TRANSIENT ANALYSIS

Transient circuit analysis is used to find the currents and voltages in a circuit containing one or more capacitors and/or inductors. The word "transient" describes a quantity that is fleeting rather than permanent, and it distinguishes this branch of circuit analysis from steady-state analysis, which is concerned with the long-term or settled behavior of a circuit. Transient circuit analysis asks not just "Where will my circuit end up?" but also "How will it get there?" The charging of a battery, the discharge of a flashbulb, and the oscillation of the pointer in a voltmeter about its resting point are all examples of transient behavior which can be analyzed using the techniques of transient circuit analysis.

Before proceeding, it should be noted that the circuits to be considered in this article in fact form only a subset of the universe of circuits—they are all linear, time-invariant, and lumped. A linear circuit is one in which each element (except the independent sources that drive the circuit) is described by one or more linear equations involving its current(s) and voltage(s). For example, the resistor defined by $v = Ri$ is linear, but the diode defined by $i = I_s(e^{v/V_T} - 1)$ is nonlinear, and any circuit containing the diode is therefore nonlinear. Nonlinear circuits can exhibit highly complex behavior and cannot be handled by the techniques described in this article. A time-invariant circuit is one in which the equations defining the elements (except the independent sources) do not change with time. A lumped circuit is one which is small enough that all electromagnetic waves in the circuit propagate virtually instantaneously through the circuit, and the behavior of the circuit is unaffected by physical distances between elements. Circuits that are not lumped are handled by a special branch of circuit theory known as distributed circuit theory or transmission line theory. We will assume throughout this article that all circuits under consideration are linear, time-invariant, and lumped.

The equations describing a circuit arise from two sources: Kirchhoff's laws tell us how the elements in the circuit are interconnected, and then each element in the circuit has an individual equation (or equations) describing its behavior. If all of the circuit elements are described by algebraic equations (i.e., ones in which no derivatives appear) involving their currents and voltages, these equations can be combined with Kirchhoff's equations to give a set of algebraic equations that completely describe the circuit. These equations are linear equations in terms of the currents and/or voltages in the circuit, and they can be solved by any of the techniques of linear algebra. The power of linear algebra means that these circuits, known as *resistive circuits,* are (relatively) easy to analyze. The behavior of these circuits is quite simple: If a linear resistive circuit is driven by a 1 V battery, then changing to a 2 V battery will cause all voltages and currents in the circuit to double. There is no time delay in this response: The doubling of voltages and currents occurs at the precise instant when the 2 V battery is inserted into the circuit. If the battery is replaced by a more complicated voltage source which varies with time, each voltage and current in the circuit will also vary with time as a scaled replica of the new voltage source.

Although easy to analyze, the limited behavior of a linear resistive circuit means that such circuits are not very useful. Instead of producing a scaled replica of the signal that drives them, most circuits are required to convert a signal into a more useful form. For example, a radio receiver can receive a jumbled signal containing contributions from the myriad of stations that inhabit the airwaves and tune into a single one; the graphic equaliser on a stereo system can change the sound quality by boosting or diminishing certain frequencies; and an ignition circuit in a car is driven by a battery, but its output is a short sharp spark. These effects rely on the use of capacitors and/or inductors. These circuit elements are defined by equations involving not just their currents and voltages, but the rate of change (or derivative) of these quantities with time. Specifically, the current through a capacitor is proportional to the derivative of its voltage with respect to time, and the voltage across an inductor is proportional to the derivative of its current with respect to time. Capacitors and inductors are known as *dynamic circuit elements*, to convey the importance to them of time variation, or energy storage elements, since they are capable of storing energy for later release. Dynamic elements can be placed deliberately in a circuit, or they can be unwanted parasitic elements, modeling for example the capacitance between wires in the circuit. If a circuit contains even a single dynamic element, it is in general described no longer by a set of algebraic equations, but by one or more differential equations in which the variables are not only the voltages and currents but also the derivatives of certain of these quantities with respect to time. A dynamic circuit is one which contains at least one dynamic element. The goal of transient circuit analysis is to solve the differential equations that describe a dynamic circuit and thus to come up with expressions predicting the way in which the voltages and currents in the circuit will vary with time. It is concerned in particular with the response of the circuit to changes, such as when a source is inserted, removed, or suddenly changed in some way, or a switch is closed and the make-up of the circuit is thereby changed.

Dynamic circuits can exhibit more interesting behavior than resistive circuits, but they are also more difficult to analyze. One of the simplest dynamic circuits contains a single capacitor in series with a resistor and a constant voltage source which is switched on at some specified time. This circuit is described by an equation involving the capacitor voltage v_C and its derivative with respect to time dv_C/dt. The absence of any higher derivatives gives this equation the description "first-order." A circuit containing just a single dynamic element is described by a first-order differential equation and is called a *first-order circuit.*

The solution of a first-order differential equation will contain an unknown constant. To find this constant, it is necessary to apply some additional information about the value of the solution at a specified time instant. Since in general we are concerned with finding the response of the circuit to changes that occur at a certain time instant, we often know the state of the circuit just before the change occurs and can apply this information in order to find the unknown constant. The value of the capacitor voltage (or inductor current, if the circuit contains an inductor rather than a capacitor) just before the change occurs is known as the *initial condition*.

Solving the first-order circuit just described yields the result that the capacitor voltage plotted as a function of time is of exponential form, moving from its initial value toward the value of the constant voltage source and eventually settling there. (Certain assumptions have been made here and are discussed in the next section.) This is intuitively plausible—once the voltage source has been inserted the resistor voltage and capacitor voltage must sum to equal the voltage of the source. If the capacitor voltage does not initially equal that of the source, the voltage difference must be developed across the resistor by a current flowing through it. This current charges the capacitor, bringing its voltage closer to that of the source, and the net effect is to cause the capacitor voltage to approach that of the voltage source. This circuit is reminiscent of a simple battery charger, with the battery voltage increasing over time to equal that of a source.

Already in this simple circuit we can see how dynamic circuits behave in ways that would be impossible for a resistive circuit. If the circuit described above had been resistive, all voltages and currents would have been scaled versions of the source. In this circuit, however, the capacitor voltage takes on a form quite unlike that of the source: It varies exponentially with time, whereas the source is constant. The action of the resistor and capacitor has processed the source signal, with the capacitor voltage resisting the sudden change when the source was inserted, but retaining the steady behavior of the source. The resistor voltage, on the other hand, captures the change in the source very well, but eventually dies away to include nothing of the steady behavior of the input. This behavior is an example of the filtering behavior of this simple resistor–capacitor combination, which is useful in a variety of communications applications.

The exponential nature of the voltage observed in this simple circuit is not unusual; in fact, as we shall see, exponential functions appear in various guises in the solution to linear differential equations. Possibly the most widely known example of an exponential function appears in the analysis of radioactive decay, where the rate of decay of a substance is proportional to the amount of the substance present, and so the amount remaining decays exponentially to zero at a rate depending on the half-life of the substance.

In general, a circuit which contains two dynamic elements gives rise to a second-order differential equation (containing the second derivative of the variable with respect to time) and is termed a *second-order circuit.* If all sources in the circuit are dc (constant) sources, this equation can be solved by application of standard theory of linear differential equations, with the aid of two initial conditions, one for each dynamic element. Instead of the single exponential transient of the first-order circuit, this circuit contains two exponential transients which are added to give the overall transient. The relationship (via complex numbers) between the exponential and sinusoidal functions can give rise to a new type of behavior arising from these transients. If the arguments of these exponentials are complex, as may turn out to be the case, then they can be added to give a transient which oscillates sinusoidally. In most circuits the magnitude of this oscillation decays exponentially with time. A common example of such a decaying oscillation is produced when a tuning fork is struck or a child's swing given a single push. If there are no losses in the circuit (not a practical requirement), the oscillation could persist indefinitely without decaying; and if the circuit is unstable, it is possible that the oscillation can actually grow.

While it is possible to analyze simple first- and second-order dynamic circuits by applying standard theory of differential equations, such solution becomes rapidly more difficult when the order of the circuit increases or when the sources become more complicated. When faced with such a problem, one might look enviously back at the much simpler process of solving a resistive circuit. In fact it is possible to apply techniques of resistive circuit analysis to dynamic circuits with the aid of a variety of transforms. A transform is a method of changing a problem into a different form, solving it in the new form (where the solution is easier to obtain) and then changing the solution back to the original form. For example, a student unfamiliar with binary arithmetic, when asked to add two binary numbers, might convert the numbers to decimal form (presumably with the aid of a table), add the decimal numbers, and then convert back to binary. The transforms to be applied in this context change a system of differential equations to a system of algebraic equations which are significantly easier to solve.

The most important and most widely used of these transforms in circuit analysis is the Laplace transform. A second transform, the Fourier transform, is particularly useful in analyzing circuits designed for applications in communication systems. These transforms convert a set of differential equations involving the time variable into a set of algebraic equations involving a new variable called the frequency (in the case of the Fourier transform) or the complex frequency (in the case of the Laplace transform). Application of these transforms allows us to analyze a circuit by transforming it into an equivalent form in the frequency domain, where its equations are purely algebraic, analyzing the circuit in this frequency domain using the techniques of linear algebra, and then applying the transform in reverse to convert the result of this analysis into a function of time.

Once again, Laplace transform analysis shows up the special role of the exponential function (and its complex cousin the sinusoid) in the behavior of circuits. Every dynamic circuit favors certain exponential (including sinusoidal) modes of behavior whose rate of decay (and frequency of oscillation, if applicable) is governed by the so-called natural frequencies of the circuit. These natural frequencies tell us whether the currents and voltages in a circuit will, of their own accord, tend to exhibit exponential or oscillatory decay, constant behavior or steady oscillation, exponential or oscillatory growth, or some combination thereof. When an input signal is applied to the circuit, the currents and voltages may contain components controlled by the natural frequencies as well as a component controlled by the input. In practical circuits it is desirable that the output should depend on the input; and the prospect of an oscillation or exponential growing in the circuit, swamping out the effect of the input and wreaking havoc with the circuit components, is clearly a designer's nightmare. This effect is similar to that demonstrated by sound systems when a microphone is placed in the path of a loudspeaker and an unwanted tone appears and swamps the desired signal. Fortunately, Laplace transform techniques allow us to analyze a system to determine if this effect is possible. An asymptotically stable system is one in which all exponential transients die away, leaving only the effect of the input signal.

The effects of transients are seen in a huge range of electronic and electrical engineering applications, from the transmission of tiny pulses between parts of a communication system to the behavior of an electrical network struck by lightning. The techniques described in this article provide the reader with the ability to understand and analyze transient behavior in a wide variety of circuits.

TIME-DOMAIN ANALYSIS

Natural Response and Step Response of a First-order Circuit

Consider the circuit shown in Fig. 1(a). Until the time $t = 0$, the switch S is in position 1, and the resistor R and capacitor C are connected in a loop. At time $t = 0$ the switch is moved to position 2, connecting the dc voltage source E in series with R and C. We assume that the switch closes instantaneously and that it presents a short circuit between the terminals which it connects. Mathematically, we say that the voltage

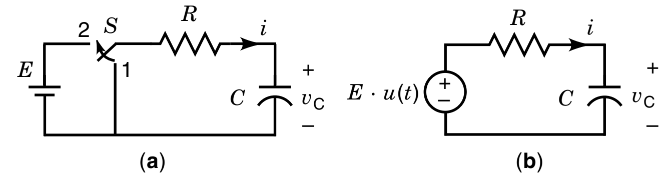

Figure 1. (a) The switch S moves from position 1 to position 2 at time $t = 0$, so the voltage applied to the RC series combination is 0 for $t < 0$ and E thereafter. The switch–voltage-source combination is represented in (b) by the single voltage source $E \cdot u(t)$.

applied to the RC series combination is $E \cdot u(t)$, where $u(t)$ is the unit step function given by

$$u(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ 1 & \text{for} \quad t \geq 0 \end{cases} \tag{1}$$

The circuit of Fig. 1(a) can, therefore, also be drawn in the form shown in Fig. 1(b).

The analysis of this circuit for $t \geq 0$ will require knowledge of the initial voltage across the capacitor just after the switch is thrown—that is, $v_C(0^+)$, where $0^+ = \lim_{\epsilon \to 0} \epsilon$. We generally know, or can find from analysis of a previous regime, $v_C(0^-)$, the voltage at the instant just before the switch is thrown ($0^- = \lim_{\epsilon \to 0} -\epsilon$). If the capacitor current is finite, $v_C(0^+)$ must equal $v_C(0^-)$, and we can refer to both as $v_C(0)$. Similarly if the voltage across an inductor is finite its current waveform must be continuous. We will assume these continuity conditions throughout this analysis. The alternative case, where the capacitor current or inductor voltage can be infinite, is not practical but turns out to be mathematically interesting and useful in analysis. It can be handled by an extension of our analysis in this section (see Ref. 1 for details), but we will postpone consideration of this possibility until the later section on the Laplace transform where it can be handled more conveniently.

For $t \geq 0$, Kirchhoff's voltage law gives the equation

$$v_C(t) + i(t)R = E$$

or, applying the constitutive relation $i(t) = C\, dv_C(t)/dt$ for the capacitor,

$$RC\frac{dv_C(t)}{dt} + v_C(t) = E \tag{2}$$

This is a first-order differential equation in the capacitor voltage v_C, and so this circuit is referred to as a first-order circuit. It can be solved by a number of methods to give an expression for v_C as a function of time. One such method is to recast the equation in the form

$$\frac{d(v_C(t) - E)}{dt} = -\frac{1}{RC}(v_C(t) - E)$$

This equation is of the familiar form

$$\frac{dx(t)}{dt} = ax(t)$$

which has the solution (see Ref. 2)

$$x(t) = x(0)e^{at}$$

where $x(0)$ is the value of x at time $t = 0$. This initial condition must be known if the equation is to be solved for $x(t)$. Thus Eq. (2) has the solution

$$v_C(t) - E = (v_C(0) - E)e^{-t/RC} \tag{3a}$$

or

$$v_C(t) = v_C(0)e^{-t/RC} + E(1 - e^{-t/RC}) \tag{3b}$$

The response of the series RC circuit with zero initial capacitor voltage to the application of a voltage source given by the unit step function is known as the *step response* of the series RC circuit. (Note that we will use the word "response" to signify any current or voltage in the circuit, or any set thereof, including for example the set of all currents and voltages. Throughout this article the variable or variables which constitute the response in any given instance will be clear from the context in which the word is used.)

It is clear from Eq. (3a) that the difference between v_C and E varies exponentially with time, and when the product RC is positive (a condition that will be assumed to hold unless otherwise stated) this difference tends to zero as t tends to infinity. v_C is plotted as a function of time in Fig. 2, where, as expected, v_C is seen to converge exponentially to E. The rate of this convergence is governed by the value of RC, which is termed the time constant of the waveform and denoted by the symbol τ. The smaller the time constant, the faster the rate of convergence. After one time constant has elapsed (i.e., at $t = \tau$), $v_C(t) - E$ has decreased to $e^{-1} = 36.8\%$ of its value at $t = 0$, and at time $t = 5\tau$ this difference has decreased to $e^{-5} = 0.7\%$ of its initial value. Although v_C does not reach E within any finite time (unless, of course, it started out at E), after five time constants have elapsed the difference between v_C and E has been reduced to less than 1% of its initial value. The time constant is a useful measure of the response speed of a first-order circuit. For more general circuits, the *rise time* is used as a measure of response time. This is defined as the time taken for the step response to rise from 10% to 90% of the steady-state value. For the first-order circuit analyzed in this section, the rise time can be found to be $\tau \ln 9 \approx 2.2\tau$.

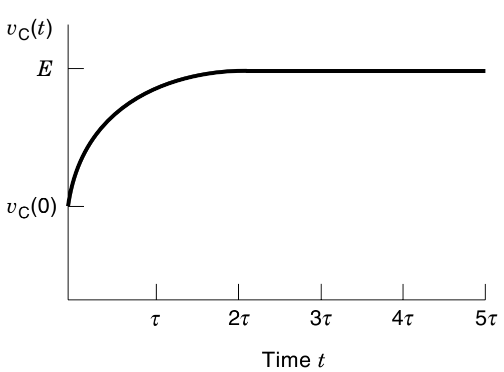

Figure 2. The capacitor voltage in the circuit of Fig. 1 varies exponentially from its starting value $v_C(0)$ to its steady-state value E, with time constant $\tau = RC$.

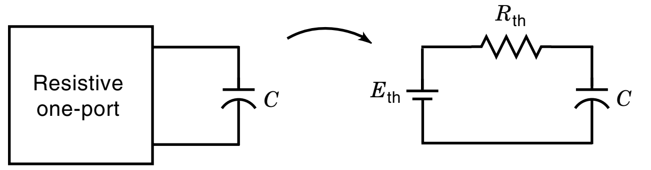

Figure 3. A circuit consisting of a single capacitor in an otherwise resistive circuit is simplified by replacing the resistive one-port seen by the capacitor by its Thévenin equivalent.

The value E to which the capacitor voltage converges is termed the steady-state value of this voltage. It is the only value of capacitor voltage at which the circuit can settle; in other words, it is only when $v_C = E$ that all currents and voltages in the circuit cease to vary with time. Clearly, when a quantity ceases to vary with time, its derivative with respect to time is zero, and so the steady-state of value of v_C can be found directly from the differential equation (2) by setting the term dv_C/dt to zero (or, in circuit terms, replacing the capacitor by an open circuit), yielding the equation $v_C = E$, as expected. The overall waveform $v_C(t)$ is the sum of this steady-state component and a second component which dies away with time. This second component is known as the transient component (or just the transient). The exponential form of the transient in this circuit is, as we will see later, particularly common in linear circuits and other linear systems.

Note, however, that the procedure just outlined yields the value of v_C at which the circuit variables (currents and voltages) can remain constant, but it does not guarantee that the circuit will actually converge to this state. For example, if $RC < 0$, Eq. (3a) implies that v_C will diverge exponentially away from E and the circuit has no steady-state response. (The only exception to this divergence is when $v_C(0) = E$, in which case it will theoretically remain fixed at E for all time. The word "theoretically" is important: In practice, any noise in the circuit that causes v_C to differ even infinitesimally from E will result in its diverging exponentially from E.) This distinction relates to the issue of the stability of equilibria of differential equations (2).

Another useful view of the solution waveform (3b) for $v_C(t)$ is that it is composed of two components: one caused by the initial condition $v_C(0)$, and the other caused by the voltage source E. If $E = 0$ the response (3) reduces to $v_C(t) = v_C(0)e^{-t/RC}$, which is termed the natural or unforced response of the circuit. This is a viewpoint to which we will return later.

Any circuit consisting of a single capacitor in an otherwise resistive circuit containing only dc sources is generally analyzed by transforming it to single-loop form by means of a Thévenin transformation (3), as shown in Fig. 3. The analysis described above is then applicable, where E is the Thévenin equivalent voltage source and R is the Thévenin equivalent resistance. (The small number of circuits that do not have a Thévenin equivalent can be handled separately.)

Before leaving the single-loop first-order circuit of Fig. 1 we note that the analysis of this section can be used to find the response of a first-order circuit to a voltage source that is piecewise-constant—that is, constant over certain time intervals with discontinuous jumps between these constant levels. One important such waveform is the pulse

$$p(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ E & \text{for} \quad 0 \le t < t_0 \\ 0 & \text{for} \quad t \ge t_0 \end{cases}$$

The response of the first-order RC circuit to this source waveform is found by an extension of the analysis just performed. For $0 \le t < t_0$ the analysis proceeds as before and $v_C(t)$ is given by Eq. (3c):

$$v_C(t) = v_C(0)e^{-t/RC} + E(1 - e^{t/RC}) \qquad \text{for} \quad 0 \le t < \quad t_0 \qquad (3c)$$

For $t \ge t_0$ the response is just the natural response found previously, the only difference being that since this phase of the analysis commences at $t = t_0$ instead of $t = 0$ the initial condition is $v_C(t_0)$ instead of $v_C(0)$. Applying this initial condition in the usual way, we find that

$$v_C(t) = v_C(t_0)e^{-(t-t_0)/RC} \qquad \text{for} \quad t \ge t_0 \qquad (4)$$

$v_C(t_0)$ is, by our assumption of bounded currents, equal to $v_C(t_0^-)$, the capacitor voltage just before the source waveform drops to zero. Since Eq. (3b) gives $v_C(t)$ for all times in the range $0 \le t < t_0$, it can be used to find that

$$v_C(t_0^-) = v_C(0)e^{-t_0/RC} + E(1 - e^{-t_0/RC})$$

Substituting this value for $v_C(t_0)$ in Eq. (4) completes the analysis of the response of the series RC circuit to the voltage pulse. This response is plotted in Fig. 4, for two different values of the time constant. The response of a circuit to a pulse is particularly important in communication systems where such pulses are used to carry information and must be clearly identifiable at the receiver. An RC combination of the type studied here often occurs in such transmission systems, formed by the output resistance of the part of the system where the signal originates and the input capacitance of the part of the system into which the signal is fed, and thus exponential distortion

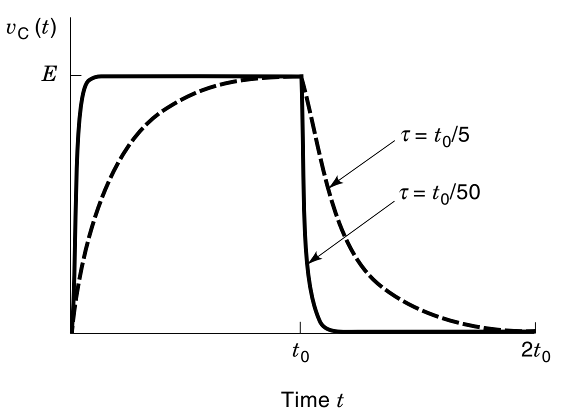

Figure 4. The response of a first-order RC circuit to a voltage pulse of amplitude E and duration t_0. The solid line shows the response if $\tau = t_0/50$, and the dashed line shows the response if $\tau = t_0/5$. Note the "smearing" of the pulse when τ is large.

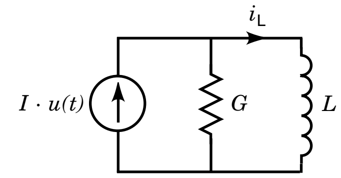

Figure 5. First-order circuit consisting of the parallel combination of current source $I \cdot u(t)$, conductance G, and inductor L.

will inevitably ensue. Clearly the "smearing" of the pulse evident in Fig. 4 when the time constant is large limits the rate at which pulses can be transmitted if they are to be separated at the receiver.

The response of the series RC circuit to any piecewise-constant source waveform is found by an extension of the analysis performed above. The circuit is analyzed using the standard method over each of the time intervals in which the source is constant, starting with the first time interval. The initial condition for the nth time interval, commencing at time $t = t_n$, is found by evaluating the response from the previous time interval at time $t = t_n^-$.

The second type of first-order circuit is one in which the single energy storage element in the circuit is an inductor rather than a capacitor and, by application of a Norton transformation (where possible), is of the form shown in Fig. 5, where the constant current source I is connected in parallel with conductance G and inductor L for $t \ge 0$. Kirchhoff's current law applied to this circuit gives the following differential equation in the inductor current i_L for $t \ge 0$:

$$GL\frac{di_L(t)}{dt} + i_L(t) = I$$

which can be solved as before to find

$$i_L(t) - I = (i_L(0) - I)e^{-t/GL}$$

or

$$i_L(t) = i_L(0)e^{-t/GL} + I(1 - e^{-t/GL})$$

Thus the inductor current waveform for the circuit of Fig. 5 takes the same form as the capacitor voltage waveform for the circuit of Fig. 1, with time constant GL and steady-state value I. This is a consequence of the fact that the circuit of Fig. 5 is the dual of that of Fig. 1. The response to a piecewise-constant source waveform can be found by applying the method previously described for the series RC circuit.

Natural Response of a Second-order Circuit

The circuit in Fig. 6 consists of a resistor and two energy storage elements—a capacitor and an inductor. Kirchhoff's voltage law gives the equation

$$v_C(t) + L\frac{di_L(t)}{dt} + Ri_L(t) = 0$$

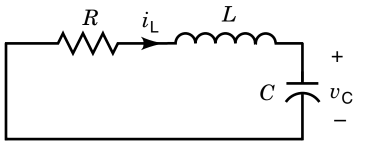

Figure 6. Second-order circuit consisting of resistor R, capacitor C, and inductor L.

which on application of the relation $i_L(t) = C\,dv_C(t)/dt$ becomes

$$LC\,\frac{d^2v_C(t)}{dt^2} + RC\frac{dv_C(t)}{dt} + v_C(t) = 0 \tag{5}$$

This is a second-order differential equation, and so the circuit is termed a second-order circuit. The exponential waveform

$$v_C(t) = Ae^{st}$$

is a solution to Eq. (5) provided that

$$LCs^2 + RCs + 1 = 0$$

which yields

$$s = \frac{-R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

If these two values, s_1 and s_2, are distinct (i.e., $s_1 \neq s_2$), then the general solution of Eq. (5) is of the form

$$v_C(t) = A_1e^{s_1t} + A_2e^{s_2t} \tag{6}$$

Since there are no sources in the circuit, this is the natural or unforced response of the series RLC circuit. The constants A_1 and A_2 will be determined by applying the initial conditions $v_C(0)$ and $i_L(0)$ and solving the resulting simultaneous equations:

$$v_C(0) = A_1 + A_2$$

$$i_L(0) = C\frac{dv_C}{dt}\bigg|_{t=0} = CA_1s_1 + CA_2s_2$$

We will now consider the nature of the natural or unforced voltage waveform represented by Eq. (6). We will use the following shorthand form for s_1 and s_2:

$$s_1 = -\alpha + \sqrt{\alpha^2 - \omega_0^2} \quad \text{and} \quad s_2 = -\alpha - \sqrt{\alpha^2 - \omega_0^2}$$

where

$$\alpha = \frac{R}{2L}$$

and

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

We will assume for now that $\alpha \geq 0$.

The first case to be considered is the case where $\omega_0^2 < \alpha^2$ and s_1 and s_2 are real and distinct. In this case the circuit is said to be *overdamped* and the response $v_C(t)$ is the sum of two exponentials with time constants $1/|s_1|$ and $1/|s_2|$. An example of an overdamped response is plotted in Fig. 7(a).

The second case occurs when $\omega_0^2 > \alpha^2$ and s_1 and s_2 are complex conjugates of the form $-\alpha \pm j\omega_d$, where ω_d =

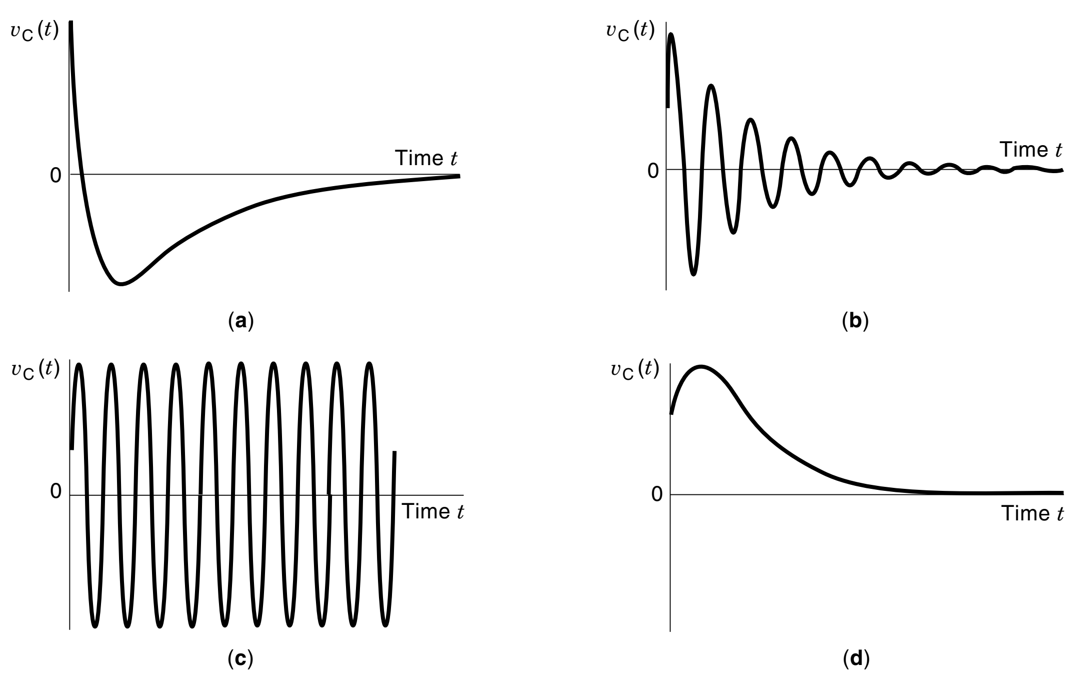

Figure 7. Examples of the natural response of the series RLC circuit: (a) overdamped, (b) underdamped, (c) underdamped and lossless, and (d) critically damped.

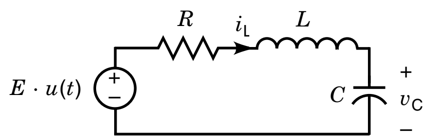

Figure 8. Second-order circuit consisting of resistor R, capacitor C, inductor L, and voltage source $E \cdot u(t)$.

$\sqrt{\omega_0^2 - \alpha^2}$. In this case the circuit is said to be *underdamped.* Equation (6) remains valid, but can be expressed more clearly in the form

$$v_C(t) = e^{-\alpha t}\left[(A_1 + A_2)\cos\omega_d t + j(A_1 - A_2)\sin\omega_d t\right]$$

A_1 and A_2 are complex conjugates, and so the coefficients $B_1 = (A_1 + A_2)$ and $B_2 = j(A_1 - A_2)$ are real and can once again be found from the initial conditions. The underdamped response takes the form of an oscillation of frequency ω_d multiplied by an exponential envelope $e^{-\alpha t}$. If $\alpha > 0$, the amplitude of the oscillation decreases exponentially with time, with the rate of this decrease, known as *damping,* controlled by α. If $\alpha = 0$, the response is an oscillation of constant amplitude and frequency $\omega_d = \omega_0 = 1/\sqrt{LC}$. This is the case of the well-known LC oscillator, which arises when $R = 0$ and there is no dissipation in the circuit. The underdamped response is plotted in Figs. 7(b) and 7(c) for the two cases $\alpha > 0$ and $\alpha = 0$. Note that the underdamped response is always characterized by oscillation, sometimes termed *ringing.*

If $\omega_0^2 = \alpha^2$, then $s_1 = s_2 = -\alpha = -R/2L$. In this case the general solution of Eq. (5) is no longer given by Eq. (6) but instead by

$$v_C(t) = (D_1 + D_2 t)e^{-\alpha t}$$

and is said to be *critically damped* (2). The constants D_1 and D_2 are once again found by application of the initial conditions. An example of a critically damped response is plotted in Fig. 7(d).

Step Response of a Second-order Circuit

The circuit in Fig. 8 is identical to that of Fig. 6 but for the addition of the voltage source E at $t = 0$. Applying Kirchhoff's voltage law for $t \geq 0$ gives the equation

$$v_C(t) + L\frac{di_L(t)}{dt} + Ri_L(t) = E$$

which on application of the relation $i_L(t) = C\, dv_C(t)/dt$ becomes

$$LC\frac{d^2 v_C(t)}{dt^2} + RC\frac{dv_C(t)}{dt} + v_C(t) = E \qquad (7)$$

To solve this equation, we apply the fact that the general solution to a differential equation is the sum of two components, which are known in mathematics as the *homogeneous solution* and a *particular solution* (2). The homogeneous solution is the solution to the differential equation obtained when all input terms (i.e., all terms not involving the variable or its derivatives) are set to zero. In circuit terms, this is just the response obtained when all independent voltage and current sources are removed—that is, the natural or unforced response. A particular solution is any solution to the differential equation. This decomposition may seem to be of no particular benefit, since it states that to solve the differential equation one must obtain a solution to the differential equation. The benefit lies in the ability to choose a particularly simple form for the particular solution, which can then be extended to yield the general solution by the addition of the homogeneous solution. The simplest particular solution is the constant solution which is obtained by setting all derivatives to zero.

The particular solution to Eq. (7) obtained by setting all derivatives to zero is $v_C(t) = E$. Adding this solution to the homogeneous solution which has already been found in Eq. (6) yields the general solution, which is of the form

$$v_C(t) = A_1 e^{s_1 t} + A_2 e^{s_2 t} + E \qquad \text{if } \omega_0^2 < \alpha^2 \quad \text{(overdamped)} \qquad (8a)$$

$$v_C(t) = e^{-\alpha t}[B_1 \cos\omega_d t + B_2 \sin\omega_d t] + E \qquad \text{if } \omega_0^2 > \alpha^2 \quad \text{(underdamped)} \qquad (8b)$$

$$v_C(t) = (D_1 + D_2 t)e^{-\alpha t} + E \qquad \text{if } \omega_0^2 = \alpha^2 \quad \text{(critically damped)} \qquad (8c)$$

The appropriate constants A_1 and A_2, B_1 and B_2, and D_1 and D_2, are found by applying the initial conditions. If the initial conditions are zero, Eq. (8) represent the step response of the series RLC circuit and is plotted in Fig. 9.

Depending on the system in which a circuit is to be used, different demands may be made of its step response. In some applications, for example, there may be a requirement that the voltage reach its steady-state value as soon as possible, while in others it may be necessary that the voltage never exceed its steady-state value by more than some specified percentage, to avoid driving circuit elements into saturation. A number of figures of merit have been defined to characterize the step response of a circuit in order to test its suitability for a given application (1). The rise time has already been defined. The *settling time* is the time beyond which the step response does not differ from its steady-state value by more than 2%. The *delay time* is the time taken for the step response to reach 50% of its steady-state value. The *overshoot* is defined as the difference between the peak value and the steady-state value of the step response, expressed as a percentage of the steady-state value.

LAPLACE TRANSFORM CIRCUIT ANALYSIS

The analyses described previously have found the circuit variables as a function of time by directly solving the differential equations that describe the circuit. While such a procedure is reasonably straightforward for first- and second-order circuits with simple source waveforms, it becomes significantly more difficult as the order of the circuit increases and as the source waveforms become more complex. It is desirable, therefore, to have a more powerful method of finding a solution. In the special case where all sources in the circuit are sinusoidal of the same frequency, the transformation of circuit variables into phasor or complex number form (3–5) allows the circuit to be handled using purely algebraic equations instead of differential equations. While extremely useful in certain circumstances, this is not a general circuit analysis method: It can handle only sinusoidal sources, it is applicable only if the circuit is stable, it finds only the steady-state component of the waveform, and it does not allow consideration of initial capacitor voltages and inductor currents.

The Laplace Transform

A more general transform than the phasor transform is the Laplace transform, named after the French mathematician Pierre-Simon Laplace (1749–1827). (See Refs. 3–6.) This transform method retains the fundamental advantage of the phasor transform, which is the ability to transform a system of differential equations into a system of algebraic equations, but has the additional advantages of being able to (a) handle a much broader class of source waveforms (including all that are of any practical interest), (b) accommodate initial conditions, and (c) yield solutions that incorporate both transient and steady-state components without requiring that the circuit be stable.

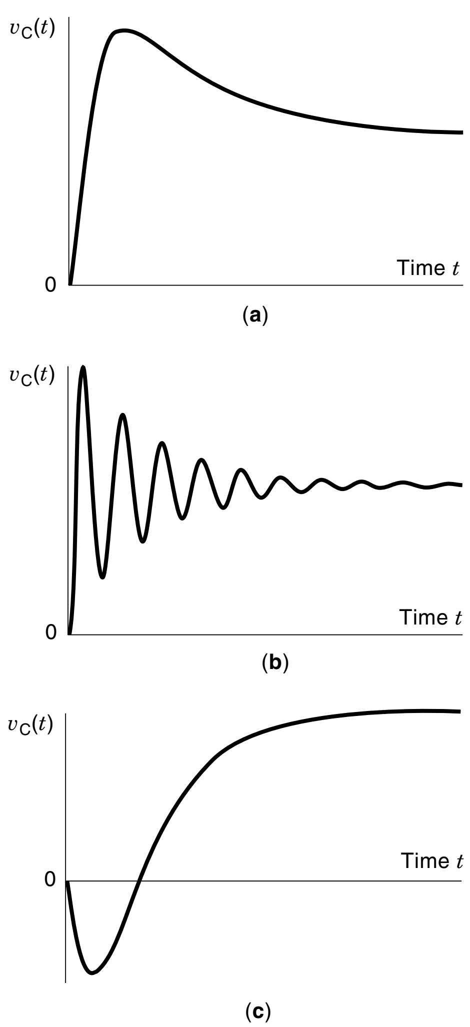

Figure 9. Examples of the step response of the series RLC circuit: (a) overdamped, (b) underdamped, and (c) critically damped.

The Laplace transform is discussed in the article on LINEAR SYSTEMS, and we will merely summarize its properties here. Given a function of time $f(t)$, its Laplace transform is

$$F(s) = \mathcal{L}\{f(t)\} = \int_{0^-}^{\infty} f(t)e^{-st}\,dt \tag{9}$$

where the variable s is complex and is termed the (complex) frequency. Thus the Laplace transform converts a function $f(t)$ from the time domain into a function $F(s)$ in the frequency domain. There exist functions which do not have a Laplace transform, since the integral in Eq. (9) fails to converge, but all functions of interest in circuit theory have a Laplace transform. Since the interval of integration is from 0^- to ∞, the transform defined by Eq. (9) is sometimes called the one-sided Laplace transform, to distinguish it from another version in which the integration is from $-\infty$ to ∞, but we will not need to draw this distinction here and will refer to it simply as the Laplace transform. The lower limit of integration of 0^- is chosen in order to accommodate functions with infinite spikes at $t = 0$. Such functions will prove extremely useful in our analysis.

Some of the properties of the Laplace transform that make it so useful in circuit analysis are the following (3,6), where $F(s)$ denotes the Laplace transform of $f(t)$, $F_1(s)$ the Laplace transform of $f_1(t)$, and $F_2(s)$ the Laplace transform of $f_2(t)$:

Uniqueness: $f_1(t) = f_2(t)$ for all $t \geq 0 \Leftrightarrow F_1(s) = F_2(s)$ [More precisely, if $F_1(s) = F_2(s)$, then

$$\int_{0^-}^{\infty} |f_1(t) - f_2(t)|\,dt = 0,$$

but for our purposes it will suffice to assume that $F_1(s) = F_2(s) \Rightarrow f_1(t) = f_2(t)$ for all $t \geq 0$.]

Linearity: $\mathcal{L}\{k_1 f_1(t) + k_2 f_2(t)\} = k_1 F_1(s) + k_2 F_2(s)$, where k_1 and k_2 are scalars.

Differentiation: $\mathcal{L}\left\{\dfrac{d}{dt} f(t)\right\} = sF(s) - f(0^-)$

Integration: $\mathcal{L}\left\{\displaystyle\int_{0^-}^{t} f(\tau)\,d\tau\right\} = \dfrac{1}{s}F(s)$

Time shift: $\mathcal{L}\{f(t-\tau)u(t-\tau)\} = e^{-s\tau}F(s)$, where $\tau > 0$ and $u(t)$ is the unit step function given by Eq. (1).

Frequency shift: $\mathcal{L}\{e^{-\alpha t} f(t)\} = F(s+\alpha)$

The first three of these properties are particularly important. The uniqueness property guarantees that if a system of differential equations is solved by transforming to the frequency domain, solving in the frequency domain and transforming back to the time domain, the solution obtained will be the same as would have been obtained if the solution had been carried out entirely in the time domain. The linearity property guarantees that a system of linear equations in the time domain will remain linear in the frequency domain, allowing powerful linear analysis techniques to be applied in both domains. The differentiation property allows differentiation in the time domain to be replaced by multiplication in the frequency domain, together with the addition of an term related to the initial condition. It is this property that allows

Table 1. Laplace Transforms of Some Important Functions

$f(t)$	$F(s) = \mathscr{L}(f(t))$
$\delta(t)$	1
$u(t)$	$\frac{1}{s}$
t^n	$n!\frac{1}{s^{n+1}}, n = 1, 2, \ldots$
e^{-at}	$\frac{1}{s+a}$
$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$
$\cos \omega t$	$\frac{s}{s^2+\omega^2}$

a system of differential equations in the time domain to be replaced by a system of algebraic equations in the frequency domain, which can be solved by a variety of powerful and elegant techniques. The Laplace transforms of some important functions are given in Table 1, in which $\delta(t)$ is the delta function defined by

$$\begin{aligned} \delta(t) &= 0 \qquad \text{for}\, t \neq 0 \\ \int_{-\infty}^{\infty} \delta(t) &= 1 \end{aligned} \tag{10}$$

There are three steps to be taken in solving a set of differential equations using Laplace transform analysis: (1) The system of differential equations in the time domain is transformed to a set of algebraic equations in the frequency domain; (2) this set of algebraic equations is solved in the frequency domain, using standard linear techniques; and (3) the solution is transformed from the frequency domain back to the time domain. Step 1 involves application of the definition of the Laplace transform (9) together with certain of its properties (notably the differentiation property). Step 2 involves standard techniques from linear algebra. Step 3 involves the application of the inverse Laplace transform, which converts a function $F(s)$ in the frequency domain to a function of time $f(t) = \mathscr{L}^{-1}(F(s))$ in such a way that $\mathscr{L}(f(t)) = F(s)$. Note that the function $f(t)$ is unique only for $t \geq 0$, since two functions of time which differ for $t < 0$ but are identical for $t \geq 0$ will have the same Laplace transform.

The Inverse Laplace Transform

There is a closed-form equation for the inverse Laplace transform (see Ref. 6 for details), but it is rather difficult to apply (involving contour integration) and is rarely used in circuit analysis applications (although it is sometimes used for numerical inversion of the Laplace transform). Instead, the inverse Laplace transform of a function is generally found by writing the function as the sum of simpler functions, each of whose inverse Laplace transform is known. A technique that is particularly useful here is the *partial fraction expansion* (2,6). This is a technique which allows the decomposition of a function $F(s)$ which is the ratio of two real polynomials in s into the sum of simpler terms. It is assumed that the degree of the numerator of $F(s)$ is less than that of the denominator; if this is not the case, then $F(s)$ can be expressed in the form $F(s) = r(s) + \hat{n}(s)/\hat{d}(s)$, where $r(s)$ is a polynomial in s and the degree of $\hat{n}(s)$ is less than that of $\hat{d}(s)$. The inverse Laplace transform of $r(s)$ can be found from Table 1, leaving only the component $\hat{n}(s)/\hat{d}(s)$ to be handled by the partial fraction expansion. Thus, without loss of generality, we can assume that the degree of the numerator of $F(s)$ is less than that of the denominator. The first step in the partial fraction expansion is the factorization of the denominator polynomial:

$$F(s) = \frac{n(s)}{d(s)} = \frac{n(s)}{(s-p_1)^{\alpha_1}(s-p_2)^{\alpha_2}\ldots(s-p_m)^{\alpha_m}}$$

The quantities p_i, the zeros of the denominator $d(s)$ of $F(s)$, are known as the poles of $F(s)$, and the multiplicity of the pole p_i is the number of times α_i it appears as a zero of $d(s)$. A pole of multiplicity one is called a simple pole. If all poles are simple then

$$\begin{aligned} F(s) &= \frac{n(s)}{(s-p_1)(s-p_2)\ldots(s-p_m)} \\ &= \frac{k_1}{s-p_1} + \frac{k_2}{s-p_2} + \ldots + \frac{k_m}{s-p_m} \end{aligned}$$

where $k_i = [(s-p_i)F(s)]_{s=p_i}$.

K_i is termed the *residue* of $F(s)$ at the pole p_i. If $F(s)$ has a pole of multiplicity α_j at p_j, the partial fraction expansion takes the form

$$\begin{aligned} F(s) &= \frac{n(s)}{(s-p_j)^{\alpha_j}\tilde{d}(s)} \\ &= \frac{k_{j1}}{s-p_j} + \frac{k_{j2}}{(s-p_j)^2} + \cdots + \frac{k_{j\alpha_j}}{(s-p_j)^{\alpha_j}} + \frac{\tilde{n}(s)}{\tilde{d}(s)} \end{aligned}$$

where

$$k_{ji} = \left.\left\{\frac{1}{(\alpha_j - i)!}\frac{d^{\alpha_j - i}}{ds^{\alpha_j - i}}\left[(s-p_j)^{\alpha_j}F(s)\right]\right\}\right|_{s=p_j}$$

Since the numerator and denominator of $F(s)$ are real polynomials in s, poles appear in complex conjugate pairs, as do their residues. This allows the combination of any complex term in the expansion with its conjugate to give a real term.

The inverse Laplace transform of each of the terms in the partial fraction expansion is known:

$$\mathscr{L}^{-1}\left\{\frac{k_{j\alpha_j}}{(s-p_j)^{\alpha_j}}\right\} = k_{j\alpha_j}\frac{t^{\alpha_j - 1}}{(\alpha_j - 1)!}e^{p_j t}$$

In this way it is possible to find the inverse Laplace transform of any function consisting of the ratio of two polynomials in s by decomposing the function via the partial fraction expansion and taking the inverse Laplace transform of each of the constituent functions. This method relies fundamentally on the uniqueness and linearity properties of the Laplace transform. Clearly the method applies only to a restricted range of functions, those which can be expressed as the ratio of two polynomials in s. As will be seen, however, functions of this type are particularly important in circuit analysis, and so this is not a significant limitation.

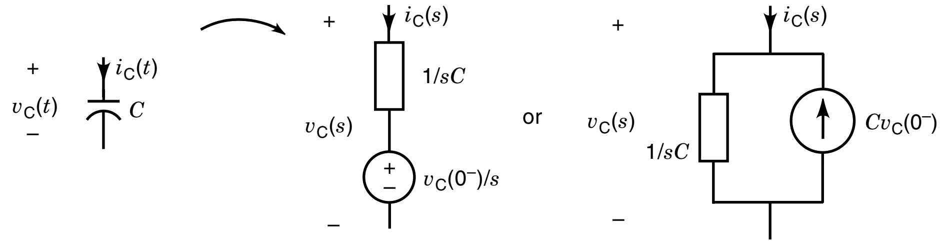

Figure 10. Transformation of a capacitor with initial voltage $v_C(0^-)$ into the frequency domain.

Laplace Transform Circuit Analysis

The first step in the Laplace transform analysis of a circuit is the transformation of the circuit from the time domain to the frequency domain. All branch voltages $v(t)$ and currents $i(t)$ which appear as variables in the differential equations describing the circuit will appear in the transformed equations as variables $V(s) = \mathcal{L}\{v(t)\}$ and $I(s) = \mathcal{L}\{i(t)\}$. Independent voltage and current sources are transformed from known functions of time $v_s(t)$ and $i_s(t)$ to known functions of frequency $V_s(s) = \mathcal{L}\{v_s(t)\}$ and $I_s(s) = \mathcal{L}\{i_s(t)\}$. A resistor is described in the time domain by the linear equation $v(t) = Ri(t)$ and so is defined in the transformed circuit by the relation $V(s) = RI(s)$. Similarly, the linear equations describing all resistive two-ports (including ideal transformers, gyrators, and controlled sources), and indeed resistive n-ports, are unchanged in the transformation from time domain to frequency domain. The capacitor is defined in the time domain by the equation

$$i_C(t) = C\frac{dv_C(t)}{dt}$$

Applying the differentiation property of the Laplace transform yields the frequency-domain equation for the capacitor:

$$I_C(s) = sCV_C(s) - Cv_C(0^-)$$

Thus the capacitor C with initial voltage $v_C(0^-)$ appears in the transformed circuit as the parallel combination of the independent current source $Cv_C(0^-)$ and the linear element defined by the relation $V(s) = (1/sC)I(s)$. This second element can be thought of as a generalized resistance (known as an *impedance*) $1/sC$ and throughout the analysis in the frequency domain can be handled as if it were a resistance. Figure 10 shows the transformation of a capacitor from the time domain into the parallel combination of an impedance and an independent current source in the frequency domain or, by Thévenin's theorem, into the series combination of an impedance and an independent voltage source.

In a similar manner, the inductor defined in the time domain by the relation

$$v_L(t) = L\frac{di_L(t)}{dt}$$

is defined in the frequency domain by the relation

$$V_L(s) = sLI_L(s) - Li_L(0^-)$$

Thus, as shown in Fig. 11, the inductor appears in the transformed circuit as the series combination of an impedance sL and voltage source $Li_L(0^-)$ or the parallel combination of the same impedance and current source $i_L(0^-)/s$. Note that once again this circuit transformation could have been obtained from Fig. 10 by application of the principle of duality. Coupled inductors can be transformed in a similar manner.

When all of the elements in the circuit have been transformed into the frequency domain, the first step of the analysis process is complete. The second step is to analyze the circuit in the frequency domain, employing any of a wide variety of techniques such as loop current analysis, node voltage analysis, modified nodal analysis, or sparse tableau analysis. The analysis of a circuit in the frequency domain is described in the article on FREQUENCY-DOMAIN CIRCUIT ANALYSIS and also in most circuit theory textbooks, such as Refs. 3–5. The third step is then to transform the results of the analysis back to the time domain via the inverse Laplace transform.

Example 1 The circuit of Fig. 1 can be transformed into the Laplace transform domain, yielding the circuit of Fig. 12. Analysis in the frequency domain, followed by partial fraction expansion, yields the following result:

$$V_C(s) = \frac{v_C(0^-)}{s+\frac{1}{RC}} + \frac{E\frac{1}{RC}}{s\left(s+\frac{1}{RC}\right)} = \frac{v_C(0^-)}{s+\frac{1}{RC}} + \frac{E}{s} - \frac{E}{s+\frac{1}{RC}}$$

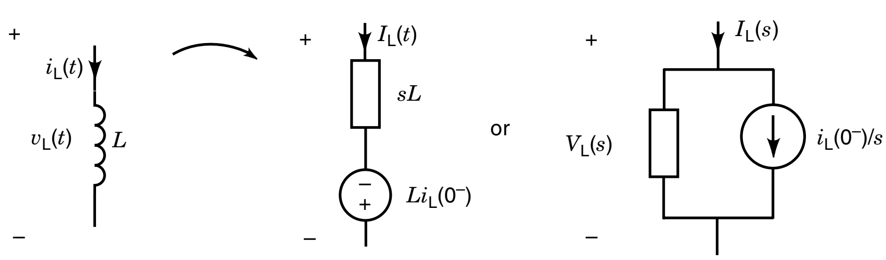

Figure 11. Transformation of an inductor with initial current $i_L(0^-)$ into the frequency domain.

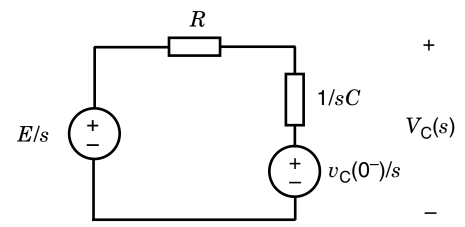

Figure 12. Laplace transform of the circuit of Fig. 1.

The inverse Laplace transform is then applied to find

$$v_C(t) = v_C(0^-)e^{-t/RC} + E(1 - e^{-t/RC}) \qquad \text{for } t \geq 0$$

which agrees with the time-domain analysis performed earlier.

NATURAL RESPONSE AND ZERO-STATE RESPONSE

When converted into the frequency domain, a circuit contains independent sources of two types. The first are the transformed versions of the independent sources from the time domain. These sources drive the circuit in the time domain and are often termed the inputs to the circuit, borrowing a viewpoint from system theory. The second group of independent sources in the frequency-domain circuit are those that are introduced during the transformation of energy storage elements and account for the initial conditions in the circuit—that is, the capacitor voltages and inductor currents at time $t = 0^-$. We will call these sources the initial condition generators, to distinguish them from those sources that represent the independent sources from the time domain. By superposition, the response of the circuit to these sources (by which we mean any current or voltage in the circuit, or any collection thereof) is the sum of two components: one due to the independent sources acting alone, with the initial condition generators removed, and the other due to the initial condition generators acting alone, with the independent sources removed. Since these two components of the response arise from different mechanisms, it is often useful to treat them separately. The component of the response due to the independent sources, with the initial conditions set to zero, is called the *zero-state response,* and the component due to the initial conditions, with the independent sources set to zero, is the *natural* or *unforced response* (also called the *zero-input response*).

Natural Response and Natural Frequencies

We will consider first the natural response of a circuit. Application of any of the standard frequency-domain analysis techniques will yield a matrix equation of the form

$$\mathbf{M}(s)\mathbf{X}(s) = \mathbf{U}(s)$$

where $\mathbf{M}(s)$ is a matrix each element of which is a polynomial in s; $\mathbf{X}(s)$ is a vector containing some subset of the unknown branch voltages, branch currents, node voltages, and loop currents; and $\mathbf{U}(s)$ is a vector, each nonzero element of which is a linear combination of the initial condition generators (3). If the circuit has a unique solution, that solution is given by

$$\mathbf{X}(s) = \mathbf{M}^{-1}(s)\mathbf{U}(s) = \frac{1}{\det(\mathbf{M}(s))}\text{Adj}(\mathbf{M}(s)).\mathbf{U}(s)$$

where the existence of a unique solution guarantees that the determinant $\det(\mathbf{M}(s))$ is not identically zero (2). We assume, unless otherwise stated, that all zeros $p_1, p_2, \ldots, p_m$ of $\det(\mathbf{M}(s))$ are simple. Each component $X_i(s)$ of the vector $\mathbf{X}(s)$ is the ratio of two polynomials in s, and so the partial fraction expansion can be applied to yield the expression

$$X_i(s) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \cdots + \frac{k_m}{s - p_m}$$

The time-domain response $x_i(t)$ is, therefore,

$$x_i(t) = k_1 e^{p_1 t} + k_2 e^{p_2 t} + \cdots + k_m e^{p_m t}$$

for $t \geq 0$. [If some of the zeros of $\det(\mathbf{M}(s))$ have multiplicity greater than one, the time response will contain terms of the form $t^{\alpha}e^{p_i t}$.]

Clearly the zeros p_i of $\det(\mathbf{M}(s))$ play a crucial role in determining the natural response of the circuit. These quantities are known as the *natural frequencies of the circuit.* The number of natural frequencies in a circuit is less than or equal to the number of energy storage elements in the circuit. The contribution of each natural frequency to the natural response depends on its location in the complex plane. A natural frequency at zero contributes a constant term to the natural response. A real and positive natural frequency p_i contributes a term $k_i e^{p_i t}$ that grows exponentially with time. A real and negative natural frequency p_i contributes a term $k_i e^{p_i t}$ that decays exponentially with time. Complex natural frequencies occur in conjugate pairs, and their contributions add to make a real contribution to the response waveform. If the natural frequencies in question lie on the imaginary axis at $\pm j\omega$, their composite contribution to the time response is of the form $k_i e^{j\omega t} + \bar{k}_i e^{-j\omega t} = 2|k_i| \cos(\omega t + \angle k_i)$, an oscillation of constant amplitude. If the complex natural frequencies lie in the right half-plane at $\alpha \pm j\omega$, their composite contribution is of the form $k_i e^{(\alpha + j\omega)t} + \bar{k}_i e^{(\alpha - j\omega)t} = 2|k_i|e^{\alpha t} \cos(\omega t + \angle k_i)$, an oscillation whose amplitude grows exponentially with time. Finally, if the complex natural frequencies lie in the left half-plane at $\alpha \pm j\omega$, their composite contribution is of the form $k_i e^{(\alpha + j\omega)t} + \bar{k}_i e^{(\alpha - j\omega)t} = 2|k_i|e^{\alpha t} \cos(\omega t + \angle k_i)$, an oscillation whose amplitude decays exponentially with time. (If some of the natural frequencies have multiplicity greater than one, their contribution to the time response will be more complicated, with polynomials times exponentials in place of exponentials, but can be handled by an extension of the above analysis.)

The above discussion leads to the important conclusion that if all natural frequencies of a circuit lie in the open left half-plane (i.e., if their real parts are less than 0), then for any set of initial conditions the natural or zero-input response of the circuit decays to zero as $t \to \infty$. This decay may be oscillatory, depending on the presence of complex natural frequencies. A circuit is said to be asymptotically stable or exponentially stable if all of its natural frequencies lie in the open left half-plane. If any natural frequency lies in the open right half-plane, then the initial conditions can cause certain currents and voltages to grow exponentially with time, which is clearly undesirable. Obviously in a real circuit this growth cannot continue indefinitely as the circuit elements will eventually cease to function, possibly in dramatic fashion. Also obvious is the fact that this behavior cannot occur in a circuit made up entirely of passive elements, since the exponential growth requires that energy be supplied to the circuit by an active element such as a controlled source or negative resistance.

While the natural frequencies determine the possible natural modes of behavior of a circuit, the actual response that will be observed in a circuit with zero input depends on the values of the initial conditions. Certain sets of initial conditions will excite one mode only, which means that all circuit variables will exhibit the same exponential or oscillatory behavior, but for most sets the response will be the combination of various modes. Also, all modes will not necessarily be observed in any given circuit variable—it may be that certain variables are not susceptible to the influence of one or more natural frequencies.

Example 2 To find the natural frequencies of the circuit of Fig. 13, the voltage source can be set to zero and the resulting circuit can be analyzed in the frequency domain by any of the usual methods. In this case, node voltage analysis is possible, yielding the matrix equation

$$\begin{pmatrix} \dfrac{1}{R_1} + sC_1 & 0 \\ -g_m & \dfrac{1}{R_2} + sC_2 \end{pmatrix} \begin{pmatrix} V_1(s) \\ V_2(s) \end{pmatrix} = \begin{pmatrix} C_1 v_1(0^-) \\ C_2 v_2(0^-) \end{pmatrix}$$

The natural frequencies are the values of s for which the determinant of the matrix in this equation is zero, and therefore they equal $-1/R_1C_1$ and $-1/R_2C_2$. Solving explicitly for $V_1(s)$ and $V_2(s)$ we find that

$$V_1(s) = \frac{(v_1(0^-)}{s + \dfrac{1}{R_1C_1}} \quad \text{and} \quad V_2(s) = \frac{\dfrac{g_m}{C_2} v_1(0^-) + \left(s + \dfrac{1}{R_1C_1}\right) v_2(0^-)}{\left(s + \dfrac{1}{R_1C_1}\right)\left(s + \dfrac{1}{R_2C_2}\right)}.$$

Thus the voltage v_1 (natural or zero-input component) exhibits only the behavior controlled by the natural frequency at $-1/R_1C_1$. and is unaffected by the natural frequency at $-1/R_2C_2$.

The Zero-State Response and Transfer Functions

The zero-state response of a circuit is its response to one or more independent sources (inputs) with all initial capacitor voltages and inductor currents set to zero. It suffices to consider the response to a single input, since superposition can then be applied to calculate the response due to multiple inputs. Application of any of the standard frequency-domain analysis techniques to a single-input zero-state circuit will yield a matrix equation of the form

$$\mathbf{M}(s)\mathbf{X}(s) = \mathbf{U}(s)$$

where $\mathbf{M}(s)$ is a matrix, each element of which is a polynomial in s; $\mathbf{X}(s)$ is a vector containing some subset of the unknown branch voltages, branch currents, node voltages, and loop currents; and $\mathbf{U}(s)$ is a vector, each nonzero element of which is a term involving the independent source, say $V_s(s)$ (although the theory applies equally to the case where the input is a current source). It follows from linear algebra (2,3) that

$$X_i(s) = \frac{n(s)}{\det(\mathbf{M}(s))} V_s(s) = H(s)V_s(s) \tag{11}$$

where $n(s)$ is a polynomial in s and $\det(\mathbf{M}(s))$ is not identically zero, by our standing assumption of unique solvability. Thus the zero-state response to a source $v_s(t)$ is obtained by multiplying its Laplace transform $v_s(s)$ by the appropriate function $H(s) = n(s)/\det(\mathbf{M}(s))$ and taking the inverse Laplace transform to return to the time domain. This function is known as a *transfer function* or *network function.* Note that the poles of a transfer function are zeros of $\det(\mathbf{M}(s))$ and are therefore natural frequencies of the circuit. However, not all natural frequencies need show up as poles of a given transfer function, due to cancellations with numerator terms.

Once again we see that the natural frequencies play a crucial role in determining the response of the circuit—even, as in this case, when the initial conditions are zero. From Eq. (11) the poles of $X_i(s)$ will be some subset (determined by numerator cancellations) of the poles of $V_s(s)$ and the natural frequencies. $x_i(t)$ will in general, therefore, contain terms related to the input together with exponential, constant, or oscillatory terms governed once again by the natural frequencies. If the circuit is asymptotically stable, the contributions governed by the natural frequencies will die away, leaving only the component governed by the input.

The simplest application of Eq. (11) occurs when $V_s(s) = 1$—that is, when the independent source $v_s(t)$ is the delta function or impulse function defined by Eq. (10). Although this function is physically unrealizable, it proves extremely useful in circuit and system analysis. When $V_s(t) = \delta(t)$, $x_i(s) = H(s) \cdot \mathscr{L}\{\delta(y)\} = H(s)$, and so the zero-state response is $x_i(t) = h(t) = \mathscr{L}^{-1}\{H(s)\}$. The zero-state response to an impulse function is known as the *impulse response,* and so we have found that the Laplace transform of the impulse response equals the transfer function. The expression

$$x_i(s) = H(s) \cdot \mathscr{L}\{v_s(t)\}$$

gives the frequency-domain response of the system with transfer function $H(s)$ to an input $v_s(t)$ and can be expressed in the time domain as

$$x_i(t) = h(t)^* v_s(t) = \int_{0_-}^{t^+} h(t-\tau) v_s(\tau)\, d\tau$$

where $h(t) = \mathscr{L}^{-1}\{H(s)\}$ is the impulse response and * is called the convolution operator (3,6).

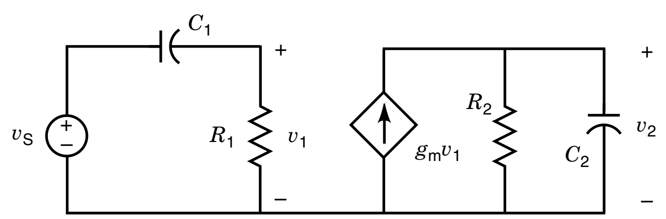

Figure 13. Circuit to be analyzed in Examples 2 and 3.

If the input is the unit step function $u(t)$, which has Laplace transform $1/s$, then $x_i(s) = H(s) \cdot \mathscr{L}\{u(t)\} = H(s)/s$ and so the step response is $x_i(t) = \mathscr{L}^{-1}\{H(s)/s\}$. It is easy to see that the impulse response is the derivative of the step response.

Example 3 The transfer function $V_2(s)/V_s(s)$ of the circuit of Fig. 13 is

$$H(s) = \frac{\frac{g_m}{C_2}s}{\left(s + \frac{1}{R_1C_1}\right)\left(s + \frac{1}{R_2C_2}\right)} = \frac{\frac{g_m}{C_2}}{R_2C_2 - R_1C_1}\left[\frac{R_2C_2}{\left(s + \frac{1}{R_1C_1}\right)} - \frac{R_1C_1}{\left(s + \frac{1}{R_2C_2}\right)}\right]$$

and so the impulse response is

$$h(t) = \frac{g_mR_2}{R_2C_2 - R_1C_1}e^{-t/R_1C_1} - \frac{g_mR_1C_1/C_2}{R_2C_2 - R_1C_1}e^{-t/R_2C_2} \qquad \text{for } t \geq 0$$

The step response is

$$\mathscr{L}^{-1}\left\{\frac{\frac{g_m}{C_2}}{\left(s + \frac{1}{R_1C_1}\right)\left(s + \frac{1}{R_2C_2}\right)}\right\} = \frac{g_mR_1C_1R_2}{R_2C_2 - R_1C_1}[-e^{-t/R_1C_1} + e^{-t/R_2C_2}] \qquad \text{for } t \geq 0$$

Note the exponential modes corresponding to the natural frequencies in both the step response and the impulse response. Note also that the impulse response is the derivative of the step response.

FOURIER TRANSFORM CIRCUIT ANALYSIS

The power of the Laplace transform in finding the transient and steady-state response of a circuit, the variety of source waveforms which it can handle, and its ability to accommodate initial conditions make it the method of choice in transient circuit analysis. Despite these advantages, another transform, closely related to the Laplace transform, is preferred in certain situations. This is the Fourier transform (4–6), named after the French mathematician Jean Baptiste Joseph Fourier (1768–1830). The close relationship between the Fourier transform of a signal and the frequency content of that signal make it particularly useful in applications such as communications and signal processing where this frequency content is of paramount importance. However, the Fourier transform is defined for a smaller class of source waveforms than the Laplace transform, and it cannot handle initial conditions. The latter condition in particular makes it poorly suited to transient circuit analysis and so we will merely give a brief discussion of its properties here, with the intention of (1) explaining why it is unsuited to transient circuit analysis and (2) providing a link to other forms of transient circuit analysis for circuits such as filters that are more usually handled using Fourier analysis.

The Fourier transform is closely related to the Fourier series (4–6), in which a periodic function with period T is decomposed into the weighted sum of sinusoids whose angular frequencies are integer multiples of $2\pi/T$. By superposition, the response of a circuit to a periodic function could be obtained by decomposing the function into the sum of sinusoids, finding the response to each of these sinusoids via phasor analysis, and summing these responses to find the overall response. The main disadvantage to this Fourier series method of analysis is that many source waveforms of interest are not periodic; and since the method is based on phasor analysis, it finds only the steady-state component of the response. The fundamental idea underlying this method, however, namely the idea of a sum of input sinusoids being processed (i.e., altered in magnitude and phase) in different ways by a circuit and then added to form the response, is a very useful one and underlies the more general Fourier transform analysis.

The Fourier transform is a generalization of the Fourier series to accommodate nonperiodic functions. A nonperiodic function can be viewed as the limit of a periodic function as the period T tends to infinity. The Fourier series of this periodic function consists of weighted sinusoids spaced in frequency at integer multiples of $2\pi/T$. As T tends to infinity, the separation of these sinusoidal frequency components tends to zero, and in the limit we have the nonperiodic function represented by a continuum or spectrum of sinusoidal components. This spectrum of sinusoidal components constitutes the Fourier transform of the function. The Fourier transform of a signal $f(t)$ is found, as in the above discussion, by taking the limit of the expression for the Fourier series of a periodic function as the period tends to infinity, which turns out to be

$$F(j\omega) = \mathscr{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt \tag{12}$$

and exists if the integral in Eq. (12) converges. Once again we say that the Fourier transform converts a function from the time domain into the frequency domain, with $F(j\omega)$ indicating the frequency content of the signal at frequency ω. If $f(t) = 0$ for $t < 0$ and the above integral converges, the Fourier transform of f is just the Laplace transform with $j\omega$ substituted for s. Given $F(j\omega)$, the function $f(t)$ such that $F(j\omega) = \mathscr{F}\{f(t)\}$ is found by application of the inverse Fourier transform

$$f(t) = \mathscr{F}^{-1}\{F(j\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{j\omega t}\,d\omega$$

One important feature of the Fourier transform is the differentiation property, which states that differentiation in the time domain is equivalent to multiplication by $j\omega$ in the frequency domain. Thus the Fourier transform can, like the Laplace transform, be used to transform a system of differential equations in the time domain to a system of algebraic equations in the frequency domain.

In Fourier transform analysis a circuit is transformed into the frequency domain by replacing all independent sources by their Fourier transforms, replacing each inductor L by an impedance $j\omega L$ (and replacing any time-domain coupling M between inductors by the frequency-domain coupling $j\omega M$), re-

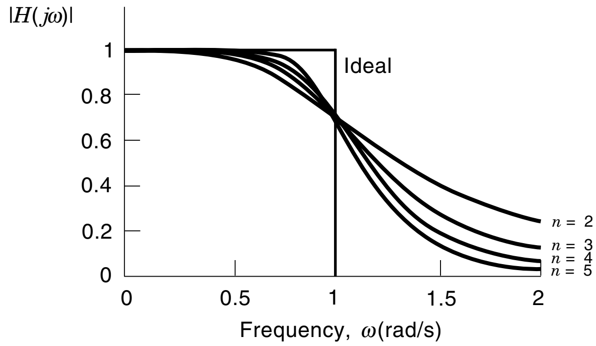

Figure 14. Amplitude response of the ideal low-pass filter with cutoff frequency at 1 rad/s, together with the amplitude responses of the normalized Butterworth filters of order $n = 2, 3, 4$, and 5. Note that the approximation more closely matches the ideal as the order of the filter increases.

placing each capacitor C by an impedance $1/j\omega C$, and leaving resistive components unchanged. Note the lack of any initial condition generators; this is a consequence of the fact that the lower limit of integration in the definition of the Fourier transform is $-\infty$ rather than 0^-. Analysis in the frequency domain proceeds as described in the article on FREQUENCY-DOMAIN CIRCUIT ANALYSIS or in Refs. 4 and 5, using the standard tools, and the frequency-domain response is converted back to the time domain by application of the inverse Fourier transform. Once again there is a transfer function—in this case a function of frequency $H(j\omega)$—relating input and output in the frequency domain. Note that the response obtained through Fourier transform analysis is the zero-state response only, since the method contains no provision for handling initial conditions.

Given a circuit with input $\sin(\omega_0 t)$ and transfer function $H(j\omega)$ [which in general is complex and, for the circuits in which we are interested, has the property that $H(-j\omega)$ is the complex conjugate of $H(j\omega)$], the output is obtained by taking the inverse Fourier transform of $H(j\omega)\cdot\mathscr{F}\{\sin(\omega_0 t)\}$, which turns out to be $|H(j\omega_0)|\sin(\omega_0 t + \angle H(j\omega_0))$. In other words, the sinusoidal input appears at the output as a sinusoid of the same frequency, with amplitude multiplied by the magnitude of the transfer function at that frequency and phase incremented by the phase of the transfer function at that frequency. If the input to the circuit is more general, it can be viewed as the finite or infinite sum of sinusoids, which will be altered in magnitude and phase by the action of the circuit and then recombined to form the output of the circuit. The magnitude and phase of the transfer function will generally vary with frequency, and when plotted versus frequency they are called the amplitude (or magnitude) response and phase response plots.

Frequency-selective circuits which pass certain ranges of frequencies from input to output while blocking other ranges are known as *filters* (7). For example, an ideal low-pass filter would pass to the output all frequency components of its input up to a certain cutoff frequency and would pass no higher-frequency components. This ideal low-pass filter cannot be realized and is therefore approximated by a variety of functions such as the Butterworth and Chebyshev approximations. Figure 14 plots the amplitude response of the ideal low-pass filter with cutoff frequency at 1 rad/s, together with the amplitude responses of the normalized Butterworth filters of orders 2, 3, 4, and 5. The amplitude response of each of these Butterworth filters is 0.7071 or -3 dB at $\omega = 1$ rad/s, which is to say that their 3 dB bandwidth is 1 rad/s.

In a communication system designed to transmit pulses, the step response of a filter is crucial. Too slow a rise time leads to neighboring pulses in a pulse train being smeared over one another, rendering them indistinguishable at the output. Too high an overshoot can drive circuit elements into saturation. The step response of a filter can be found by Fourier transform methods, by taking the inverse Fourier transform of the function $H(j\omega)\mathscr{F}\{u(t)\}$, but there is in general no reason to prefer the Fourier transform over the Laplace transform in this situation, and it is usual to take instead the inverse Laplace transform of the function $H(s)\mathscr{L}\{u(t)\}$. For example, the normalized third-order Butterworth low-pass filter has transfer function $H(s) = 1/(s^3 + 2s^2 + 2s + 1)$ and so its step response is $\mathscr{L}^{-1}\{1/s(s^3 + 2s^2 + 2s + 1)\}$, which can be found by the partial fraction decomposition to be $1 - e^{-t} - (2/\sqrt{3})e^{-t/2}\sin(\sqrt{3}/2)t$ for $t > 0$. Figure 15 plots the step response of the normalized Butterworth filters of orders 2, 3, 4, and 5, as obtained by application of the Laplace transform. It can be seen that as the order increases (and the amplitude response more closely approximates the ideal) the overshoot, settling time, and delay time of the filters all increase, but the rise time is approximately constant.

The procedure outlined above can be used to find the exact step response of a filter, allowing a designer to compare the suitability of various filters in pulse transmission applications. Designers should also have an intuitive understanding of the relationship between amplitude response and transient response of a filter. A low-pass filter allows low frequencies to pass to the output, but blocks high frequencies. Thus when the input is a step function, the output will preserve the steady-state constant behavior of the input, but will act to block the high frequencies involved in the transition from 0 to 1. This can be seen in Fig. 15, where the high-order filters that are most effective at blocking high frequencies are least effective in capturing the discontinuity in the input. We now recognize the RC circuit of Fig. 1, with the output voltage taken across the capacitor, as a low-pass filter. If the output voltage were taken across the resistor, we would have a high-pass filter, whose step response captures the initial discontinuity in the step, but then falls away to zero due to its inability to pass dc. Readers interested in a more detailed discus-

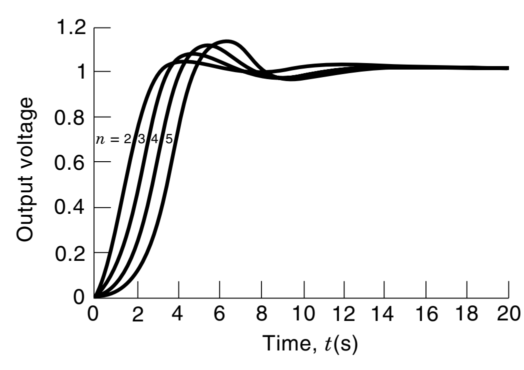

Figure 15. Step response of the normalized Butterworth filters or orders $n = 2, 3, 4$, and 5.

sion of the relationship between frequency response and transient response of filters are referred to Refs. 1 and 7.

HAZARDS FOR THE UNWARY

Computer Simulation of Transient Circuit Performance

Circuit simulation programs such as SPICE (8) are now ubiquitous, and it is important that users understand the operation of these programs so that their results can be interpreted. Our focus here is on the methods by which circuit simulators approach transient circuit behavior. In order to obtain an approximate solution to a differential equation, a circuit simulator approximates all derivatives in the equation by discrete-time approximations. Transient circuit simulation proceeds in three steps: (1) The time interval of interest, consisting of a continuum of time values, is broken up into a set of small individual time steps: (2) the differential equation is approximated by an algebraic equation over each time step, converting it into a form which the computer can readily solve; and finally (3) the solutions over each of these time steps are pieced together to form an approximation to the solution of the differential equation over the entire time interval. The key issue here is the nature of the simplifying approximation to the derivative. There are a number of these approximations, known as *numerical integration methods.* For example, the forward Euler approximation replaces the derivative dv_C/dt at time t_k by the approximation $v_C(t_{k+1}) - v_C(t_k)/(t_{k+1} - t_k)$, which is exact if the solution $v_C(t)$ is a straight line. In practice, the solution will rarely be of this form; but if the time step $t_{k+1} - t_k$ is short enough, then the linear approximation is a reasonable one and the solution computed by the numerical integration will in general be a reasonable approximation to the actual solution. There are three other numerical integration methods commonly used in circuit simulators: the backward Euler approximation, the trapezoidal rule, and Gear's methods. The two Euler approximations are first order, giving exact results if the solution is in fact a straight line; the trapezoidal rule is second order, and it gives exact results if the solution is a quadratic; and Gear's methods are of any order, with the second-order method most widely used.

A poor choice of numerical integration method, or a poor choice of time step for a given method, can result in an approximate solution which differs wildly from the exact solution. Unfortunately for the designer, these erroneous approximations often appear plausible, displaying behavior of the types seen throughout this article. For example, poor use of a simulator can cause the natural response of a stable first-order circuit appear to exhibit damped oscillation, sustained oscillation, or even growing oscillation. Alternatively, numerical integration can add artificial damping to a response, producing for example a damped oscillatory response in a lossless *LC* oscillator. Designers who make use of circuit simulators in their study of transient responses should be aware of these hazards, and they are referred to Ref. 8 for further details.

Nonlinear Circuits

The theory and techniques of transient linear circuit analysis are powerful and elegant and form part of the tool kit of all electronic and electrical engineers. As has been seen throughout this article, the behavior of the linear circuits to which this analysis is applied is actually rather limited. This is not to say that these circuits are not useful—quite the reverse. The power of transient circuit analysis (and other forms of linear circuit analysis), coupled with the tremendous variety of uses to which linear circuits can be applied, may tend to give the impression that all circuits behave in a reasonably simple fashion and that it is only by adding a complex signal (such as noise) that complex behavior can be observed in a circuit. This is not the case. The transient and steady-state behavior of nonlinear circuits can be extraordinarily complex, even in the absence of an input signal. An appreciation of the complexity of nonlinear systems, together with an improved ability to analyze and understand it, has been developed by mathematicians, engineers, and scientists from various disciplines since the 1960s, with terms such as "chaos" entering the lexicon and popular culture. In circuit theory this work was pioneered by Chua and his co-workers, and readers interested in venturing from the comparatively tame world of linear circuit analysis into the fascinating world of nonlinear circuits are referred to the seminal paper (9) and to the article NONLINEAR DYNAMIC PHENOMENA IN CIRCUITS.

BIBLIOGRAPHY

1. F. F. Kuo, *Network Analysis and Synthesis,* 2nd ed., New York: Wiley, 1966.
2. E. Kreyszig, *Advanced Engineering Mathematics,* 6th ed., New York: Wiley, 1988.
3. L. O. Chua, C. A. Desoer, and E. S. Kuh, *Linear and Nonlinear Circuits,* New York: McGraw-Hill, 1987.
4. R. C. Dorf and J. A. Svoboda, *Introduction to Electric Circuits,* 3rd ed., New York: Wiley, 1996.
5. J. W. Nilsson and S. A. Riedel, *Electric Circuits,* 5th ed., Reading, MA: Addison-Wesley, 1996.
6. A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and Systems,* London: Prentice-Hall, 1983.
7. L. P. Huelsman, *Active and Passive Analog Filter Design,* Singapore: McGraw-Hill, 1993.
8. K. S. Kundert, *The Designer's Guide to Spice and Spectre,* Boston: Kluwer, 1995.
9. T. Matsumoto, L. O. Chua, and M. Komuro, The double scroll, *IEEE Trans. Circuits Syst.,* **32**: 797–818, 1985.

ORLA FEELY
University College Dublin

TRANSIENT INTERMODULATION MEASUREMENT. See INTERMODULATION MEASUREMENT.
TRANSIENTS OF ELECTRICAL MACHINES. See ELECTRIC MACHINE ANALYSIS AND SIMULATION.
TRANSIENTS, OVERVOLTAGE. See OVERVOLTAGE PROTECTION.
TRANSIENT STABILITY. See POWER SYSTEM TRANSIENTS.
TRANSISTOR, BIPOLAR PERMEABLE. See BIPOLAR PERMEABLE TRANSISTOR.
TRANSISTOR RELIABILITY. See POWER DEVICE RELIABILITY.

TRANSISTORS, BIPOLAR. See Bipolar transistors.

TRANSISTORS, CHARGE INJECTION. See Charge injection devices.

TRANSISTORS, POWER. See Power devices.

TRANSISTORS, STATIC INDUCTION. See Static induction transistors.

TRANSISTORS, THIN FILM. See Thin film devices; Thin film transistors.