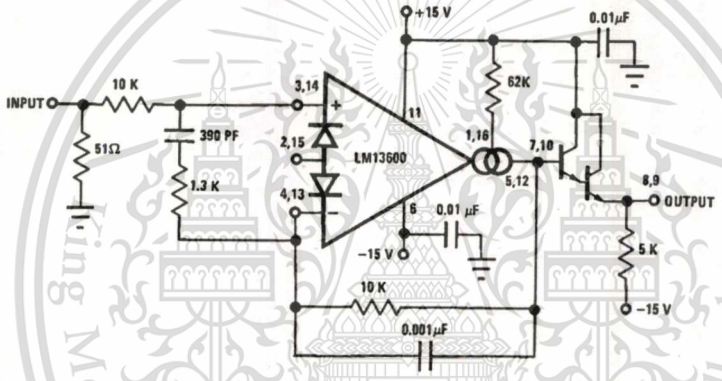
TL/H/7980–3

Typical Performance Characteristics (Continued)



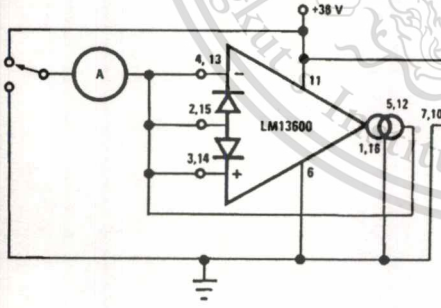
TL/H/7980–4

Unity Gain Follower

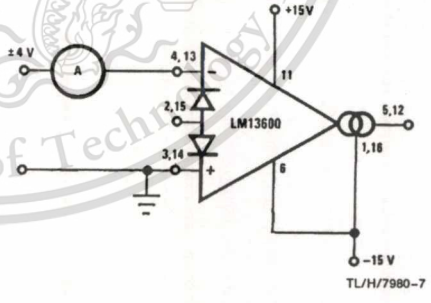


TL/H/7980–5

Leakage Current Test Circuit



Differential Input Current Test Circuit



TL/H/7980–7

TL/H/7980–6

Circuit Description

The differential transistor pair Q_4 and Q_5 form a transconductance stage in that the ratio of their collector currents is defined by the differential input voltage according to the transfer function:

$$V_{IN} = \frac{kT}{q} \ln \frac{I_5}{I_4} \quad (1)$$

where V_{IN} is the differential input voltage, kT/q is approximately 26 mV at 25°C and I_5 and I_4 are the collector currents of transistors Q_5 and Q_4 respectively. With the exception of Q_3 and Q_{13}, all transistors and diodes are identical in size. Transistors Q_1 and Q_2 with Diode D_1 form a current mirror which forces the sum of currents I_4 and I_5 to equal I_{ABC}:

$$I_4 + I_5 = I_{ABC} \quad (2)$$

where I_{ABC} is the amplifier bias current applied to the gain pin.

For small differential input voltages the ratio of I_4 and I_5 approaches unity and the Taylor series of the ln function can be approximated as:

$$\frac{kT}{q} \ln \frac{I_5}{I_4} \approx \frac{kT}{q} \frac{I_5 - I_4}{I_4} \quad (3)$$

$$I_4 \approx I_5 \approx \frac{I_{ABC}}{2}$$

$$V_{IN} \left[\frac{I_{ABC}\, q}{2kT} \right] = I_5 - I_4 \quad (4)$$

Collector currents I_4 and I_5 are not very useful by themselves and it is necessary to subtract one current from the other. The remaining transistors and diodes form three current mirrors that produce an output current equal to I_5 minus I_4 thus:

$$V_{IN} \left[\frac{I_{ABC}\, q}{2kT} \right] = I_{OUT} \quad (5)$$

The term in brackets is then the transconductance of the amplifier and is proportional to I_{ABC}.

Linearizing Diodes

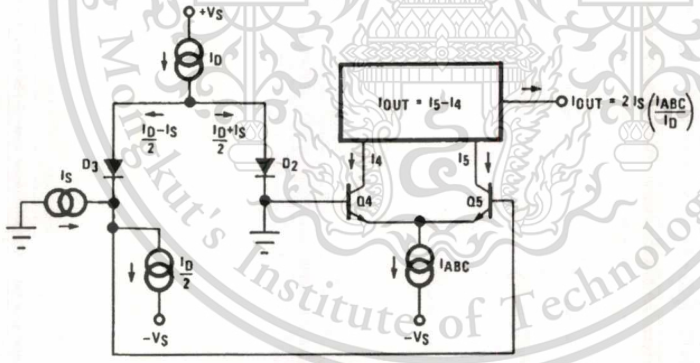
For differential voltages greater than a few millivolts, Equation 3 becomes less valid and the transconductance becomes increasingly nonlinear. *Figure 1* demonstrates how the internal diodes can linearize the transfer function of the amplifier. For convenience assume the diodes are biased with current sources and the input signal is in the form of current I_S. Since the sum of I_4 and I_5 is I_{ABC} and the difference is I_{OUT}, currents I_4 and I_5 can be written as follows:

$$I_4 = \frac{I_{ABC}}{2} - \frac{I_{OUT}}{2}, \; I_5 = \frac{I_{ABC}}{2} + \frac{I_{OUT}}{2}$$

Since the diodes and the input transistors have identical geometries and are subject to similar voltages and temperatures, the following is true:

$$\frac{kT}{q} \ln \frac{\frac{I_D}{2} + I_S}{\frac{I_D}{2} - I_S} = \frac{kT}{q} \ln \frac{\frac{I_{ABC}}{2} + \frac{I_{out}}{2}}{\frac{I_{ABC}}{2} - \frac{I_{out}}{2}}$$

$$\therefore I_{out} = I_S \left(\frac{2I_{ABC}}{I_D} \right) \quad \text{for } |I_S| < \frac{I_D}{2} \quad (6)$$



TL/H/7980–8

FIGURE 1. Linearizing Diodes

Linearizing Diodes (Continued)

Notice that in deriving Equation 6 no approximations have been made and there are no temperature-dependent terms. The limitations are that the signal current not exceed $I_D/2$ and that the diodes be biased with currents. In practice, replacing the current sources with resistors will generate insignificant errors.

Controlled Impedance Buffers

The upper limit of transconductance is defined by the maximum value of I_{ABC} (2 mA). The lowest value of I_{ABC} for which the amplifier will function therefore determines the overall dynamic range. At very low values of I_{ABC}, a buffer which has very low input bias current is desirable. An FET follower satisfies the low input current requirement, but is somewhat non-linear for large voltage swing. The controlled impedance buffer is a Darlington which modifies its input bias current to suit the need. For low values of I_{ABC}, the buffer's input current is minimal. At higher levels of I_{ABC}, transistor Q_3 biases up Q_{12} with a current proportional to I_{ABC} for fast slew rate. When I_{ABC} is changed, the DC level of the Darlington output buffer will shift. In audio applications where I_{ABC} is changed suddenly, this shift may produce an audible "pop". For these applications the LM13700 may produce superior results.

Applications—Voltage Controlled Amplifiers

Figure 2 shows how the linearizing diodes can be used in a voltage-controlled amplifier. To understand the input biasing, it is best to consider the 13 kΩ resistor as a current source and use a Thevenin equivalent circuit as shown in *Figure 3*. This circuit is similar to *Figure 1* and operates the same. The potentiometer in *Figure 2* is adjusted to minimize the effects of the control signal at the output.

For optimum signal-to-noise performance, I_{ABC} should be as large as possible as shown by the Output Voltage vs. Amplifier Bias Current graph. Larger amplitudes of input signal also improve the S/N ratio. The linearizing diodes help here by allowing larger input signals for the same output distortion as shown by the Distortion vs. Differential Input Voltage graph. S/N may be optimized by adjusting the magnitude of the input signal via R_{IN} *(Figure 2)* until the output distortion is below some desired level. The output voltage swing can then be set at any level by selecting R_L.

Although the noise contribution of the linearizing diodes is negligible relative to the contribution of the amplifier's internal transistors, I_D should be as large as possible. This minimizes the dynamic junction resistance of the diodes (r_e) and maximizes their linearizing action when balanced against R_{IN}. A value of 1 mA is recommended for I_D unless the specific application demands otherwise.

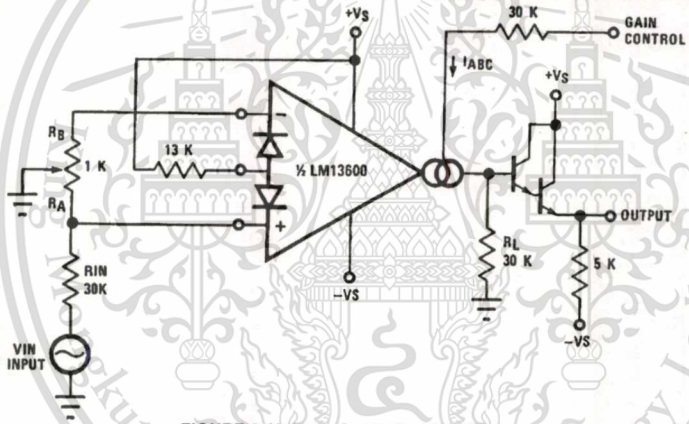


FIGURE 2. Voltage Controlled Amplifier TL/H/7980–9

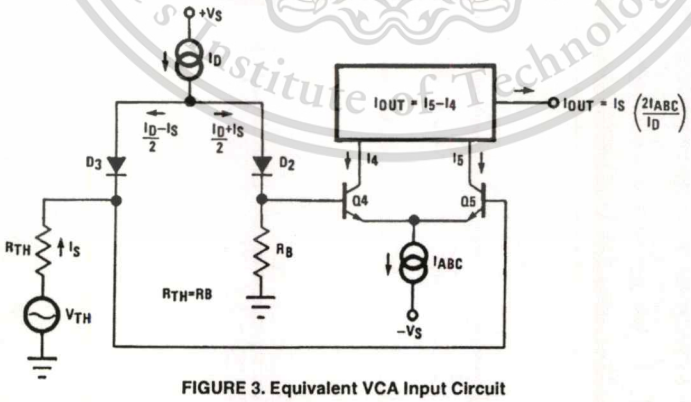


FIGURE 3. Equivalent VCA Input Circuit TL/H/7980–10

Stereo Volume Control

The circuit of Figure 4 uses the excellent matching of the two LM13600 amplifiers to provide a Stereo Volume Control with a typical channel-to-channel gain tracking of 0.3 dB. R_P is provided to minimize the output offset voltage and may be replaced with two 510Ω resistors in AC-coupled applications. For the component values given, amplifier gain is derived for *Figure 2* as being:

$$\frac{V_O}{V_{IN}} = 940 \times I_{ABC}$$

If V_C is derived from a second signal source then the circuit becomes an amplitude modulator or two-quadrant multiplier as shown in *Figure 5*, where:

$$I_O = \frac{-2I_S}{I_D}(I_{ABC}) = \frac{-2I_S}{I_D}\frac{V_{IN2}}{R_C} - \frac{2I_S}{I_D}\frac{(V^- + 1.4V)}{R_C}$$

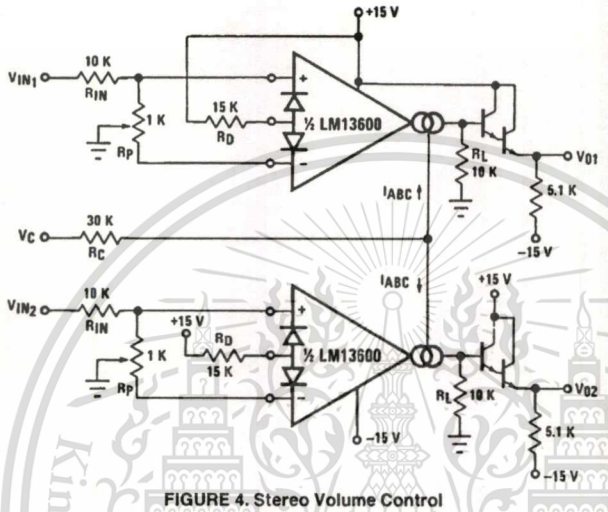


FIGURE 4. Stereo Volume Control

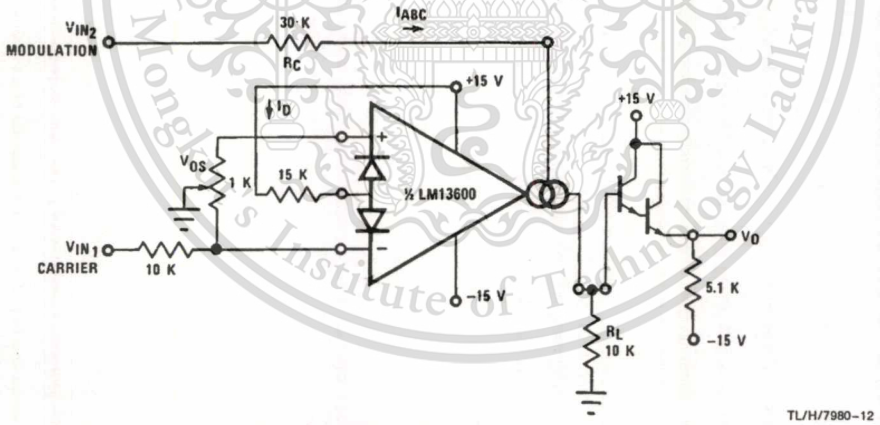


FIGURE 5. Amplitude Modulator

Stereo Volume Control (Continued)

The constant term in the above equation may be cancelled by feeding $I_S \times I_D R_C/2\ (V^- + 1.4V)$ into I_O. The circuit of *Figure 6* adds R_M to provide this current, resulting in a four-quadrant multiplier where R_C is trimmed such that $V_O = 0V$ for $V_{IN2} = 0V$. R_M also serves as the load resistor for I_O.

Noting that the gain of the LM13600 amplifier of *Figure 3* may be controlled by varying the linearizing diode current I_D as well as by varying I_{ABC}, *Figure 7* shows an AGC Amplifier using this approach. As V_O reaches a high enough amplitude (3 V_{BE}) to turn on the Darlington transistors and the linearizing diodes, the increase in I_D reduces the amplifier gain so as to hold V_O at that level.

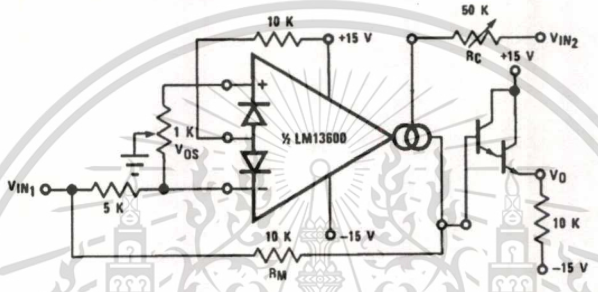
Voltage Controlled Resistors

An Operational Transconductance Amplifier (OTA) may be used to implement a Voltage Controlled Resistor as shown in *Figure 8*. A signal voltage applied at R_X generates a V_{IN} to the LM13600 which is then multiplied by the g_m of the amplifier to produce an output current, thus:

$$R_X = \frac{R + R_A}{g_m R_A}$$

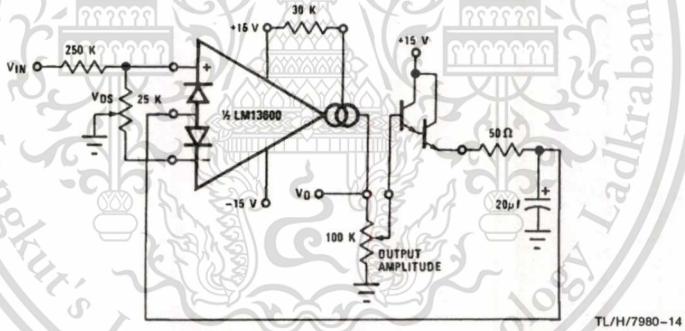
where $g_m \approx 19.2\ I_{ABC}$ at 25°C. Note that the attenuation of V_O by R and R_A is necessary to maintain V_{IN} within the linear range of the LM13600 input.

Figure 9 shows a similar VCR where the linearizing diodes are added, essentially improving the noise performance of the resistor. A floating VCR is shown in *Figure 10*, where each "end" of the "resistor" may be at any voltage within the output voltage range of the LM13600.



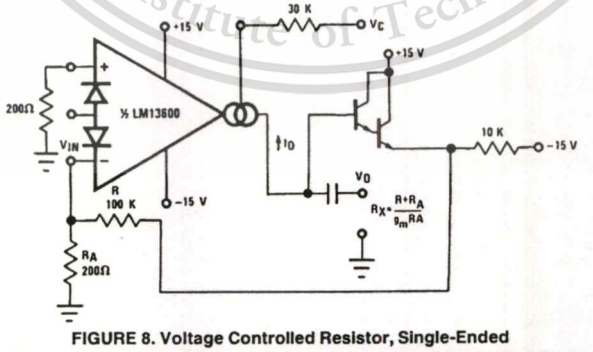
TL/H/7980–13

FIGURE 6. Four-Quadrant Multiplier



TL/H/7980–14

FIGURE 7. AGC Amplifier



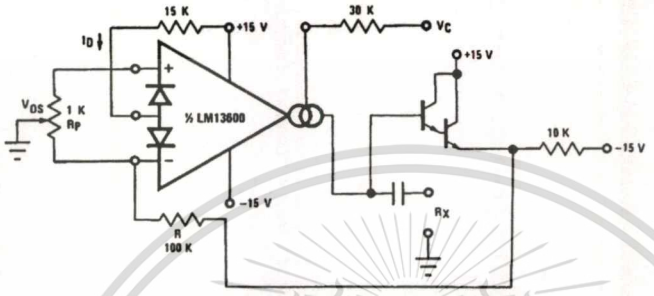
TL/H/7980–15

FIGURE 8. Voltage Controlled Resistor, Single-Ended

Voltage Controlled Filters

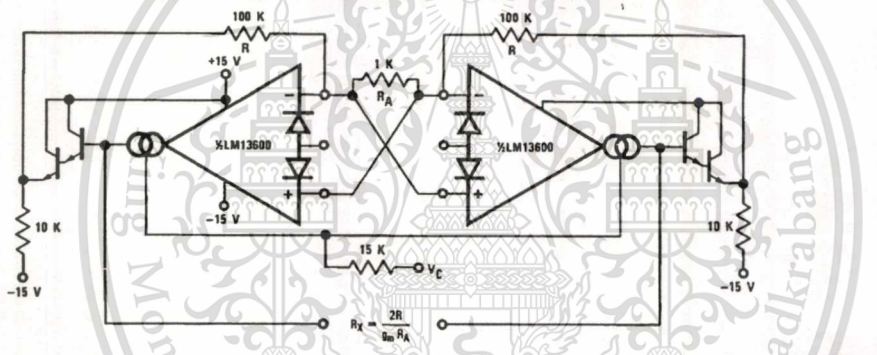
OTA's are extremely useful for implementing voltage controlled filters, with the LM13600 having the advantage that the required buffers are included on the I.C. The VC Lo-Pass Filter of *Figure 11* performs as a unity-gain buffer amplifier at frequencies below cut-off, with the cut-off frequency being the point at which X_C/g_m equals the closed-loop gain of (R/R_A). At frequencies above cut-off the circuit provides a single RC roll-off (6 dB per octave) of the input signal amplitude with a −3 dB point defined by the given equation, where g_m is again 19.2 × I_{ABC} at room temperature. *Figure 12* shows a VC High-Pass Filter which operates in much the same manner, providing a single RC roll-off below the defined cut-off frequency.

Additional amplifiers may be used to implement higher order filters as demonstrated by the two-pole Butterworth Lo-Pass Filter of *Figure 13* and the state variable filter of *Figure 14*. Due to the excellent g_m tracking of the two amplifiers and the varied bias of the buffer Darlingtons, these filters perform well over several decades of frequency.



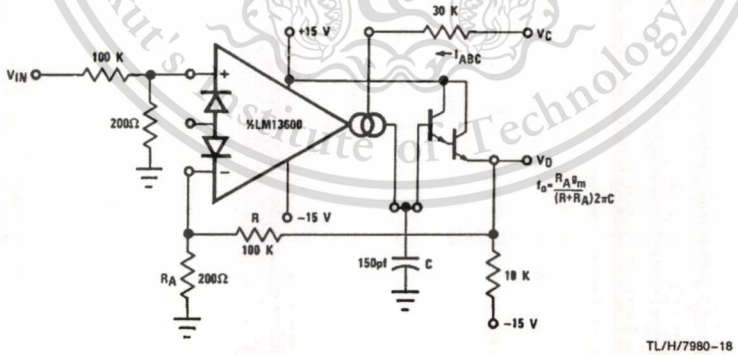
TL/H/7980–16

FIGURE 9. Voltage Controlled Resistor with Linearizing Diodes



TL/H/7980–17

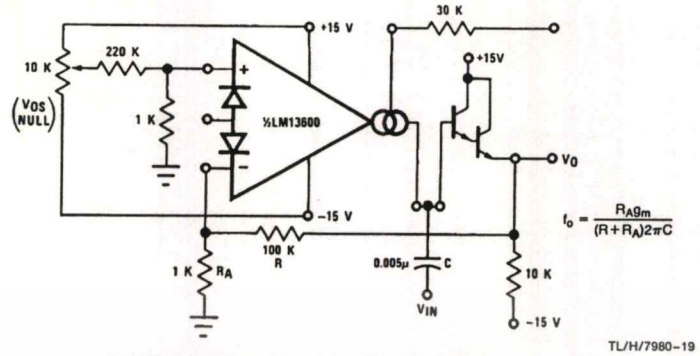
FIGURE 10. Floating Voltage Controlled Resistor



TL/H/7980–18

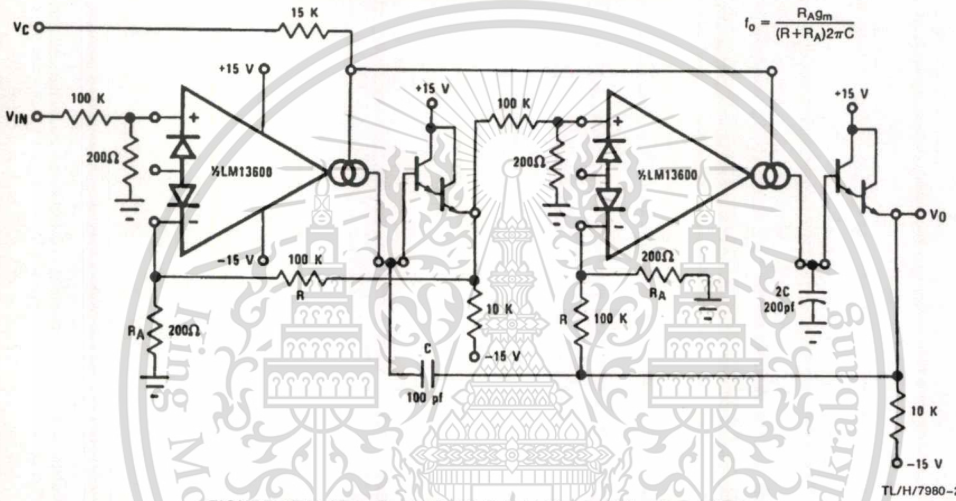
FIGURE 11. Voltage Controlled Low-Pass Filter

Voltage Controlled Filters (Continued)



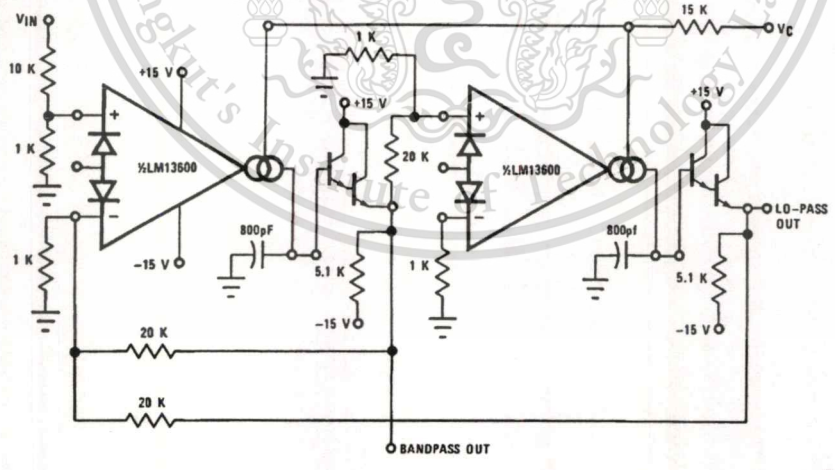
TL/H/7980–19

FIGURE 12. Voltage Controlled Hi-Pass Filter



TL/H/7980–20

FIGURE 13. Voltage Controlled 2-Pole Butterworth Lo-Pass Filter



TL/H/7980–21

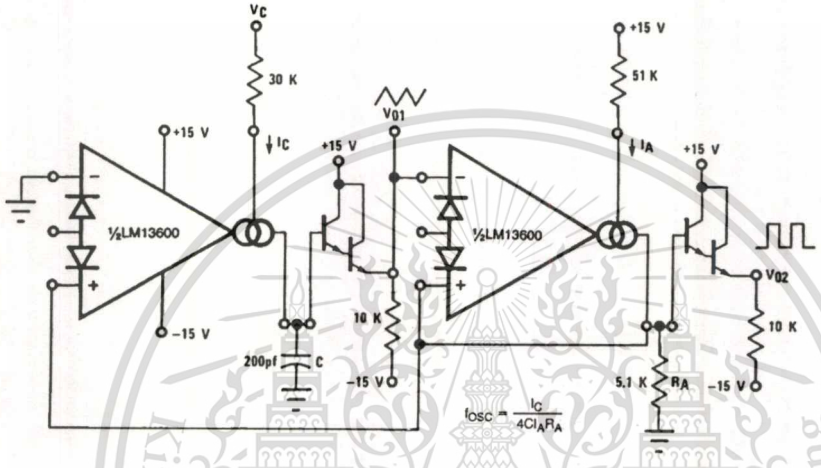
FIGURE 14. Voltage Controlled State Variable Filter

Voltage Controlled Oscillators

The classic Triangular/Square Wave VCO of *Figure 15* is one of a variety of Voltage Controlled Oscillators which may be built utilizing the LM13600. With the component values shown, this oscillator provides signals from 200 kHz to below 2 Hz as I_C is varied from 1 mA to 10 nA. The output amplitudes are set by $I_A \times R_A$. Note that the peak differential input voltage must be less than 5V to prevent zenering the inputs.

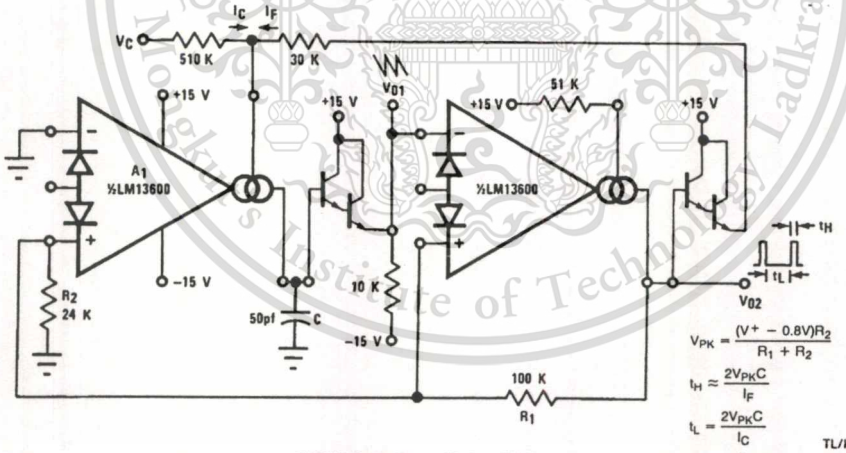
A few modifications to this circuit produce the ramp/pulse VCO of *Figure 16*. When V_{O2} is high, I_F is added to I_C to increase amplifier A1's bias current and thus to increase the charging rate of capacitor C. When V_{O2} is low, I_F goes to zero and the capacitor discharge current is set by I_C.

The VC Lo-Pass Filter of *Figure 11* may be used to produce a high-quality sinusoidal VCO. The circuit of *Figure 16* employs two LM13600 packages, with three of the amplifiers configured as lo-pass filters and the fourth as a limiter/inverter. The circuit oscillates at the frequency at which the loop phase-shift is 360° or 180° for the inverter and 60° per filter stage. This VCO operates from 5 Hz to 50 kHz with less than 1% THD.



TL/H/7980–22

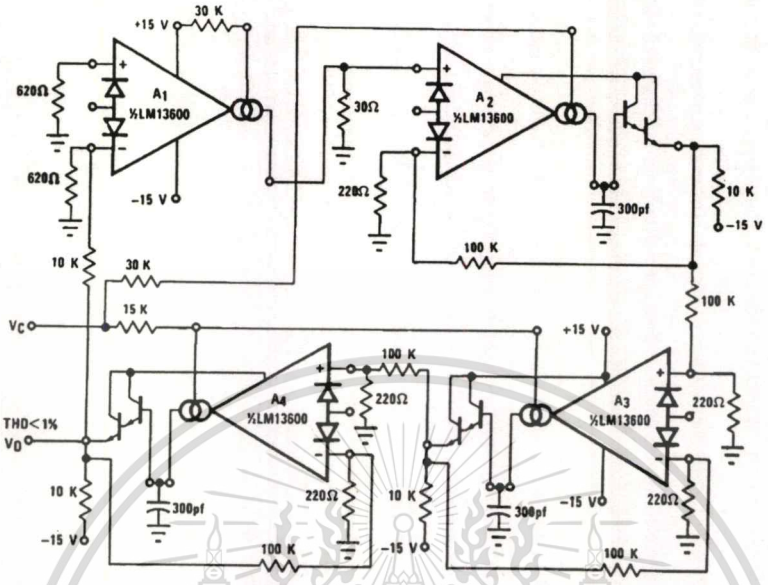
FIGURE 15. Triangular/Square-Wave VCO



TL/H/7980–23

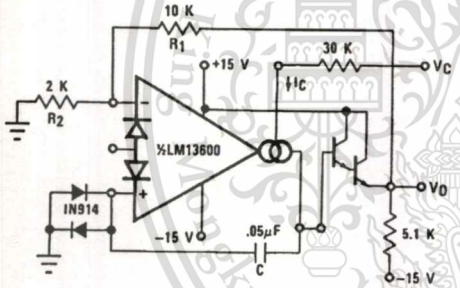
FIGURE 16. Ramp/Pulse VCO

Voltage Controlled Oscillators (Continued)



TL/H/7980-24

FIGURE 17. Sinusoidal VCO



TL/H/7980-25

FIGURE 18. Single Amplifier VCO

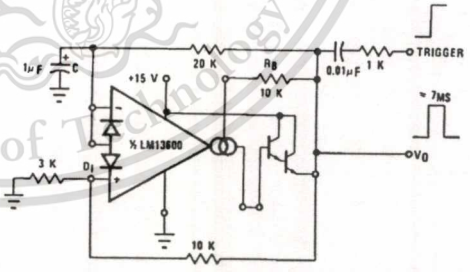
Figure 18 shows how to build a VCO using one amplifier when the other amplifier is needed for another function.

Additional Applications

Figure 19 presents an interesting one-shot which draws no power supply current until it is triggered. A positive-going trigger pulse of at least 2V amplitude turns on the amplifier through R_B and pulls the non-inverting input high. The amplifier regenerates and latches its output high until capacitor C charges to the voltage level on the non-inverting input. The output then switches low, turning off the amplifier and discharging the capacitor. The capacitor discharge rate is increased by shorting the diode bias pin to the inverting input so than an additional discharge current flows through D_I when the amplifier output switches low. A special feature of this timer is that the other amplifier, when biased from V_O, can perform another function and draw zero stand-by power as well.

The operation of the multiplexer of *Figure 20* is very straightforward. When A1 is turned on it holds V_O equal to V_{IN1} and when A2 is supplied with bias current then it controls V_O. C_C and R_C serve to stabilize the unity-gain configuration of amplifiers A1 and A2. The maximum clock rate is limited to about 200 kHz by the LM13600 slew rate into 150 pF when the (V_{IN1}-V_{IN2}) differential is at its maximum allowable value of 5V.

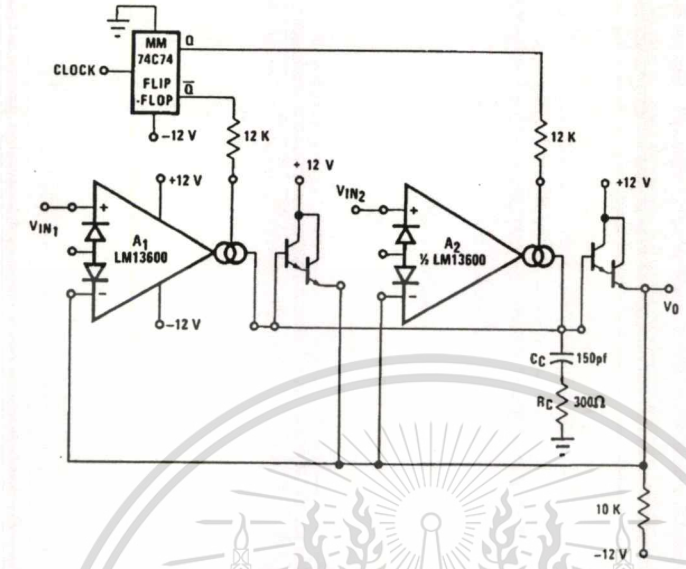
The Phase-Locked Loop of *Figure 21* uses the four-quadrant multiplier of *Figure 6* and the VCO of *Figure 18* to produce a PLL with a ±5% hold-in range and an input sensitivity of about 300 mV.



TL/H/7980-26

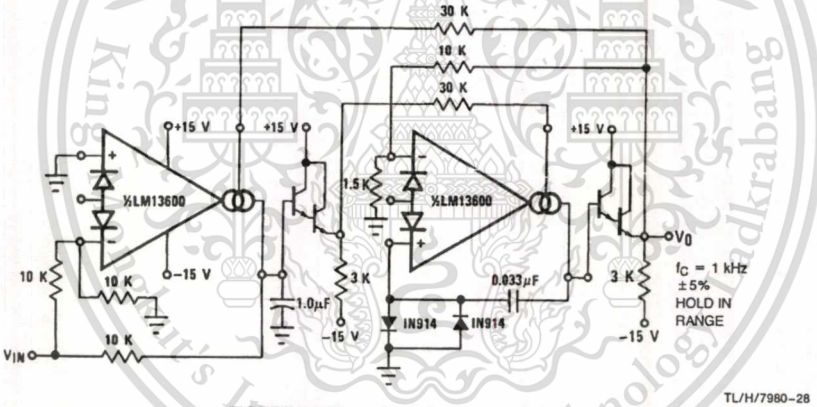
FIGURE 19. Zero Stand-By Power Timer

Additional Applications (Continued)



TL/H/7980–27

FIGURE 20. Multiplexer



TL/H/7980–28

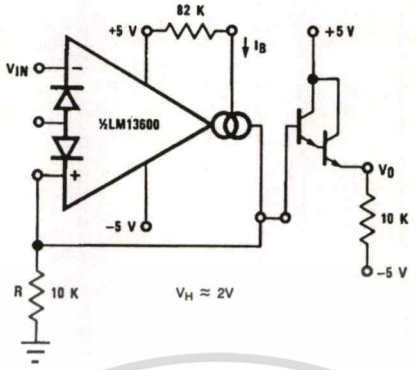
FIGURE 21. Phase Lock Loop

The Schmitt Trigger of *Figure 22* uses the amplifier output current into R to set the hysteresis of the comparator; thus $V_H = 2 \times R \times I_B$. Varying I_B will produce a Schmitt Trigger with variable hysteresis.

Figure 23 shows a Tachometer or Frequency-to-Voltage converter. Whenever A1 is toggled by a positive-going input, an amount of charge equal to $(V_H - V_L)\ C_t$ is sourced into C_f and R_f. This once-per-cycle charge is then balanced by the current of V_O/R_f. The maximum f_{IN} is limited by the amount of time required to charge C_t from V_L to V_H with a current of I_B, where V_L and V_H represent the maximum low and maximum high output voltage swing of the LM13600. D1 is added to provide a discharge path for C_t when A1 switches low.

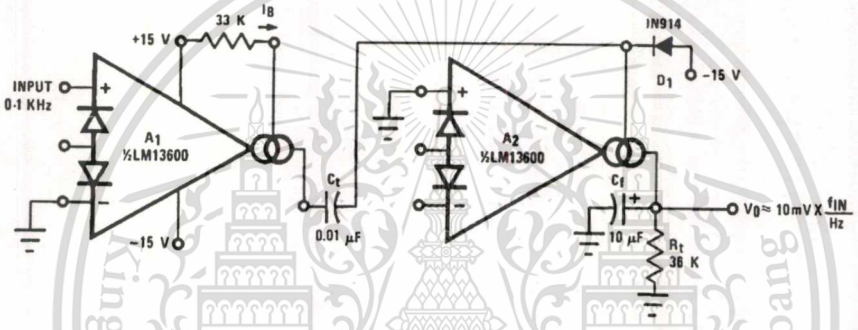
The Peak Detector of *Figure 24* uses A2 to turn on A1 whenever V_{IN} becomes more positive than V_O. A1 then charges storage capacitor C to hold V_O equal to $V_{IN}PK$. One precaution to observe when using this circuit: the Darlington transistor used must be on the same side of the package as A2 since the A1 Darlington will be turned on and off with A1. Pulling the output of A2 low through D1 serves to turn off A1 so that V_O remains constant.

Additional Applications (Continued)



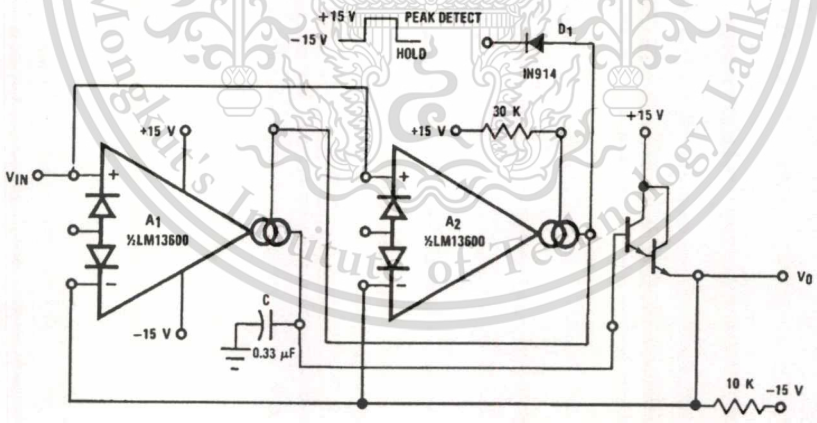
TL/H/7980–29

FIGURE 22. Schmitt Trigger



TL/H/7980–30

FIGURE 23. Tachometer



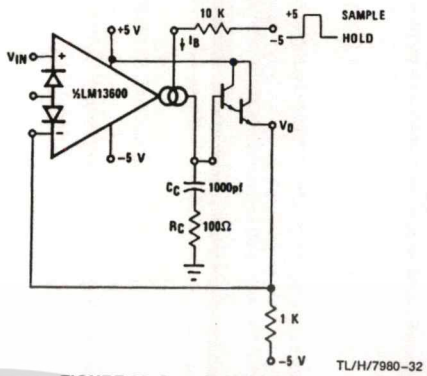
TL/H/7980–31

FIGURE 24. Peak Detector and Hold Circuit

Additional Applications (Continued)

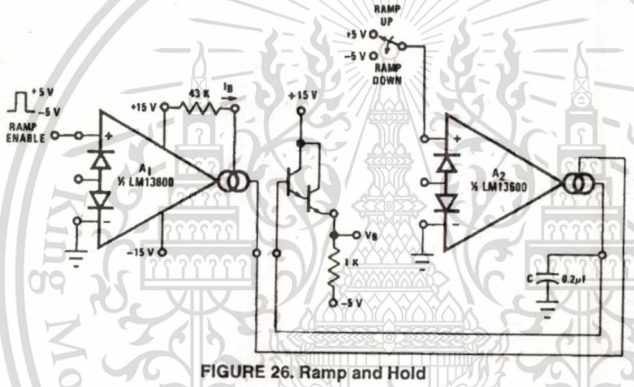
The Sample-Hold circuit of *Figure 25* also requires that the Darlington buffer used be from the other (A2) half of the package and that the corresponding amplifier be biased on continuously. The Ramp-and-Hold of *Figure 26* sources I_B into capacitor C whenever the input to A1 is brought high, giving a ramp-rate of about 1 V/ms for the component values shown.

The true-RMS converter of *Figure 27* is essentially an automatic gain control amplifier which adjusts its gain such that the AC power at the output of amplifier A1 is constant. The output power of amplifier A1 is monitored by squaring amplifier A2 and the average compared to a reference voltage with amplifier A3. The output of A3 provides bias current to the diodes of A1 to attenuate the input signal. Because the output power of A1 is held constant, the RMS value is constant and the attentuation is directly proportional to the RMS value of the input voltage. The attenuation is also proportional to the diode bias current. Amplifier A4 adjusts the ratio of currents through the diodes to be equal and therefore the voltage at the output of A4 is proportional to the RMS value of the input voltage. The calibration potentiometer is set such that V_O reads directly in RMS volts.



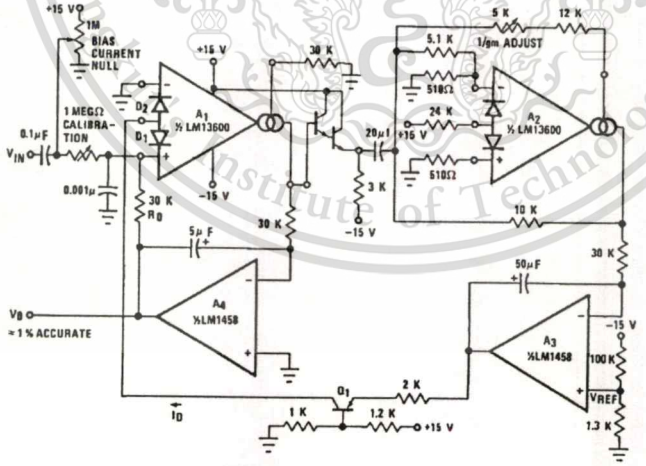
TL/H/7980–32

FIGURE 25. Sample-Hold Circuit



TL/H/7980–33

FIGURE 26. Ramp and Hold



TL/H/7980–34

FIGURE 27. True RMS Converter

Additional Applications (Continued)

The circuit of *Figure 28* is a voltage reference of variable temperature coefficient. The 100 kΩ potentiometer adjusts the output voltage which has a positive TC above 1.2V, zero TC at about 1.2V and negative TC below 1.2V. This is accomplished by balancing the TC of the A2 transfer function against the complementary TC of D1.

The log amplifier of *Figure 29* responds to the ratio of currents through buffer transistors Q3 and Q4. Zero temperature dependence for V_{OUT} is ensured because the TC of the A2 transfer function is equal and opposite to the TC of the logging transistors Q3 and Q4.

The wide dynamic range of the LM13600 allows easy control of the output pulse width in the Pulse Width Modulator of *Figure 30*.

For generating I_{ABC} over a range of 4 to 6 decades of current, the system of *Figure 31* provides a logarithmic current out for a linear voltage in.

Since the closed-loop configuration ensures that the input to A2 is held equal to 0V, the output current of A1 is equal to $I_3 = -V_C/R_C$.

The differential voltage between Q1 and Q2 is attenuated by the R1, R2 network so that A1 may be assumed to be operating within its linear range. From equation (5), the input voltage to A1 is:

$$V_{IN}1 = \frac{-2kTI_3}{qI_2} = \frac{2kTV_C}{qI_2R_C}$$

The voltage on the base of Q1 is then

$$V_B1 = \frac{(R_1+R_2)\,V_{IN}1}{R_1}$$

The ratio of the Q1 and Q2 collector currents is defined by:

$$V_B1 = \frac{kT}{q}\ln\frac{I_{C2}}{I_{C1}} \approx \frac{kT}{q}\ln\frac{I_{ABC}}{I_1}$$

Combining and solving for I_{ABC} yields:

$$I_{ABC} = I_1 \exp\left[\frac{2(R_1+R_2)\,V_C}{R_1I_2R_C}\right]$$

This logarithmic current can be used to bias the circuit of *Figure 4* provide a temperature independent stereo attenuation characteristic.

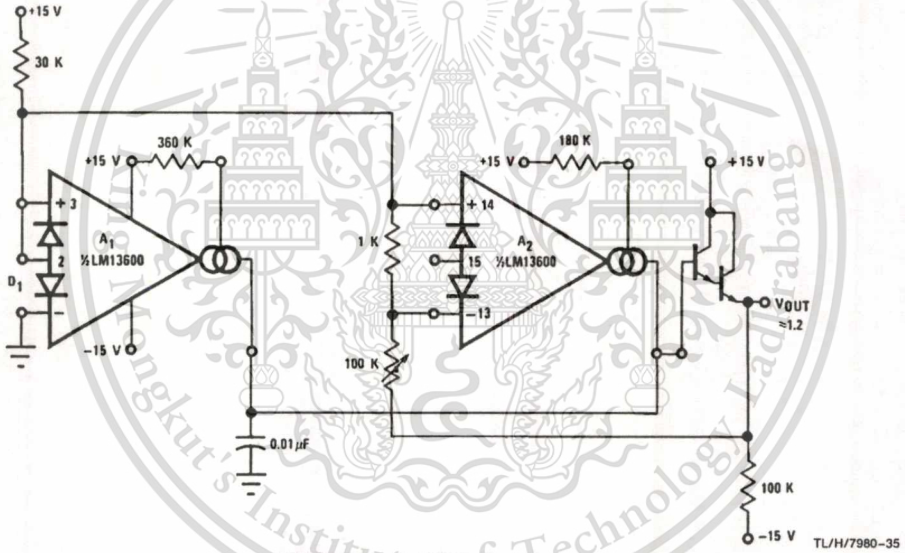


FIGURE 28. Delta VBE Reference

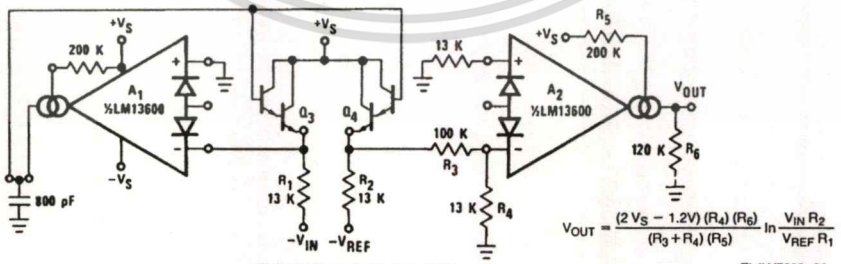
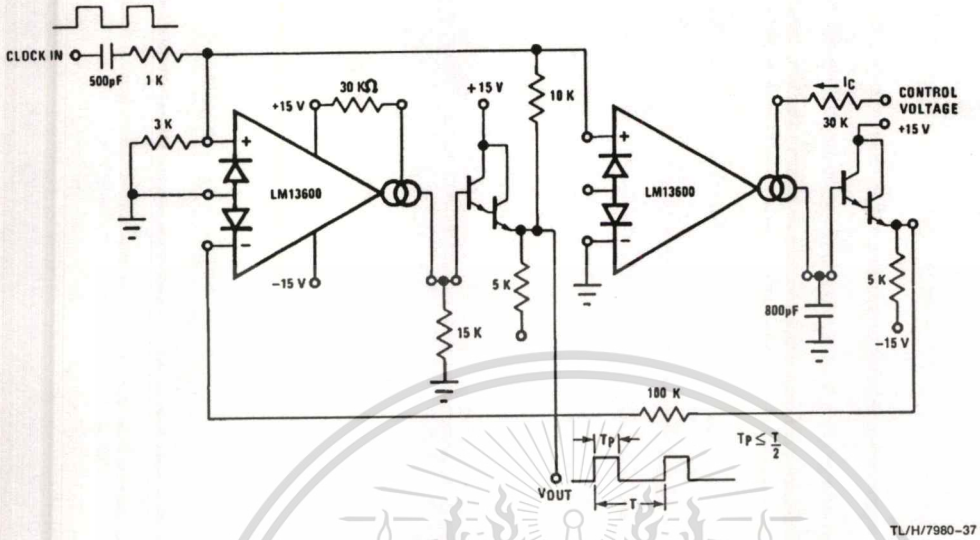


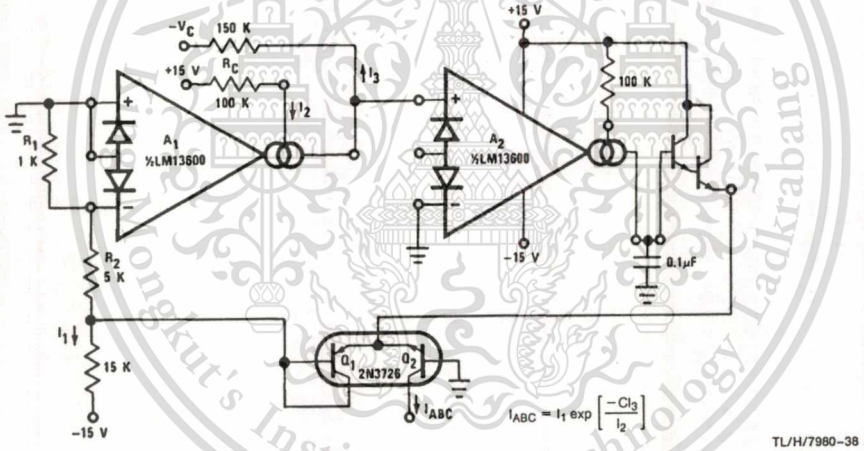
FIGURE 29. Log Amplifier

Additional Applications (Continued)



TL/H/7980–37

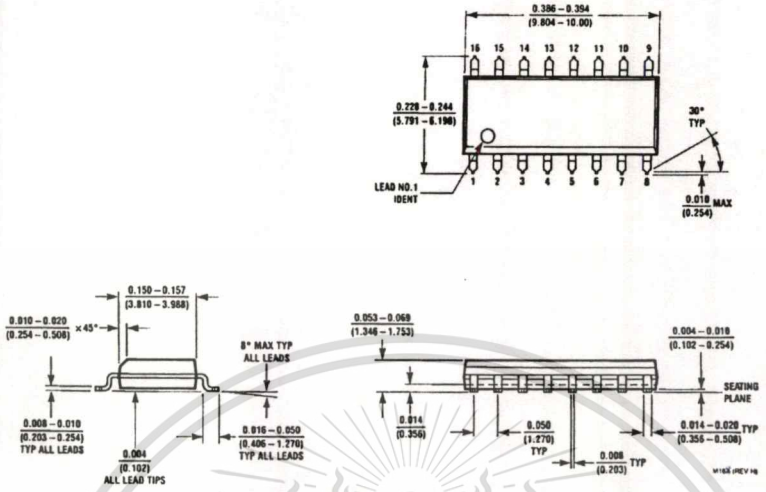
FIGURE 30. Pulse Width Modulator



TL/H/7980–38

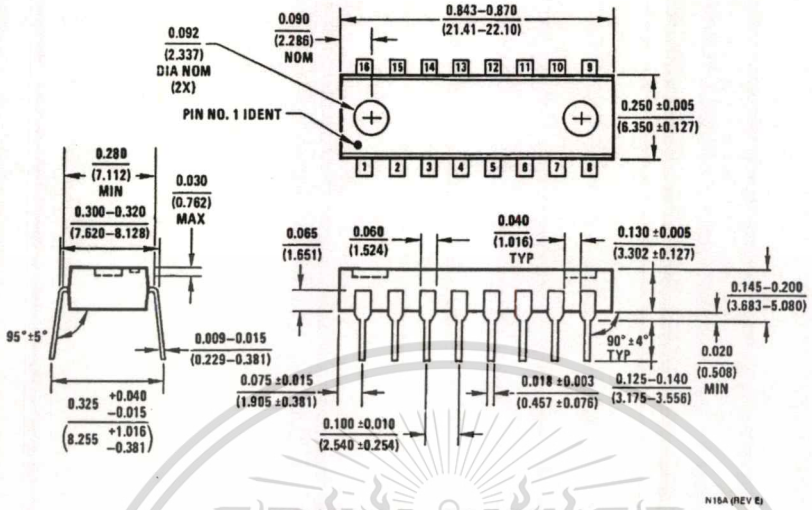
FIGURE 31. Logarithmic Current Source

Physical Dimensions inches (millimeters)



S.O. Package (M)
Order Number LM13600M
NS Package Number M16A

Physical Dimensions inches (millimeters) (Continued)



Molded Dual-In-Line Package (N)
Order Number LM13600N or LM13600AN
NS Package Number N16A

LIFE SUPPORT POLICY

NATIONAL'S PRODUCTS ARE NOT AUTHORIZED FOR USE AS CRITICAL COMPONENTS IN LIFE SUPPORT DEVICES OR SYSTEMS WITHOUT THE EXPRESS WRITTEN APPROVAL OF THE PRESIDENT OF NATIONAL SEMICONDUCTOR CORPORATION. As used herein:

1. Life support devices or systems are devices or systems which, (a) are intended for surgical implant into the body, or (b) support or sustain life, and whose failure to perform, when properly used in accordance with instructions for use provided in the labeling, can be reasonably expected to result in a significant injury to the user.
2. A critical component is any component of a life support device or system whose failure to perform can be reasonably expected to cause the failure of the life support device or system, or to affect its safety or effectiveness.

National Semiconductor Corporation
1111 West Bardin Road
Arlington, TX 76017
Tel: 1(800) 272-9959
Fax: 1(800) 737-7018

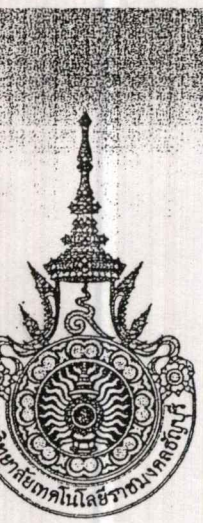
National Semiconductor Europe
Fax: (+49) 0-180-530 85 86
Email: cnjwge@tevm2.nsc.com
Deutsch Tel: (+49) 0-180-530 85 85
English Tel: (+49) 0-180-532 78 32
Français Tel: (+49) 0-180-532 93 58
Italiano Tel: (+49) 0-180-534 16 80

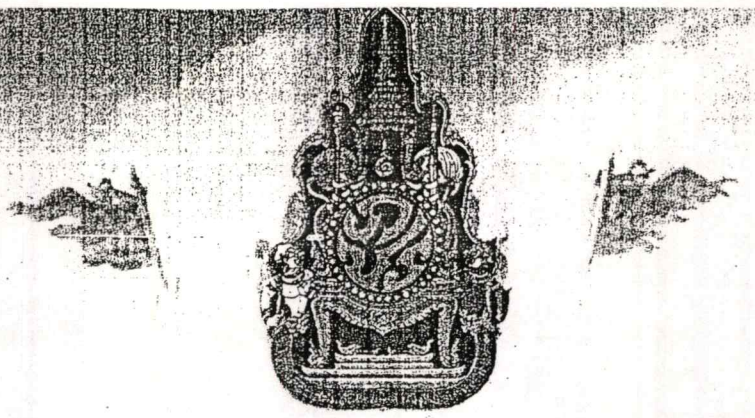
National Semiconductor Hong Kong Ltd.
13th Floor, Straight Block,
Ocean Centre, 5 Canton Rd.
Tsimshatsui, Kowloon
Hong Kong
Tel: (852) 2737-1600
Fax: (852) 2736-9960

National Semiconductor Japan Ltd.
Tel: 81-043-299-2309
Fax: 81-043-299-2408

Appendix D

Related Publication





The EECON-29 Conference Joins the Sixtieth Anniversary
Celebration of His Majesty's Accession to the Throne

29th Electrical Engineering Conference

การประชุมวิชาการทางวิศวกรรมไฟฟ้าครั้งที่ ๒๙

Volume II

ไฟฟ้าสื่อสาร (CM)
อิเล็กทรอนิกส์ (EL)
การประมวลผลสัญญาณดิจิตอล (DS)
ระบบควบคุมและการวัดคุม (CT)
งานวิจัยที่เกี่ยวข้องกับวิศวกรรมไฟฟ้า (GN)

๙-๑๐ พฤศจิกายน ๒๕๔๙

ณ โรงแรมแอมบาสซาเดอร์ ซิตี้ จอมเทียน พัทยา จังหวัดชลบุรี

ดำเนินการโดย ภาควิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

A Design of an OTA-based Sinusoidal Nonlinear Oscillator

Poramate Pranayanuntana and Pongrapee Kaewsaiha
Department of Control Engineering, Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
3 Moo 2 Chalongkrung Rd. Ladkrabang Bangkok 10520, Thailand
Phone 0-6667-3205, E-mail: kpporama@kmitl.ac.th, p_kaewsaiha@hotmail.com

Abstract

This paper describes a systematic approach to designing a sinusoidal nonlinear oscillator which uses an operational transconductance amplifier (OTA) as a nonlinear element to achieve the desired specifications on amplitude and frequency of oscillation. The describing function method is used for predicting the existence of limit cycles and, more generally, for analyzing the magnitude stabilization phenomena. The results are simulated using MAPLE and MATLAB SIMULINK.

Keywords: OTA, nonlinear oscillator, describing function

1. Introduction

A nonlinear oscillator is a nonlinear system that can display oscillation of fixed amplitude and fixed period without external excitation. The oscillations of this kind are called limit cycles, or self-excited oscillations. An advantage of nonlinear oscillator over the linear one is due to a magnitude stabilization phenomenon that keeps the amplitude of oscillation constant and not depending on the circuit's initial condition.

The describing function method [2, 3] is a first-order version of the method of harmonic balance that used to find periodic solutions for nonlinear systems by fitting a truncated Fourier series. This method applies to a system shown in Figure 1, with the linear system $G(s)$ having low-pass or band-pass filtering characteristics and $\psi(\cdot)$ as a time-invariant nonlinear element. Traditionally, the solution is found using graphical constructions involving the Nyquist locus. The presented design technic employs a numerical integration ability of mathematical softwares, for example, MAPLE or MATLAB to obtain the describing function $D(a)$ used in oscillator circuit design.

Also we give a systematic approach to synthesizing an OTA-based nonlinear oscillator given specifications on amplitude and period (or frequency) of oscillation. An equivalent feedback configuration of an oscillator circuit with a linear dynamic system $G(s)$ as a feedforward element and a memoryless nonlinearity $\psi(\cdot)$ as a feedback element, as shown in Figure 1, is used in the design.

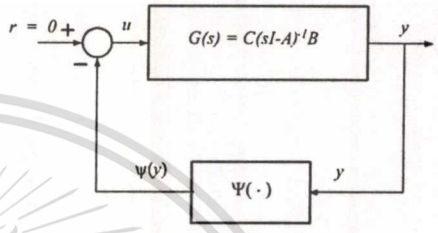


Figure 1. Feedback connection

2. Approach and Methods

The study needs to meet some basic requirements. In order to develop the basic version of the describing function method, the system has to satisfy the following assumption:

1. There is only a single nonlinear component
2. The nonlinear component is time-invariant
3. Corresponding to a sinusoidal input $y=a\sin(\omega_0 t)$ only the fundamental component $w_1(t)$ in the output $w(t) = \psi(a\sin(\omega_0 t))$ has to be considered
4. The nonlinearity is odd

The definition of the describing function, conditions for existence and stability of limit cycles, and nonlinearity of OTA are quoted. Then, a parameter calculation, a design example and simulation results are given.

2.1. Periodic Function

Consider a sinusoidal input to the nonlinear element, of amplitude a and frequency ω_0. The output of the nonlinear component $w(t) = \psi(y(t))$ in Figure 1 is always periodic: Corresponding to a sinusoidal input $y = a\sin\omega t$, the output becomes $\psi(a\sin(\omega_0(t + 2\pi/\omega_0))) = \psi(a\sin(\omega_0 t))$. Using Fourier series, this periodic function can be expanded as

$$w(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)]. \quad (1)$$

Since the set $\{\frac{1}{\sqrt{2}}, cos(\omega_0 t), sin(\omega_0 t), cos(2\omega_0 t), sin(2\omega_0 t), \ldots\}$ is an orthonormal basis with respect to the inner product $<f(t), g(t)> := \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)g(t)dt$,

CT15

The coefficients can be found as a_0=$<f(t), 1>$, a_n=$<f(t), cos(n\omega_0 t)>$, b_n=$<f(t), sin(n\omega_0 t)>$

Furthermore, since $G(s)$ used has band-pass or low-pass characteristics, this implies that only the fundamental component $w_1(t)$ is needed to be considered, namely

$$\begin{aligned} w(t) \approx w_1(t) &= a_1 \cos(\omega_0 t) + b_1 \sin(\omega_0 t) \\ &= M \sin(\omega_0 t + \phi), \end{aligned} \quad (2)$$

where $M(a, \omega_0) = \sqrt{a_1^2 + b_1^2}$ and $\phi(a, \omega_0) = \tan^{-1}(a_1/b_1)$, which indicate that the fundamental component corresponding to a sinusoidal input is a sinusoidal at the same frequency. In complex representation, this sinusoidal can be written as $w_1 = Me^{j(\omega_0 t + \phi)} = (b_1 + ja_1)e^{j\omega_0 t}$.

2.2. The Describing Function Method

The describing function of the nonlinear element is defined to be the complex ratio of the fundamental component of the output of the nonlinear element by the input sinusoid, i.e.,

$$D(a, \omega_0) = \frac{Me^{j(\omega_0 t + \phi)}}{ae^{j\omega_0 t}} = \frac{M}{a} e^{j\phi} = \frac{1}{a}(b_1 + ja_1). \quad (3)$$

For the case of single-valued nonlinearities, the describing function D is a function of the gain a only, i.e., $D(a, \omega_0) = D(a)$. By assumption 4, we obtain

$$D(a) = \frac{b_1}{a} = \frac{< w(t), sin(\omega_0 t) >}{a} \quad (4)$$

since $w(t)$ is odd function.

2.3. Existence of Limit Cycles

In order for a self-sustained oscillation of amplitude a and frequency ω_0 in the system of Figure 1 to exist, the variables in the loop must satisfy the following relations:

$$\begin{aligned} w &= D(a)y, \\ y &= -G(j\omega_0)w. \end{aligned}$$

Therefore, we have $y = -G(j\omega_0)D(a)y$. Because $y \neq 0$, this implies

$$G(j\omega_0)D(a) + 1 = 0 \quad (5)$$

which can be written as

$$G(j\omega_0) = -\frac{1}{D(a)}. \quad (6)$$

Plot of both the frequency response function $G(j\omega)$ (varying ω) and the negative inverse describing function $(-1/D(a))$ (varying a) in the complex plane are investigated. If the two curves intersect, then there exist limit cycles, and the values of a and ω corresponding to the intersection points are the solutions of (6). If the curves intersect n times, then the system has n possible limit cycles. Which one is actually reached depends on the initial conditions. Note that for single-valued nonlinearities, D, is real and therefore the plot of $(-1/D)$ always lies on the real axis.

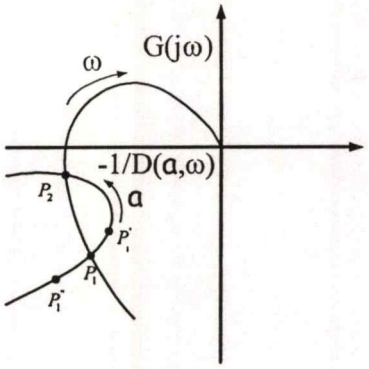


Figure 2. Detection of limit cycles

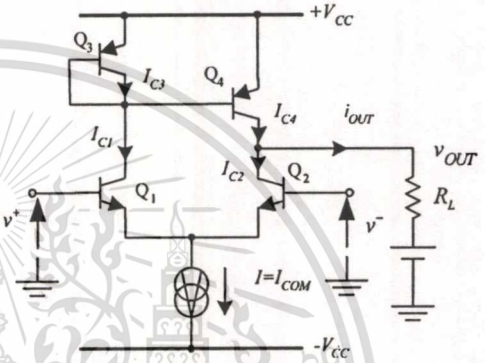


Figure 3. OTA equivalent circuit.

2.4. Stability of Limit Cycles

Each intersection point of the curve $G(j\omega)$ and the curve $(-1/D(a))$ corresponds to a limit cycle. If points near the intersection and along the increasing-a side of the curve $(-1/D(a))$ are not encircled by the curve $G(j\omega)$, then the corresponding limit cycle is stable. Otherwise, the limit cycle is unstable.

In Figure 2, there exist two limit cycles. Points P_1 and P_2 correspond to the two existing limit cycles of a system. P_1 corresponds to the unstable limit cycle, while P_2 refers to the stable one. Refer to [2, 3] for detailed treatment of the describing function method.

2.5. OTA Nonlinear Behavior

The simplified equivalent circuit diagram is given in Figure 3. According to [1] and [4], i_{OUT} of the OTA equivalent circuit in Figure 3 is

$$i_{OUT} = I\frac{\exp\left(\frac{qv}{kT}\right) - 1}{\exp\left(\frac{qv}{kT}\right) + 1} = I \tanh\left(\frac{v}{2V_T}\right) \quad (7)$$

where $v = v^+ - v^-$, and $V_T = \frac{kT}{q}$.

3. Calculation of the system parameters

For the linear system in Figure 1 having band-pass filter characteristics of the form

$$G(s) = \frac{-\omega_0 s}{s^2 + 2\alpha s + \omega_0^2} \quad (8)$$

CT15

where ω_0 is the resonant frequency in rad/sec and 2α is the bandwidth in rad/sec of the band-pass filter. If a self-excited oscillation exists the frequency of oscillation will be $\omega = \omega_0$. Therefore the condition of oscillation (condition for existence of limit cycle) in (6) becomes

$$\frac{1}{Q} = \frac{2\alpha}{\omega_0} = \frac{2}{a\pi}\int_0^{\pi} \psi(a\sin\theta)\sin\theta\, d\theta \qquad (9)$$

where Q is the quality factor of the band-pass filter. The far right-hand side of (9) is the describing function of the function $\psi(\cdot)$, and for a fixed $\psi(\cdot)$ it depends only on the magnitude of oscillation a, from (4), as follows:

$$D(a) := \frac{2}{a\pi}\int_0^{\pi} \psi(a\sin\theta)\sin\theta\, d\theta. \qquad (10)$$

For known values of α and ω_0, (9) is an equation in only one unknown variable, a, and with the aid of modern computers, one can solve (9) numerically for a or plot $D(a)$ against a. In the next section, we will give an example showing how to design an OTA-based sinusoidal nonlinear oscillator of specified amplitude and frequency of oscillation.

4. Example

Given the specifications on the frequency of oscillation $\omega_0 = 120$ rad/s and on the amplitude of oscillation $a = 0.3$ volts, then follows the following steps in designing an OTA-based sinusoidal nonlinear oscillator of the feedback configuration in Figure 1.

First, consider the OTA nonlinear behavior in (7). With $I = 1mA$, we have

$$i_{OUT} = 0.001\tanh(20v)$$

since $V_T \approx 25$ mV. Then we can realize $\psi(y) = \frac{1}{50}\tanh(20y)$ with an additional amplifier of gain 20 V/A at the output of the OTA.

From (10), plot $D(a)$ with respect to a, for $\psi(y) = \frac{1}{50}\tanh(20y)$, as shown below in Figure 4.

With $\omega_0 = 120$ rad/sec and $a = 0.3$ V, then $Q = \frac{1}{D(0.3)} = \frac{1}{0.08388617} = 11.92091526$, where $D(0.3) = 0.08388617$ was evaluated numerically from (10) by MAPLE program, so was the bandwidth $2\alpha = \omega_0/Q = 10.06634116$. Therefore the transfer function of our band-pass filter in (8) becomes

$$G(s) = \frac{-120s}{s^2 + 10.06634116s + 14400} \qquad (11)$$

The condition of oscillation in (6) can be written as

$$Q = \frac{\omega_0}{2\alpha} = -G(j\omega_0) = \frac{1}{D(a)} \qquad (12)$$

and in order for the limit cycle to exist it is required that

$$0 < \frac{2\alpha}{\omega_o} = \frac{1}{Q} < D_{max} = D(0), \qquad (13)$$

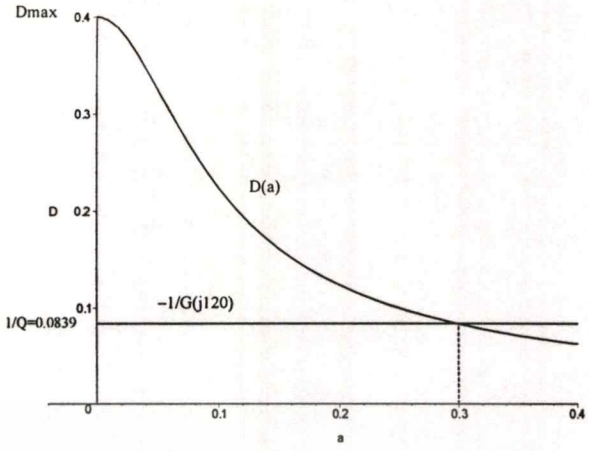


Figure 4. $D(a)$ for $\psi(y) = \frac{1}{50}\tanh(20y)$.

and in the case here, (13) is satisfied with Q high enough for $G(s)$ to be very selective.

Note: Condition (13) confirms an existence of a limit cycle which can also be seen as in the graph of $D(a)$ in Figure 4. If the desired value of amplitude of oscillation a yields a too low value of Q in (12) or if D_{max} is not greater than $1/Q$ in (13) then either an OTA circuit have to be redesigned or an amplifier may be required at the input or/and output of $\psi(\cdot)$.

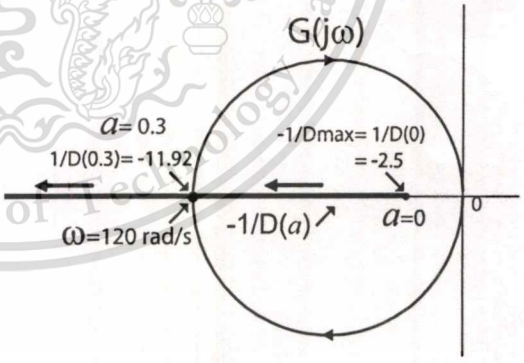


Figure 5. Plots of $G(j\omega)$ and $-1/D(a)$

The plot of $G(j\omega)$ and $-\frac{1}{D(a)}$ of the obtained sinusoidal nonlinear oscillator is shown in Figure 5. Both graphs in Figure 5 intersect at $\omega = 120$ rad/sec and $a = 0.3$ volts, at this point $G(j120) \approx -11.9209$. The corresponding limit cycle is stable since points near the intersection and along the increasing-a side of the curve $-1/D(a)$ are not encircled by the curve $G(j\omega)$. Simulation using MATLAB SIMULINK was done as follows:

CT15

5. MATLAB SIMULINK Simulation

The state and output equation of the transfer function $G(s)$ in (11) are as follow:

$$\dot{x} = \begin{bmatrix} -10.07 & 14400 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u \quad (14)$$

$$y = \begin{bmatrix} -120 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (15)$$

The phase plane plot of $x(t)$ is shown in Figure 6.

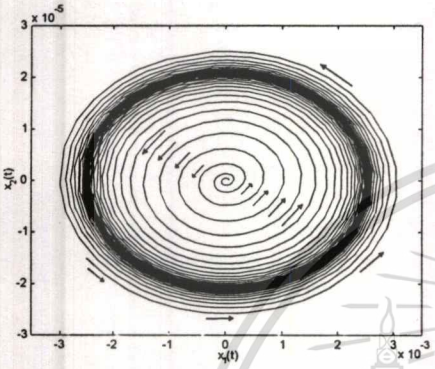


Figure 6. Phase plane plot of $x = [x_1, x_2]^T$.

The limit cycle in Figure 6 shows that $x_1(t)$ has amplitude of 0.0025 volts, therefore after multiplied by the gain of -120, the amplitude of $y(t)$ becomes 0.3 volts as expected.

The graphs of the output $y(t)$ for two different initial conditions are also plotted against time t. Figure 7 shows the graph of $y(t)$ with initial condition $y(0) = 0.25$, and Figure 8 shows the graph of $y(t)$ with initial condition $y(0) = -0.005$. Both graphs show the same frequency of oscillation $\omega_0 = 120$ rad/sec (about 23 cycles in 1.2 seconds) and also show the magnitude stabilization phenomenon of output amplitude of oscillation $a = 0.3$ volts. This feature make a sinusoidal nonlinear oscillator superior to the linear ones.

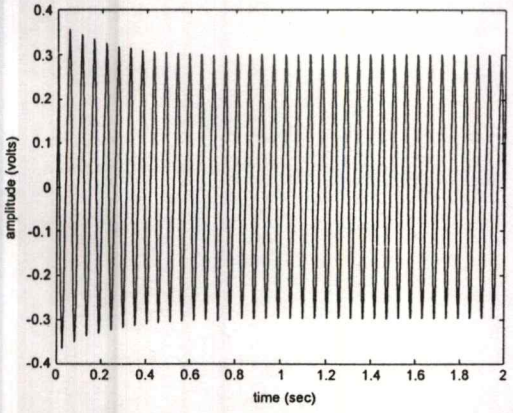


Figure 7. Plot of $y(t)$ with $y(0) = 0.25$.

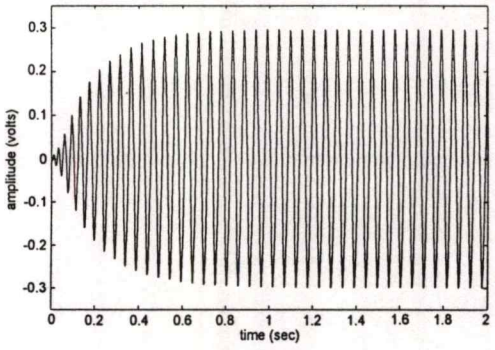


Figure 8. Plot of $y(t)$ with $y(0) = -0.005$.

6. Conclusion

For a high Q band-pass filter $G(s)$, the describing function method gives a very accurate forecast of the existence of oscillations and their stabilized amplitudes. This describing function method gives an insight into analyzing and designing a sinusoidal nonlinear oscillator circuit. Together with the desired frequency of oscillation ω_0 and conditions as (12) and (13) are concerned in choosing $G(s)$ that guarantees the existence of a limit cycle. The characteristics of the describing functions $D(a)$ of the nonlinear element $\psi(\cdot)$ in the system of Figure 1 gives information about stability of the limit cycle. The ability to plot $D(a)$ benefits in designing while the ability to solve $D(a)$ numerically is vital in the determination of the amplitude of oscillation, both were possible due to availability of mathematical softwares such as MAPLE, MATLAB, etc. The proposed design procedure guides us directly to what is needed to be done in order to get the designed circuit to oscillate at desired amplitude and frequency of oscillation.

References

[1] J. Bayard, M. Ayachi, "OTA- or CFOA-Based LC Sinusoidal Oscillators – Analysis of the Magnitude Stabilization Phenomenon," *IEEE Trans. on Circuits and Systems-1: Fundamental Theory and Applications*, Vol. 49, No. 8, August 2002.

[2] J.-J. E. Slotine, W. Li, *Applied Nonlinear Control*, Prentice-Hall, Inc. 1991.

[3] H. K. Khalil, *Nonlinear Systems*, New Jersey : Prentice-Hall, Inc. 2002.

[4] P. Pranayanuntana, K. Anuntahirunrat, C. Fongsamut, P. Kaewsaiha, "The Describing Function Method and the Analysis of the Magnitude Stabilization Phenomenon in a Nonlinear OSC," *The Proceedings of the International Symposium on Communications and Information Technologies 2005 (ISCIT 2005)*, Beijing, October 12-14, 2005.

[5] M. Stoll, *Introduction to Real Analysis*, Addison Wesley Longman, Inc. 2001.

CT15

BIOGRAPHY

Mr. Pongrapee Kaewsaiha was born in Nov 27, 1983 at Bangkok Thailand. He recieved the Bachelor's Degree of Engineering in Mechatronics Engineering, Faculty of Engineering, King Monkut's Institute of Technology Ladkrabang, 2004. His research related to nonlinear control systems.