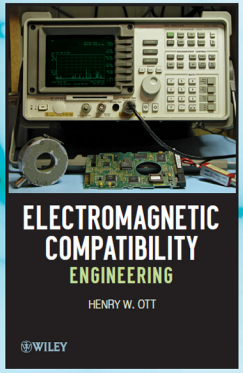
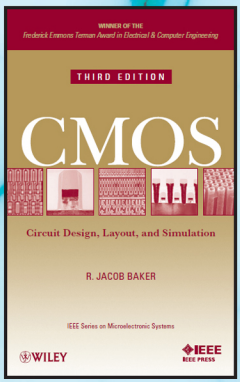
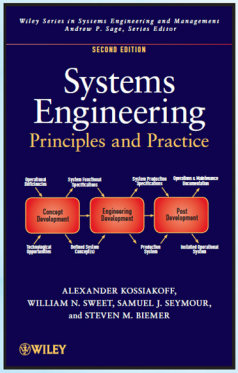
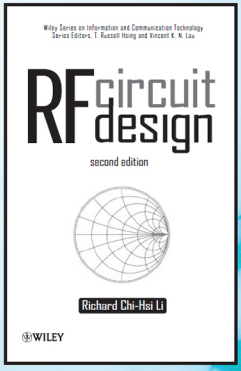


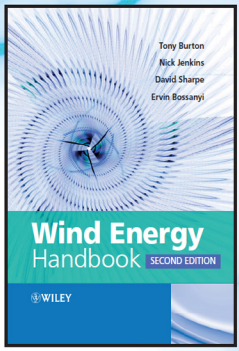


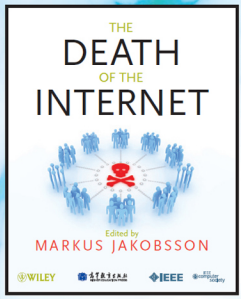


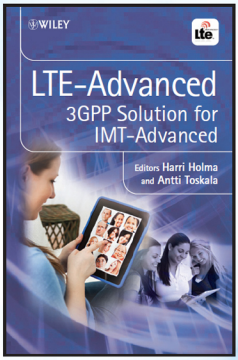


ELECTRICAL ENGINEERING

eBOOK SAMPLER







WILEY

Introduction

The 2013 Electrical Engineering eBook Sampler includes select material from seven Electrical Engineering titles. Chapters and the Table of Contents have been selected from 7 of Wiley's key titles in Electrical Engineering. If you would like to read more from these books, you can purchase the full book or e-book at your favorite online retailer.

CMOS: Circuit Design, Layout, and Simulation, 3rd Edition

RF Circuit Design, 2nd Edition

Electromagnetic Compatibility Engineering

Systems Engineering Principles and Practice, 2nd Edition

LTE Advanced: 3GPP Solution for IMT-Advanced

The Death of the Internet

Wind Energy Handbook, 2nd Edition

Click here to view our complete collection of Electrical Engineering titles!

WINNER OF THE

Frederick Emmons Terman Award in Electrical & Computer Engineering

THIRD EDITION

CMOS

Circuit Design, Layout, and Simulation

R. JACOB BAKER

IEEE Series on Microelectronic Systems

WILEY

IEEE

IEEE PRESS

Contents

Chapter

1

Introduction to CMOS Design

This chapter provides a brief introduction to the CMOS (complementary metal oxide semiconductor) integrated circuit (IC) design process (the design of "chips"). CMOS is used in most very large scale integrated (VLSI) or ultra-large scale integrated (ULSI) circuit chips. The term "VLSI" is generally associated with chips containing thousands or millions of metal oxide semiconductor field effect transistors (MOSFETs). The term "ULSI" is generally associated with chips containing billions, or more, MOSFETs. We'll avoid the use of these descriptive terms in this book and focus simply on "digital and analog CMOS circuit design."

We'll also introduce circuit simulation using SPICE (simulation program with integrated circuit emphasis). The introduction will be used to review basic circuit analysis and to provide a quick reference for SPICE syntax.

1.1 The CMOS IC Design Process

The CMOS circuit design process consists of defining circuit inputs and outputs, hand calculations, circuit simulations, circuit layout, simulations including parasitics, reevaluation of circuit inputs and outputs, fabrication, and testing. A flowchart of this process is shown in Fig. 1.1. The circuit specifications are rarely set in concrete; that is, they can change as the project matures. This can be the result of trade-offs made between cost and performance, changes in the marketability of the chip, or simply changes in the customer's needs. In almost all cases, major changes after the chip has gone into production are not possible.

This text concentrates on custom IC design. Other (noncustom) methods of designing chips, including field-programmable-gate-arrays (FPGAs) and standard cell libraries, are used when low volume and quick design turnaround are important. Most chips that are mass produced, including microprocessors and memory, are examples of chips that are custom designed.

The task of laying out the IC is often given to a layout designer. However, it is extremely important that the engineer can lay out a chip (and can provide direction to the layout designer on how to layout a chip) and understand the parasitics involved in the

layout. Parasitics are the stray capacitances, inductances, pn junctions, and bipolar transistors, with the associated problems (breakdown, stored charge, latch-up, etc.). A fundamental understanding of these problems is important in precision/high-speed design.

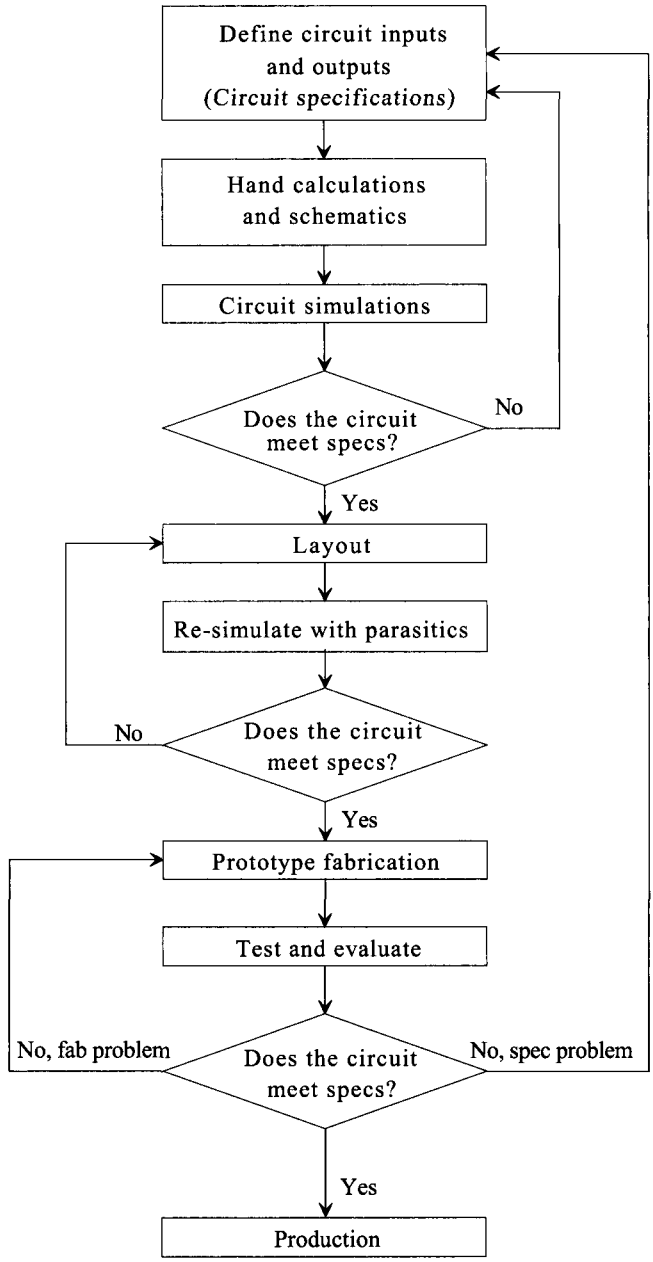
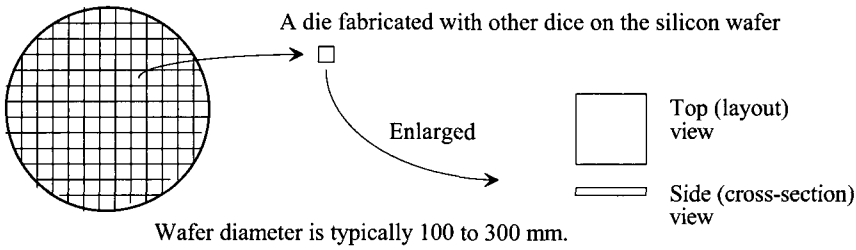


Figure 1.1 Flowchart for the CMOS IC design process.

1.1.1 Fabrication

CMOS integrated circuits are fabricated on thin circular slices of silicon called wafers. Each wafer contains several (perhaps hundreds or even thousands) of individual **chips or "die"** (Fig. 1.2). For production purposes, each die on a wafer is usually identical, as seen in the photograph in Fig. 1.2. Added to the wafer are test structures and process monitor plugs (sections of the wafer used to monitor process parameters). The most common wafer size (diameter) in production at the time of this writing is 300 mm (12 inch).



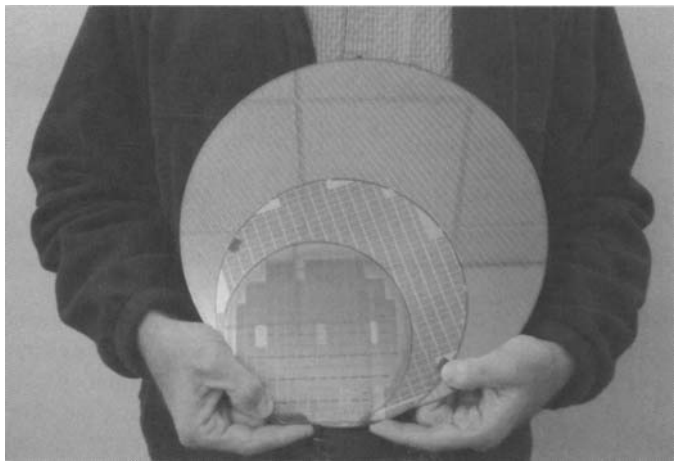


Figure 1.2 CMOS integrated circuits are fabricated on and in a silicon wafer. Shown are 150, 200, and 300 mm diameter wafers. Notice the reflection of ceiling tiles in the 300 mm wafer.

The ICs we design and lay out using a layout program can be fabricated through MOSIS (http://mosis.com) on what is called a **multiproject wafer**; that is, a wafer that is comprised of chip designs of varying sizes from different sources (educational, private, government, etc.). MOSIS combines multiple chips on a wafer to split the fab cost among several designs to keep the cost low. MOSIS subcontracts the fabrication of the chip designs (multiproject wafer) out to one of many commercial manufacturers (vendors). MOSIS takes the wafers it receives from the vendors, after fabrication, and cuts them up to isolate the individual chip designs. The chips are then packaged and sent to the originator. A sample package (40-pin ceramic) from a MOSIS-submitted student design is seen in Fig. 1.3. Normally a cover (not shown) keeps the chip from being exposed to light or accidental damage.

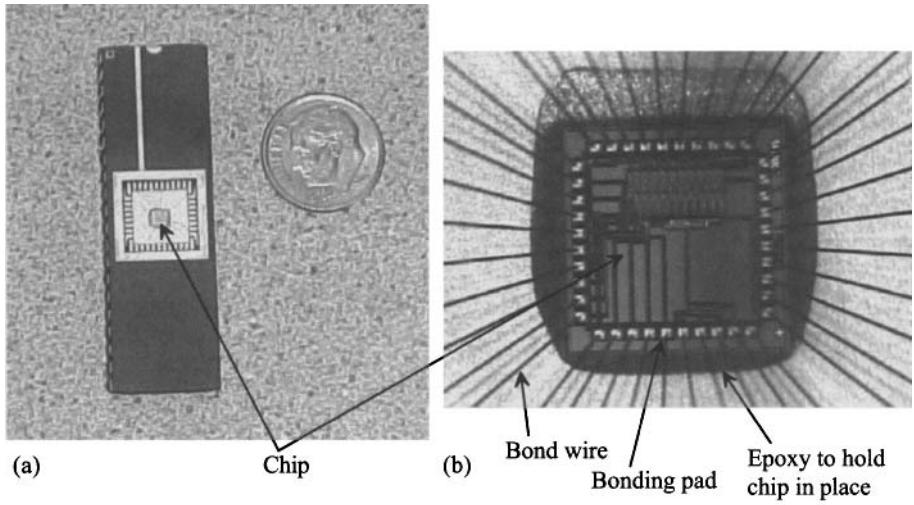


Figure 1.3 How a chip is packaged (a) and (b) a closer view.

Note, in Fig. 1.3, that the chip's electrical signals are transmitted to the pins of the package through wires. These wires (called "bond wires") electrically bond the chip to the package so that a pin of the chip is electrically connected (shorted) to a piece of metal on the chip (called a bonding pad). The chip is held in the cavity of the package with an epoxy resin ("glue") as seen in Fig. 1.3b.

The ceramic package used in Fig. 1.3 isn't used for most mass-produced chips. Most chips that are mass produced use plastic packages. Exceptions to this statement are chips that dissipate a lot of heat or chips that are placed directly on a printed circuit board (where they are simply "packaged" using a glob of resin). Plastic packaged (encapsulated) chips place the die on a lead frame (Fig. 1.4) and then encapsulate the die and lead frame in plastic. The plastic is melted around the chip. After the chip is encapsulated, its leads are bent to the correct position. This is followed by printing information on the chip (the manufacturer, the chip type, and the lot number) and finally placing the chip in a tube or reel for shipping to a company that makes products that use the chips. Example products might include chips that are used in cell phones, computers, microwave ovens, printers.

Layout and Cross Sectional Views

The view that we see when laying out a chip is the top, or layout, view of the die. However, to understand the parasitics and how the circuits are connected together, it's important to understand the chip's cross-sectional view. Since we will often show a layout view followed by a cross-sectional view, let's make sure we understand the difference and how to draw a cross-section from a layout. Figure 1.5a shows the layout (top) view of a pie. In (b) we show the cross-section of the pie (without the pie tin) at the line indicated in (a). To "lay-out" a pie we might have layers called: crust, filling, caramel, whipped-cream, nuts, etc. We draw these layers to indicate how to assemble the pie (e.g., where to place nuts on the top). Note that *the order we draw the layers doesn't matter*. We could draw the nuts (on the top of the pie) first and then the crust. When we fabricate the pie, the order does matter (the crust is baked before the nuts are added).

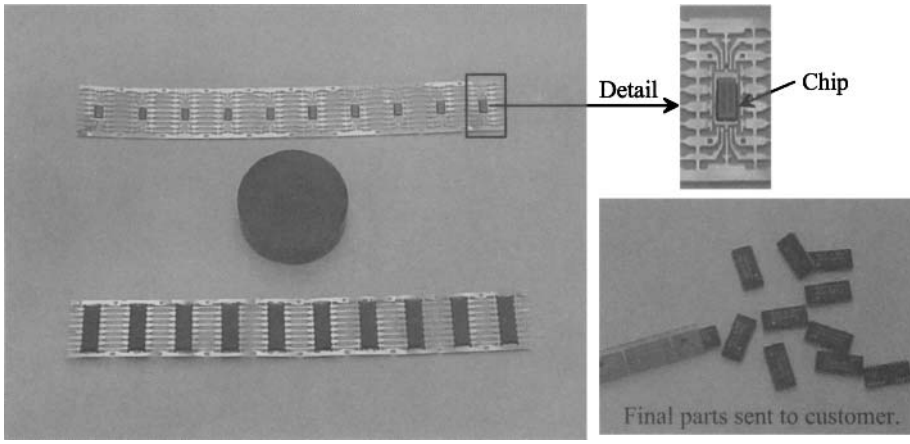


Figure 1.4 Plastic packages are used (generally) when the chip is mass produced.

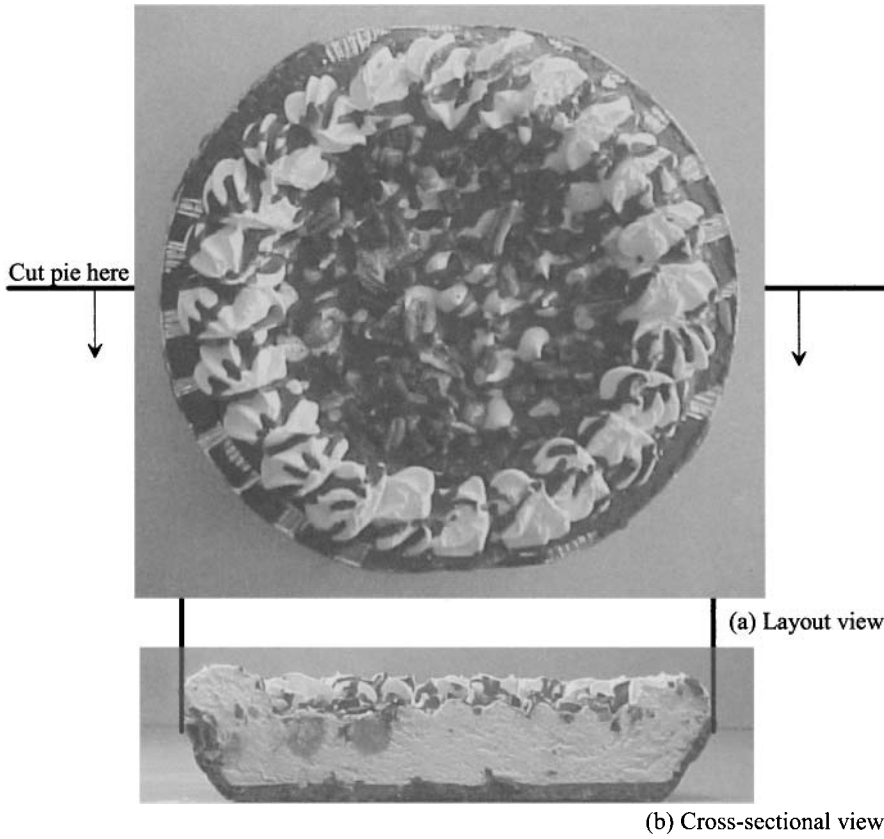


Figure 1.5 Layout and cross sectional view of a pie (minus pie tin).

1.2 CMOS Background

CMOS circuit design (the idea and basic concepts) was invented in 1963 by Frank Wanlass while at Fairchild Semiconductor, see US Patent 3,356,858, [5]. The idea that a circuit could be made with discrete complementary MOS devices, an NMOS (n-channel MOSFET) transistor (Fig. 1.6) and a PMOS (p-channel) transistor (Fig. 1.7) was quite novel at the time given the immaturity of MOS technology and the rising popularity of the bipolar junction transistor (BJT) as a replacement for the vacuum tube.

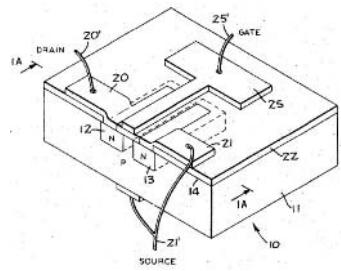


Figure 1.6 Discrete NMOS device from US Patent 3,356,858 [5]. Note the metal gate and the connection to the MOSFET's body on the bottom of the device. Also note that the source and body are tied together.

The CMOS Acronym

Note in Figs. 1.6 and 1.7 the use of a metal gate and the connection to the MOSFET's body on the bottom of the transistor (these are discrete devices). As we'll see later in the book (e.g., Fig. 4.3) the gate material used in a modern MOSFET is no longer metal but rather polysilicon. Strictly speaking, modern technology is not CMOS then but rather CPOS (complementary-polysilicon-oxide-semiconductor). US Patent 3,356,858 refers to the use of insulated field effect transistors (IFETs). The acronym IFET is perhaps, even today, a more appropriate descriptive term than MOSFET. Others (see the footnote on page 154) have used the term IGFET (insulated-gate-field-effect-transistor) to describe the devices. We'll stick to the ubiquitous terms MOSFET and CMOS since they are standard terms that indicate devices, design, or technology using complementary field effect devices.

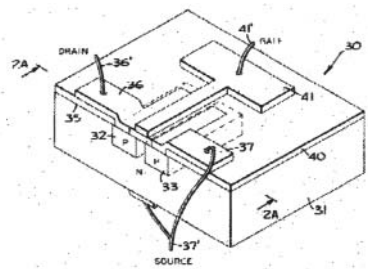


Figure 1.7 Discrete PMOS device from US Patent 3,356,858 [5].

CMOS Inverter

Figure 1.8 shows the schematic of a CMOS inverter. Note the use of a modified bipolar symbol for the MOSFET (see Fig. 4.14 and the associated discussion). Also note that the connections of the sources (the terminals with arrows) and drains are backwards from most circuit design and schematic drawing practices. Current flows from the top of the schematic to the bottom, and the arrow indicates the direction of current flow.

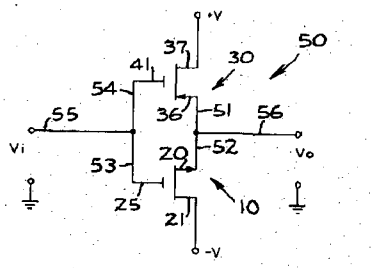


Figure 1.8 Inverter schematic from US Patent 3,356,858 [5].

When the input voltage, V_i, is $-V$ (the negative supply rail), the output, V_o, goes to $+V$ (the positive supply voltage). The NMOS device (bottom) shuts off and the PMOS device (top) turns on. When the input goes to $+V$, the output goes to $-V$ turning on the NMOS and turning off the PMOS. So if a logic 0 corresponds to $-V$ and a logic 1 to $+V$, the circuit performs the logical inversion operation. This topology has several advantages over digital circuits implemented using BJTs including an output swing that goes to the power supply rails, very low static power dissipation, and no storage time delays (see Sec. 2.4.3).

The First CMOS Circuits

In 1968 a group led by Albert Medwin at RCA made the first commercial CMOS integrated circuits (the 4000 series of CMOS logic gates). At first CMOS circuits were a low-power, but slower, alternative to BJT logic circuits using TTL (transistor-transistor logic) digital logic. During the 1970s, the makers of watches used CMOS technology because of the importance of long battery life. Also during this period, MOS technology was used for computing processor development, which ultimately led to the creation of the personal computer market in the 1980s and the use of internet, or web, technology in the 1990s. It's likely that the MOS transistor is the most manufactured device in the history of mankind.

Currently more than 95% of integrated circuits are fabricated in CMOS. For the present, and foreseeable future, CMOS will remain the dominant technology used to fabricate integrated circuits. There are several reasons for this dominance. CMOS ICs can be laid out in a small area. They can handle very high operating speeds while dissipating relatively low power. Perhaps the most important aspect of CMOS's dominance is its manufacturability. CMOS circuits can be fabricated with few defects. Equally important, the cost to fabricate in CMOS has been kept low by shrinking devices (scaling) with each new generation of technology. This also, for digital circuits, is significant because in many cases the same layout can be used from one fabrication size (process technology node) to the next via simple scaling.

Analog Design in CMOS

While initially CMOS was used exclusively for digital design, the constant push to lower costs and increase the functionality of ICs has resulted in it being used for analog-only, analog/digital, and mixed-signal (chips that combine analog circuits with digital signal processing) designs. The main concern when using CMOS for an analog design is matching. Matching is a term used to describe how well two identical transistors' characteristics match electrically. How well circuits "match" is often the limitation in the quality of a design (e.g., the clarity of a monitor, the accuracy of a measurement, etc.).

1.3 An Introduction to SPICE

The simulation program with an integrated circuit emphasis (SPICE) is a ubiquitous software tool for the simulation of circuits. In this section we'll provide an overview of SPICE. In addition, we'll provide some basic circuit analysis examples for quick reference or as a review. Note that the reader should review the links at CMOSedu.com for SPICE download and installation information. In addition, the examples from the book are available at this website. Note that all SPICE engines use a text file (a netlist) for simulation input.

Generating a Netlist File

We can use, among others, the Window's notepad or wordpad programs to create a SPICE netlist. SPICE likes to see files with "*.cir, *.sp, or *.spi" (among others) extensions. To save a file with these extensions, place the file name and extension in quotes, as seen in Fig. 1.9. If quotes are not used, then Windows may tack on ".txt" to the filename. This can make finding the file difficult when opening the netlist in SPICE.

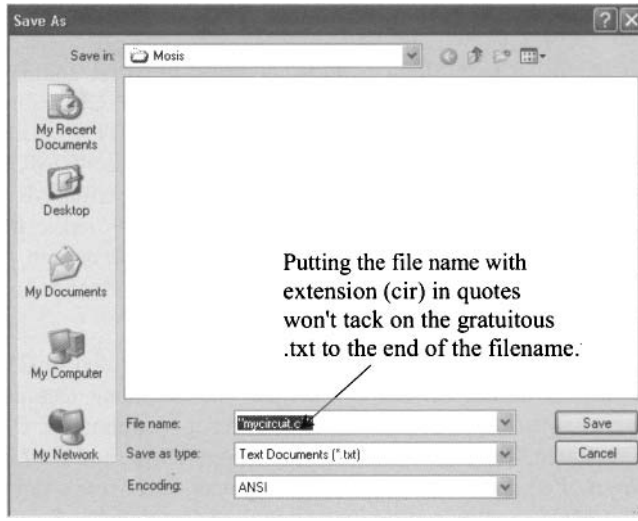


Figure 1.9 Saving a text file with a ".cir" extension.

Operating Point

The first SPICE simulation analysis we'll look at is the .op or operating point analysis. An operating point simulation's output data is not graphical but rather simply a list of node voltages, loop currents, and, when active elements are used, small-signal AC parameters. Consider the schematic seen in Fig. 1.10. The SPICE netlist used to simulate this circuit may look like the following (again, remember, that all of these simulation examples are available for download at CMOSedu.com):

```
*** Figure 1.10 CMOS: Circuit Design, Layout, and Simulation ***

*#destroy all
*#run
*#print all

.op

Vin    1     0     DC    1
R1     1     2     1k
R2     2     0     2k

.end
```

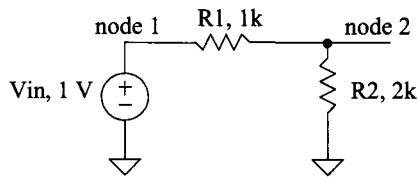


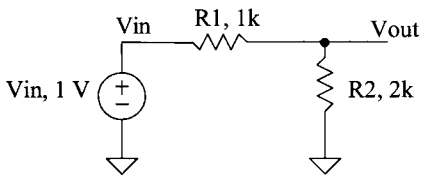
Figure 1.10 Operation point simulation for a resistive divider.

The first line in a netlist is a title line. SPICE ignores the first line (important to avoid frustration!). A comment line starts with an asterisk. SPICE ignores lines that start with a * (in most cases). In the netlist above, however, the lines that start with *# are command lines. These command lines are used for control in some SPICE simulation programs. In other SPICE programs, these lines are simply ignored. The commands in this netlist destroy previous simulation data (so we don't view the old data), run the simulation, and then print the simulation output data. SPICE analysis commands start with a period. Here we are performing an operating point analysis. Following the .op, we've specified an input voltage source called Vin (voltage source names must start with a V, resistor names must start with an R, etc.). connected from node 1 to ground (ground always has a node name of 0 [zero]). We then have a 1k resistor from node 1 to node 2 and a 2k resistor from node 2 to ground. Running the simulation gives the following output:

```
v(1) = 1.000000e+00
v(2) = 6.666667e-01
vin#branch = -3.33333e-04
```

The node voltages, as we would expect, are 1 V and 667 mV, respectively. The current flowing through Vin is 333 μA. Note that SPICE defines positive current flow as from the + terminal of the voltage source to the – terminal (hence, the current above is negative).

It's often useful to use names for nodes that have meaning. In Fig. 1.11, we replaced the names node 1 and 2 with Vin and Vout. Vin corresponds to the input voltage source's name. This is useful when looking at a large amount of data. Also seen in Fig. 1.11 is the modified netlist.




```
*** Figure 1.11 CMOS ***
*#destroy all
*#run
*#print all
.op
Vin Vin 0 DC 1
R1 Vin Vout 1k
R2 Vout 0 2k
.end
```

Figure 1.11 Operation point simulation for a resistive divider.

Transfer Function Analysis

The transfer function analysis can be used to find the DC input and output resistances of a circuit as well as the DC transfer characteristics. To give an example, let's replace, in the netlist seen above, .op with

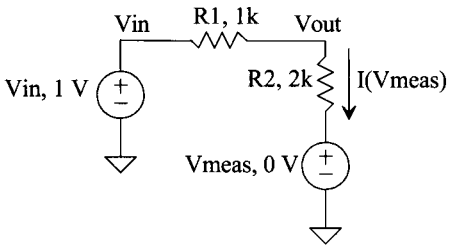
```
.TF     V(Vout,0)       Vin
```

The output is defined as the voltage between nodes Vout and 0 (ground). The input is a source (here a voltage source). When we run the simulation with this command line, we get an output of

```
transfer_function = 6.666667e-01
output_impedance_at_v(vout,0) = 6.666667e+02
vin#input_impedance = 3.000000e+03
```

As expected, the "gain" of this voltage divider is 2/3, the input resistance is 3k (1k + 2k), and the output resistance is 667 Ω (1k||2k).

As another example of the use of the .tf command consider adding the 0 V voltage source to Fig. 1.11, as seen in Fig. 1.12. Adding a 0 V source to a circuit is a common method to measure the current in an element (we plot or print I(Vmeas) for example).




```
*** Figure 1.12 CMOS ***
*#destroy all
*#run
*#print all
.TF I(Vmeas) Vin
Vin Vin 0 DC 1
R1 Vin Vout 1k
R2 Vout Vmeas 2k
Vmeas Vmeas 0 DC 0
.end
```

Figure 1.12 Measuring the transfer function in a resistive divider when the output variable is the current through R2 and the input is Vin.

Here, in the .tf analysis, we have defined the output variable as a current, I(Vmeas) and the input as the voltage, Vin. Running the simulation, we get an output of

```
transfer_function = 3.333333e-04
vin#input_impedance = 3.000000e+03
vmeas#output_impedance = 1.000000e+20
```

The gain is I(Vmeas)/Vin or 1/3k (= 333 μmhos), the input resistance is still 3k, and the output resistance is now an open (Vmeas is removed from the circuit).

The Voltage-Controlled Voltage Source

SPICE can be used to model voltage-controlled voltage sources (VCVS). Consider the circuit seen in Fig. 1.13. The specification for a VCVS starts with an E in SPICE. The netlist for this circuit is

```
*** Figure 1.13 CMOS: Circuit Design, Layout, and Simulation ***

*#destroy all
*#run
*#print all

.TF     V(Vout,0) Vin

Vin     Vin     0       DC      1
R1      Vb      0       3k
R2      Vt      Vout    1k
R3      Vout    0       2k
E1      Vt      Vb      Vin     0       23

.end
```

The first two nodes (Vt and Vb), following the VCVS name E1, are the VCVS outputs (the first node is the + output). The second two nodes (Vin and ground) are the controlling nodes. The gain of the VCVS is, in this example, 23. The voltage between Vt and Vb is 23·Vin. Running this simulation gives an output of

```
transfer_function = 7.666667e+00
output_impedance_at_v(vout,0) = 1.333333e+03
vin#input_impedance = 1.000000e+20
```

Notice that the input resistance is infinite.

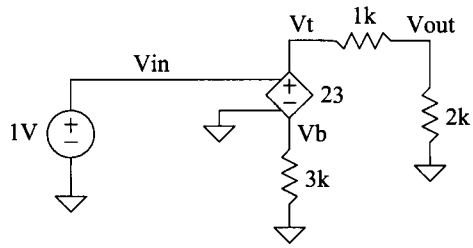
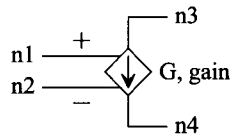


Figure 1.13 Example using a voltage-controlled voltage source.

An Ideal Op-Amp

We can implement a (near) ideal op-amp in SPICE with a VCVS or with a voltage-controlled current source (VCCS), Fig. 1.14. It turns out that using a VCCS to implement an op-amp in SPICE results, in general, in better simulation convergence. The input voltage, the difference between nodes n1 and n2 in Fig. 1.14, is multiplied by the transconductance G (units of amps/volts or mhos) to cause a current to flow between n3 and n4. Note that the input resistance of the VCCS, the resistance seen at n1 and n2, is infinite.



Voltage-Controlled Current Source (VCCS)
G1 n3 n4 n1 n2 G

Figure 1.14 Voltage-controlled current source in SPICE.

Figure 1.15 shows the implementation of an ideal op-amp in SPICE along with an example circuit. The open-loop gain of the op-amp is a million (the product of the VCCS's transconductance with the 1-ohm resistor). Note how we've flipped the polarity of the (SPICE model of the) op-amp's input to ensure a rising voltage on the noninverting input (+ input) causes Vout to increase. The closed-loop gain is –3 (if this isn't obvious then the reader should revisit sophomore circuits before going too much further in the book).

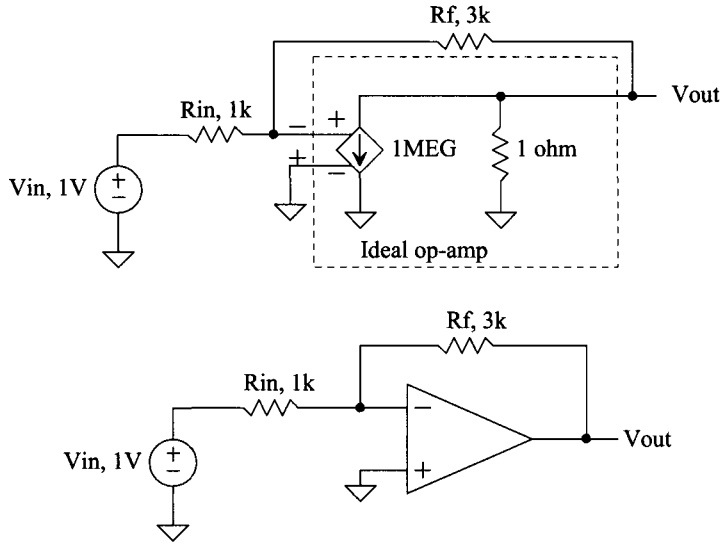


Figure 1.15 An op-amp simulation example.

The Subcircuit

In a simulation we may want to use a circuit, like an op-amp, more than once. In these situations we can generate a subcircuit and then, in the main part of the netlist, call the circuit as needed. Below is the netlist for simulating, using a transfer function analysis, the circuit in Fig. 1.15 where the op-amp is specified using a subcircuit call.

```
*** Figure 1.15 CMOS: Circuit Design, Layout, and Simulation ***

*#destroy all
*#run
*#print all

.TF     V(Vout,0) Vin

Vin     Vin     0       DC      1
Rin     Vin     Vm      1k
Rf      Vout    Vm      3k

X1      Vout    0       vm      Ideal_op_amp

.subckt Ideal_op_amp    Vout    Vp      Vm
G1      Vout    0       Vm      Vp      1MEG
RL      Vout    0       1
.ends
.end
```

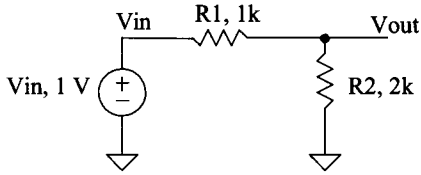
Notice that a subcircuit call begins with the letter X. Note also how we've called the noninverting input (the + input) Vp and not V+ or +. Some SPICE simulators don't like + or – symbols used in a node's name. Further note that a subcircuit ends with .ends (end subckt). Care must be exercised with using either .end or .ends. If, for example, a .end is placed in the middle of the netlist all of the SPICE netlist information following this .end is ignored.

The output results for this simulation are seen below. Note how the ideal gain is –3 where the simulated gain is –2.99999. Our near-ideal op-amp has an open-loop gain of one million and thus the reason for the slight discrepancy between the simulated and calculated gains. Also note how the input resistance is 1k, and the output resistance, because of the feedback, is essentially zero.

```
transfer_function = -2.99999e+00
output_impedance_at_v(vout,0) = 3.999984e-06
vin#input_impedance = 1.000003e+03
```

DC Analysis

In both the operating point and transfer function analyses, the input to the circuit was constant. In a DC analysis, the input is varied and the circuit's node voltages and currents (through voltage sources) are simulated. A simple example is seen in Fig. 1.16. Note how we are now plotting, instead of printing, the node voltages. We could also plot the current through Vin (plot Vin#branch). The .dc command specifies that the input source, Vin, should be varied from 0 to 1 V in 1 mV steps. The x-axis of the simulation results seen in the figure is the variable we are sweeping, here Vin. Note that, as expected, the slope of the Vin curve is one (of course) and the slope of Vout is 2/3 (= Vout/Vin).




```
*** Figure 1.16 CMOS ***
*#destroy all
*#run
*#plot Vin Vout
.dc Vin 0 1 1m
Vin Vin 0 DC 1
R1 Vin Vout 1k
R2 Vout 0 2k
.end
```

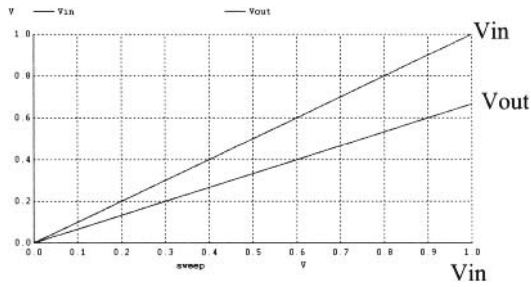
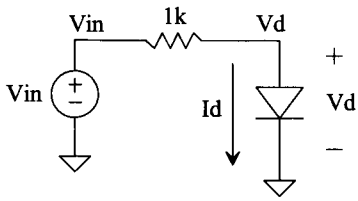


Figure 1.16 DC analysis simulation for a resistive divider.

Plotting IV Curves

One of the simulations that is commonly performed using a DC analysis is plotting the current-voltage (IV) curves for an active device (e.g., diode or transistor). Examine the simulation seen in Fig. 1.17. The diode is named D1. (Diodes must have names that start with a D.) The diode's anode is connected to node Vd, while its cathode is connected to




```
*** Figure 1.17 CMOS ***
*#destroy all
*#run
*#let ID=-Vin#branch
*#plot ID
.dc Vin 0 1 1m
Vin Vin 0 DC 1
R1 Vin Vd 1k
D1 Vd 0 mydiode
.model mydiode D
.end
```

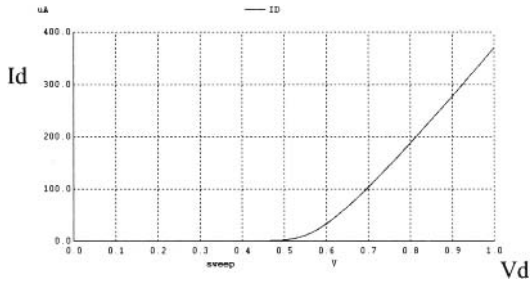


Figure 1.17 Plotting the current-voltage curve for a diode.

ground. This is our first introduction to the .model specification. Here our diode's model name is mydiode. The .model parameter D seen in the netlist simply indicates a diode model. We don't have any parameters after the D in this simulation, so SPICE uses default parameters. The interested reader is referred to Table 2.1 on page 47 for additional information concerning modeling diodes in SPICE. Note, again, that SPICE defines positive current through a voltage source as flowing from the + terminal to the – terminal (hence why we defined the diode current the way we did in the netlist).

Dual Loop DC Analysis

An outer loop can be added to a DC analysis, Fig. 1.18. In this simulation we start out by setting the base current to 5 μA and sweeping the collector-emitter voltage from 0 to 5 V in 1 mV steps. The output data for this particular simulation is the trace, seen in Fig. 1.18, with a label of "Ib=5u." The base current is then increased by 5 μA to 10 μA, and the collector-emitter voltage is stepped again (resulting in the trace labeled "Ib=10u". This continues until the final iteration when Ib is 25 μA. Other examples of using a dual-loop DC analysis for MOSFET IV curves are found in Figs. 6.11, 6.12, and 6.13.

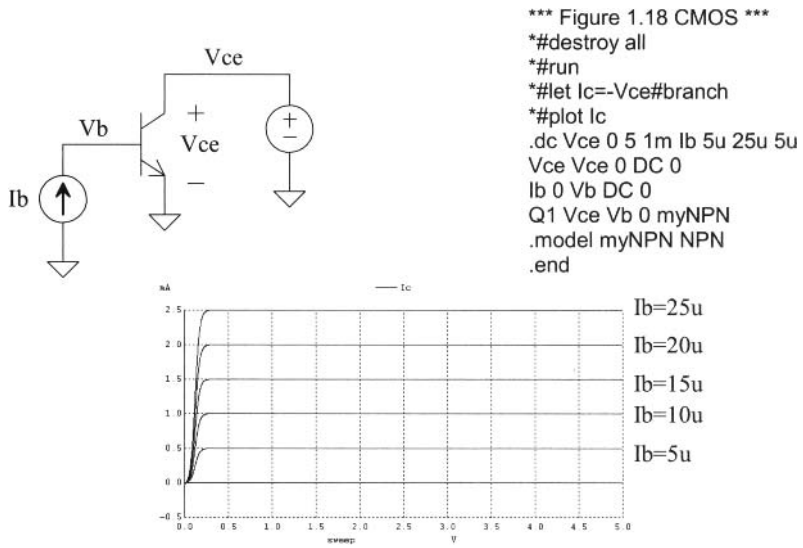


Figure 1.18 Plotting the current-voltage curves for an NPN BJT.

Transient Analysis

The form of the transient analysis statement is

```
.tran tstep tstop <tstart> <tmax> <uic>
```

where the terms in < > are optional. The tstep term indicates the (suggested) time step to be used in the simulation. The parameter tstop indicates the simulation's stop time. The starting time of a simulation is always time equals zero. However, for very large (data) simulations, we can specify a time to start saving data, tstart. The tmax parameter is used to specify the maximum step size. If the plots start to look jagged (like a sinewave that isn't smooth), then tmax should be reduced.

A SPICE transient analysis simulates circuits in the time domain (as in an oscilloscope, the x-axis is time). Let's simulate, using a transient analysis, the simple circuit seen back in Fig. 1.11. A simulation netlist may look like (see output in Fig. 1.19):

```
*** Figure 1.19 CMOS: Circuit Design, Layout, and Simulation ***

*#destroy all
*#run
*#plot vin vout

.tran 100p 100n

Vin     Vin     0       DC      1
R1      Vin     Vout    1k
R2      Vout    0       2k

.end
```

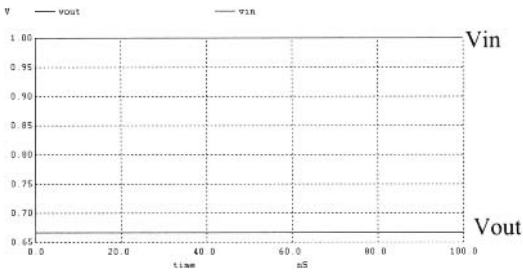


Figure 1.19 Transient simulation for the circuit in Fig. 1.11.

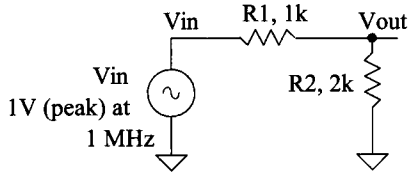
The SIN Source

To illustrate a simulation using a sinewave, examine the schematic in Fig 1.20. The statement for a sinewave in SPICE is

```
SIN Vo Va freq <td> <theta>
```

The parameter Vo is the sinusoid's offset (the DC voltage in series with the sinewave). The parameter Va is the peak amplitude of the sinewave. Freq is the frequency of the sinewave, while td is the delay before the sinewave starts in the simulation. Finally, theta is used if the amplitude of the sinusoid has a damped nature. Figure 1.20 shows the netlist corresponding to the circuit seen in this figure and the simulation results.

Some key things to note in this simulation: (1) MEG is used to specify 10^6. Using "m" or "M" indicates milli or 10^{-3}. The parameter 1MHz indicates 1 milliHertz. Also, f indicates femto or 10^{-15}. A capacitor value of 1f doesn't indicate one Farad but rather 1 femto Farad. (2) Note how we increased the simulation time to 3 μs. If we had a simulation time of 100 ns (as in the previous simulation), we wouldn't see much of the sinewave (one-tenth of the sinewave's period). (3) The "SIN" statement is used in a transient simulation analysis. The SIN specification is **not** used in an AC analysis (discussed later).




```
*** Figure 1.20 ***
*#destroy all
*#run
*#plot vin vout
.tran 1n 3u
Vin Vin 0 DC 0 SIN 0 1 1MEG
R1 Vin Vout 1k
R2 Vout 0 2k
.end
```

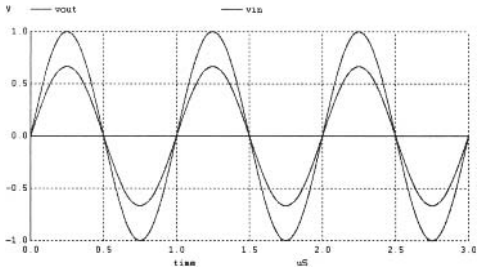


Figure 1.20 Simulating a resistive divider with a sinusoidal input.

An RC Circuit Example

To illustrate the use of a .tran simulation let's determine the output of the RC circuit seen in Fig. 1.21 and compare our hand calculations to simulation results. The output voltage can be written in terms of the input voltage by

$$V_{out} = V_{in} \cdot \frac{1/j\omega C}{1/j\omega C + R} \text{ or } \frac{V_{out}}{V_{in}} = \frac{1}{1 + j\omega RC} \tag{1.1}$$

Taking the magnitude of this equation gives

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{1}{\sqrt{1 + (2\pi fRC)^2}} \tag{1.2}$$

and taking the phase gives

$$\angle\frac{V_{out}}{V_{in}} = -\tan^{-1}\frac{2\pi fRC}{1} \tag{1.3}$$

From the schematic the resistance is 1k, the capacitance is 1 μF, and the frequency is 200 Hz. Plugging these numbers into Eqs. (1.1) – (1.3) gives $\left|\frac{V_{out}}{V_{in}}\right| = 0.623$ and $\angle\frac{V_{out}}{V_{in}} =$ −0.898 radians or −51.5 degrees. With a 1 V peak input then our output voltage is 623 mV (and as seen in Fig. 1.21, it is). Remembering that phase shift is simply an indication of time delay at a particular frequency,

$$\angle \text{ (radians)} = \frac{t_d}{T} \cdot 2\pi \text{ or } \angle \text{ (degrees)} = \frac{t_d}{T} \cdot 360 = t_d \cdot f \cdot 360 \tag{1.4}$$

The way to remember this equation is that the time delay, t_d, is a percentage of the period (T), t_d/T, multiplied by either 2π (radians) or 360 (degrees). For the present example, the time delay is 715 μs (again, see Fig. 1.21). Note that the minus sign indicates that the output is lagging (occurring later in time) the input (the input leads the output).

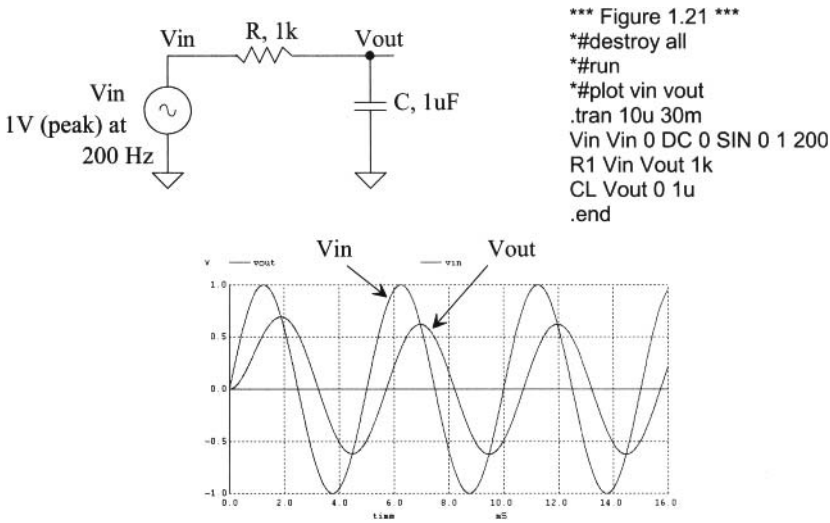


Figure 1.21 Simulating the operation of an RC circuit using a .tran analysis.

Another RC Circuit Example

As one more example of simulating the operation of an RC circuit consider the circuit seen in Fig. 1.22. Combining the impedances of C1 and R, we get

$$Z = \frac{R/j\omega C_1}{R + 1/j\omega C_1} = \frac{R}{1 + j\omega RC_1} \tag{1.5}$$

The transfer function for this circuit is then

$$\frac{V_{out}}{V_{in}} = \frac{1/j\omega C_2}{1/j\omega C_2 + Z} = \frac{1 + j\omega RC_1}{1 + j\omega R(C_1 + C_2)} \tag{1.6}$$

The magnitude of this transfer function is

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{\sqrt{1 + (2\pi fRC_1)^2}}{\sqrt{1 + (2\pi fR \cdot (C_1 + C_2))^2}} \tag{1.7}$$

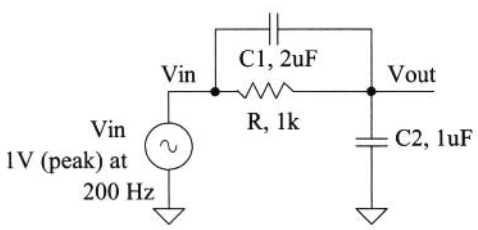
and the phase response is

$$\angle\frac{V_{out}}{V_{in}} = \tan^{-1}\frac{2\pi fRC_1}{1} - \tan^{-1}\frac{2\pi fR(C_1 + C_2)}{1} \tag{1.8}$$

Plugging in the numbers from the schematic gives a magnitude response of 0.6 (which matches the simulation results) and a phase shift of – 0.119 radians or – 6.82 degrees. The amount of time the output is lagging the input is then

$$t_d = \frac{T \cdot \angle}{360} = \frac{\angle}{f \cdot 360} = \frac{-6.82}{200 \cdot 360} = -95\ \mu s \tag{1.9}$$

which is confirmed with the simulation results seen in Fig. 1.22.




```
*** Figure 1.22 ***
*#destroy all
*#run
*#plot vin vout
.tran 10u 30m
Vin Vin 0 DC 0 SIN 0 1 200
R1 Vin Vout 1k
C1 Vin Vout 2u
C2 Vout 0 1u
.end
```

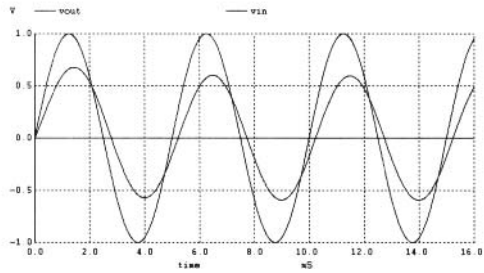


Figure 1.22 Another RC circuit example.

AC Analysis

When performing a transient analysis (.tran) the x-axis is time. We can determine the frequency response of a circuit (the x-axis is frequency) using an AC analysis (.ac). An AC analysis is specified in SPICE using

```
.ac dec nd fstart fstop
```

The dec indicates that the x-axis should be plotted in decades. We could replace dec with lin (linear plot on the x-axis) or oct (octave). The term nd indicates the number of points per decade (say 100), while fstart and fstop indicate the start and stop frequencies (note that fstart cannot be zero, or DC, since this isn't an AC signal). The netlist used to simulate the AC response of the circuit in Fig. 1.21 follows. The simulation output is seen in Fig. 1.23, where we've pointed out the response at 200 Hz (the frequency used in Fig. 1.21 and used for calculations on page 17).

```
*** Figure 1.23 CMOS: Circuit Design, Layout, and Simulation ***

*#destroy all
*#run
*#plot db(vout/vin)
*#set units=degrees
*#plot ph(vout/vin)

.ac dec 100 1 10k

Vin     Vin     0       DC      0       SIN 0 1 200     AC 1
R1      Vin     Vout    1k
CL      Vout    0       1u

.end
```

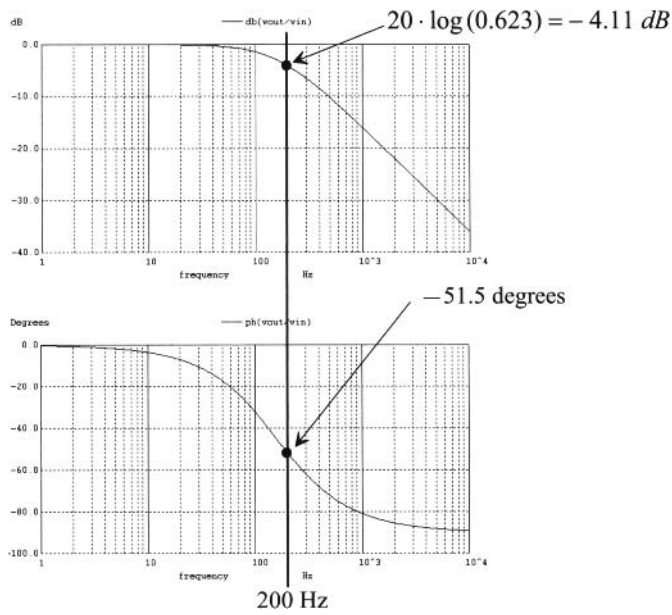


Figure 1.23 AC simulation for the RC circuit in Fig. 1.21.

Note in this netlist that the SIN specification in Vin has nothing to do with an AC analysis (it's ignored for an AC analysis). For the AC analysis, we added, to the statement for Vin, the term AC 1 (specifying that the magnitude or peak of the AC signal is 1). We can add a phase shift of 45 degrees by using AC 1 45 in the statement.

Decades and Octaves

In the simulation results seen in Fig. 1.23 we used decades. When we talk about decades we either are multiplying or dividing by 10. One decade above 23 MHz is 230 MHz, while one decade below 1.2 kHz is 120 Hz.

When we talk about octaves, we talk about either multiplying or dividing by 2. One octave above 23 MHz is 46 MHz while one octave below 1.2 kHz is 600 Hz. Two octaves above 23 MHz is (multiply by 4) 92 MHz.

Decibels

When the magnitude response of a transfer function decreases by 10, it is said it goes down by −20 dB (divide by 10, $20 \cdot \log(0.1) = -20\,dB$). When the magnitude response increases by 10, it goes up by 20 dB (multiply by 10). For the frequency response in Fig. 1.23 (above 159 Hz, the −3 dB frequency, or here when the magnitude response is 0.707), the response is rolling off at −20 dB/decade. What this means is that if we increase the frequency by 10 the magnitude response decreases by 10. We could also say the response is rolling off at −6 dB/octave above 159 Hz (for every increase in frequency by 2 the magnitude response drops by a factor of 2). If a magnitude response is rolling off at −40 dB/decade, then for every increase in frequency by 10 the magnitude drops by 100. Similarly if a response rolls off at −12 dB/octave, for every doubling in frequency our response drops by 4. Note that −6 dB/octave is the same rate as −20 dB/decade.

Pulse Statement

The SPICE pulse statement is used in transient simulations to specify pulses or clock signals. This statement has a format given by

```
pulse vinit vfinal td tr tf pw per
```

The pulse's initial voltage is vinit while vfinal is the pulse's final (or pulsed) value, td is the delay before the pulse starts, tr and tf are the rise and fall times, respectively, of the pulse (noting that when these are set to zero the step size used in the transient simulation is used), pw is the pulse's width; and per is the period of the pulse. Figure 1.24 provides an example of a simulation that uses the pulse statement. A section of the netlist used to generate the waveforms in this figure follows.

```
.tran 100p 30n

Vin    Vin    0      DC    0      pulse 0 1 6n 0 0 3n 10n
R1     Vin    Vout   1k
C1     Vout   0      1p
```

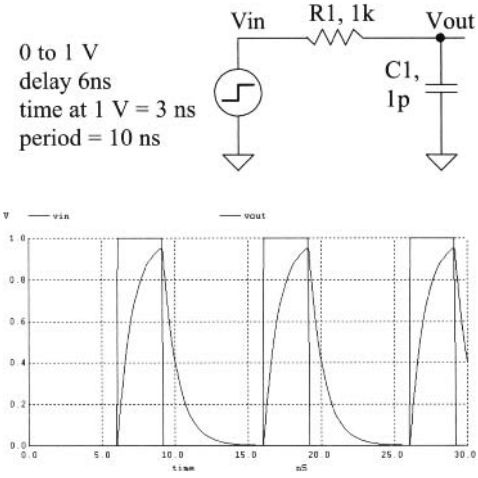


Figure 1.24 Simulating the step response of an RC circuit using a pulsed source voltage.

Finite Pulse Rise Time

Notice, in the simulation results seen in Fig. 1.24, that the rise and fall times of the input pulse are not 0 as specified in the pulse statement but rather 100 ps as specified by the suggested maximum step size in the .tran statement. Figure 1.25 shows the simulation results if we change the pulse statement to

```
Vin    Vin    0      DC    0      pulse 0 1 6n 10p 10p 3n 10n
```

where we've specified 10 ps rise and fall times. Note that in some SPICE simulators you must specify a maximum step size in the .tran statement. You could do this in the .tran statement above by using .tran 10p 30n 0 **10p** (where the 10p is the maximum step size and the simulation starts saving data at 0.)

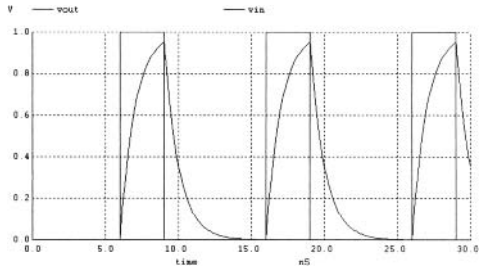


Figure 1.25 Specifying a rise time in the pulse statement to avoid slow rise times (rise times set by the maximum step size in the .tran statement.)

Step Response

The pulse statement can also be used to generate a step function

```
Vin Vin 0 DC 0 pulse 0 1 2n 10p
```

We've reduced the delay to 2n and have specified (only) a rise time for the pulse. Since the pulse width isn't specified, the pulse transitions and then stays high for the extent of the simulation. Figure 1.26 shows the step response for the RC circuit seen in Fig. 1.24.

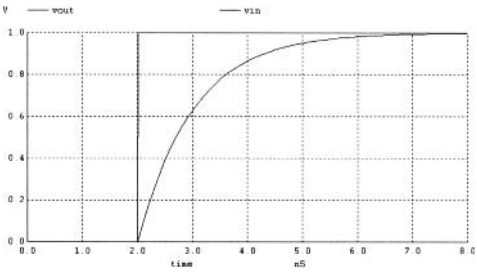


Figure 1.26 Step response of an RC circuit.

Delay and Rise Time in RC Circuits

From the RC circuit review on page 50 we can write the delay time, the time it takes the pulse to reach 50% of its final value in an RC circuit, using

$$t_d \approx 0.7RC \tag{1.10}$$

and the rise time (or fall time) as

$$t_r \approx 2.2RC \tag{1.11}$$

Using the RC in Fig. 1.24 (1 ns), we get a (calculated) delay time of 700 ps and a rise time of 2.2 ns. These numbers are verified in Fig. 1.26. To show that the pulse statement can be used for other amplitude steps consider resimulating the circuit in Fig. 1.24 (see Fig. 1.27) with an input pulse that transitions from –1 to –2 V (note how the delay and transition times remain unchanged. The SPICE pulse statement is now

```
Vin Vin 0 DC 0 pulse -1 -2 2n 10p
```

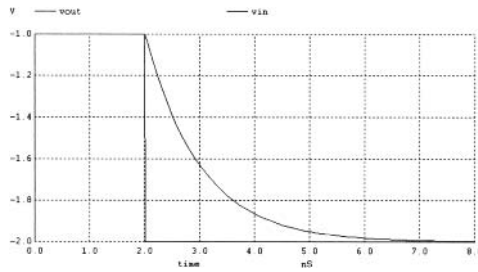


Figure 1.27 Another step response (negative going) of an RC circuit.

Piece-Wise Linear (PWL) Source

The piece-wise linear (PWL) source specifies arbitrary waveform shapes. The SPICE statement for a PWL source is

```
pwl t1 v1 t2 v2 t3 v3 ... <rep>
```

To provide an example using a PWL voltage source, examine Fig. 1.28. The input waveform in this simulation is specified using

```
pwl 0 0.5 3n 1 5n 1 5.5n 0 7n 0
```

At 0 ns, the input voltage is 0.5 V. At 3 ns the input voltage is 1 V. Note the linear change between 0 and 3 ns. Each pair of numbers, the first the time and the second the voltage (or current if a current source is used) represent a point on the PWL waveform. Note that in some simulators the specification for a PWL source may be quite long. In these situations a text file is specified that contains the PWL for the simulation.

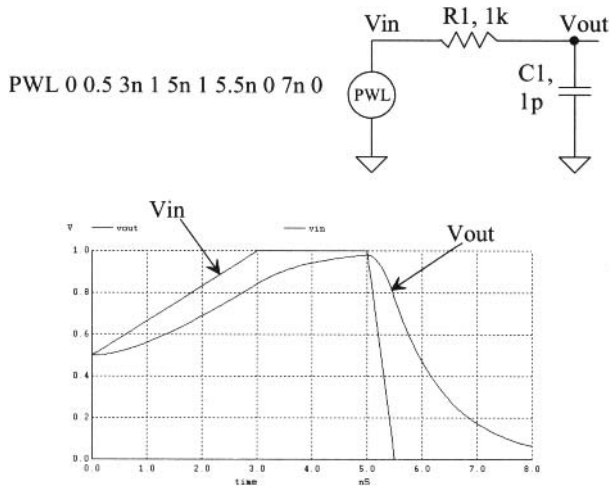


Figure 1.28 Using a PWL source to drive an RC circuit.

Simulating Switches

A switch can be simulated in SPICE using the following (for example) syntax

```
s1 node1 node2 controlp controlm switmod
.model switmod sw ron=1k
```

The name of a switch must start with an s. The switch is connected between node1 and node2, as seen in Fig. 1.29. When the voltage on node controlp is greater than the voltage on node controlm, the switch closes. The switch is modeled using the .model statement. As seen above, we are setting the series resistance of the switch to 1k.

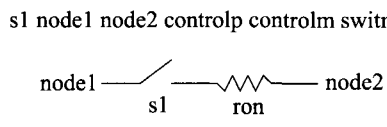
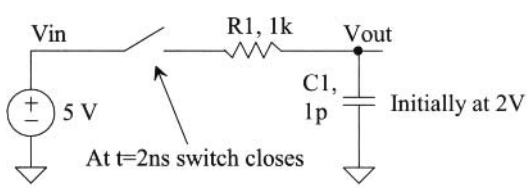


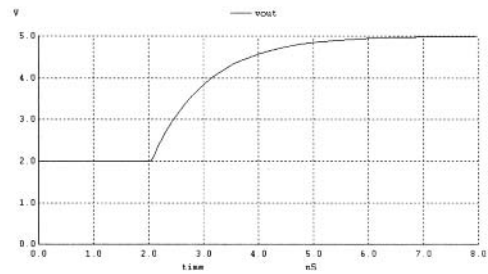
Figure 1.29 Modeling a switch in SPICE.

Initial Conditions on a Capacitor

An example of a circuit that uses both a switch and an initial voltage on a capacitor is seen in Fig. 1.30. Notice, in the netlist, that we have added UIC to the end of the .tran statement. This addition makes SPICE "use initial conditions" or skip an initial operating point calculation. Also note that to set the initial voltage across the capacitor we simply added IC=2 to the end of the statement for a capacitor. To set a node to a voltage (that may have a capacitor connected to it or not), we can add, for example,

```
.ic v(vout)=2
```



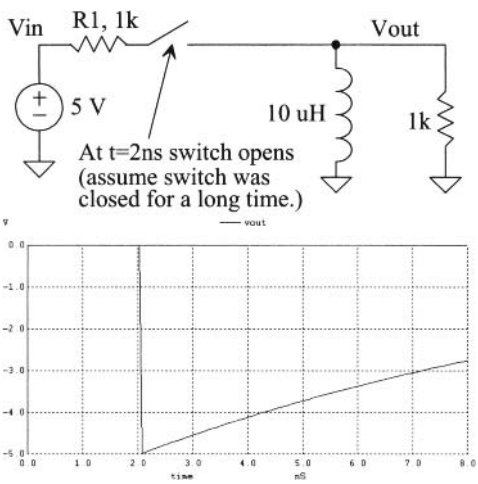



```
*** Figure 1.30 ***

*#destroy all
*#run
*#plot vout

.tran 100p 8n UIC

Vclk clk 0 pulse -1 1 2n
Vin Vin 0 DC 5
S1 Vin Vouts clk 0 switmodel
R1 Vouts Vout 1k
C1 Vout 0 1p IC=2

.model switmodel sw ron=0.1

.end
```

Figure 1.30 Using initial conditions and a switch in an RC circuit simulation.

Initial Conditions in an Inductor

Consider the circuit seen in Fig. 1.31. Here we assume that the switch has been closed for a long period of time so that the circuit reaches steady-state. The inductor shorts the output to ground and the current flowing in the inductor is 5 mA. To simulate this initial condition, we set the current in the inductor using the IC statement as seen in the netlist (remembering to include the UIC in the .tran statement). At 2 ns after the simulation starts, we open the switch (the control voltage connections are switched from the previous simulations). Since we know we can't change the current through an inductor instantaneously (the inductor wants to keep pulling 5 mA), the voltage across the inductor will go from 0 to −5 V. The inductor will pull the 5 mA of current through the 1k resistor connected to the output node. Note that we select the transient simulation time by looking at the time constant, L/R, of the circuit (here 10 ns).




```
*** Figure 1.31 ***

*#destroy all
*#run
*#plot vout

.tran 100p 8n UIC

Vclk clk 0 pulse -1 1 2n
Vin Vin 0 DC 5
S1 Vin Vouts 0 clk switmodel
R1 Vouts Vout 1k
R2 Vout 0 1k
L1 Vout 0 10u IC=5m

.model switmodel sw ron=0.1

.end
```

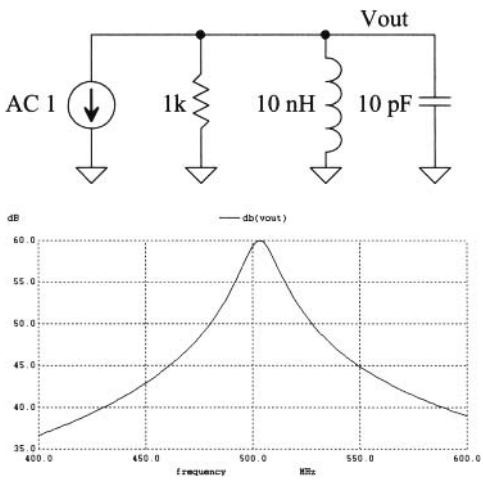
Figure 1.31 Using initial conditions in an inductive circuit.

Q of an LC Tank

Figure 1.32 shows a simulation useful in determining the quality factor or Q of a parallel LC circuit (a tank, used in communication circuits among others). The current source and resistor may model a transistor. The resistor can also be used to model the losses in the capacitor or inductor. Quality factor for a resonant circuit is defined as the ratio of the energy stored in the tank to the energy lost. Our circuit definition for Q is the ratio of the center (resonant) frequency to the bandwidth of the response at the 3 dB points. We can write an equation for this circuit definition of Q as

$$Q = \frac{f_{center}}{BW} = \frac{f_{center}}{f_{3dBhigh} - f_{3dBlow}} \tag{1.12}$$

The center frequency of the circuit in Fig. 1.32 is roughly 503 MHz, while the upper 3 dB frequency is 511.2 MHz and the lower 3 dB frequency is 494.8 MHz. The Q is roughly 30. Note the use of linear plotting in the ac analysis statement.




```
*** Figure 1.32 ***

*#destroy all
*#run
*#plot db(vout)

.AC lin 100 400MEG 600MEG

Iin Vout 0 DC 0 AC 1
R1 Vout 0 1k
L1 Vout 0 10n
C1 Vout 0 10p

.end
```

Figure 1.32 Determining the Q, or quality factor, of an LC tank.

Frequency Response of an Ideal Integrator

The frequency response of the integrator seen in Fig. 1.33 can be determined knowing the op-amp keeps the inverting input terminal at the same potential as the non-inverting input (here ground). The current through the resistor must equal the current through the capacitor so

$$\frac{V_{in}}{R} + \frac{V_{out}}{1/j\omega C} = 0 \tag{1.13}$$

or

$$\frac{V_{out}}{V_{in}} = \frac{-1}{j\omega RC} = \frac{-(1 + j \cdot 0)}{0 + j\omega RC} \tag{1.14}$$

The magnitude of the integrator's transfer function is

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{\sqrt{(-1)^2 + (-0)^2}}{\sqrt{(0)^2 + (2\pi fRC)^2}} = \frac{1}{2\pi RCf} \tag{1.15}$$

while the phase shift through the integrator is

$$\angle\frac{V_{out}}{V_{in}} = \tan^{-1}\frac{-0}{-1} - \tan^{-1}\frac{2\pi RCf}{0} = -90° \tag{1.16}$$

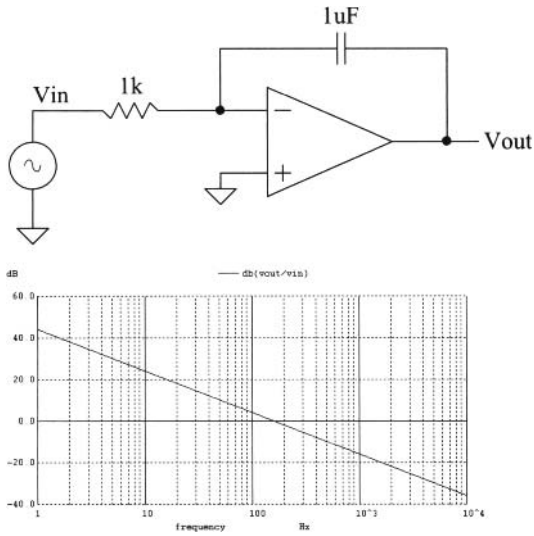
Note that the gain of the integrator approaches infinity as the frequency decreases towards DC while the phase shift is constant.

Unity-Gain Frequency

It's of interest to determine the frequency where the magnitude of the transfer function is unity (called the unity-gain frequency, f_{un}.) Using Eq. (1.15), we can write

$$\left|\frac{V_{out}}{V_{in}}\right| = 1 = \frac{1}{2\pi RCf_{un}} \rightarrow f_{un} = \frac{1}{2\pi RC} \tag{1.17}$$

Using the values seen in the schematic, the unity-gain frequency is 159 Hz (as verified in the SPICE simulation seen in Fig. 1.33).




```
*** Figure 1.33 ***

*#destroy all
*#run
*#plot db(vout/vin)
*#set units=degrees
*#plot ph(vout/vin)

.ac dec 100 1 10k

Vin Vin 0 DC 1 AC 1
Rin Vin vm 1k
Cf Vout vm 1u

X1 Vout 0 vm Ideal_op_amp
.subckt Ideal_op_amp Vout Vp Vm
G1 Vout 0 Vm Vp 1MEG
RL Vout 0 1
.ends
.end
```

Figure 1.33 An integrator example.

Time-Domain Behavior of the Integrator

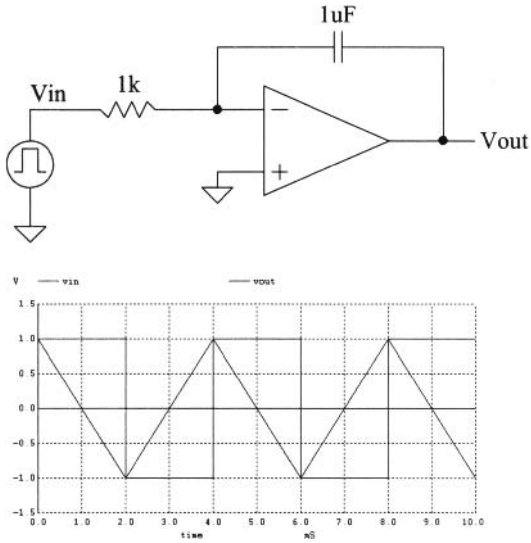
The time-domain behavior of the integrator can be characterized, again, by equating the current in the resistor with the current in the capacitor

$$V_{out} = \frac{1}{C}\int \frac{V_{in}}{R} \cdot dt \tag{1.18}$$

If our input is a constant voltage, then the output is a linear ramp increasing (if the input is negative) or decreasing (if the input is positive) with time. If the input is a squarewave, with zero mean then the output will look like a triangle wave. Using the values seen in Fig. 1.33 for the time-domain simulation seen in Fig. 1.34, we can estimate that if a 1 V signal is applied to the integrator the output voltage will have a slope of

$$V_{out}(t) = \frac{V_{in}}{RC} = \frac{1}{1\ ms} \tag{1.19}$$

or 1 V/ms slope. This equation can be used to design a sawtooth waveform generator from an input squarewave. Note, however, there are several practical concerns. To begin, we set the output of the integrator, using the .ic statement, to ground at the beginning of the simulation. In a real circuit this may be challenging (one method is to add a reset switch across the capacitor). Another issue, discussed later in the book, is the op-amp's offset voltage. This will cause the outputs to move towards the power supply rails even with no input applied. Finally, notice that putting a + in the first column treats the SPICE code as if it were continued from the previous line.




```
*** Figure 1.34 ***

*#destroy all
*#run
*#plot vout vin

.tran 10u 10m
.ic v(vout)=0

Vin Vin 0 DC 1
+ pulse -1 1 0 1u 1u 2m 4m
Rin Vin vm 1k
Cf Vout vm 1u

X1 Vout 0 vm Ideal_op_amp
.subckt Ideal_op_amp Vout Vp Vm
G1 Vout 0 Vm Vp 1MEG
RL Vout 0 1
.ends
.end
```

Figure 1.34 Time-domain integrator example.

Convergence

A netlist that doesn't simulate isn't converging numerically. *Assuming* that the circuit contains no connection errors, there are basically three parameters that can be adjusted to help convergence: ABSTOL, VNTOL, and RELTOL.

ABSTOL is the absolute current tolerance. Its default value is 1 pA. This means that when a simulated circuit gets within 1 pA of its "actual" value, SPICE assumes that the current has converged and moves onto the next time step or AC/DC value. VNTOL is the node voltage tolerance, default value of 1 μV. RELTOL is the relative tolerance parameter, default value of 0.001 (0.1 percent). RELTOL is used to avoid problems with simulating large and small electrical values in the same circuit. For example, suppose the default value of RELTOL and VNTOL were used in a simulation where the actual node voltage is 1 V. The RELTOL parameter would signify an end to the simulation when the node voltage was within 1 mV of 1 V (1V·RELTOL), while the VNTOL parameter signifies an end when the node voltage is within 1 μV of 1 V. SPICE uses the larger of the two, in this case the RELTOL parameter results, to signify that the node has converged.

Increasing the value of these three parameters helps speed up the simulation and assists with convergence problems at the price of reduced accuracy. To help with convergence, the following statement can be added to a SPICE netlist:

```
.OPTIONS ABSTOL=1uA VNTOL=1mV RELTOL=0.01
```

To (hopefully) force convergence, these values can be increased to

```
.OPTIONS ABSTOL=1mA VNTOL=100mV RELTOL=0.1
```

Note that in some high-gain circuits with feedback (like the op-amp's designed later in the book) decreasing these values can actually help convergence.

Some Common Mistakes and Helpful Techniques

The following is a list of helpful techniques for simulating circuits using SPICE.

1. The first line in a SPICE netlist must be a comment line. SPICE ignores the first line in a netlist file.
2. One megaohm is specified using 1MEG, not 1M, 1m, or 1 MEG.
3. One farad is specified by 1, not 1f or 1F. 1F means one femto-Farad or 10^{-15} farads.
4. Voltage source names should always be specified with a first letter of V. Current source names should always start with an I.
5. Transient simulations display time data; that is, the x-axis is time. A jagged plot such as a sinewave that looks like a triangle wave or is simply not smooth is the result of not specifying a maximum print step size.
6. Convergence with a transient simulation can usually be helped by adding a UIC (use initial conditions) to the end of a .tran statement.
7. A simulation using MOSFETs must include the scale factor in a .options statement unless the widths and lengths are specified with the actual (final) sizes.
8. In general, the body connection of a PMOS device is connected to *VDD*, and the body connection of an n-channel MOSFET is connected to ground. This is easily checked in the SPICE netlist.
9. Convergence in a DC sweep can often be helped by avoiding the power supply boundaries. For example, sweeping a circuit from 0 to 1 V may not converge, but sweeping from 0.05 to 0.95 will.
10. In any simulation adding .OPTIONS RSHUNT=1E8 (or some other value of resistor) can be used to help convergence. This statement adds a resistor in parallel with every node in the circuit (see the WinSPICE manual for information concerning the GMIN parameter). Using a value too small affects the simulation results.

ADDITIONAL READING

[1] R. J. Baker, *CMOS Mixed-Signal Circuit Design*, 2nd ed., John-Wiley and Sons, 2009. ISBN 978-0470290262

[2] S. M. Sandler and C. Hymowitz, *SPICE Circuit Handbook*, McGraw-Hill, 2006. ISBN 978-0071468572

[3] K. Kundert, *The Designer's Guide to SPICE and Spectre*, Springer, 1995. ISBN 978-0792395713

[4] A. Vladimirescu, *The SPICE Book*, John-Wiley and Sons, 1994. ISBN 978-0471609261

[5] F. M. Wanlass, "Low Standby-Power Complementary Field Effect Transistor," US Patent 3,356,858, filed June 18, 1963, and issued December 5, 1967.

PROBLEMS

1.1 What would happen to the transfer function analysis results for the circuit in Fig. 1.11 if a capacitor were added in series with R1? Why? What about adding a capacitor in series with R2?

1.2 Resimulate the op-amp circuit in Fig. 1.15 if the open-loop gain is increased to 100 million while, at the same time, the resistor used in the ideal op-amp is increased to 100 Ω. Does the output voltage move closer to the ideal value?

1.3 Simulate the op-amp circuit in Fig. 1.15 if Vin is varied from −1 to +1V. Verify, with hand calculations, that the simulation output is correct.

1.4 Regenerate IV curves, as seen in Fig. 1.18, for a PNP transistor.

1.5 Resimulate the circuit in Fig. 1.20 if the sinewave doesn't start to oscillate until 1 μs after the simulation starts.

1.6 At what frequency does the output voltage, in Fig. 1.21, become half of the input voltage? Verify your answer with SPICE

1.7 Determine the output of the circuit seen in Fig. 1.22 if a 1k resistor is added from the output of the circuit to ground. Verify your hand calculations using SPICE.

1.8 Using an AC analysis verify the time domain results seen in Fig. 1.22.

1.9 If the capacitor in Fig. 1.24 is increased to 1 μF simulate, similar to Fig. 1.26 but with a longer time scale, the step response of the circuit. Compare the simulation results to the hand-calculated values using Eqs. (1.10) and (1.11).

1.10 Using a PWL source (instead of a pulse source), regenerate the simulation data seen in Fig. 1.26.

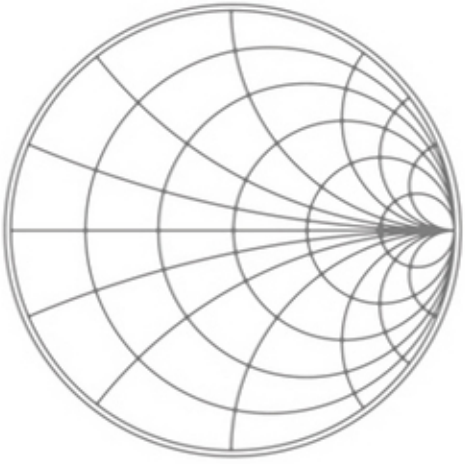
1.11 Using the values seen in Fig. 1.32, for the inductor and capacitor determine the Q of a series resonant LC tank with a resistor value of 10 ohms. Note that the resistor is in series with the LC and that an input voltage source should be used (the voltage across the LC tank goes to zero at resonance.)

1.12 Suppose the input voltage of the integrator in Fig. 1.34 is zero and that the op-amp has a 10 mV input-referred offset voltage. If the input-referred offset voltage is modeled using a 10 mV voltage source in series with the non-inverting (+) op-amp input then estimate the output voltage of the op-amp in the time-domain. Assume that at $t = 0$ $V_{out} = 0$. Verify your answer with SPICE.

Wiley Series on Information and Communication Technology
Series Editors, T. Russell Hsing and Vincent K. N. Lau

RF circuit design

second edition



Richard Chi-Hsi Li

WILEY

Contents

1

DIFFERENCE BETWEEN RF AND DIGITAL CIRCUIT DESIGN

1.1 CONTROVERSY

For many years, there has been continued controversy between digital and RF circuit designers, some of which are given below:

- RF circuit designers emphasize impedance matching, whereas digital circuit designers are indifferent to it.
- RF circuit designers are concerned with frequency response, whereas digital circuit designers are interested in the waveform, or "eye's diagram." In other words, RF circuit designers prefer to work in the frequency domain, whereas digital circuit designers like to work in the time domain.
- As a consequence of the above, in a discussion of the budget for equipment, RF circuit designers like to purchase good network analyzers, whereas digital circuit designers prefer to buy the best oscilloscopes.
- RF circuit designers use the unit of dB_W, whereas digital circuit designers insist on using dB_V.
- Not only are the design methodologies different, so are their respective jargons. Digital circuit designers talk about AC bypass capacitors or DC blocking capacitors, but RF circuit designers rename those as "zero" capacitors.

It almost seems as if they were two different kinds of aliens from different planets. Even in some conferences or publications, these two kinds of "aliens" argue with

RF Circuit Design, Second Edition. Richard Chi Hsi Li.

each other. Each tries to prove that their design methodology is superior to the others'. Eventually, nobody is the winner.

Let us outline the main controversies in the following.

1.1.1 Impedance Matching

The phrase "impedance matching" comes out of RF circuit designers' mouths almost everyday. They were told by their supervisors that impedance matching is a "must" skill in circuit design. On the other hand, such terminology is never heard among digital circuit designers. Their supervisors tell them, "ignore that 'foreign language' just focus on the 'eye diagram,' or waveform."

It is not just digital circuit designers who ignore impedance matching. Even some RF circuit designers "discovered" something new in their "advanced" RFIC (RF integrated circuit) design. While it was necessary to take care of impedance matching in RF module design or in RF blocks built by discrete parts, where the incident and reflected power in the circuit really existed, they thought it unnecessary to take care of impedance matching in an RFIC circuit design because the size of an IC die is so small as to render distinguishing the incident and reflective power or voltage redundant or meaningless. In agreement with their assertions, in the IC realm the design methodology for the RF circuit should be more or less the same as that for the digital circuit. Since then, they have been designing RF circuit blocks with the same method as used for digital circuit blocks. All the individual RF blocks are simply crowded together since "impedance matching between the individual blocks is not necessary." Their design methodology for RF blocks is specially named as the "Combo" or "Jumbo" design. Theoretically, they thought that all kinds of circuitry must obey Ohm's law and follow KCL (Kirchhoff's current law) and KVL (Kirchhoff's voltage law) rules without exception. So, why is the difference of design methodology? From their viewpoint, it seems unnecessary to divide the circuit design team into an RF and a digital circuit design group accordingly.

RF circuit designers would be very happy if impedance matching was unnecessary because impedance matching is the most difficult task in RF circuit design, especially in RFIC design for the UWB (ultrawide-band) system. Unfortunately, design experience indicates that the Combo or Jumbo design philosophy is absolutely wrong. For instance, without impedance matching, a LNA (low-noise amplifier) becomes a noisy attenuator or an oscillator in an RFIC chip. Without impedance matching, a mixer would become a "real" mixer indeed, blending all desired signals and undesired interference or noise together!

The key point to stop the controversy is whether the concept of voltage or power reflection is available in RF or digital circuitry. Should the reflection of voltage or power not exist in a practical circuitry, the idea of a Combo or Jumbo design could be a correct design methodology. On the contrary, if the reflection of voltage or power exists in a practical circuitry, impedance matching would be important for power transportation or manipulation in a circuitry, and then the idea of Combo or Jumbo design would be an incorrect design methodology.

As a matter of fact, the existence of power or voltage reflection can be deduced from a rough analysis of an RF block. For example, without impedance matching, the insertion loss of an LC passive filter could be significant. However, if the Q values of the inductors or capacitors are high, the LC passive filter itself should not conceivably produce a loss of power. This significant insertion loss demonstrates that quite a lot of power is reflected from the filter or load to the source. On the other hand, power or

voltage reflection is not related to the size of the block but to the impedance matching status between the source and the load. A simple example could illustrate the validity of such an assertion: light is reflected from a mirror in the same way no matter whether the light source is far from or very close to the mirror.

1.1.2 Key Parameter

There is a true story from a start-up company researching and developing a wireless communication system.

In spite of different opinions and various comments among his engineering teams, the engineering director asked both his RF and digital circuit design teams to work together for the system design of a communication system. He ruled that *voltage* must be taken as the key parameter to measure the performance of every block, including digital and RF blocks. In other words, the goal of the input and the output in every block, no matter RF or digital, must be specified with the voltage value. This engineering director hates the RF circuit designers' incessant "gossip" about power and impedance.

The engineers did try very hard to follow his instructions. There seemed to be no problem for the digital circuit blocks. However, the engineers were confused and did not know how to specify the goals for RF blocks by voltage instead of power.

By the RF engineers' understanding, all the parameters including G (power gain), NF (noise figure), IP_3 (3rd order intercept point), and IP_2 (2nd order intercept point) applied in RF circuit design were expressed by power but not voltage. In order to follow the director's instructions, they spent a lot of time to convert all the parameters from power to voltage, since power was the traditional unit and was read by most equipment. Sometimes, the conversion was meaningless or uncertain. For instance, by the unit of voltage, CNR (carrier-to-noise ratio) at the input of the demodulator was significantly dependent on the output impedance of the stage before the demodulator and the input impedance of the demodulator. Especially when the output impedance of the stage before the demodulator and the input impedance of the demodulator were different from each other, the conversion becomes impossible. Even more awkwardly, members of audience who attended the presentation meeting held by this system design team could not understand why the values appearing in the system plan were surprisingly higher or lower than those from other companies. Eventually, after they learnt of the extraordinary instructions given by the engineering director, the part of the audience equipped with calculators at hand could not but convert those values back from voltage to power!

Among this system design team, selection of a common key parameter for both RF and digital circuit designs became a hot topic. People argued with each other without result, while the director still insisted on his original instructions. After a couple of weeks, the system design still hung in the air and, finally, for unknown reasons the plan for the system design was dropped quietly. Some RF circuit designers felt upset and left the company despite the director's exhortations: "Nothing is Impossible!"

As a matter of fact, system design for a communication system must be divided into two portions: the digital portion and the RF portion. Yes, the key parameter in the digital circuit design is voltage or current. By means of voltage or current, all the intermediate parameters can be characterized. However, the key parameter in RF circuit design must be power or impedance. By means of power and impedance, all the intermediate parameters in a RF circuit block can be characterized. Impedance matching ensures the best performance of power transportation or manipulation in RF circuit blocks; therefore, impedance can be taken as the key parameter in RF circuit design.

Why? The answer can be found in the following sections.

1.1.3 Circuit Testing and Main Test Equipment

In addition to the arguments about impedance matching and the key parameters, the difference between digital and RF circuit design can also be found in circuit testing and test equipment.

In a digital test laboratory, the test objective is always voltage, and occasionally current. There are many pieces of test equipment available in a digital test laboratory; however, the main test equipment is the oscilloscope. The oscilloscope can sense the voltage at any node in the circuitry and display its eye diagram or a waveform on screen, which characterizes the performance of a digital circuit intuitionally. In general, digital circuit designers prefer to analyze the circuitry in the time domain because the speed of response is important to the performance of a digital circuit block.

In an RF test laboratory, the test objective is always power. Most RF test equipment, such as the spectrum analyzer, noise meter, signal generator, and so on, measure the parameters of an RF circuit block in terms of power but not voltage. The main test equipment is the network analyzer. The performance of an RF circuit block can be characterized mainly by its frequency response on network analyzer screen, which is expressed by power gain or loss, in decibels. The RF circuit designer prefers to analyze the circuitry in the frequency domain because coverage of bandwidth is important to the performance of an RF block.

In the test laboratory, testing a digital circuit block is somewhat easier than testing an RF circuit block. In testing for a digital circuit block, the probe of an oscilloscope is usually a sensor with high impedance. It does not disturb the circuit performance when it touches a node in the circuitry.

On the other hand, while using a network analyzer, the circuit designers may worry about the difference of circuit performance before and after the tested equipment is connected to the desired test node, because the input and output impedance of the equipment is low, usually 50 Ω. In most cases, it certainly will disturb the circuit performance.

Instead of voltage testing, the RF circuit designer is concerned with power testing. All power testing must be conducted under a good impedance matching condition so the test equipment must be well calibrated. Unlike the testing for a digital circuit block by an oscilloscope, a buffer connected between the desired test node and the input of the network analyzer is not allowed because all the power tests for the RF block must be conducted under the condition of impedance matching.

So far, the different methodology between RF and digital circuit design has been introduced only in terms of the three main aspects above. More differences exist but will not be listed. We are going to focus on the explanation of where these differences come from.

1.2 DIFFERENCE OF RF AND DIGITAL BLOCK IN A COMMUNICATION SYSTEM

1.2.1 Impedance

The input and output impedance of an RF circuitry are usually pretty low. In most cases, they are typically 50 Ω. On the contrary, the input and output impedances in a digital circuitry are usually quite high. For example, the input and output impedances of an Op-Amp (operating amplifier) are mostly higher than 10 kΩ.

The lower impedance in an RF circuitry is beneficial to deliver power to a block or a part. It is well known that the power of a signal delivered to a block or a part with impedance Z can be expressed by

$$P = vi = \frac{v^2}{Z}, \tag{1.1}$$

where

P = the power delivered to a block or a part,
v = the AC or RF voltage across the block or the part,
i = the AC or RF current flowing through the block or part, and
Z = the impedance of the block or the part.

For a given value of power, v^2 is proportional to Z. This implies that, in order to deliver a given power to a block or a part, a higher voltage must be provided if its impedance is high. On the contrary, a lower voltage across the block is enough to deliver the same given power to a block or a part if its impedance is low. From the viewpoint of either cost or engineering design of the circuit, the application of a lower voltage is much better than that of a higher voltage. It is one of the reasons why the input and output impedance in the RF blocks are intentionally assigned to be low because only a lower voltage is needed in order to deliver the same given power to a block or part with low impedance.

However, it is just the opposite for a digital signal. The higher impedance in digital circuitry is beneficial to the voltage swing in a digital block or part. For a given current, a higher impedance can have a higher voltage swing across a block or a part, and then the signal can ON/OFF the device more effectively, because

$$v = iZ. \tag{1.2}$$

The question is: why is RF circuitry focused on the power while digital circuitry is concerned about voltage?

1.2.2 Current Drain

In RF circuit blocks, the current drains are usually in the order of milliamperes while in digital circuit blocks they are usually in the order of microamperes. That is, the difference of the current drain's magnitude between RF and digital circuit blocks is approximately 1000 times.

In RF circuit blocks, it is desirable to increase the power of the RF signal as much as possible. This implies that higher current drains are preferred in RF circuit blocks because they are beneficial to deliver power to the block or the part for a given voltage.

In digital circuit blocks, it is desirable to reduce the power of the digital signal as much as possible. This implies that lower current drains are preferred in digital circuit blocks as long as the voltage swing is high enough.

Again, the question is: why is RF circuitry focused on power while digital circuitry is concerned with voltage? The answer can be found in the following section.

1.2.3 Location

In a communication system, the demodulator is a remarkable demarcation in the receiver. As shown in Figure1.1, before the demodulator, the blocks operate in the range of radio

frequency so that they are called *RF blocks*. They are sometimes called the *RF front end* in the receiver, where the RF circuit design is conducted. After demodulation, the blocks operate in the range of intermediate frequency or in the low digital data rate and are categorized as baseband blocks, or the digital/analog section. They are sometimes called the *back end* in the receiver, where digital/analog circuit design is conducted. The demodulator is a critical block in which both digital and RF design technology are needed.

The order of blocks in the transmitter is just opposite. Before the modulator, the blocks operate in the range of intermediate frequency or in the low digital data rate and are categorized as baseband blocks, or the digital/analog section. They are sometimes called the *front end* in the transmitter, where digital/analog circuit design is conducted. After the modulator, the blocks operate in the range of radio frequency so that they are called *RF blocks* and sometimes the *RF back end* in the transmitter, where RF circuit design is conducted. The modulator is also a critical block, in which both digital and RF design technology are needed.

A common feature can be seen from Figure 1.1. In either the receiver or the transmitter, the circuit portion close to the antenna side contains RF blocks and the portion farther from antenna side contains digital/analog circuit blocks.

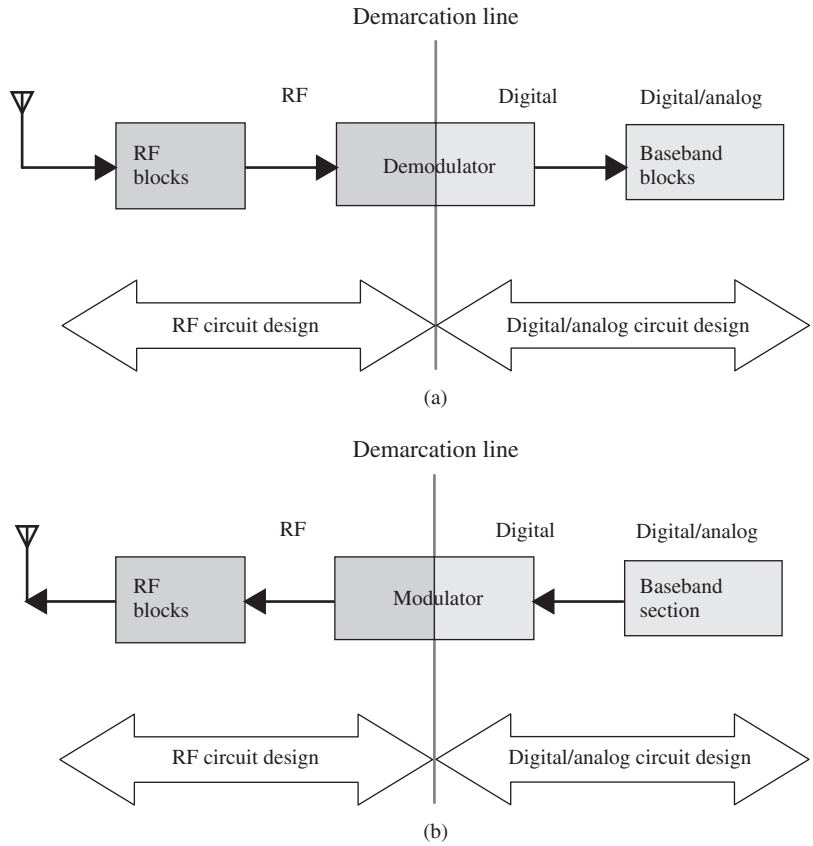


Figure 1.1. Demarcation line in a communication system. (a) Receiver. (b) Transmitter.

In the receiver, the received modulated carrier is usually very weak. After it is power-magnified by the LNA and its frequency mixed down by the mixer, the modulated carrier can be demodulated only if its power is strong enough to suppress the noise power at the input of the demodulator. Typically, the ratio of RF signal power to the noise power at the input of the demodulator is required to be more than 10 dB. It is therefore required that the RF signal be power-transported or power-operated before demodulation. After the demodulator, the digital-type message is demodulated from the RF to the base band. The digital signal is not required to be "power-transported" but is only "status-maintained" or voltage-transported between local blocks for digital signal processing. The voltage represents the status of the signal. For the sake of power saving, the power of the signal is reduced as much as possible. This answers the question why voltage transportation or manipulation or the "status" transportation or manipulation is a logical task in the digital circuit design.

Similarly, in the transmitter, the digital signal is only required to reach the "modulation-effective status level" before the modulator. This implies that the power or voltage of the input digital signal to the modulator could be as low as possible as long as the input voltage or power reaches a level by which the carrier can be effectively modulated. In this case, the digital signal is transported or manipulated between the local circuit blocks and is not required to be power-transported but only voltage-transported. However, the modulated carrier after the modulator must be power-magnified and delivered to the antenna so that the modulated carrier is powerful enough to propagate to a receiver located a long distance from the transmitter.

1.3 CONCLUSIONS

From the discussion in Section 1.2.3, it can be concluded that the power transportation or manipulation, but not voltage transportation or manipulation, is required in the RF blocks before the demodulator in a receiver and after the modulator in the transmitter. The voltage transportation or manipulation, but not power transportation or manipulation, is required in the digital blocks after the demodulator in a receiver or before the modulator in a transmitter.

This is the answer to the question in Sections 1.2.2 and 1.2.3: why is RF circuitry focused on the power while the digital circuitry is concerned about voltage?

It can be seen that the controversy between digital and RF designers is not necessary. The difference in circuit design methodology arises from the different circuit design tasks. The digital circuit is designed for voltage transportation or manipulation, while the RF circuit is designed for power transportation or manipulation.

1.4 NOTES FOR HIGH-SPEED DIGITAL CIRCUIT DESIGN

In digital communication, the synonym of "high speed" is high data rate.

In the case of high digital data rates, the tasks of both types of circuits are unchanged: the RF circuit still works for the power while digital circuit still works for the *status* transportation or operation. However, the design methodology in high data rate digital circuit designs is close to that of the RF circuit design methodology, because of the following reasons:

1. The remarkable variation of a digital circuit design from low speed to high speed is the change of the input and output impedance of a digital block. In the digital circuit design for low speeds, high input and output impedance of a digital block can be obtained from high input and output impedance of a transistor. The input and output capacitance of a transistor can impact on the input or output impedance so that they impact on the rising or dropping time. In the digital circuit design for high speeds, the input and output impedance of a transistor is reduced. The low input and output impedance leads to the necessity of impedance matching in the digital circuit design. Impedance matching is beneficial not only to power transportation but also to voltage transportation. Without impedance matching, the reflected voltage will appear and interfere with the incident voltage. This brings about additional attenuation, additional jitter, an additional cross talk, and eventually an additional bit error to the digital signal.
2. The traditional layout scheme is no longer reliable in the case of digital circuits with high data rates. For example, in the traditional layout for digital circuits with low data rates, the runners are always lined up in parallel so that the entire layout looks nice and neat. However, in the layout for digital circuits with high data rates, the runners lined up in parallel could cause appreciable coplanar capacitance and are coupled with each other, and, consequently, it may cause interference or cross talk. The layout for high-speed digital circuitry must be taken as seriously as for the RF circuitry.
3. The traditional AC grounding scheme for a digital circuit with low data rate is no longer reliable in the case of a digital circuit with high data rate. For high-speed digital circuitry, the AC grounding must be taken as seriously as for an RF circuitry.
4. In a digital circuitry with high data rate, isolation may become a serious problem. Usually, a digital signal is a rectangular pulse while an RF signal is sinusoidal. The former contains a wide-band spectrum while the latter contains a narrow-band spectrum. This implies that a digital circuit with a high data rate is easier to be interfaced with inside or outside interference sources because its frequency spectrum is much wider than that of a digital circuit with low data rate. Isolation between blocks becomes important and should be taken as seriously as for an RF circuitry.

FURTHER READING

Li RC-H. Key Issues in RF/RFIC Circuit Design. Beijing: Higher Education Press; 2005.

Sakurai T., Low power digital circuit design, 31th European Solid-State Circuits Conference (ESSCIRC), 2004, Leuven, Belgium, 11–18.

Thomas H. Lee, The Design of CMOS Radio-Frequency, Integrated Circuit, Cambridge University Press, 1998.

Turnbull JR, Transistor switching circuits. International Solid State Circuits Conference, Digest of Technical Papers; 1963 Feb 21; University of Pennsylvania-University Museum; 1963. pp. 58–59.

Xiao E, Yuan JS. RF circuit design in reliability. IEEE Radio Frequency Integrated Circuits Symposium; Philadelphia (PA); 2003. pp. 575–578.

EXERCISES

1. Specify the main difference in task between RF and digital circuit designs.
2. Compare and comment on the task between RF and digital circuit designs when the digital data rate is much less than the RF carrier frequency.
3. What is the difference in digital circuit design between low and high data rate cases?
4. High impedance is beneficial to voltage transportation or manipulation while low impedance is beneficial to power transportation or manipulation. Why?
5. Why is impedance the main parameter in RF circuit design?
6. Why is the normalized impedance regulated as 50 Ω?
7. Assuming that the input capacitance of a MOSFET (metal–oxide–semiconductor field-effect transistor) is 0.15915 pF $= 1/(2\pi)$ pF, what is the input impedance for an RF signal if its operating frequency is 10, 100, 1000, and 10,000 MHz, respectively? Also, what is the input impedance for a digital signal to be considered if its digital data rate is 10, 100, 1000, 10,000 Mb/s, respectively, and if its second and third harmonics are as important as the main frequency corresponding to the digital data rate?
8. Why is power the transportation type of an RF signal while status (voltage or current) is the transportation type for a digital signal?
9. Why is impedance matching important in RF but not in digital circuit design when $R \ll f_{\text{RF}}$?
10. Why is impedance matching important not only in RF but also in digital circuit design when $R \approx f_{\text{RF}}$?

ANSWERS

1. The main difference of task between RF and digital circuit design can be tabulated as follows:

Item	RF Circuit Design	Digital Circuit Design
Transportation type	Power	Status (voltage or current)

2. When the data rate is low, the difference between RF and digital circuit can be tabulated as follows:

Item	RF Module/RFIC	Digital Circuit (Low Data Rate)
Impedance	Low (50 Ω typically)	High (infinity ideally)
Current	High (mA)	Low (μA)
Location in a Communication System		
• *Rx*	Front end (before demodulation)	Back end (after demodulation)
• *Tx*	Back end (after modulation)	Front end (before modulation)
Transportation type	Power (W)	Status (voltage or current)
Impedance matching	Important	Unimportant (usually)

3. In the low digital data rate case, the input or output impedance in the digital circuit is usually high. However, in the high data rate case, the input or output impedance

in the digital circuit is not high because of the existence of the input or output capacitance of the device. Impedance matching becomes effective in the voltage transportation. The voltage will be pumped up when the load resistance is greater than the source resistance, that is, $R_L > R_S$.

4. Low impedance is beneficial to power transportation or manipulation because,

$$P = vi = \frac{v^2}{Z}$$

for a definite voltage, the power becomes high if the impedance is low (say, 50 Ω). On the contrary, high impedance is beneficial to status transportation because

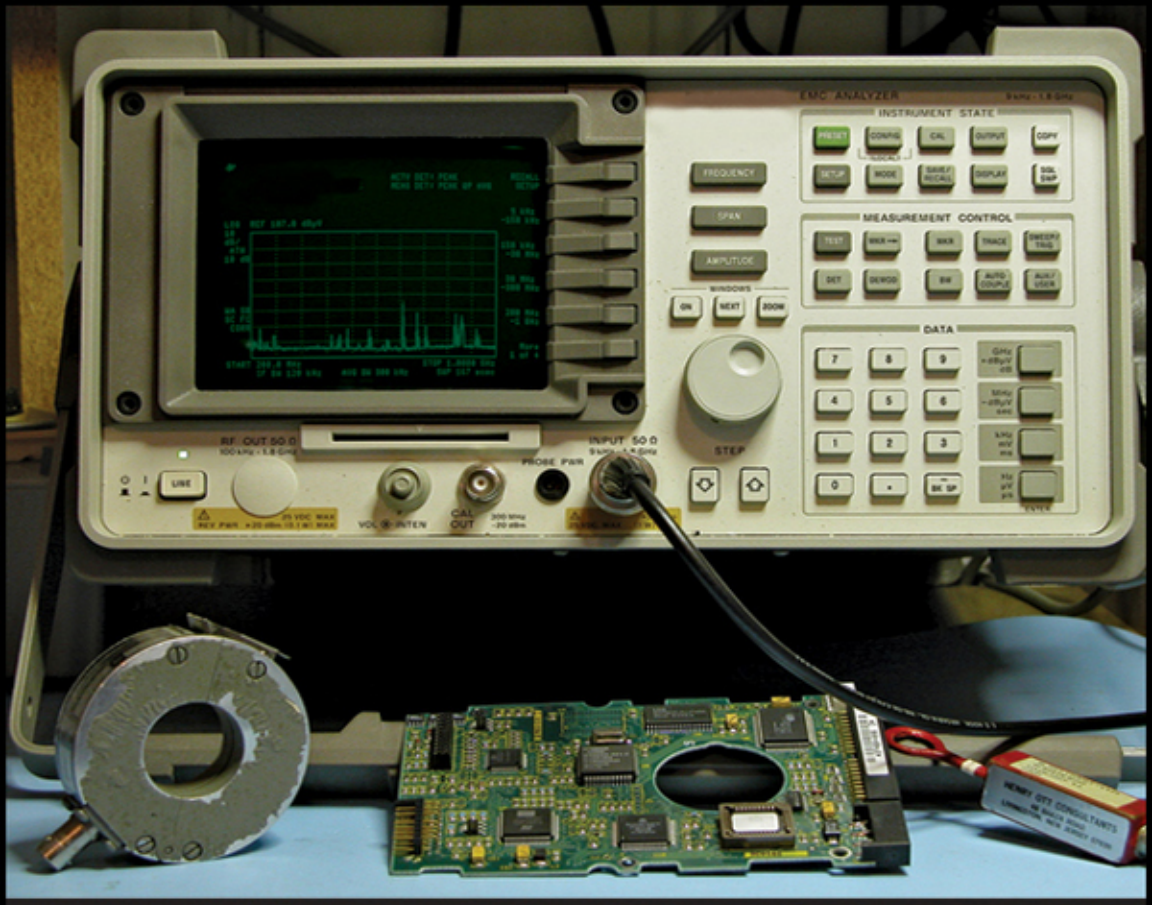
$$v = iZ$$

for a definite current, and the voltage swing becomes large if the impedance is high.

5. The main task of the RF circuit is to transport or to manipulate the power of signal. This is the main reason that the impedance becomes the main parameter to be taken care of in an RF circuit design. The performance of power transportation or manipulation is mainly determined by the status of impedance matching.
6. For a coaxial cable with an air dielectric, in order to maximize the power-handling capability, one could increase the diameter of the inner wire. However, there are two competing effects: the breakdown voltage is increased but the characteristic impedance Z_o would be incresed also, which would tend to reduce the power deliverable to a load. The maximum of the power-handling capability can be reached if the diameters of the inner wire and the outer conductor are well-selected so that a Z_o of 30 Ω is approached. On the other hand, in order to minimize the attenuation of a coaxial cable, there are also two competing effects due to the increase of the inner wire diameter: both of resistance per unit length and the characteristic impedance Z_o would be reduced at the same time, which would tend to an uncertainity of the attenuation. The minimum of the attenuation can be reached if the diameters of the inner wire and the outer conductor are well-selected so that a Z_o of 77 Ω is approached...'' Since 77 Ω gives us minimum loss and 30 Ω gives us maximum power-handling capability, a reasonable compromise is a round average value, Z_o of 50 Ω.
7. The input impedance for an RF signal is 100 kΩ, 10 kΩ, 1000 Ω, and 100 Ω, respectively.
   The input impedance for a digital signal is 33.3 kΩ, 3.3 kΩ, 333.3 Ω, and 33.3 Ω, respectively, if its second and third harmonics are as important as the main frequency corresponding to the digital data rate.
8. In the receiver, RF and digital circuits are located before and behind the demodulator, respectively. The performance of RF circuits must be good in power transportation so that the threshold of the CNR power ratio at the input of demodulator can be reached or exceeded and consequently the demodulation becomes available. On the contrary, the task of the digital circuits after the demodulator is to transport or manipulate the status (voltage or current) of digital signals.
   In the transmitter, digital and RF circuits are located before and behind the modulator, respectively. The task of the digital circuits before the modulator is to transport or manipulate the status (voltage or current) of digital signals. On the contrary, the performance of RF circuits after modulator must be good in power transportation

so that the CNR power ratio at the output of modulator can be radiated to remote places where the receiver antenna is located.

9. When $R \ll f_{RF}$, impedance matching is important in RF but not in digital circuit design because the carrier frequency f_{RF} of RF signal is high, whereas the frequency components of digital signals are low. The input/output impedances in RF circuits are low, while the input/output impedances in digital circuits are high. Impedance matching is important when the input/output impedances of a circuitry are low, but it is not important when the input/output impedances of a circuitry are high.
10. When $R \approx f_{RF}$, impedance matching is not only important in RF but also in digital circuit design because both the carrier frequency f_{RF} of RF signal and the frequency components of digital signals are high. Both of the input/output impedances in RF circuits and the input/output impedances in digital circuits are low. Impedance matching is important when the input/output impedances of a circuitry are low.



ELECTROMAGNETIC COMPATIBILITY ENGINEERING

HENRY W. OTT

WILEY

Contents

1 Electromagnetic Compatibility

1.1 INTRODUCTION

The widespread use of electronic circuits for communication, computation, automation, and other purposes makes it necessary for diverse circuits to operate in close proximity to each other. All too often, these circuits affect each other adversely. Electromagnetic interference (EMI) has become a major problem for circuit designers, and it is likely to become even more severe in the future. The large number of electronic devices in common use is partly responsible for this trend. In addition, the use of integrated circuits and large-scale integration has reduced the size of electronic equipment. As circuitry has become smaller and more sophisticated, more circuits are being crowded into less space, which increases the probability of interference. In addition, clock frequencies have increased dramatically over the years—in many cases to over a gigahertz. It is not uncommon today for personal computers used in the home to have clock speeds in excess of 1 GHz.

Today's equipment designers need to do more than just make their systems operate under ideal conditions in the laboratory. Besides that obvious task, products must be designed to work in the "real world," with other equipment nearby, and to comply with government electromagnetic compatibility (EMC) regulations. This means that the equipment should not be affected by external electromagnetic sources and should not itself be a source of electromagnetic noise that can pollute the environment. Electromagnetic compatibility should be a major design objective.

1.2 NOISE AND INTERFERENCE

Noise is any electrical signal present in a circuit other than the desired signal. This definition excludes the distortion products produced in a circuit due to nonlinearities. Although these distortion products may be undesirable, they are not considered noise unless they are coupled into another part of the circuit. It follows that a desired signal in one part of a circuit can be considered to be noise when coupled to some other part of the circuit.

Electromagnetic Compatibility Engineering, by Henry W. Ott

Noise sources can be grouped into the following three categories: (1) intrinsic noise sources that arise from random fluctuations within physical systems, such as thermal and shot noise; (2) man-made noise sources, such as motors, switches, computers, digital electronics, and radio transmitters; and (3) noise caused by natural disturbances, such as lightning and sunspots.

Interference is the undesirable effect of noise. If a noise voltage causes improper operation of a circuit, it is interference. Noise cannot be eliminated, but interference can. Noise can only be reduced in magnitude, until it no longer causes interference.

1.3 DESIGNING FOR ELECTROMAGNETIC COMPATIBILITY

Electromagnetic compatibility (EMC) is the ability of an electronic system to (1) function properly in its intended electromagnetic environment and (2) not be a source of pollution to that electromagnetic environment. The electromagnetic environment is composed of both radiated and conducted energy. EMC therefore has two aspects, emission and susceptibility.

Susceptibility is the capability of a device or circuit to respond to unwanted electromagnetic energy (i.e., noise). The opposite of susceptibility is *immunity*. The immunity level of a circuit or device is the electromagnetic environment in which the equipment can operate satisfactorily, without degradation, and with a defined margin of safety. One difficulty in determining immunity (or susceptibility) levels is defining what constitutes performance degradation.

Emission pertains to the interference-causing potential of a product. The purpose of controlling emissions is to limit the electromagnetic energy emitted and thereby to control the electromagnetic environment in which other products must operate. Controlling the emission from one product may eliminate an interference problem for many other products. Therefore, it is desirable to control emission in an attempt to produce an electromagnetically compatible environment.

To some extent, susceptibility is self-regulating. If a product is susceptible to the electromagnetic environment, the user will become aware of it and may not continue to purchase that product. Emission, however, tends not to be self-regulating. A product that is the source of emission may not itself be affected by that emission. To guarantee that EMC is a consideration in the design of all electronic products, various government agencies and regulatory bodies have imposed EMC regulations that a product must meet before it can be marketed. These regulations control allowable emissions and in some cases define the degree of immunity required.

EMC engineering can be approached in either of two ways: one is the *crisis approach*, and the other is the *systems approach*. In the crisis approach, the designer proceeds with a total disregard of EMC until the functional design is finished, and testing—or worse yet—field experience suggests that a problem

exists. Solutions implemented at this late stage are usually expensive and consist of undesirable "add ons." This is often referred to as the "Band Aid" approach.

As equipment development progresses from design to testing to production, the variety of noise mitigation techniques available to the designer decreases steadily. Concurrently, cost goes up. These trends are shown in Fig. 1-1. Early solutions to interference problems, therefore, are usually the best and least expensive.

The systems approach considers EMC throughout the design; the designer anticipates EMC problems at the beginning of the design process, finds the remaining problems in the breadboard and early prototype stages, and tests the final prototypes for EMC as thoroughly as possible. This way, EMC becomes an integral part of the electrical, mechanical, and in some cases, software/firmware design of the product. As a result, EMC is designed into—and not added onto—the product. This approach is the most desirable and cost effective.

If EMC and noise suppression are considered for one stage or subsystem at a time, when the equipment is initially being designed, the required mitigation techniques are usually simple and straightforward. Experience has shown that when EMC is handled this way, the designer should be able to produce equipment with 90% or more of the potential problems eliminated prior to initial testing.

A system designed with complete disregard for EMC will almost always have problems when testing begins. Analysis at that time, to find which of the many possible noise path combinations are contributing to the problem, may not be simple or obvious. Solutions at this late stage usually involve the addition of extra components that are not integral parts of the circuit. Penalties paid include the added engineering and testing costs, as well as the cost of the

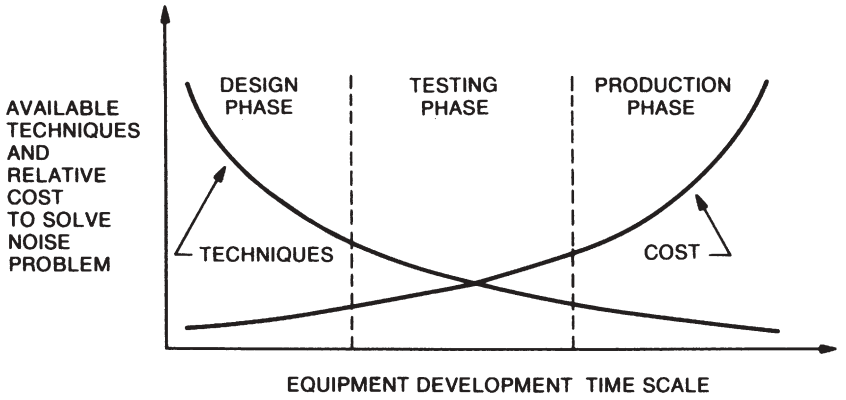


FIGURE 1-1. As equipment development proceeds, the number of available noise-reduction techniques goes down. At the same time, the cost of noise reduction goes up.

mitigation components and their installation. There also may be size, weight, and power dissipation penalties.

1.4 ENGINEERING DOCUMENTATION AND EMC

As the reader will discover, much of the information that is important for electromagnetic compatibility is not conveyed conveniently by the standard methods of engineering documentation, such as schematics, and so on. For example, a ground symbol on a schematic is far from adequate to describe where and how that point should be connected. Many EMC problems involve parasitics, which are not shown on our drawings. Also, the components shown on our engineering drawings have remarkably ideal characteristics.

The transmission of the standard engineering documentation alone is therefore insufficient. Good EMC design requires cooperation and discussion among the complete design team, the systems engineer, the electrical engineer, the mechanical engineer, the EMC engineer, the software/firmware designer, and the printed circuit board designer.

In addition, many computer-assisted design (CAD) tools do not include sufficient, if any, EMC considerations. EMC considerations therefore must often be applied manually by overriding the CAD system. Also, you and your printed circuit designer often have different objectives. Your objective is, or should be, to design a system that works properly and meets EMC requirements. Your printed circuit board (PCB) designer has the objective of doing what ever has to be done to fit all the components and traces on the board regardless of the EMC implications.

1.5 UNITED STATES' EMC REGULATIONS

Added insight into the problem of interference, as well as the obligations of equipment designers, manufacturers, and users of electronic products, can be gained from a review of some of the more important commercial and military EMC regulations and specifications.

The most important fact to remember about EMC regulations is that they are "living documents" and are constantly being changed. Therefore, a 1-year-old version of a standard or regulation may no longer be applicable. When working on a new design project, always be sure to have copies of the most recent versions of the applicable regulations. These standards may actually even change during the time it takes to design the product.

1.5.1 FCC Regulations

In the United States, the Federal Communications Commission (FCC) regulates the use of radio and wire communications. Part of its responsibility

concerns the control of interference. Three sections of the FCC Rules and Regulations* have requirements that are applicable to nonlicensed electronic equipment. These requirements are contained in Part 15 for radio frequency devices; Part 18 for industrial, scientific, and medical (ISM) equipment; and Part 68 for terminal equipment connected to the telephone network.

Part 15 of the FCC Rules and Regulations sets forth technical standards and operational requirements for radio frequency devices. *A radio-frequency device is any device that in its operation is capable of emitting radio-frequency energy by radiation, conduction, or other means (§ 2.801)*. The radio-frequency energy may be emitted intentionally or unintentionally. Radio-frequency (rf) energy is defined by the FCC as any electromagnetic energy in the frequency range of 9 kHz to 3000 GHz (§15.3(u)). The Part 15 regulations have a twofold purpose: (1) to provide for the operation of low-power transmitters without a radio station license and (2) to control interference to authorized radio communications services that may be caused by equipment that emits radio-frequency energy or noise as a by-product to its operation. Digital electronics fall into the latter category.

Part 15 is organized into six parts. Subpart A—General, Subpart B—Unintentional Radiators, Subpart C—Intentional Radiators, Subpart D—Unlicensed Personal Communications Devices, Subpart E—Unlicensed National Information Infrastructure Devices, and Subpart F—Ultra-Wideband Operation. Subpart B contains the EMC Regulations for electronic devices that are not intentional radiators.

Part 18 of the FCC Rules and Regulations sets forth technical standards and operational conditions for ISM equipment. ISM equipment is defined as any device that uses radio waves for industrial, scientific, medical, or other purposes (including the transfer of energy by radio) and that is neither used nor intended to be used for radio communications. Included are medical diathermy equipment, industrial heating equipment, rf welders, rf lighting devices, devices that use radio waves to produce physical changes in matter, and other similar noncommunications devices.

Part 68 of the FCC Rules and Regulations provides uniform standards for the protection of the telephone network from harm caused by connection of terminal equipment [including private branch exchange (PBX) systems] and its wiring, and for the compatibility of hearing aids and telephones to ensure that persons with hearing aids have reasonable access to the telephone network. Harm to the telephone network includes electrical hazards to telephone company workers, damage to telephone company equipment, malfunction of telephone company billing equipment, and degradation of service to persons other than the user of the terminal equipment, his calling or called party.

In December 2002, the FCC released a Report and Order (Docket 99-216) privatizing most of Part 68, with the exception of the requirements on hearing

* Code of Federal Regulations, Title 47, Telecommunications.

aid compatibility. Section 68.602 of the FCC rules authorized the Telecommunications Industry Association (TIA) to establish the Administrative Council for Terminal Attachments (ACTA) with the responsibility of defining and publishing technical criteria for terminal equipment connected to the U.S. public telephone network. These requirements are now defined in TIA-968. The legal requirement for all terminal equipment to comply with the technical standards, however, remains within Part 68 of the FCC rules. Part 68 requires that terminal equipment connected directly to the public switched telephone network meet both the criteria of Part 68 and the technical criteria published by ACTA.

Two approval processes are available to the manufacturer of telecommunications terminal equipment, as follows: (1) The manufacturer can provide a Declaration of Conformity (§68.320) and submit it to ACTA, or (2) the manufacturer can have the equipment certified by a Telecommunications Certifying Body (TCB) designated by the Commission (§68.160). The TCB must be accredited by the National Institute of Standards and Technology (NIST).

1.5.2 FCC Part 15, Subpart B

The FCC rule with the most general applicability is Part 15, Subpart B because it applies to virtually all digital electronics. In September 1979, the FCC adopted regulations to control the interference potential of digital electronics (at that time called "computing devices"). These regulations, "Technical Standards for Computing Equipment" (Docket 20780); amended Part 15 of the FCC rules relating to restricted radiation devices. The regulations are now contained in Part 15, Subpart B of Title 47 of the Code of Federal Regulations. Under these rules, limits were placed on the maximum allowable radiated emission and on the maximum allowable conducted emission on the alternating current (ac) power line. These regulations were the result of increasing complaints to the FCC about interference to radio and television reception where digital electronics were identified as the source of the interference. In this ruling the FCC stated the following:

> Computers have been reported to cause interference to almost all radio services, particularly those services below 200 MHz,* including police, aeronautical, and broadcast services. Several factors contributing to this include: (1) digital equipment has become more prolific throughout our society and are now being sold for use in the home; (2) technology has increased the speed of computers to the point where the computer designer is now working with radio frequency and electromagnetic interference (EMI) problems—something he didn't have to contend with 15 years ago; (3) modern production economics has replaced the steel cabinets which shield or reduce radiated emanations with plastic cabinets which provide little or no shielding.

* Remember this was 1979.

In the ruling, the FCC defined a digital device (previously called a computing device) as follows:

> An unintentional radiator (device or system) that generates and uses timing signals or pulses at a rate in excess of 9000 pulses (cycles) per second and uses digital techniques; inclusive of telephone equipment that uses digital techniques or any device or system that generates and uses radio frequency energy for the purpose of performing data processing functions, such as electronic computations, operations, transformations, recording, filing, sorting, storage, retrieval or transfer (§ 15.3(k)).

Computer terminals and peripherals, which are intended to be connected to a computer, are also considered to be digital devices.

This definition was intentionally broad to include as many products as possible. Thus, if a product uses digital circuitry and has a clock greater than 9 kHz, then it is a digital device under the FCC definition. This definition covers most digital electronics in existence today.

Digital devices covered by this definition are divided into the following two classes:

Class A: A digital device that is marketed for use in a commercial, industrial, or business environment (§ 15.3(h)).

Class B: A digital device that is marketed for use in a residential environment, notwithstanding use in commercial, business, and industrial environments (§ 15.3(i)).

Because Class B digital devices are more likely to be located in closer proximity to radio and television receivers, the emission limits for these devices are about 10 dB more restrictive than those for Class A devices.

Meeting the technical standards contained in the regulations is the obligation of the manufacturer or importer of a product. To guarantee compliance, the FCC requires the manufacturer to test the product for compliance before the product can be *marketed* in the United States. The FCC defines marketing as shipping, selling, leasing, offering for sale, importing, and so on (§ 15.803(a)). Until a product complies with the rules, it cannot legally be advertised or displayed at a trade show, because this would be considered an offer for sale. To advertise or display a product legally prior to compliance, the advertisement or display must contain a statement worded as follows:

> This device has not been authorized as required by the rules of the Federal Communications Commission. This device is not, and may not be, offered for sale or lease, or sold or leased, until authorization is obtained (§ 2.803(c)).

For personal computers and their peripherals (a subcategory of Class B), the manufacturer can demonstrate compliance with the rules by a Declaration of Conformity. A Declaration of Conformity is a procedure where the manufacturer makes measurements or takes other steps to ensure that the equipment

complies with the applicable technical standards (§ 2.1071 to 2.1077). Submission of a sample unit or representative test data to the FCC is not required unless specifically requested.

For all other products (Class A and Class B—other than personal computers and their peripherals), the manufacturer must verify compliance by testing the product before marketing. *Verification* is a self-certification procedure where nothing is submitted to the FCC unless specifically requested by the Commission, which is similar to a declaration of conformity (§ 2.951 to 2.956). Compliance is by random sampling of products by the FCC. The time required to do the compliance tests (and to fix the product, and redo the test if the product fails) should be scheduled into the product's development timetable. Precompliance EMC measurements (see Chapter 18) can help shorten this time considerably.

Testing must be performed on a sample that is representative of production units. This usually means an early production or preproduction model. Final compliance testing must therefore be one of the last items in the product development timetable. This is no time for unexpected surprises! If a product fails the compliance test, then changes at this point are difficult, time consuming, and expensive. Therefore, it is desirable to approach the final compliance test with a high degree of confidence that the product will pass. This can be done if (1) proper EMC design principles (as described in this book) have been used throughout the design and (2) preliminary pre-compliance EMC testing as described in Chapter 18 was performed on early models and subassemblies.

It should be noted that the limits and the measurement procedures are interrelated. The derived limits were based on specified test procedures. Therefore, compliance measurements must be made following the procedure outlined by the regulations (§ 15.31). The FCC specifies that for digital devices, measurements to show compliance with Part 15, must be performed following the procedures described in measurement standard ANSI C63.4–1992 titled "Methods of Measurement of Radio-Noise Emissions from Low-Voltage Electrical and Electronic Equipment in the Range of 9 kHz to 40 GHz," excluding Section 5.7, Section 9, and Section 14 (§ 15.31(a)(6)).*

The test must be made on a complete system, with all cables connected and configured in a reasonable way that tends to maximize the emission (§ 15.31(i)). Special authorization procedures are provided in the case of central processor unit (CPU) boards and power supplies that are used in personal computers and sold separately (§ 15.32).

* Section 5.7 pertains to the use of an artificial hand to support handheld devices during testing. Section 9 pertains to measuring radio-noise power using an absorbing clamp in lieu of radiated emission measurements for certain restricted frequency ranges and certain types of equipment. Section 14 pertains to relaxing the radiated and/or conducted emission limits for short duration ($\leq$200 ms) transients.

1.5.3 Emissions

The FCC Part 15 EMC Regulations limit the maximum allowable conducted emission, on the ac power line in the range of 0.150 to 30 MHz, and the maximum radiated emission in the frequency range of 30 MHz to 40 GHz.

1.5.3.1 Radiated Emissions. For radiated emissions, the measurement procedure specifies an open area test site (OATS) or equivalent measurement made over a ground plane with a tuned dipole or other correlatable, linearly polarized antenna. This setup is shown in Fig. 1-2. ANSI C63.4 allows for the use of an alternative test site, such as an absorber-lined room, provided it meets specified site attenuation requirements. However, a shielded enclosure without absorber lining may not be used for radiated emission measurements.

The specified receive antenna in the 30- to- 1000-MHz range is a tuned dipole, although other linearly polarized broadband antennas may also be used. However, in case of a dispute, data taken with the tuned dipole will take precedence. Above 1000 MHz, a linearly polarized horn antenna shall be used.

Table 1-1 lists the FCC radiated emission limits (§ 15.109) for a Class A product when measured at a distance of 10 m. Table 1-2 lists the limits for a Class B product when measured at a distance of 3 m.

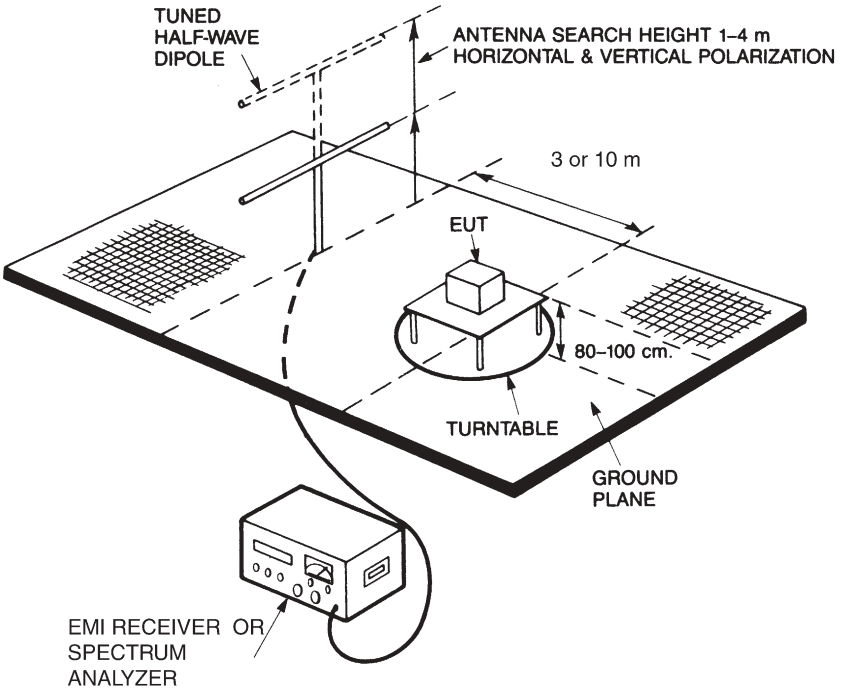


FIGURE 1-2. Open area test site (OATS) for FCC radiated emission test. The equipment under test (EUT) is on the turntable.

TABLE 1-1. FCC Class A Radiated Emission Limits Measured at 10 m.

Frequency (MHz)	Field Strength (μV/m)	Field Strength (dB μV/m)
30–88	90	39.0
88–216	150	43.5
216–960	210	46.5
>960	300	49.5

TABLE 1-2. FCC Class B Radiated Emission Limits Measured at 3 m.

Frequency (MHz)	Field Strength (μV/m)	Field Strength (dB μV/m)
30–88	100	40.0
88–216	150	43.5
216–960	200	46.0
>960	500	54.0

TABLE 1-3. FCC Class A and Class B Radiated Emission Limits Measured at 10 m.

Frequency (MHz)	Class A Limit (μV/m)	Class B Limit (dB μV/m)
30–88	39.0	29.5
88–216	43.5	33.0
216–960	46.5	35.5
>960	49.5	43.5

A comparison between the Class A and Class B limits must be done at the same measuring distance. Therefore, if the Class B limits are extrapolated to a 10-m measuring distance (using a 1/d extrapolation), the two sets of limits can be compared as shown in Table 1-3. As can be observed, the Class B limits are more restrictive by about 10 dB below 960 MHz and 5 dB above 960 MHz. A plot of both FCC Class A and Class B radiated emission limits over the frequency range of 30 MHz to 1000 MHz (at a measuring distance of 10 m) is shown in Fig. 1-5.

The frequency range over which radiated emission tests must be performed is from 30 MHz up to the frequency listed in Table 1-4, which is based on the highest frequency that the equipment under test (EUT) generates or uses.

1.5.3.2 Conducted Emissions. Conducted emission regulations limit the voltage that is conducted back onto the ac power line in the frequency range of 150 kHz to 30 MHz. Conducted emission limits exist because regulators believes

TABLE 1-4. Upper Frequency Limit for Radiated Emission Testing.

Maximum Frequency Generated or Used in the EUT (MHz)	Maximum Measurement Frequency (GHz)
<108	1
108–500	2
500–1000	5
>1000	5^{th} Harmonic or 40 GHz, whichever is less

TABLE 1-5. FCC/CISPR Class A Conducted Emission Limits.

Frequency (MHz)	Quasi-peak (dB μV)	Average (dB μV)
0.15–0.5	79	66
0.5–30	73	60

TABLE 1-6. FCC/CISPR Class B Conducted Emission Limits.

Frequency (MHz)	Quasi-peak (dB μV)	Average (dB μV)
0.15–0.5	66–56[a]	56–46[a]
0.5–5	56	46
5–30	60	50

[a]Limit decreases linearly with log of frequency.

that at frequencies below 30 MHz, the primary cause of interference with radio communications occurs by conducting radio-frequency energy onto the ac power line and subsequently radiating it from the power line. Therefore, conducted emission limits are really radiated emission limits in disguise.

The FCC conducted emission limits (§ 15.107) are now the same as the International Special Committee on Radio Interference (CISPR, from its title in French) limits, used by the European Union. This is the result of the Commission amending its conducted emission rules in July 2002 to make them consistent with the international CISPR requirements.

Tables 1-5 and 1-6 show the Class A and Class B conducted emission limits, respectively. These voltages are measured common-mode (hot to ground and neutral to ground) on the ac power line using a 50-Ω/50-μH line impedance stabilization network (LISN) as specified in the measurement procedures.* Figure 1-3 shows a typical FCC conducted emission test setup.

* The circuit of an LISN is shown in Fig. 13-2.

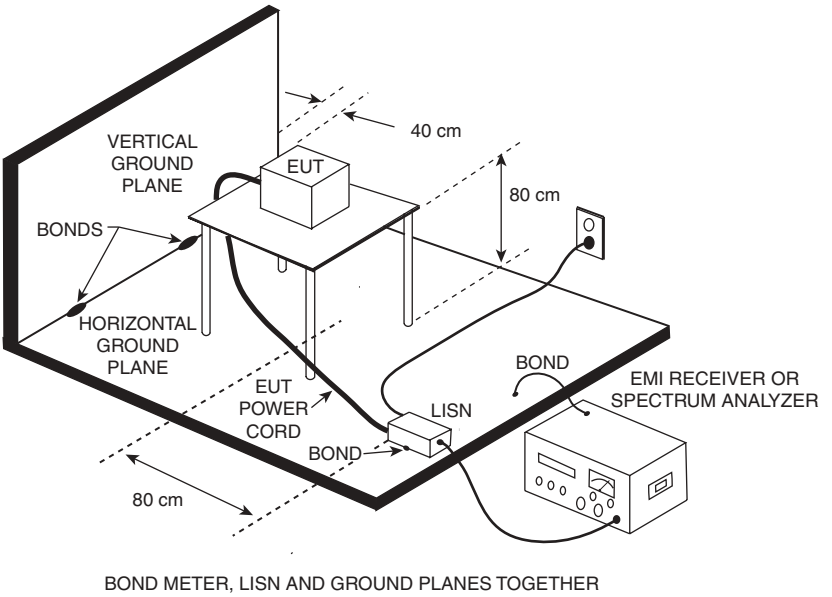


FIGURE 1-3. Test setup for FCC conducted emission measurements.

A comparison between Tables 1-5 and 1-6 shows that the Class B quasi-peak conducted emission limits are from 13 dB to 23 dB more stringent than the Class A limits. Note also that both peak and average measurements are required. The peak measurements are representative of noise from narrowband sources such as clocks, whereas the average measurements are representative of broadband noise sources. The Class B average conducted emission limits are from 10 to 20 dB more restrictive than the Class A average limits.

Figure 1-4 shows a plot of both the average and the quasi-peak FCC/CISPR conducted emission limits.

1.5.4 Administrative Procedures

The FCC rules not only specify the technical standards (limits) that a product must satisfy but also the administrative procedures that must be followed and the measuring methods that must be used to determine compliance. Most administrative procedures are contained in Part 2, Subpart I (Marketing of Radio Frequency Devices), Subpart J (Equipment Authorization Procedures), and Subpart K (Importation of Devices Capable of Causing Harmful Interference) of the FCC Rules and Regulations.

Not only must a product be tested for compliance with the technical standards contained in the regulations, but also it must be labeled as compliant (§ 15.19), and information must be provided to the user (§ 15.105) on its interference potential.

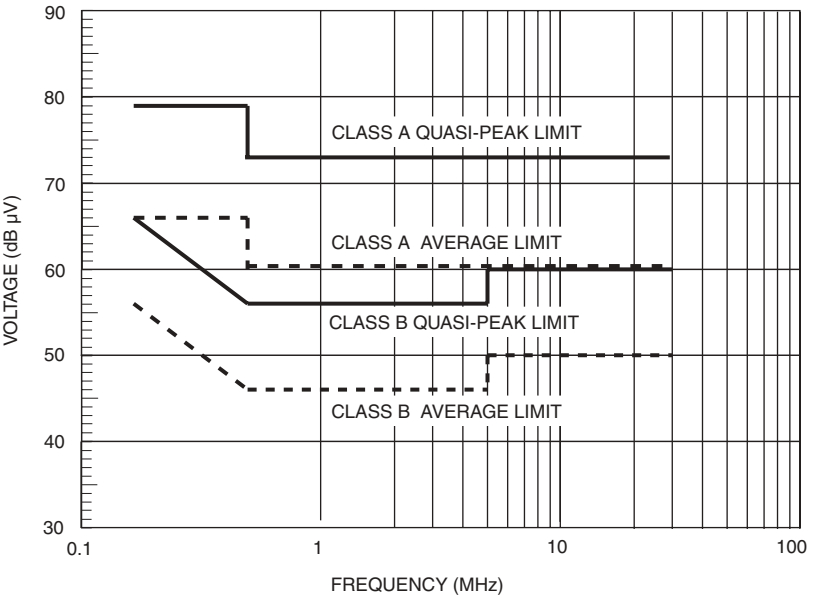


FIGURE 1-4. FCC/CISPR conducted emission limits.

In addition to the technical standards mentioned above, the rules also contain a noninterference requirement, which states that if use of the product causes harmful interference, the *user* may be required to cease operation of the device (§ 15.5). Note the difference in responsibility between the technical standards and the noninterference requirement. Although meeting the technical standards (limits) is the responsibility of the manufacturer or importer of the product, satisfying the noninterference requirement is the responsibility of the user of the product.

In addition to the initial testing to determine compliance of a product, the rules also specify that the manufacturer or importer is responsible for the continued, or ongoing, compliance of subsequently manufactured units (§ 2.953, 2.955, 2.1073, 2.1075).

If a change is made to a compliant product, the manufacturer has the responsibility to determine whether that change has an effect on the compliance of the product. The FCC has cautioned manufacturers (Public Notice 3281, April 7, 1982) to note that:

> Many changes, which on their face seem insignificant, are in fact very significant. Thus a change in the layout of a circuit board, or the addition or removal or even rerouting of a wire, or even a change in the logic will almost surely change the emission characteristics of the device. Whether this change in characteristics is enough to throw the product out of compliance can best be determined by retesting.

As of this writing (September 2008), the FCC has exempted eight subclasses of digital devices (§ 15.103) from meeting the technical standards of the rules. These are as follows:

1. Digital devices used exclusively in a transportation vehicle such as a car, plane, or boat.
2. Industrial control systems used in an industrial plant, factory, or public utility.
3. Industrial, commercial, or medical test equipment.
4. Digital devices exclusively used in an appliance such as a microwave oven, dishwasher, clothes dryer, air conditioner, and so on.
5. Specialized medical devices generally used at the direction or under the supervision of a licensed health care practitioner, whether used in a patient's home or a health care facility. Note, medical devices marketed through retail channels for use by the general public, are not exempted.
6. Devices with power consumption not exceeding 6 nW, for example, a digital watch.
7. Joystick controllers or similar devices (such as a mouse) that contain no digital circuitry. Note, a simple analog to digital converter integrated circuit (IC) is allowed in the device.
8. Devices in which the highest frequency is below 1.705 MHz and that does not operate from the ac power line, or contain provisions for operation while connected to the ac power line.

Each of the above exempted devices is, however, still subject to the noninterference requirement of the rules. If any of these devices actually cause harmful interference in use, the user must stop operating the device or in some way remedy the interference problem. The FCC also states, although not mandatory, it is strongly recommended that the manufacturer of an exempted device endeavor to have that device meet the applicable technical standards of Part 15 of the rules.

Because the FCC has purview over many types of electronic products, including digital electronics, design and development organizations should have a complete and current set of the FCC rules applicable to the types of products they produce. These rules should be referenced during the design to avoid subsequent embarrassment when compliance demonstration is required.

The complete set of the FCC rules is contained in the Code of Federal Regulations, Title 47 (Telecommunications)—Parts 0 to 300. They consist of five volumes and are available from the Superintendent of Documents, U.S. Government Printing Office. The FCC rules are in the first volume that contains Parts 0 to 19 of the Code of Federal Regulations. A new edition is published in the spring of each year and contains all current regulations codified as of October 1 of the previous year. The Regulations are also available online at the FCC's website, www.fcc.gov.

When changes are made to the FCC regulations, there is a transition period before they become official. This transition period is usually stated as x-number of days after the regulation is published in the Federal Register.

1.5.5 Susceptibility

In August 1982, the U.S. Congress amended the Communications Act of 1934 (House Bill #3239) to give the FCC authority to regulate the susceptibility of home electronics equipment and systems. Examples of home electronics equipment are radio and television sets, home burglar alarm and security systems, automatic garage door openers, electronic organs, and stereo/high-fidelity systems. Although this legislation is aimed primarily at home entertainment equipment and systems, it is not intended to prevent the FCC from adopting susceptibility standards for devices that are also used outside the home. To date, however, the FCC has not acted on this authority. Although it published an inquiry into the problem of Radio Frequency Interference to Electronic Equipment in 1978 (General Docket No. 78-369), the FCC relies on self-regulation by industry. Should industry become lax in this respect, the FCC may move to exercise its jurisdiction.

Surveys of the electromagnetic environment (Heirman 1976, Janes 1977) have shown that a field strength greater than 2 V/m occurs about 1% of the time. Because no legal susceptibility requirements exist for commercial equipment in the United States, a reasonable minimum immunity level objective might be 2 to 3 V/m. Clearly products with susceptibility levels of less than 1 V/m are not well designed and are very likely to experience interference from rf fields during their life span.

In 1982, the government of Canada released an Electromagnetic Compatibility Advisory Bulletin (EMCAB-1) that defined three levels, or grades, of immunity for electronic equipment, and stated the following:

1. Products that meet GRADE 1 (1 V/m) are likely to experience performance degradation.
2. Products that meet GRADE 2 (3 V/m) are unlikely to experience degradation.
3. Products that meet GRADE 3 (10 V/m) should experience performance degradation only under very arduous circumstances.

In June 1990, an updated version of EMCAB-1 was issued by Industry Canada. This updated version concludes that products located in populated areas can be exposed to field strengths that range from 1 V/m to 20 V/m over most of the frequency band.

1.5.6 Medical Equipment

Most medical equipment (other than what comes under the Part 18 Rules) is exempt from the FCC Rules. The Food and Drug Administration (FDA), not

the FCC, regulates medical equipment. Although the FDA developed EMC standards, as early as 1979 (MDS-201-0004, 1979), they have never officially adopted them as mandatory. Rather, they depend on their inspectors' guideline document to assure that medical devices are properly designed to be immune to electromagnetic interference (EMI). This document, *Guide to Inspections of Electromagnetic Compatibility Aspects of Medical Devices Quality Systems*, states the following:

> At this time the FDA does not require conformance to any EMC standards. However, EMC should be addressed during the design of new devices, or redesign of existing devices.

However, the FDA is becoming increasingly concerned about the EMC aspects of medical devices. Inspectors are now requiring assurance from manufacturers that they have addressed EMC concerns during the design process, and that the device will operate properly in its intended electromagnetic environment. The above-mentioned Guide encourages manufacturers to use IEC 60601-1-2 Medical Equipment, Electromagnetic Compatibility Requirements and Tests as their EMC standard. IEC 60601-1-2 provides limits for both emission and immunity, including transient immunity such as electrostatic discharge (ESD).

As a result, in most cases, IEC 60601-1-2 has effectively become the unofficial, de facto, EMC standard that has to be met for medical equipment in the United States.

1.5.7 Telecom

In the United States, telecommunications central office (network) equipment is exempt from the FCC Part 15 Rules and Regulations as long as it is installed in a dedicated building or large room owned or leased by the telephone company. If it is installed in a subscriber's facility, such as an office or commercial building, the exemption does not apply and the FCC Part 15 Rules are applicable.

Telecordia's (previously Bellcore's) GR-1089 is the standard that usually applies to telecommunications network equipment in the United States. GR-1089 covers both emission and susceptibility, and it is somewhat similar to the European Union's EMC requirements. The standard is often referred to as the NEBS requirements. NEBS stands for New Equipment Building Standard. The standard is derived from the original AT&T Bell System internal NEBS standard.

These standards are not mandatory legal requirements but are contractual between the buyer and the seller. As such, the requirements can be waived or not applied in some cases.

1.5.8 Automotive

As stated, much (although not all) of the electronics built into transportation vehicles are exempt from EMC regulation, such as the FCC Part 15 Rules, in the United States (§ 15.103). This does not mean that vehicle systems do not have legal EMC requirements. In many regions of the world, there are legislated requirements for vehicle electromagnetic emissions and immunity. The legislated requirements are typically based on many internationally recognized standards, including CISPR, International Organization for Standardization (ISO), and the Society of Automotive Engineers (SAE). Each of these organizations has published several EMC standards applicable to the automotive industry. Although these standards are voluntary, the automotive manufacturers either rigorously apply them or use these standards as a reference in the development of their own corporate requirements. These developed corporate requirements may include both component and vehicle level items and are often based upon the customer satisfaction goals of the manufacturer—therefore, they almost have the effect of mandatory standards.

For example, SAE J551 is a vehicle-level EMC standard, and SAE J1113 is a component-level EMC standard applicable to individual electronic modules. Both standards cover emissions and immunity and are somewhat similar to the military EMC standards.

The resulting vehicle EMC standards cover both emissions and immunity and are some of the toughest EMC standards in the world, partly because of the combination of types of systems on vehicles and their proximity to each other. These systems include high-voltage discharges (such as spark ignition systems) located near sensitive entertainment radio receiver systems, wiring for inductive devices such as motors and solenoids in the same wiring harness as data communication lines, and with the newer "hybrid vehicles" high-current motor drive systems that operate at fast switching speeds. The radiated emission standards are typically 40 dB more stringent than the FCC Class B limits. Radiated immunity tests are specified up to an electric field strength of 200 V/m (or in some cases higher) as compared with 3 or 10 V/m for most non-automotive commercial immunity standards.

In the European Union, vehicles and electronic equipment intended for use in these vehicles are exempt from the EMC Directive (204/108/EC), but they do fall within the scope of the automotive directive (95/54/EC) that contains EMC requirements.

1.6 CANADIAN EMC REQUIREMENTS

The Canadian EMC regulations are similar to those of the United States. The Canadian regulations are controlled by Industry Canada. Table 1-7 lists the Canadian EMC standards applicable to various types of products. These standards can be accessed from the Industry Canada web page (www.ic.gc.ca).

TABLE 1-7. Canadian EMC Test Standards.

Equipment Type	Standard
Information technology equipment (ITE)[a]	ICES–003
Industrial, Scientific & Medical Equipment (ISM)	ICES–001
Terminal Equipment Connected to the Telephone Network	CS–03

[a]Digital Equipment.

The ITE and ISM standards can be accessed from the Industry Canada home page by following the following links: A-Z Index/Spectrum Management and Telecommunications/Official Publications/Standards/Interference-Causing Equipment Standards (ICES). The telecom standard can be accessed from the Industry Canada home page by following the following links: A-Z Index/Spectrum Management and Telecommunications/Official Publications/Standards/Terminal Equipment-Technical Specifications List.

The methods of measurement and actual limits for ITE are contained in CAN/CSA-CEI/IEC CISPR 22:02, Limits and Methods of Measurement of Radio Disturbance Characteristics of Information Technology Equipment.

To reduce the burden on U.S. and Canadian manufacturers, the United States and Canada have a mutual recognition agreement whereby each country agrees to accept test reports from the other country for equipment authorization purposes (FCC Public Notice 54795, July 12, 1995).

1.7 EUROPEAN UNION'S EMC REQUIREMENTS

In May 1989, the European Union (EU) published a directive (89/336/EEC) relating to electromagnetic compatibility, which was to be effective January 1, 1992. However, the European Commission underestimated the task of implementing the directive. As a result, the European Commission amended the directive in 1992 allowing for a 4-year transition period and requiring full implementation of the EMC directive by January 1, 1996.

The European EMC directive differs from the FCC regulations by including immunity requirements in addition to emission requirements. Another difference is that the directive, without exception, covers all electrical/electronic equipment. There are no exemptions—the EMC directive even covers a light bulb. The directive does, however, exclude equipment that is covered by another directive with EMC provisions, such as the automotive directive. Another example would be medical equipment, which comes under the medical directive (93/42/EEC) not the EMC directive.

1.7.1 Emission Requirements

As stated, the EU's conducted emission requirements are now the same as the FCC's (see Tables 1-5 and 1-6 as well as Fig. 1-4). The radiated emission

TABLE 1-8. CISPR Radiated Emission Limits at 10 m.

Frequency (MHz)	Class A Limit (dB μV/m)	Class B Limit (dB μV/m)
30–230	40	30
230–1000	47	37

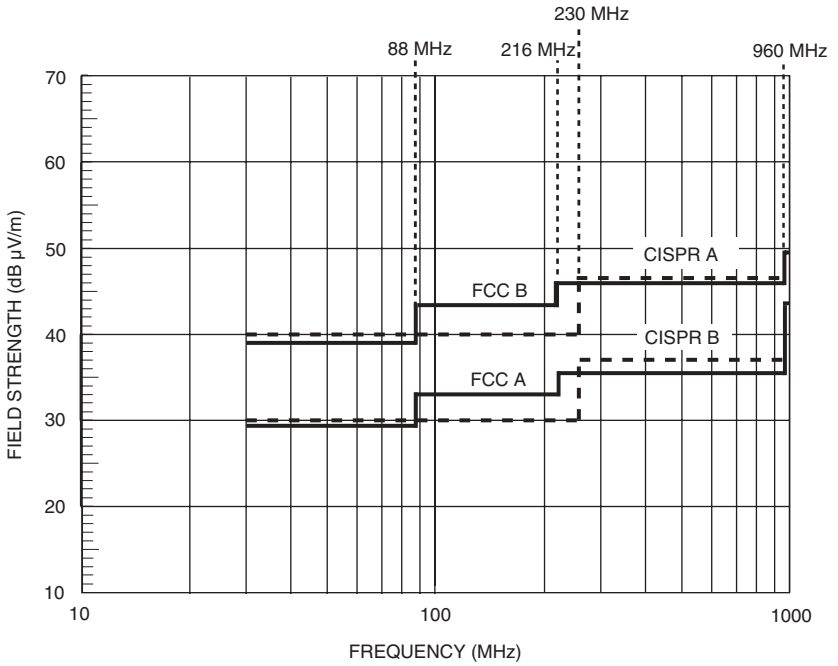


FIGURE 1-5. Comparison of FCC and CISPR radiated emission limits, measured at a distance of 10 m.

standards are similar but not exactly the same. Table 1-8 shows the European Union's Class A and Class B radiated emission limits when measured at 10 m.

Figure 1-5 compares the EU's radiated emission standard with the current FCC standard over the frequency range of 30 MHz to 1000 MHz. The FCC Class B limits have been extrapolated to a 10-m measuring distance for this comparison. As can be observed the European (CISPR) limits are more restrictive in the frequency range from 88 to 230 MHz. Below 88 MHz and above 230 MHz the CISPR and FCC limits are virtually the same (within 0.5 dB of each other). However, the EU has no radiated emission limit above 1 GHz, whereas the FCC limits, under some circumstances (see Table 1-4), go up to 40 GHz.

Table 1-9 is a composite worst-case combination of the FCC and CISPR radiated emission limits when measured at 10 m.

TABLE 1-9. Composite Worst-Case Radiated Emission Limits for Commercial Products, Measured at a Distance of 10 m.

Frequency (MHz)	Class A Limit (dB μV/m)	Class B Limit (dB μV/m)
30–230	39	29.5
230–1000	46.5	35.5
>1000	49.5	43.5

1.7.2 Harmonics and Flicker

The EU has two additional emission requirements that relate to power quality issues—harmonics and flicker. These regulations apply to products that draw an input current of 16 A per phase or less and are intended to be connected to the public ac power distribution system. The FCC has no similar requirement.

The harmonic requirement (EN 61000-3-2) limits the harmonic content of the current drawn by the product from the ac power line, (see Table 18-3). The generation of harmonics is the result of the nonlinear behavior of the loads connected to the ac power line. Common nonlinear loads include switched-mode power supplies, variable-speed motor drives, and electronic ballasts for fluorescent lamps.

A major source of harmonics is a full-wave rectifier connected directly to the ac power line and followed by a large-value capacitor input filter. Under these circumstances, current is only drawn from the power line when the input voltage exceeds that on the filter capacitor. As a result, current is drawn from the power line only on the peaks of the ac voltage waveform (see Fig. 13-4). The resultant current waveshape is rich in odd harmonics (third, fifth, seventh, etc.). Total harmonic distortion (THD) values of 70% to 150% are not uncommon under these circumstances.

The number of harmonics present is determined by the rise and fall time of the current pulse, and their magnitude by the current wave shape. Most switching power supplies (the exception is very low-power supplies) and variable-speed motor drives cannot meet this requirement without some kind of passive or active power factor correction circuitry.

To alleviate this problem, the ac input current pulse must be spread out over a larger portion of a cycle to reduce the harmonic content. Normally the THD of the current pulse must be reduced to 25% or less to be compliant with the EU regulations.

The flicker requirements (EN 61000-3-3) limit the transient ac power line current drawn by the product; see Table 18-4. The purpose of this requirement is to prevent lights from flickering, because it is perceived as being disturbing to people. The regulations are based on not providing a noticeable change in the illumination of a 60-W incandescent lamp powered off the same ac power supply as the equipment under test.

Because of the finite source impedance of the power line, the changing current requirements of equipment connected to the line produces corresponding voltage fluctuations on the ac power line. If the voltage variation is large enough, it will produce a perceptible change in lighting illumination. If the load changes are of sufficient magnitude and repetition rate, the resulting flickering of lights can be irritating and disturbing.

To determine an applicable limit, many people were subjected to light flicker to determine the irritability threshold. When the flicker rate is low (<1 per minute), the threshold of irritability is when the ac line voltage changes by 3%. People are most sensitive to light flicker when the rate is around 1000 times per minute. At a rate of 1000 times per minute, a 0.3% voltage change is just as irritating as a 3% change at less than one change per minute. Above about 1800 changes per minute, light flicker is no longer perceived.

Most EMC emission requirements are based on the magnitude of a measured parameter not exceeding a specified amount (the limit). However, flicker tests are different in that they require many measurements to be made and then a statistical analysis to be performed on the measured data to determine whether the limit is exceeded.

For most equipment, this requirement is not a problem because they naturally do not draw large transient currents off the ac power line. However, the requirement can be a problem for products that suddenly switch on heaters that draw large currents, or motors under a heavy load. An example would be when an air conditioner compressor or a large heater in a copy machine is suddenly switched on.

1.7.3 Immunity Requirements

The EU's immunity requirements cover radiated and conducted immunity, as well as transient immunity that include ESD, electrical fast transient (EFT), and surge.

The EFT requirement simulates noise generated by inductively switched loads on the ac power line. As a contactor is opened to an inductive load, an arc is formed that extinguishes and restarts many times. The surge requirement is intended to simulate the effect of a nearby lightning pulse.

In addition, the EU has susceptibility requirements that cover ac voltage dips, sags, and interruptions.

For additional information on these transient immunity and power line disturbance requirements, see Sections 14.3 and 14.4.

1.7.4 Directives and Standards

The European regulations consist of directives and standards. The directives are very general and are the legal requirements. The standards provide one way, but not the only way, to comply with the directive.

The EMC Directive 2004/108/EC (which superceded the original EMC Directive 89/336/EEC) defines the *essential requirements* for a product to be marketed in the EU. They are as follows:

1. The equipment must be constructed to ensure that any electromagnetic disturbance it generates allows radio and telecommunication equipment and other apparatus to function as intended.
2. The equipment must be constructed with an inherent level of immunity to externally generated electromagnetic disturbances.

These are the *only legal requirements* with respect to EMC and the requirements are vague. The directive provides for two methods of demonstrating compliance with its requirements. The most commonly used is by a declaration of conformity; the other option is the use of a technical construction file.

If a product is tested to and complies with the applicable EMC standards it is presumed to meet the requirements of the directive, and the manufacturer can produce a declaration of conformity attesting to that fact.

A declaration of conformity is a self-certification process in which the responsible party, manufacturer or importer, must first determine the applicable standards for the product, test the product to the standards, and issue a declaration declaring compliance with those standards and the EMC directive. The declaration of conformity can be a single-page document but must contain the following:

- Application of which council directives (all applicable directives)
- Standards used (including date of standard) to determine conformity
- Product name and model number, also serial numbers if applicable
- Manufacturer's name and address
- A dated declaration that the product conforms to the directives
- A signature by a person empowered to legally bind the manufacturer

The technical construction file approach to demonstrating conformity is unique to the European Union. The technical construction file is often used where no harmonized standards exist for the product and the manufacturer does not think that the generic standards are appropriate. In this case, the manufacturer produces a technical file to describe the procedures and tests used to ensure compliance with the EMC directive. The manufacturer can develop its own EMC specifications and test procedures. The manufacturer can decide how, where, when, or if, the product is tested for EMC. An independent *competent body*, however, must approve the technical construction file. The competent bodies are appointed by the individual states of the European Union, and the European Commission publishes a list of them in the Official Journal of the European Union. The competent body must agree that, using the manufacturer's procedures and tests, the product satisfies the essential

requirements of the EMC directive. This approach is acceptable, because in the European Union, the EMC directive is the legal document that must be satisfied, not the standards. In most other jurisdictions, the standards are the legal documents that must be complied with.

Products whose compliance with the EMC directive has been demonstrated by one of the above procedures shall be labeled with the *CE mark*. The CE mark consists of the lower case letters "ce" in a specified, distinctive font. Affixing the CE mark to a product indicates conformity to *all* applicable directives, not just the EMC directive. Other applicable directives might be, the safety directive, the toy directive, the machinery directive, and so on.

Two types of standards exist in the European Union: product specific and generic.* Product-specific standards always take precedence over generic standards. However, if no applicable product-specific standard exists for a product, the generic standards are then applicable. Emission and immunity requirements for a product are usually covered by different standards. Currently, over 50 different standards are associated with the EMC directive. Table 1-10 lists some of the more commonly applicable product-specific standards, as well as the four generic EMC standards. If a product-specific standard does not exist in a category, then the requirement defaults to the appropriate generic standard.

The EU's standards writing organization CENELEC (the European Committee for Electro-Technical Standardization) has been given the task of drawing up the corresponding technical specifications meeting the essential requirements of the EMC directive, compliance with which will provide a presumption of conformity with the essential requirements of the EMC

TABLE 1-10. European Union's EMC Test Standards.

Equipment Type	Emission	Immunity
Product Specific Standards		
Information Technology Equipment (ITE)	EN 55022	EN 55024
Industrial, Scientific & Medical Equipment (ISM)	EN 55011	–
Radio & Television Receivers	EN 55013	EN 55020
Household Appliances/Electric Tools	EN 55014-1	EN 55014-2
Lamps & Luminaries	EN 55015	EN 61547
Adjustable Speed Motor Drives	EN 61800-3	EN 61800-3
Medical Equipment[a]	EN 60601-1-2	EN 60601-1-2
Generic Standards		
Residential, Commercial, Light Industrial Environment	EN 61000-6-3	EN 61000-6-1
Heavy Industrial Environment	EN 61000-6-4	EN 61000-6-2

[a]Covered by the Medical Directive (93/42/EEC), not the EMC Directive

* A third type of standard also exists, which is a basic standard. Basic standards are usually test or measurement procedures and are referenced by the product-specific or generic standards.

directive. Such specifications are referred to as harmonized standards. Most CENELEC standards are derived from International Electro-Technical Committee (ITC) or CISPR standards—ITC for immunity standards and CISPR for emission standards. The CENELEC standards, or European Norms (EN), are not official until a reference to them is published in the "Official Journal of the European Union."

As new standards come into existence and existing standards are modified, as regularly happens, a transition period, usually of 2 years is specified in the standard. During the transition period, either the old standard or the new standard can be used to demonstrate compliance with the EMC directive.

The latest information on the EMC Directive 2004/108/EC and the harmonized standards can be obtained on the following website: http://europa.eu.int/comm/enterprise/newapproach/standardization/harmstds/reflist/emc.html.

In light of the large breadth and scope of the EMC Directive and the variety of products covered, the European Commission in 1997 felt it necessary to publish a 124-page guideline to the interpretation of the EMC directive to be used by manufacturers, test laboratories, and other parties affected by the directive (European Commission, 1997). This guideline was intended to clarify matters and procedures relating to the interpretation of the EMC Directive. It also clarified the application of the Directive to components, subassemblies, apparatus, systems, and installations, as well as the application of the Directive to spare parts, used, and repaired apparatus.

1.8 INTERNATIONAL HARMONIZATION

It would be desirable to have one international EMC standard for allowable emission and immunity of electronic products, instead of many different national standards. This would allow a manufacturer to design and test a product to one standard that would be acceptable worldwide. Figure 1-6 depicts a typical commercial product and shows the different types of EMC requirements, both emission and immunity, that it might have to meet in a harmonized world market.

Even more important than a single uniform EMC standard is a single uniform EMC test procedure. If the test procedure is the same, then an EMC test could be performed once and the results compared against many different standards (limits) to determine compliance with each regulation. When the test procedures are different, however, the product must be retested for each standard, which is a costly and time-consuming task.

The most likely vehicle for accomplishing harmonization is the European Union's EMC standards, which are based on the CISPR standards. CISPR was formed in 1934 to determine measurement methods and limits for radio-frequency interference to facilitate international trade. CISPR has no regulatory authority, but its standards, when adopted by governments, become

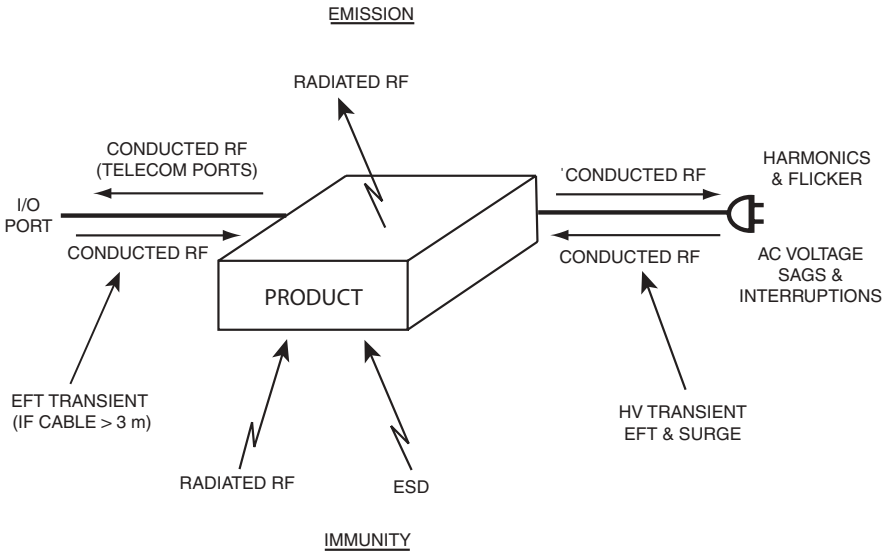


FIGURE 1-6. Typical composite worldwide commercial EMC requirements.

national standards. In 1985 CISPR adopted a new set of emission standards (Publication 22) for Information Technology Equipment (computer and digital electronics). The European Union has adopted the CISPR standard as the basis for their emission requirements. As a voting member of CISPR, the United States voted in favor of the new standard. This action puts considerable pressure on the FCC to adopt the same standards.

In 1996, the FCC modified its Part 15 Rules to allow manufacturers to use a Declaration of Conformity as a compliance procedure for personal computers and their peripherals, which is similar to that used by the EU's EMC regulations. As stated, the FCC also has adopted the CISPR limits for conducted emission.

1.9 MILITARY STANDARDS

Another important group of EMC standards are those issued by the U.S. Department of Defense and are applicable to military and aerospace equipment. In 1968, the Department of Defense consolidated the multitude of different EMC standards from the various branches of the service into two universally applicable standards. MIL-STD-461 specified the limits that had to be met, and MIL-STD-462 specified the test methods and procedures for making the tests contained in MIL-STD-461. These standards are more

stringent than the FCC regulations, and they cover immunity as well as emissions in the frequency range of 30 Hz to 40 GHz.

Over the years, these standards have gone through revisions that ranged from MIL-STD-461A in 1968 to MIL-STD-461E in 1999. In 1999, MIL-STD-461D (Limits) and MIL-STD-462D (Test Procedures) were merged into one standard MIL-STD-461E that covered both limits and test procedures.*

Unlike commercial standards, MIL-STDs are not legal requirements; rather, they are contractual requirements. As such, test limits can be negotiated and waivers are possible. Earlier versions are still applicable to current products because the requirements are contractual, not legal. Normally whatever version the original procurement contract specified is still applicable.†

The test procedures specified in the military standards are often different than those specified by commercial EMC standards, which makes a direct comparison of the limits difficult. For radiated emissions the military standard specifies enclosed chamber (shielded room) testing, whereas the FCC and the EU rules require open-area testing. For conducted emission testing, the military standards originally measured current, whereas the commercial standards measure voltage.

As more was learned about EMC testing and its accuracy, the military has come under some criticism for some of its test procedures. As a result, the military has adopted some of the commercial test procedures. For example, MIL-STD-461E specifies the use of a LISN and the measurement of voltage rather than current for conducted emission testing. Also MIL-STD-461E requires that some absorber material must be used on the walls of chambers used for emission and immunity testing to make the chamber at least partially anechoic.

Table 1-11 is a list of the emission and immunity requirements established by MIL-STD-461E. Tests are required for both radiated and conducted emissions as well as for radiated, conducted, and high-voltage transient susceptibility.

The military standards are application specific, often with different limits for different environments (such as Army, Navy, aerospace, etc.). Some requirements listed in Table 1-11 are applicable to only certain environments and not to others. Table 1-12 lists the applicability of the requirements to the various environments.

1.10 AVIONICS

The commercial avionics industry has its own set of EMC standards, which are similar to those of the military. These standards apply to the entire spectrum of commercial aircraft, which includes light general aviation aircraft, helicopters,

* On December 10, 2007, MIL-STD 461F was released.

† By contrast, when a commercial standard is revised or modified, all newly manufactured products must comply with the new limits by the end of the specified transition period.

TABLE 1-11. Emission and Susceptibility Requirements of MIL-STD-461E.

Requirement	Description
CE101	Conducted Emissions, Power Leads, 30 Hz to 10 kHz
CE102	Conducted Emissions, Power Leads, 10 kHz to 10 MHz
CE106	Conducted Emissions, Antenna Terminals, 10 kHz to 40 GHz
CS101	Conducted Susceptibility, Power Leads, 30 Hz to 50 kHz
CS103	Conducted Susceptibility, Antenna Port, Inter–modulation, 15 kHz to 10 GHz
CS104	Conducted Susceptibility, Antenna Port, Rejection of Undesired Signals, 30 Hz to 20 GHz
CS105	Conducted Susceptibility, Antenna Port, Cross-modulation, 30 Hz to 20 GHz
CS109	Conducted Susceptibility, Structure Current, 60 Hz to 100 kHz
CS114	Conducted Susceptibility, Bulk Current Injection, 10 kHz to 40 MHz
CS115	Conducted Susceptibility, Bulk Current Injection, Impulse Excitation
CS116	Conducted Susceptibility, Damped Sinusoidal Transients, Cables and Power Leads, 10 kHz to 100 MHz
RE101	Radiated Emission, Magnetic Field, 30 Hz to 100 kHz
RE102	Radiated Emission, Electric Field, 10 kHz to 18 GHz
RE103	Radiated Emission, Antenna Spurious and Harmonic Outputs, 10 kHz to 40 GHz
RS101	Radiated Susceptibility, Magnetic Field, 30 Hz to 100 kHz
RS103	Radiated Susceptibility, Electric Field, 10 kHz to 40 GHz
RS105	Radiated Susceptibility, Transient Electromagnetic Field

TABLE 1-12. Requirement Applicability Matrix, MIL-STD-461E.

Equipment Installed In, On, or Launched From the Following Platforms or Installations	CE101	CE102	CE106	CS101	CS103	CS104	CS105	CS109	CS114	CS115	CS116	RE101	RE102	RE103	RS101	RS103	RS105
Surface Ships	N	A	L	A	S	S	S	N	A	L	A	A	A	L	A	A	L
Submarines	A	A	L	A	S	S	S	L	A	L	A	A	A	L	A	A	L
Aircraft, Army, & Flight Line	A	A	L	A	S	S	S	N	A	A	A	A	A	L	A	A	L
Aircraft, Navy	L	A	L	A	S	S	S	N	A	A	A	L	A	L	L	A	L
Aircraft, Air Force	N	A	L	A	S	S	S	N	A	A	A	N	A	L	N	A	N
Space Systems & Launch Eq.	N	A	L	A	S	S	S	N	A	A	A	N	A	L	N	A	N
Ground, Army	N	A	L	A	S	S	S	N	A	A	A	N	A	L	L	A	N
Ground, Navy	N	A	L	A	S	S	S	N	A	A	A	N	A	L	A	A	L
Ground, Air Force	N	A	L	A	S	S	S	N	A	A	A	N	A	L	N	A	N

A = applicable, L = limited applicability as specified in the standard, S = applicable only if specified in procurement document, N = not applicable.

and jumbo jets. The Radio Technical Commission for Aeronautics (RTCA) produces these standards for the avionics industry. The current version is RTCA/DO-160E *Environmental Conditions and Test Procedures For Airborne Equipment* and was issued in December 2004. Sections 15 through 23 and Section 25 cover EMC issues.

Like the military standard, DO-160E is a contractual, not legal, requirement, so its terms may be negotiable.

1.11 THE REGULATORY PROCESS

We are all probably familiar with the phrase *ignorance of the law is no defense*. How then do governments make their commercial EMC regulations public, so that we all presumably know of their existence? In most countries, regulations are made public by publication, or being referenced, in the "Official Journal" of that country. In the United States, the official journal is the *Federal Register*; in Canada, it is the *Canada Gazette*; and in the European Union, it is the *Official Journal of the European Union*.

Once a regulation is published, or referenced, in the official journal, *its official*, and everyone is presumed to know of its existence.

1.12 TYPICAL NOISE PATH

A block diagram of a typical noise path is shown in Fig. 1-7. As shown, three elements are necessary to produce an interference problem. First, there must be a *noise source*. Second, there must be a *receptor* circuit that is susceptible to the noise. Third, there must be a *coupling channel* to transmit the noise from the source to the receptor. In addition, the characteristics of the noise must be such that it is emitted at a *frequency* that the receptor is susceptible, an *amplitude* sufficient to affect the receptor, and a *time* the receptor is susceptible to the noise. A good way to remember the important noise characteristics is with the acronym FAT.

The first step in analyzing a noise problem is to define the problem. This is done by determining what is the noise source, what is the receptor, what is the coupling channel, and what are the FAT characteristics of the noise. It follows that there are three ways to break the noise path: (1) the characteristics of the noise can be changed at the source, (2) the receptor can be made insensitive to the noise, or (3) the transmission through the coupling channel

FIGURE 1-7. Before noise can be a problem, there must be a noise source, a receptor, and a coupling channel.

can be eliminated or minimized. In some cases, the noise suppression techniques must be applied to two or to all three parts of the noise path.

In the case of an emission problem, we are most likely to attack the source of the emissions by changing its characteristics—its frequency, amplitude, or time. For a susceptibility problem, we are most likely to direct our attention to modifying the receptor to increase its immunity to the noise. In many cases, modifying the source or receptor is not practical, which then leaves us with only the option of controlling the coupling channel.

As an example, consider the circuit shown in Fig. 1-8. It shows a shielded direct current (dc) motor connected to its motor-drive circuit. Motor noise is interfering with a low-level circuit in the same equipment. Commutator noise from the motor is conducted out of the shield on the leads going to the drive circuit. From the leads, noise is radiated to the low-level circuitry.

In this example, the noise source consists of the arcs between the brushes and the commutator. The coupling channel has two parts: conduction on the motor leads and radiation from the leads. The receptor is the low-level circuit. In this case, not much can be done about the source or the receptor. Therefore, the interference must be eliminated by breaking the coupling channel. Noise conduction out of the shield or radiation from the leads must be stopped, or both steps may be necessary. This example is discussed more fully in Section 5.7.

1.13 METHODS OF NOISE COUPLING

1.13.1 Conductively Coupled Noise

One of the most obvious, but often overlooked, ways to couple noise into a circuit is on a conductor. A wire run through a noisy environment may pick up

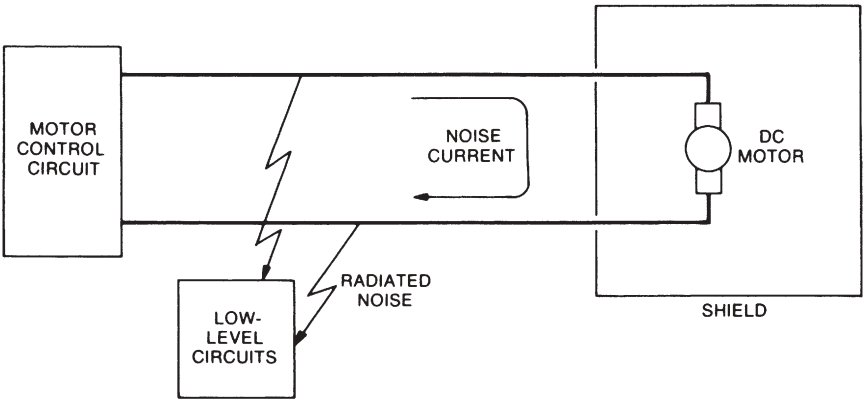


FIGURE 1-8. In this example, the noise source is the motor, and the receptor is the low-level circuit. The coupling channel consists of conduction on the motor leads and radiation from the leads.

noise and then conduct it to another circuit. There it causes interference. The solution is to prevent the wire from picking up the noise or to remove the noise from it by filtering before it interferes with the susceptible circuit.

The major example in this category is noise conducted into a circuit on the power supply leads. If the designer of the circuit has no control over the power supply, or if other equipment is connected to the power supply, it becomes necessary to decouple or filter the noise from the wires before they enter the circuit. A second example is noise coupled into or out of a shielded enclosure by the wires that pass through the shield.

1.13.2 Common Impedance Coupling

Common impedance coupling occurs when currents from two different circuits flow through a common impedance. The voltage drop across the impedance observed by each circuit is influenced by the other circuit. This type of coupling usually occurs in the power and/or ground system. The classic example of this type of coupling is shown in Fig. 1-9. The ground currents 1 and 2 both flow through the common ground impedance. As far as circuit 1 is concerned, its ground potential is modulated by ground current 2 flowing in the common ground impedance. Some noise, therefore, is coupled from circuit 2 to circuit 1, and vice versa, through the common ground impedance.

Another example of this problem is illustrated in the power distribution circuit shown in Fig. 1-10. Any change in the supply current required by circuit 2 will affect the voltage at the terminals of circuit 1 because of the common impedances of the power supply lines and the internal source impedance of the power supply. A significant improvement can be obtained by connecting

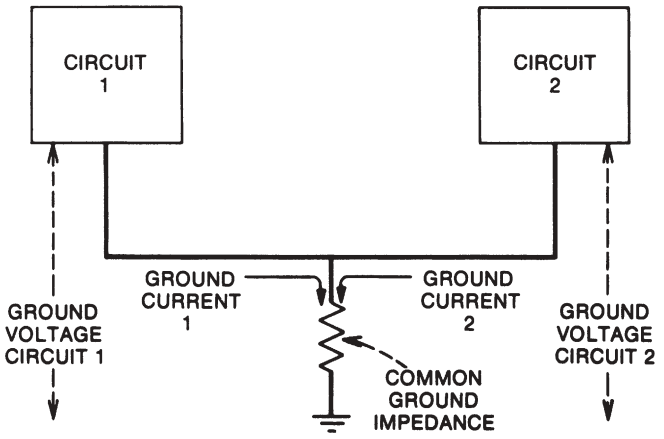


FIGURE 1-9. When two circuits share a common ground, the ground voltage of each one is affected by the ground current of the other circuit.

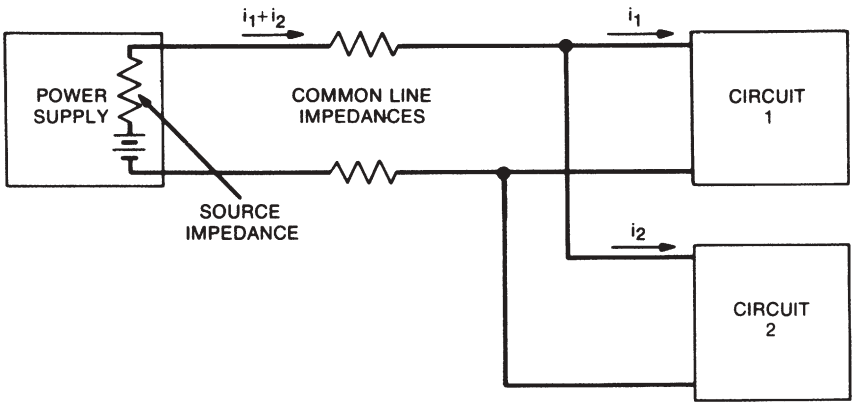


FIGURE 1-10. When two circuits share a common power supply, current drawn by one circuit affects the voltage at the other circuit.

the leads from circuit 2 directly to the power supply output terminals, thus bypassing the common line impedance. However, some noise coupling through the power supply's internal impedance will remain.

1.13.3 Electric and Magnetic Field Coupling

Radiated electric and magnetic fields provide another means of noise coupling. All circuit elements, including conductors, radiate electromagnetic fields whenever a charge is moved. In addition to this unintentional radiation, there is the problem of intentional radiation from sources such as broadcast stations and radar transmitters. When the receiver is close to the source (near field), electric and magnetic fields are considered separately. When the receiver is far from the source (far field), the radiation is considered as combined electric and magnetic or electromagnetic radiation.*

1.14 MISCELLANEOUS NOISE SOURCES

1.14.1 Galvanic Action

If dissimilar metals are used in the signal path in low-level circuitry, a noise voltage may appear from the galvanic action between the two metals. The presence of moisture or water vapor in conjunction with the two metals produces a chemical wet cell (galvanic couple). The voltage developed depends on the two metals used and is related to their positions in the galvanic series

* See Chapter 6 for an explanation of near field and far field.

TABLE 1-13. Galvanic Series.

	ANODIC END		
	(Most susceptible to corrosion)		
Group I	1. Magnesium		13. Nickel (active)
			14. Brass
	2. Zinc		15. Copper
	3. Galvanized steel		16. Bronze
Group II	4. Aluminum 2S	Group IV	17. Copper-nickel alloy
	5. Cadmium		18. Monel
	6. Aluminum 17ST		19. Silver solder
			20. Nickel (passive)[a]
	7. Steel		21. Stainless steel
	8. Iron		(passive)[a]
	9. Stainless steel		
Group III	(active)		22. Silver
	10. Lead-tin solder	Group V	23. Graphite
	11. Lead		24. Gold
	12. Tin		25 Platinum
			CATHODIC END
			(Least susceptibility to corrosion)

[a]Passivation by immersion in a strongly acidic solution.

shown in Table 1-13. The farther apart the metals are on this table, the larger the developed voltage. If the metals are the same, no potential difference can develop.

In addition to producing a noise voltage, the use of dissimilar metals can produce a corrosion problem. Galvanic corrosion causes positive ions from one metal to be transferred to the other one. This action gradually causes the anode material to be destroyed. The rate of corrosion depends on the moisture content of the environment and how far apart the metals are in the galvanic series. The farther apart the metals are in the galvanic series, the faster the ion transfer. An undesirable, but common, combination of metals is aluminum and copper. With this combination, the aluminum is eventually eaten away. The reaction slows down considerably, however, if the copper is coated with lead-tin solder because aluminum and lead-tin solder are closer in the galvanic series.

The following four elements are needed before galvanic action can occur:

1. Anode material (higher rank in Table 1-13)
2. Electrolyte (usually present as moisture)
3. Cathode material (lower rank in Table 1-13)
4. Conducting electrical connection between anode and cathode (usually present as a leakage path)

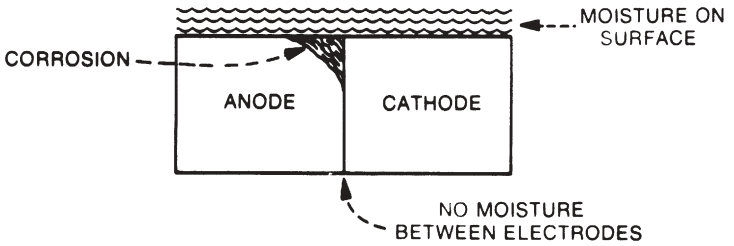


FIGURE 1-11. Galvanic action can occur if two dissimilar metals are joined and moisture is present on the surface.

Galvanic action can take place even if moisture does not get between the anode and the cathode. All that is needed is some moisture on the surface where the two metals come together, as shown in Fig. 1-11.

As observed in Table 1-13, the metals of the galvanic series are divided into five groups. When dissimilar metals must be combined, it is desirable to use metals from the same group. Usually metals from adjacent groups can be used together if the product is to be used in a fairly benign indoor environment.

Other methods of minimizing corrosion between two dissimilar metals are as follows:

- Keep the cathode material as small as possible.
- Plate one of the materials to change the group that the contact surface is in.
- Coat the surface, after joining to exclude surface moisture.

1.14.2 Electrolytic Action

A second type of corrosion is caused by electrolytic action. It is caused by a direct current flowing between two metals with an electrolyte (which could be slightly acidic ambient moisture) between them. This type of corrosion does not depend on the two metals used and will occur even if both are the same. The rate of corrosion depends on the magnitude of the current and on the conductivity of the electrolyte.

1.14.3 Triboelectric Effect

A charge can be produced on the dielectric material within a cable, if the dielectric does not maintain contact with the cable conductors. This is called the triboelectric effect. It is usually caused by mechanical bending of the cable. The charge acts as a noise voltage source within the cable. Eliminating sharp bends and cable motion minimizes this effect. A special "low-noise" cable is available in which the cable is chemically treated to minimize the possibility of charge buildup on the dielectric.

1.14.4 Conductor Motion

If a conductor is moved through a magnetic field, a voltage is induced between the ends of the wire. Because of power wiring and other circuits with high-current flow, stray magnetic fields exist in most environments. If a wire with a low-level signal is allowed to move through this field, then a noise voltage will be induced in the wire. This problem can be especially troublesome in a vibrational environment. The solution is simple: prevent wire motion with cable clamps and other tie-down devices.

1.15 USE OF NETWORK THEORY

For the exact answer to the question of how any electric circuit behaves, Maxwell's equations must be solved. These equations are functions of three space variables (x, y, z) and of time (t)—a four-dimensional problem. Solutions for any but the simplest problems are usually complex. To avoid this complexity, an approximate analysis technique called "electric circuit analysis" is used during most design procedures.

Circuit analysis eliminates the spatial variables and provides approximate solutions as a function of time (or frequency) only. Circuit analysis assumes the following:

1. All electric fields are confined to the interiors of capacitors.
2. All magnetic fields are confined to the interiors of inductors.
3. Dimensions of the circuits are small compared with the wavelength(s) under consideration.

What is really implied is that external fields, even though actually present, can be neglected in the solution of the network. Yet these external fields may not necessarily be neglected where their effect on other circuits is concerned.

For example, a 100-W power amplifier may radiate 100 mW of power. These 100 mW are completely negligible as far as the analysis and operation of the power amplifier is concerned. However, if only a small percentage of this radiated power is picked up on the input of a sensitive circuit, it may cause interference.

Even though the 100 mW of radiated emission is completely negligible to the 100-W power amplifier, a sensitive radio receiver, under the right conditions, may be capable of picking up the signal thousands of miles away.

Whenever possible, noise-coupling channels are represented as equivalent lumped component networks. For instance, a time-varying electric field that exists between two conductors can be represented by a capacitor connecting the two conductors as shown in Fig. 1-12. A time-varying magnetic field that

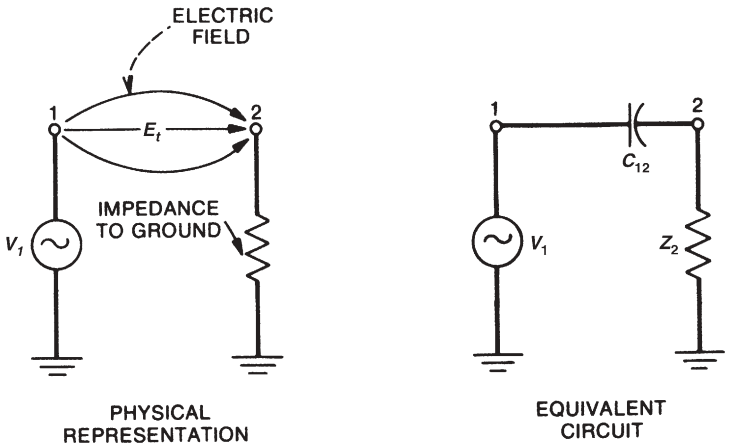


FIGURE 1-12. When two circuits are coupled by an electric field, the coupling can be represented by a capacitor.

couples two conductors can be represented by a mutual inductance between the two circuits as shown in Fig. 1-13.

For this approach to be valid, the physical dimensions of the circuits must be small compared with the wavelengths of the signals involved. Wherever appropriate, this assumption is made throughout this book.

Even when this assumption is not truly valid, the lumped component representation is still useful for the following reasons:

1. The solution of Maxwell's equations is not practical for most "real-world" noise problems because of the complicated boundary conditions.
2. Although lumped component representation will not produce the most accurate numerical answer, it does clearly show how noise depends on the parameters of the system. On the other hand, the solution of Maxwell's equations, even if possible, does not clearly show such parameter dependence.
3. To solve a noise problem, a parameter of the system must be changed, and lumped circuit analysis clearly points out the parameter dependence.

In general, the numerical values of the lumped components are extremely difficult to calculate with any precision, except for certain special geometries. One can conclude, however, that these components exist, and as will be shown, the results can be very useful even when the components are only defined in a qualitative sense.

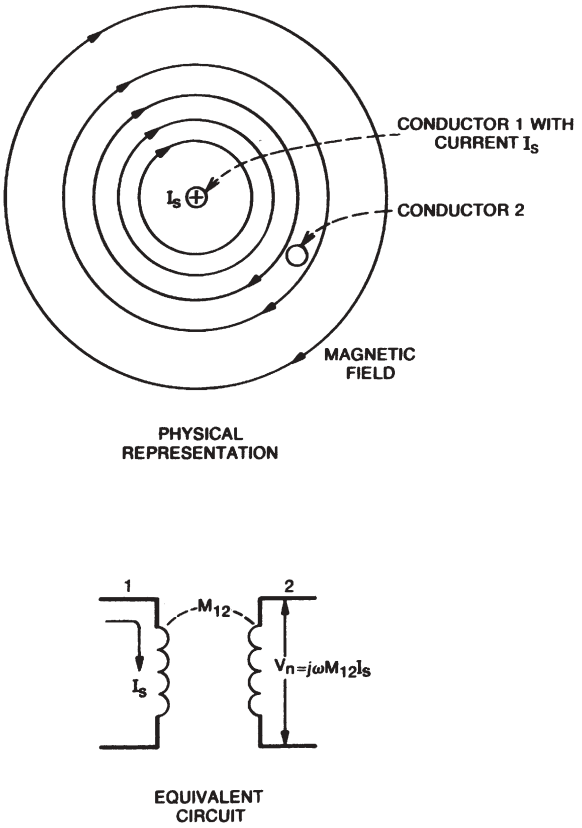


FIGURE 1-13. When two circuits are coupled by a magnetic field, the coupling can be represented as a mutual inductance.

SUMMARY

- Designing equipment that does not generate noise is as important as designing equipment that is not susceptible to noise.
- Noise sources can be grouped into the following three categories: (1) intrinsic noise sources, (2) man-made noise sources, and (3) noise caused by natural disturbances.
- To be cost effective, noise suppression must be considered early in the design.
- Electromagnetic compatibility is the ability of an electronic system to function properly in its intended electromagnetic environment.
- Electromagnetic compatibility has two aspects, emission and susceptibility.
- Electromagnetic compatibility should be designed into a product not added on at the end of the design.

- Most electronic equipment must comply with EMC regulations before being marketed.
- EMC regulations are not static but are continually changing.
- The three major EMC regulations are the FCC rules, the European Union's regulations, and the military standards.
- The following products are temporarily exempt from the FCC requirements:
  - Digital electronics in transportation vehicles
  - Industrial control systems
  - Test equipment
  - Home appliances
  - Specialized medical devices
  - Devices with power consumption not exceeding 6 nW
  - Joystick controllers or similar devices
  - Devices with clock frequencies less than 1.705 kHz, and which do not operate from the AC power line
- Virtually no products are exempt from the European Union's EMC requirements.
- Electromagnetic compatibility should be a major design objective.
- The following three items are necessary to produce an interference problem:
  - A noise source
  - A coupling channel
  - A susceptible receptor
- Three important characteristics of noise are as follows:
  - Frequency
  - Amplitude
  - Time (when does it occur)
- Metals in contact with each other must be galvanically compatible.
- Noise can be reduced in an electronic system using many techniques; a single unique solution to most noise reduction problems does not exist.

PROBLEMS

1.1 What is the difference between noise and interference?

1.2 a. Does a digital watch satisfy the FCC's definition of a digital device?
b. Does a digital watch have to meet the FCC's EMC requirements?

1.3 a. Does test equipment have to meet the technical standards of the FCC's Part 15 EMC regulations?
b. Does test equipment have to meet the non-interference requirement of the FCC's Part 15 EMC regulations?

1.4 a. Who is responsible for meeting the technical standards of the FCC's EMC regulations?

b. Who is responsible for meeting the non-interference requirement of the FCC's EMC regulations?

1.5 Are the FCC's or the European Union's Class B radiated emission limits more restrictive:

a. In the frequency range of 30 to 88 MHz?
b. In the frequency range of 88 to 230 MHz?
c. In the frequency range of 230 to 960 MHz?
d. In the frequency range of 960 to 1000 MHz?

1.6 a. Over what frequency range, below 500 MHz, does the maximum difference exist between the FCC's and the European Union's Class B radiated emission limits?

b. What is the magnitude of the maximum difference over this frequency range?

1.7 a. Over what frequency range does the FCC specify conducted emission limits?

b. Over what frequency range does the FCC specify radiated emission limits?

1.8 a. What are the essential requirements for a product to be marketed in the European Union?

b. Where are the essential requirements defined?

1.9 By what process are commercial EMC regulations made public?

1.10 What is the *major* difference between the FCC's EMC requirement and the European Union's EMC requirements?

1.11 What additional emission requirements does the European Union have that the FCC does not?

1.12 Your company is in the process of designing a new electronic widget to be marketed in the European Union. The widget will be used in both residential and commercial environments. You review the most current list of harmonized product specific EMC standards, and none of them apply to widgets. What EMC standards (specifically) should you use to demonstrate EMC compliance?

1.13 To be legally marketed in the European Union, must an electronic product be compliant with the harmonized EMC standards?

1.14 In the European Union, what are the two methods of demonstrating compliance with the EMC directive?

1.15 Which of the following EMC standards are legal requirements and which are contractual?

- FCC Part 15 B
- MIL-STD-461E

- 2004/108/EC EMC Directive
- RTCA/DO-160E for avionics
- GR-1089 for telephone network equipment
- TIA-968 for telecom terminal equipment
- SAE J551 for automobiles

1.16 What are the official journals of the following countries: the United States, Canada, and the European Union?

1.17 In the United States, does medical equipment have to meet the FCC's EMC requirements?

1.18 What are the three necessary elements to produce an interference problem?

1.19 When analyzing the characteristics of a noise source, what does the acronym FAT stand for?

1.20 a. Which of the following metals is the most susceptible to corrosion: cadmium, nickel (passive), magnesium, copper, or steel?

b. Which is the least susceptible to corrosion?

1.21 If a tin plate is bolted to a zinc casting, because of galvanic action, which metal will be corroded or eaten away?

REFERENCES

2004/108/EEC. Council Directive 2004/108/EEC Relating to Electromagnetic Compatibility and Repealing Directive 89/336/EEC, *Official Journal of the European Union*, No. L 390 December 31, 2004, pp. 24–37.

ANSI C63.4-1992. *Methods of Measurement of Radio-Noise Emissions from Low-Voltage Electrical and Electronic Equipment in the Range of 9 kHz to 40 GHz*, IEEE, July 17, 1992.

CAN/CSA-CEI/IEC CISPR 22:02. *Limits an Methods of Measurement of Radio Disturbance Characteristics of Information Technology Equipment*, Canadian Standards Association, 2002.

CISPR, Publication 22. "Limits and Methods of Measurement of Radio Interference Characteristics of Information Technology Equipment," 1997.

Code of Federal Regulations, Title 47, Telecommunications (47CFR). Parts 1, 2, 15, 18, and 68, U.S. Government Printing Office, Washington, DC.

EMCAB-1, Issue 2. Electromagnetic Compatibility Bulletin, "Immunity of Electrical/Electronic Equipment Intended to Operate in the Canadian Radio Environment (0.014–10,000 MHz)." Government of Canada, Department of Communications, August 1, 1982.

EN 61000-3-2. *Electromagnetic Compatibility (EMC)—Part 3-2: Limits—Limits for Harmonic Current Emissions (Equipment Input Current ≤ 16 A Per Phase)*, CENELEC, 2006.

EN 61000-3-3. *Electromagnetic Compatibility (EMC)—Part 3-3: Limits—Limitation of Voltage Changes, Voltage Fluctuations and Flicker in Public Low-Voltage Supply*

Systems, for Equipment with Rated Current ≤ 16 A Per Phase and Not Subject to Conditional Connection, CENELEC, 2006.

European Commission. *Guidelines on the Application of Council Directive 89/336/EEC on the Approximation of the Laws of the Member States Relating to Electromagnetic Compatibility*, European Commission, 1997.

FCC. "Commission Cautions Against Changes in Verified Computing Equipment." Public Notice No. 3281, April 7, 1982.

FCC. "United States and Canada Agree on Acceptance of Measurement Reports for Equipment Authorization," Public Notice No. 54795, July 12, 1995.

FDA. Guide to Inspections of Electromagnetic Compatibility Aspects of Medical Device Quality Systems, US Food and Drug Administration, Available at http://www.fda.gov/ora/inspect_ref/igs/elec_med_dev/emc1.html. Accessed September 2008.

GR-1089-CORE. *Electromagnetic Compatibility and Electrical Safety—Generic Criteria for Network Telecommunications Equipment*, Telcordia, November 2002.

Heirman, D. N. "Broadcast Electromagnetic Environment Near Telephone Equipment." IEEE National Telecommunications Conference, 1976.

Janes, D. E. et al. "Nonionizing Radiation Exposure in Urban Areas of the United States." *Proceedings of the 5th International Radiation Protection Association*. April 1977.

MDS-201-0004. *Electromagnetic Compatibility Standards for Medical Devices*, U.S. Department of Health Education and Welfare, Food and Drug Administration, October 1, 1979.

MIL-STD-461E. *Requirements For The Control of Electromagnetic Interference Characteristics of Subsystems and Equipment*, August 20, 1999.

MIL-STD-889B. *Dissimilar Metals*, Notice 3, May 1993.

RTCA/DO-160E. *Environmental Conditions and Test Procedures for Airborne Equipment*, Radio Technical Commission for Aeronautics (RTCA), December 7, 2004.

SAE J551. *Performance Levels and Methods of Measurement of Electromagnetic Compatibility of Vehicles and Devices (60 Hz to 18 GHz)*, Society of Automotive Engineers, June 1996.

SAE J1113. *Electromagnetic Compatibility Measurement Procedure for Vehicle Components (Except Aircraft) (60 Hz to 18 GHz)*, Society of Automotive Engineers, July 1995.

TIA-968-A. *Telecommunications Telephone Terminal Equipment Technical Requirements for Connection of Terminal Equipment to the Telephone Network*, Telecommunications Industry Association, October 1, 2002.

FURTHER READING

Cohen, T. J., and McCoy, L. G. "RFI–A New Look at an Old Problem." *QST*, March 1975.

Gruber, M. ed. *The ARRL RFI Book*, Newington, CT, American Radio Relay League, 2007.

Marshman, C. *The Guide to the EMC Directive 89/336/EEC*, IEEE Press, New York, 1992.

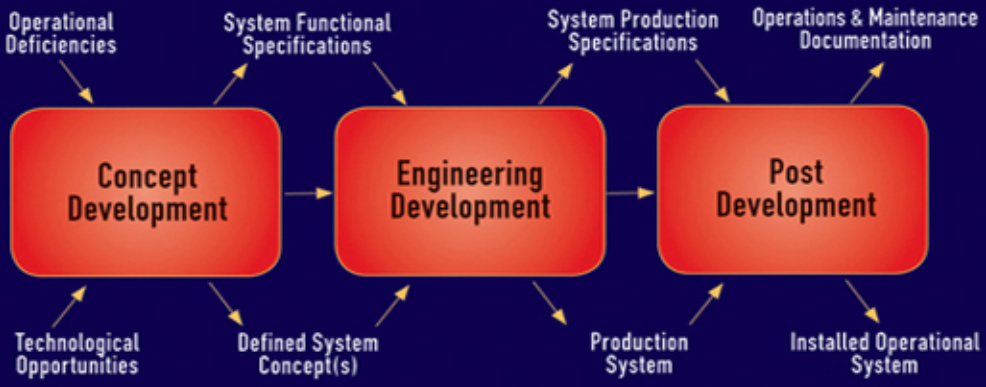
Wall, A. " Historical Perspective of the FCC Rules For Digital Devices and a Look to the Future," 2004 *IEEE International Symposium on Electromagnetic Compatibility*, August 9–13, 2004.

Wiley Series in Systems Engineering and Management
Andrew P. Sage, Series Editor

SECOND EDITION

Systems Engineering

Principles and Practice



ALEXANDER KOSSIAKOFF,
WILLIAM N. SWEET, SAMUEL J. SEYMOUR,
and STEVEN M. BIEMER

WILEY

Contents

1

SYSTEMS ENGINEERING AND THE WORLD OF MODERN SYSTEMS

1.1 WHAT IS SYSTEMS ENGINEERING?

There are many ways in which to define systems engineering. For the purposes of this book, we will use the following definition:

The function of systems engineering is to *guide* the *engineering* of *complex systems*.

The words in this definition are used in their conventional meanings, as described further below.

To guide is defined as "to lead, manage, or direct, usually based on the superior experience in pursuing a given course" and "to show the way." This characterization emphasizes the process of selecting the path for others to follow from among many possible courses—a primary function of systems engineering. A dictionary definition of engineering is "the application of scientific principles to practical ends; as the design, construction and operation of efficient and economical structures, equipment, and systems." In this definition, the terms "efficient" and "economical" are particular contributions of good systems engineering.

The word "system," as is the case with most common English words, has a very broad meaning. A frequently used definition of a system is "*a set* of *interrelated*

Systems Engineering Principles and Practice, Second Edition. Alexander Kossiakoff, William N. Sweet, Samuel J. Seymour, and Steven M. Biemer

components working *together* toward some *common* objective." This definition implies a multiplicity of interacting parts that collectively perform a significant function. The term *complex* restricts this definition to systems in which the elements are diverse and have intricate relationships with one another. Thus, a home appliance such as a washing machine would not be considered sufficiently diverse and complex to require systems engineering, even though it may have some modern automated attachments. On the other hand, the context of an *engineered* system excludes such complex systems as living organisms and ecosystems. The restriction of the term "system" to one that is complex and engineered makes it more clearly applicable to the function of systems engineering as it is commonly understood. Examples of systems requiring systems engineering for their development are listed in a subsequent section.

The above definitions of "systems engineering" and "system" are not represented as being unique or superior to those used in other textbooks, each of which defines them somewhat differently. In order to avoid any potential misunderstanding, the meaning of these terms *as used in this book* is defined at the very outset, before going on to the more important subjects of the responsibilities, problems, activities, and tools of systems engineering.

Systems Engineering and Traditional Engineering Disciplines

From the above definition, it can be seen that systems engineering differs from mechanical, electrical, and other engineering disciplines in several important ways:

1. Systems engineering is focused on the system as a whole; it emphasizes its total operation. It looks at the system from the outside, that is, at its interactions with other systems and the environment, as well as from the inside. It is concerned not only with the engineering design of the system but also with external factors, which can significantly constrain the design. These include the identification of customer needs, the system operational environment, interfacing systems, logistics support requirements, the capabilities of operating personnel, and such other factors as must be correctly reflected in system requirements documents and accommodated in the system design.
2. While the primary purpose of systems engineering is to guide, this does not mean that systems engineers do not themselves play a key role in system design. On the contrary, they are responsible for leading the formative (concept development) stage of a new system development, which culminates in the functional design of the system reflecting the needs of the user. Important design decisions at this stage cannot be based entirely on quantitative knowledge, as they are for the traditional engineering disciplines, but rather must often rely on qualitative judgments balancing a variety of incommensurate quantities and utilizing experience in a variety of disciplines, especially when dealing with new technology.
3. Systems engineering *bridges* the traditional engineering disciplines. The diversity of the elements in a complex system requires different engineering disci-

plines to be involved in their design and development. For the system to perform correctly, each system element must function properly in combination with one or more other system elements. Implementation of these interrelated functions is dependent on a complex set of physical and functional interactions between separately designed elements. Thus, the various elements cannot be engineered independently of one another and then simply assembled to produce a working system. Rather, systems engineers must guide and coordinate the design of each individual element as necessary to assure that the interactions and interfaces between system elements are compatible and mutually supporting. Such coordination is especially important when individual system elements are designed, tested, and supplied by different organizations.

Systems Engineering and Project Management

The engineering of a new complex system usually begins with an exploratory stage in which a new system concept is evolved to meet a recognized need or to exploit a technological opportunity. When the decision is made to engineer the new concept into an operational system, the resulting effort is inherently a major enterprise, which typically requires many people, with diverse skills, to devote years of effort to bring the system from concept to operational use.

The magnitude and complexity of the effort to engineer a new system requires a dedicated team to lead and coordinate its execution. Such an enterprise is called a "project" and is directed by a project manager aided by a staff. Systems engineering is an inherent part of project management—the part that is concerned with guiding the engineering effort itself—setting its objectives, guiding its execution, evaluating its results, and prescribing necessary corrective actions to keep it on course. The management of the planning and control aspects of the project fiscal, contractual, and customer relations is supported by systems engineering but is usually not considered to be part of the systems engineering function. This subject is described in more detail in Chapter 5.

Recognition of the importance of systems engineering by every participant in a system development project is essential for its effective implementation. To accomplish this, it is often useful to formally assign the leader of the systems engineering team to a recognized position of technical responsibility and authority within the project.

1.2 ORIGINS OF SYSTEMS ENGINEERING

No particular date can be associated with the origins of systems engineering. Systems engineering principles have been practiced at some level since the building of the pyramids and probably before. (The Bible records that Noah's Ark was built to a system specification.)

The recognition of systems engineering as a distinct activity is often associated with the effects of World War II, and especially the 1950s and 1960s when a number of textbooks were published that first identified systems engineering as a distinct

discipline and defined its place in the engineering of systems. More generally, the recognition of systems engineering as a unique activity evolved as a necessary corollary to the rapid growth of technology, and its application to major military and commercial operations during the second half of the twentieth century.

The global conflagration of World War II provided a tremendous spur to the advancement of technology in order to gain a military advantage for one side or the other. The development of high-performance aircraft, military radar, the proximity fuse, the German VI and V2 missiles, and especially the atomic bomb required revolutionary advances in the application of energy, materials, and information. These systems were complex, combining multiple technical disciplines, and their development posed engineering challenges significantly beyond those that had been presented by their more conventional predecessors. Moreover, the compressed development time schedules imposed by wartime imperatives necessitated a level of organization and efficiency that required new approaches in program planning, technical coordination, and engineering management. Systems engineering, as we know it today, developed to meet these challenges.

During the Cold War of the 1950s, 1960s, and 1970s, military requirements continued to drive the growth of technology in jet propulsion, control systems, and materials. However, another development, that of solid-state electronics, has had perhaps a more profound effect on technological growth. This, to a large extent, made possible the still evolving "information age," in which computing, networks, and communications are extending the power and reach of systems far beyond their previous limits. Particularly significant in this connection is the development of the digital computer and the associated software technology driving it, which increasingly is leading to the replacement of human control of systems by automation. Computer control is qualitatively increasing the complexity of systems and is a particularly important concern of systems engineering.

The relation of modern systems engineering to its origins can be best understood in terms of three basic factors:

1. *Advancing Technology,* which provide opportunities for increasing system capabilities, but introduces development risks that require systems engineering management; nowhere is this more evident than in the world of automation. Technology advances in human–system interfaces, robotics, and software make this particular area one of the fastest growing technologies affecting system design.
2. *Competition,* whose various forms require seeking superior (and more advanced) system solutions through the use of system-level trade-offs among alternative approaches.
3. *Specialization,* which requires the partitioning of the system into building blocks corresponding to specific product types that can be designed and built by specialists, and strict management of their interfaces and interactions.

These factors are discussed in the following paragraphs.

Advancing Technology: Risks

The explosive growth of technology in the latter half of the twentieth century and into this century has been the single largest factor in the emergence of systems engineering as an essential ingredient in the engineering of complex systems. Advancing technology has not only greatly extended the capabilities of earlier systems, such as aircraft, telecommunications, and power plants, but has also created entirely new systems such as those based on jet propulsion, satellite communications and navigation, and a host of computer-based systems for manufacturing, finance, transportation, entertainment, health care, and other products and services. Advances in technology have not only affected the nature of products but have also fundamentally changed the way they are engineered, produced, and operated. These are particularly important in early phases of system development, as described in Conceptual Exploration, in Chapter 7.

Modern technology has had a profound effect on the very approach to engineering. Traditionally, engineering applies known principles to practical ends. Innovation, however, produces new materials, devices, and processes, whose characteristics are not yet fully measured or understood. The application of these to the engineering of new systems thus increases the risk of encountering unexpected properties and effects that might impact system performance and might require costly changes and program delays.

However, failure to apply the latest technology to system development also carries risks. These are the risks of producing an inferior system, one that could become prematurely obsolete. If a competitor succeeds in overcoming such problems as may be encountered in using advanced technology, the competing approach is likely to be superior. The successful entrepreneurial organization will thus assume carefully selected technological risks and surmount them by skillful design, systems engineering, and program management.

The systems engineering approach to the early application of new technology is embodied in the practice of "risk management." Risk management is a process of dealing with calculated risks through a process of analysis, development, test, and engineering oversight. It is described more fully in Chapters 5 and 9.

Dealing with risks is one of the essential tasks of systems engineering, requiring a broad knowledge of the total system and its critical elements. In particular, systems engineering is central to the decision of how to achieve the best balance of risks, that is, which system elements should best take advantage of new technology and which should be based on proven components, and how the risks incurred should be reduced by development and testing.

The development of the digital computer and software technology noted earlier deserves special mention. This development has led to an enormous increase in the automation of a wide array of control functions for use in factories, offices, hospitals, and throughout society. Automation, most of it being concerned with information processing hardware and software, and its sister technology, autonomy, which adds in capability of command and control, is the fastest growing and most powerful single influence on the engineering of modern systems.

The increase in automation has had an enormous impact on people who operate systems, decreasing their number but often requiring higher skills and therefore special training. Human–machine interfaces and other people–system interactions are particular concerns of systems engineering.

Software continues to be a growing engineering medium whose power and versatility has resulted in its use in preference to hardware for the implementation of a growing fraction of system functions. Thus, the performance of modern systems increasingly depends on the proper design and maintenance of software components. As a result, more and more of the systems engineering effort has had to be directed to the control of software design and its application.

Competition: Trade-offs

Competitive pressures on the system development process occur at several different levels. In the case of defense systems, a primary drive comes from the increasing military capabilities of potential adversaries, which correspondingly decrease the effectiveness of systems designed to defeat them. Such pressures eventually force a development program to redress the military balance with a new and more capable system or a major upgrade of an existing one.

Another source of competition comes with the use of competitive contracting for the development of new system capabilities. Throughout the competitive period, which may last through the initial engineering of a new system, each contractor seeks to devise the most cost-effective program to provide a superior product.

In developing a commercial product, there are nearly always other companies that compete in the same market. In this case, the objective is to develop a new market or to obtain an increased market share by producing a superior product ahead of the competition, with an edge that will maintain a lead for a number of years. The above approaches nearly always apply the most recent technology in an effort to gain a competitive advantage.

Securing the large sums of money needed to fund the development of a new complex system also involves competition on quite a different level. In particular, both government agencies and industrial companies have many more calls on their resources than they can accommodate and hence must carefully weigh the relative payoff of proposed programs. This is a primary reason for requiring a phased approach in new system development efforts, through the requirement for justification and formal approval to proceed with the increasingly expensive later phases. The results of each phase of a major development must convince decision makers that the end objectives are highly likely to be attained within the projected cost and schedule.

On a still different basis, the competition among the essential characteristics of the system is always a major consideration in its development. For example, there is always competition between performance, cost, and schedule, and it is impossible to optimize all three at once. Many programs have failed by striving to achieve levels of performance that proved unaffordable. Similarly, the various performance parameters of a vehicle, such as speed and range, are not independent of one another; the efficiency of most vehicles, and hence their operating range, decreases at higher speeds.

Thus, it is necessary to examine alternatives in which these characteristics are allowed to vary and to select the combination that best balances their values for the benefit of the user.

All of the forms of competition exert pressure on the system development process to produce the best performing, most affordable system, in the least possible time. The process of selecting the most desirable approach requires the examination of numerous potential alternatives and the exercise of a breadth of technical knowledge and judgment that only experienced systems engineers possess. This is often referred to as "trade-off analysis" and forms one of the basic practices of systems engineering.

Specialization: Interfaces

A complex system that performs a number of different functions must of necessity be configured in such a way that each major function is embodied in a separate component capable of being specified, developed, built, and tested as an individual entity. Such a subdivision takes advantage of the expertise of organizations specializing in particular types of products, and hence is capable of engineering and producing components of the highest quality at the lowest cost. Chapter 3 describes the kind of functional and physical building blocks that make up most modern systems.

The immensity and diversity of engineering knowledge, which is still growing, has made it necessary to divide the education and practice of engineering into a number of specialties, such as mechanical, electrical, aeronautical, and so on. To acquire the necessary depth of knowledge in any one of these fields, further specialization is needed, into such subfields as robotics, digital design, and fluid dynamics. Thus, engineering specialization is a predominant condition in the field of engineering and manufacturing and must be recognized as a basic condition in the system development process.

Each engineering specialty has developed a set of specialized tools and facilities to aid in the design and manufacture of its associated products. Large and small companies have organized around one or several engineering groups to develop and manufacture devices to meet the needs of the commercial market or of the system-oriented industry. The development of interchangeable parts and automated assembly has been one of the triumphs of the U.S. industry.

The convenience of subdividing complex systems into individual building blocks has a price: that of integrating these disparate parts into an efficient, smoothly operating system. Integration means that each building block fits perfectly with its neighbors and with the external environment with which it comes into contact. The "fit" must be not only physical but also functional; that is, its design will both affect the design characteristics and behavior of other elements, and will be affected by them, to produce the exact response that the overall system is required to make to inputs from its environment. The physical fit is accomplished at intercomponent boundaries called *interfaces*. The functional relationships are called *interactions*.

The task of analyzing, specifying, and validating the component interfaces with each other and with the external environment is beyond the expertise of the individual design specialists and is the province of the systems engineer. Chapter 3 discusses further the importance and nature of this responsibility.

A direct consequence of the subdivision of systems into their building blocks is the concept of modularity. Modularity is a measure of the degree of mutual independence of the individual system components. An essential goal of systems engineering is to achieve a high degree of modularity to make interfaces and interactions as simple as possible for efficient manufacture, system integration, test, operational maintenance, reliability, and ease of in-service upgrading. The process of subdividing a system into modular building blocks is called "functional allocation" and is another basic tool of systems engineering.

1.3 EXAMPLES OF SYSTEMS REQUIRING SYSTEMS ENGINEERING

As noted at the beginning of this chapter, the generic definition of a system as a *set* of *interrelated components* working *together* as an integrated whole to achieve some common objective would fit most familiar home appliances. A washing machine consists of a main clothes tub, an electric motor, an agitator, a pump, a timer, an inner spinning tub, and various valves, sensors, and controls. It performs a sequence of timed operations and auxiliary functions based on a schedule and operation mode set by the operator. A refrigerator, microwave oven, dishwasher, vacuum cleaner, and radio all perform a number of useful operations in a systematic manner. However, these appliances involve only one or two engineering disciplines, and their design is based on well-established technology. Thus, they fail the criterion of being *complex*, and we would not consider the development of a new washer or refrigerator to involve much systems engineering as we understand the term, although it would certainly require a high order of reliability and cost engineering. Of course, home appliances increasingly include clever automatic devices that use newly available microchips, but these are usually self-contained add-ons and are not necessary to the main function of the appliance.

Since the development of new modern systems is strongly driven by technological change, we shall add one more characteristic to a system requiring systems engineering, namely, that some of its key elements use advanced technology. The characteristics of a system whose development, test, and application require the practice of systems engineering are that the system

- is an engineered product and hence satisfies a specified need,
- consists of diverse components that have intricate relationships with one another and hence is multidisciplinary and relatively complex, and
- uses advanced technology in ways that are central to the performance of its primary functions and hence involves development risk and often a relatively high cost.

Henceforth, references in this text to an *engineered* or *complex* system (or in the proper context, just *system*) will mean the type that has the three attributes noted above, that is, is an engineered product, contains diverse components, and uses advanced technology. These attributes are, of course, in addition to the generic definition stated

earlier and serve to identify the systems of concern to the systems engineer as those that require system design, development, integration, test, and evaluation. In Chapter 2, we explore the full spectrum of systems complexity and why the systems engineering landscape presents a challenge for systems engineers.

Examples of Complex Engineered Systems

To illustrate the types of systems that fit within the above definition, Tables 1.1 and 1.2 list 10 modern systems and their principal inputs, processes, and outputs.

TABLE 1.1. Examples of Engineered Complex Systems: Signal and Data Systems

System	Inputs	Process	Outputs
Weather satellite	Images	• Data storage • Transmission	Encoded images
Terminal air traffic control system	Aircraft beacon responses	• Identification • Tracking	• Identity • Air tracks • Communications
Track location system	Cargo routing requests	• Map tracing • Communication	• Routing information • Delivered cargo
Airline reservation system	Travel requests	Data management	• Reservations • Tickets
Clinical information system	• Patient ID • Test records • Diagnosis	Information management	• Patient status • History • Treatment

TABLE 1.2. Examples of Engineered Complex Systems: Material and Energy Systems

System	Inputs	Process	Outputs
Passenger aircraft	• Passengers • Fuel	• Combustion • Thrust • Lift	Transported passengers
Modern harvester combine	• Grain field • Fuel	• Cutting • Threshing	Harvested grain
Oil refinery	• Crude oil • Catalysts • Energy	• Cracking • Separation • Blending	• Gasoline • Oil products • Chemicals
Auto assembly plant	• Auto parts • Energy	• Manipulation • Joining • Finishing	Assembled auto
Electric power plant	• Fuel • Air	• Power generation • Regulation	• Electric AC power • Waste products

It has been noted that a system consists of a multiplicity of elements, some of which may well themselves be complex and deserve to be considered a system in their own right. For example, a telephone-switching substation can well be considered as a system, with the telephone network considered as a "system of systems." Such issues will be discussed more fully in Chapters 2 and 4, to the extent necessary for the understanding of systems engineering.

Example: A Modern Automobile. A more simple and familiar system, which still meets the criteria for an engineered system, is a fully equipped passenger automobile. It can be considered as a lower limit to more complex vehicular systems. It is made up of a large number of diverse components requiring the combination of several different disciplines. To operate properly, the components must work together accurately and efficiently. Whereas the operating principles of automobiles are well established, modern autos must be designed to operate efficiently while at the same time maintaining very close control of engine emissions, which requires sophisticated sensors and computer-controlled mechanisms for injecting fuel and air. Antilock brakes are another example of a finely tuned automatic automobile subsystem. Advanced materials and computer technology are used to an increasing degree in passenger protection, cruise control, automated navigation and autonomous driving and parking. The stringent requirements on cost, reliability, performance, comfort, safety, and a dozen other parameters present a number of substantive systems engineering problems. Accordingly, an automobile meets the definition established earlier for a system requiring the application of systems engineering, and hence can serve as a useful example.

An automobile is also an example of a large class of systems that require active interaction (control) by a human operator. To some degree, all systems require such interaction, but in this case, continuous control is required. In a very real sense, the operator (driver) functions as an integral part of the overall automobile system, serving as the steering feedback element that detects and corrects deviations of the car's path on the road. The design must therefore address as a critical constraint the inherent sensing and reaction capabilities of the operator, in addition to a range of associated human–machine interfaces such as the design and placement of controls and displays, seat position, and so on. Also, while the passengers may not function as integral elements of the auto steering system, their associated interfaces (e.g., weight, seating and viewing comfort, and safety) must be carefully addressed as part of the design process. Nevertheless, since automobiles are developed and delivered without the human element, for purposes of systems engineering, they may be addressed as systems in their own right.

1.4 SYSTEMS ENGINEERING AS A PROFESSION

With the increasing prevalence of complex systems in modern society, and the essential role of systems engineering in the development of systems, systems engineering as a profession has become widely recognized. Its primary recognition has come in companies specializing in the development of large systems. A number of these have estab-

lished departments of systems engineering and have classified those engaging in the process as systems engineers. In addition, global challenges in health care, communications, environment, and many other complex areas require engineering systems methods to develop viable solutions.

To date, the slowness of recognition of systems engineering as a career is the fact that it does not correspond to the traditional academic engineering disciplines. Engineering disciplines are built on quantitative relationships, obeying established physical laws, and measured properties of materials, energy, or information. Systems engineering, on the other hand, deals mainly with problems for which there is incomplete knowledge, whose variables do not obey known equations, and where a balance must be made among conflicting objectives involving incommensurate attributes. The absence of a quantitative knowledge base previously inhibited the establishment of systems engineering as a unique discipline.

Despite those obstacles, the recognized need for systems engineering in industry and government has spurred the establishment of a number of academic programs offering master's degrees and doctoral degrees in systems engineering. An increasing number of universities are offering undergraduate degrees in systems engineering as well.

The recognition of systems engineering as a profession has led to the formation of a professional society, the International Council on Systems Engineering (INCOSE), one of whose primary objectives is the promotion of systems engineering, and the recognition of systems engineering as a professional career.

Career Choices

Systems engineers are highly sought after because their skills complement those in other fields and often serve as the "glue" to bring new ideas to fruition. However, career choices and the related educational needs for those choices is complex, especially when the role and responsibilities of a systems engineer is poorly understood.

Four potential career directions are shown in Figure 1.1: financial, management, technical, and systems engineering. There are varying degrees of overlap between them despite the symmetry shown in the figure. The systems engineer focuses on the whole system product, leading and working with many diverse technical team members, following the systems engineering development cycle, conducting studies of alternatives, and managing the system interfaces. The systems engineer generally matures in the field after a technical undergraduate degree with work experience and a master of science degree in systems engineering, with an increasing responsibility of successively larger projects, eventually serving as the chief or lead systems engineer for a major systems, or systems-of-systems development. Note the overlap and need to understand the content and roles of the technical specialists and to coordinate with the program manager (PM).

The project manager or PM, often with a technical or business background, is responsible for interfacing with the customer and for defining the work, developing the plans, monitoring and controlling the project progress, and delivering the finished output to the customer. The PM often learns from on the job training (OJT) with

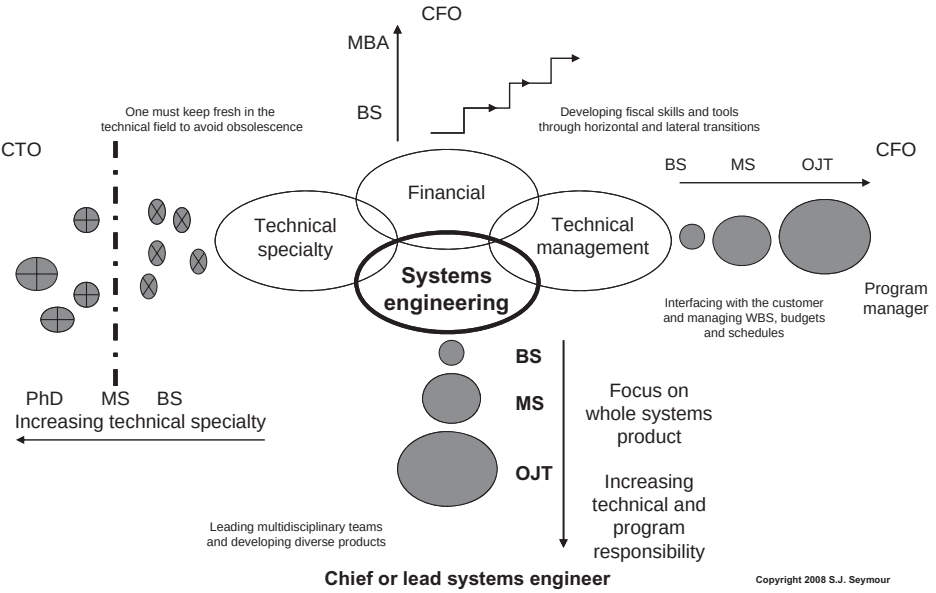


Figure 1.1. Career opportunities and growth.

projects of increasing size and importance, enhancing the toolset available with a master of science degree in technical/program management. While not exclusively true, the chief executive officer (CEO) frequently originates from the ranks of the organization's PMs.

The financial or business career path that ultimately could lead to a chief financial officer (CFO) position usually includes business undergraduate and master of business administration (MBA) degrees. Individuals progress through their careers with various horizontal and vertical moves, often with specialization in the field. There is an overlap in skill and knowledge with the PM in areas of contract and finance management.

Many early careers start with a technical undergraduate degree in engineering, science or information technology. The technical specialist makes contributions as part of a team in the area of their primary knowledge, honing skills and experience to develop and test individual components or algorithms that are part of a larger system. Contributions are made project to project over time, and recognition is gained from innovative, timely, and quality workmanship. Technical specialists need to continue to learn about their field and to stay current in order to be employable compared to the next generation of college graduates. Often advanced degrees (MS and PhDs) are acquired to enhance knowledge, capability, and recognition, and job responsibilities can lead to positions such as lead engineer, lead scientist, or chief technology officer (CTO) in an organization. The broader minded or experienced specialist often considers a career in systems engineering.

Orientation of Technical Professionals

The special relationship of systems engineers with respect to technical disciplines can be better understood when it is realized that technical people not only engage in widely different professional specialties, but their intellectual objectives, interests, and attitudes, which represent their technical orientations, can also be widely divergent. The typical scientist is dedicated to understanding the nature and behavior of the physical world. The scientist asks the questions "Why?" and "How?" The mathematician is usually primarily concerned with deriving the logical consequences of a set of assumptions, which may be quite unrelated to the real world. The mathematician develops the proposition "If A, then B." Usually, the engineer is mainly concerned with creating a useful product. The engineer exclaims "Voila!"

These orientations are quite different from one another, which accounts for why technical specialists are focused on their own aspects of science and technology. However, in most professionals, those orientations are not absolute; in many cases, the scientist may need some engineering to construct an apparatus, and the engineer may need some mathematics to solve a control problem. So, in the general case, the orientation of a technical professional might be modeled by a sum of three orthogonal vectors, each representing the extent of the individual's orientation being in science, mathematics, or engineering.

To represent the above model, it is convenient to use a diagram designed to show the composition of a mixture of three components. Figure 1.2a is such a diagram in which the components are science, mathematics, and engineering. A point at each vertex represents a mixture with 100% of the corresponding component. The composition of the mixture marked by the small triangle in the figure is obtained by finding the percentage of each component by projecting a line parallel to the baseline opposite each vertex to the scale radiating from the vertex. This process gives intercepts of 70% science, 20% mathematics, and 10% engineering for the orientation marked by the triangle.

Because the curricula of technical disciplines tend to be concentrated in specialized subjects, most students graduate with limited general knowledge. In Figure 1.2b, the circles representing the orientation of individual graduates are seen to be concentrated in the corners, reflecting their high degree of specialization.

The tendency of professional people to polarize into diverse specialties and interests tends to be accentuated after graduation, as they seek to become recognized in their respective fields. Most technical people resist becoming generalists for fear they will lose or fail to achieve positions of professional leadership and the accompanying recognition. This specialization of professionals inhibits technical communication between them; the language barrier is bad enough, but the differences in basic objectives and methods of thought are even more serious. The solution of complex interdisciplinary problems has had to depend on the relatively rare individuals who, for one reason or another, after establishing themselves in their principal profession, have become interested and involved in solving system problems and have learned to work jointly with specialists in various other fields.

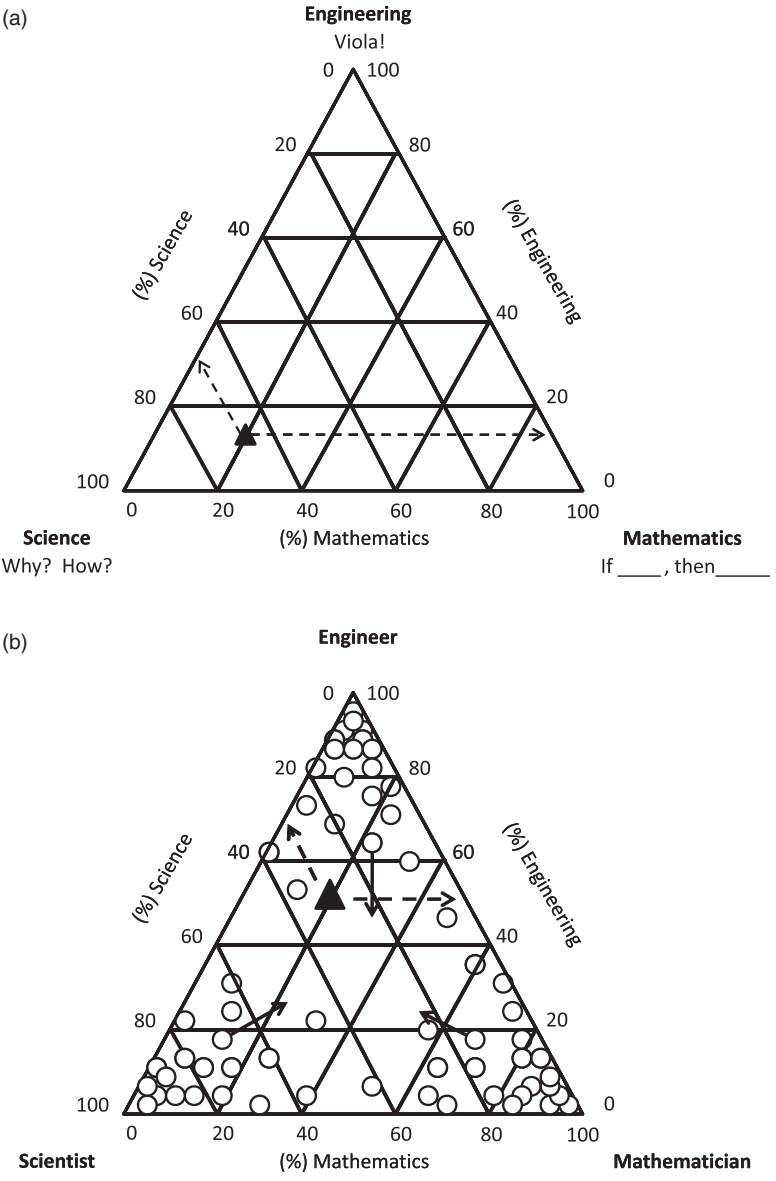


Figure 1.2. (a) Technical orientation phase diagram. (b) Technical orientation population density distribution.

The occasional evolution of technical specialists into systems engineers is symbolized in Figure 1.2b by the arrows directed from the vertices toward the center. The small black triangle corresponds to such an evolved individual whose orientation is 30% science, 50% engineering, and 20% mathematics, a balance that would be effective in the type of problem solving with which a systems engineer is typically involved. It is the few individuals who evolve into systems engineers or system architects who become the technical leaders of system development programs.

The Challenge of Systems Engineering

An inhibiting factor in becoming a professional systems engineer is that it represents a deviation from a chosen established discipline to a more diverse, complicated professional practice. It requires the investment of time and effort to gain experience and an extensive broadening of the engineering base, as well as learning communication and management skills, a much different orientation from the individual's original professional choice.

For the above reasons, an engineer considering a career in systems engineering may come to the conclusion that the road is difficult. It is clear that a great deal must be learned; that the educational experience in a traditional engineering discipline is necessary; and that there are few tools and few quantitative relationships to help make decisions. Instead, the issues are ambiguous and abstract, defying definitive solutions. There may appear to be little opportunity for individual accomplishment and even less for individual recognition. For a systems engineer, success is measured by the accomplishment of the development team, not necessarily the system team leader.

What Then Is the Attraction of Systems Engineering?

The answer may lie in the challenges of systems engineering rather than its direct rewards. Systems engineers deal with the most important issues in the system development process. They design the overall system architecture and the technical approach and lead others in designing the components. They prioritize the system requirements in conjunction with the customer to ensure that the different system attributes are appropriately weighted when balancing the various technical efforts. They decide which risks are worth undertaking and which are not, and how the former should be hedged to ensure program success.

It is the systems engineers who map out the course of the development program that prescribes the type and timing of tests and simulations to be performed along the way. They are the ultimate authorities on how the system performance and system affordability goals may be achieved at the same time.

When unanticipated problems arise in the development program, as they always do, it is the systems engineers who decide how they may be solved. They determine whether an entirely new approach to the problem is necessary, whether more intense effort will accomplish the purpose, whether an entirely different part of the system can

be modified to compensate for the problem, or whether the requirement at issue can best be scaled back to relieve the problem.

Systems engineers derive their ability to guide the system development not from their position in the organization but from their superior knowledge of the system as a whole, its operational objectives, how all its parts work together, and all the technical factors that go into its development, as well as from their proven experience in steering complex programs through a maze of difficulties to a successful conclusion.

Attributes and Motivations of Systems Engineers

In order to identify candidates for systems engineering careers, it is useful to examine the characteristics that may be useful to distinguish people with a talent for systems engineering from those who are not likely to be interested or successful in that discipline. Those likely to become talented systems engineers would be expected to have done well in mathematics and science in college.

A systems engineer will be required to work in a multidisciplinary environment and to grasp the essentials of related disciplines. It is here that an aptitude for science and engineering helps a great deal because it makes it much easier and less threatening for individuals to learn the essentials of new disciplines. It is not so much that they require in depth knowledge of higher mathematics, but rather, those who have a limited mathematical background tend to lack confidence in their ability to grasp subjects that inherently contain mathematical concepts.

A systems engineer should have a creative bent and must like to solve practical problems. An interest in the job should be greater than an interest in career advancement. Systems engineering is more of a challenge than a quick way to the top.

The following characteristics are commonly found in successful systems engineers. They

1. enjoy learning new things and solving problems,
2. like challenges,
3. are skeptical of unproven assertions,
4. are open-minded to new ideas,
5. have a solid background in science and engineering,
6. have demonstrated technical achievement in a specialty area,
7. are knowledgeable in several engineering areas,
8. pick up new ideas and information quickly, and
9. have good interpersonal and communication skills.

1.5 SYSTEMS ENGINEER CAREER DEVELOPMENT MODEL

When one has the characteristics noted above and is attracted to become a systems engineer, there are four more elements that need to be present in the work environment. As shown in Figure 1.3a, one should seek assignments to problems and tasks that are

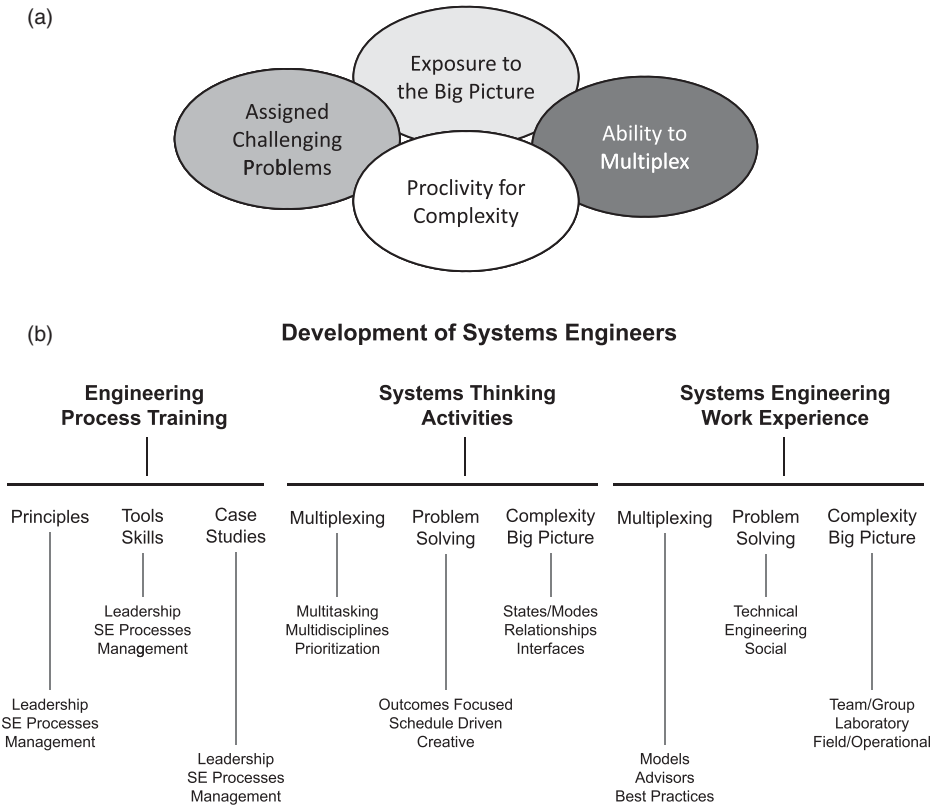


Figure 1.3. (a) Systems engineering (SE) career elements derived from quality work experiences. (b) Components of employer development of systems engineers.

very challenging and are likely to expand technical domain knowledge and creative juices. Whatever the work assignment, understanding the context of the work and understanding the big picture is also essential. Systems engineers are expected to manage many activities at the same time, being able to have broad perspectives but able to delve deeply into to many subjects at once. This ability to multiplex is one that takes time to develop. Finally, the systems engineer should not be intimidated by complex problems since this is the expected work environment. It is clear these elements are not part of an educational program and must be gained through extended professional work experience. This becomes the foundation for the systems engineering career growth model.

Employers seeking to develop systems engineers to competitively address more challenging problems should provide key staff with relevant systems engineering work experience, activities that require mature systems thinking, and opportunities for systems engineering education and training. In Figure 1.3b, it can be seen that the experience can be achieved not only with challenging problems but also with

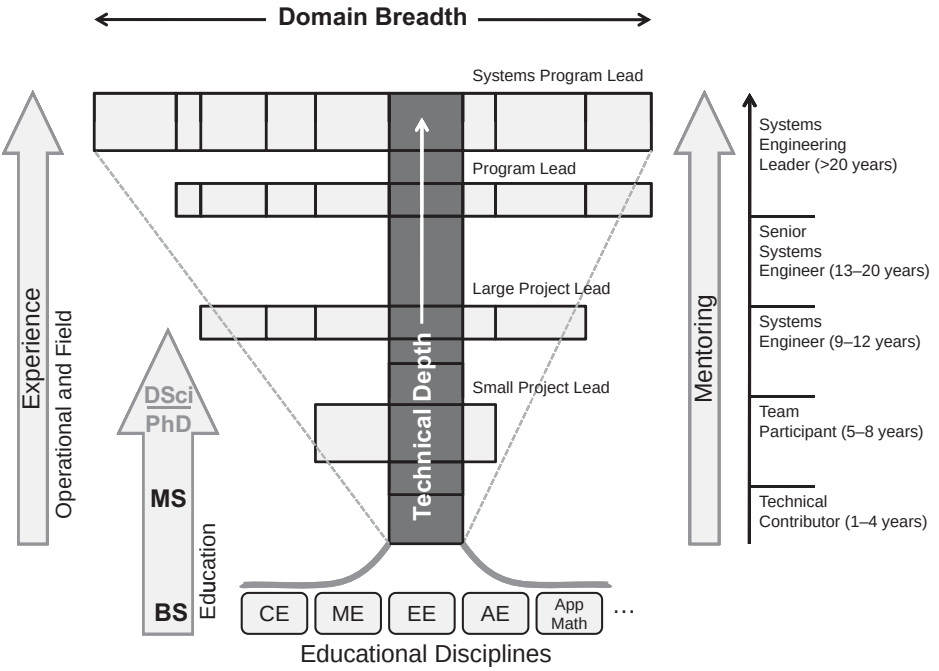


Figure 1.4. "T" model for systems engineer career development. CE, chemical engineering; ME, mechanical engineering; EE, electrical engineering; AE, aeronautical engineering; App Math, applied mathematics.

experienced mentors and real, practical exercises. While using systems thinking to explore complex problem domains, staff should be encouraged to think creatively and out of the box. Often, technically trained people rigidly follow the same processes and tired ineffective solutions. Using lessons learned from past programs and case studies creates opportunities for improvements. Formal training and use of systems engineering tools further enhance employee preparation for tackling complex issues.

Interests, attributes, and training, along with an appropriate environment, provide the opportunity for individuals to mature into successful systems engineers. The combination of these factors is captured in the "T" model for systems engineer career development illustrated in Figure 1.4. In the vertical, from bottom to top is the time progression in a professional's career path. After completion of a technical undergraduate degree, shown along the bottom of the chart, an individual generally enters professional life as a technical contributor to a larger effort. The effort is part of a project or program that falls in a particular domain such as aerodynamics, biomedicine, combat systems, information systems, or space exploration. Within a domain, there are several technical competencies that are fundamental for systems to operate or to be developed.

The T is formed by snapshots during a professional's career that illustrates in the horizontal part of the T the technical competencies at the time that were learned and used to meet the responsibilities assigned at that point in their career. After an initial

experience in one or two technical domains as technical contributor, one progresses to increasing responsibilities in a team setting and eventually to leading small technical groups. After eight or more years, the professional has acquired both sufficient technical depth and technical domain depth to be considered a systems engineer. Additional assignments lead to project and program systems engineering leadership and eventually to being the senior systems engineer for a major development program that exercises the full range of the technical competencies for the domain.

In parallel with broadening and deepening technical experience and competencies, the successful career path is augmented by assignments that involve operational field experiences, advanced education and training, and a strong mentoring program. In order to obtain a good understanding of the environment where the system under development will operate and to obtain firsthand knowledge of the system requirements, it is essential for the early systems engineer professional to visit the "field site" and operational location. This approach is important to continue throughout one's career. A wide variety of systems engineering educational opportunities are available in both classroom and online formats. As in most engineering disciplines where the student is not planning on an academic career, the master of science is the terminal degree. Courses are usually a combination of systems engineering and domain or concentration centric focused with a thesis or capstone project for the students to demonstrate their knowledge and skills on a practical systems problem. Large commercial companies also provide training in systems engineering and systems architecting with examples and tools that are specific to their organization and products. Finally, the pairing of a young professional with an experienced systems engineer will enhance the learning process.

1.6 THE POWER OF SYSTEMS ENGINEERING

If power is measured by authority over people or money, then systems engineers would appear to have little power as members of the system development team. However, if power is measured by the influence over the design of the system and its major characteristics, and over the success or failure of the system development, then systems engineers can be more powerful than project managers. The sources of this power come from their knowledge, skills, and attitude. Each of these is discussed in the following paragraphs.

The Power of Multidisciplinary Knowledge

A major system development project is a veritable "Tower of Babel." There are literally dozens of specialists in different disciplines whose collective efforts are necessary to develop and produce a successful new system. Each group of specialists has its own language, making up for the imprecision of the English language with a rich set of acronyms, which convey a very specific meaning but are unintelligible to those outside the specialty. The languages, in turn, are backed up by knowledge bases, which the specialists use to ply their trade. These knowledge bases contain descriptions of the different materials peculiar to each discipline, as well as bodies of relationships, many

of them expressed in mathematical terms, that enable the specialists to compute various characteristics of their components on the basis of design assumptions. These knowledge bases are also foreign to those outside the discipline.

Such a collection of multi-tongued participants could never succeed in collectively developing a new system by themselves, just as the citizens of Babylon could never build their tower. It is the systems engineers who provide the linkages that enable these disparate groups to function as a team. The systems engineers accomplish this feat through the power of multidisciplinary knowledge. This means that they are sufficiently literate in the different disciplines involved in their system that they can understand the languages of the specialists, appreciate their problems, and are able to interpret the necessary communications for their collective endeavor. Thus, they are in the same position as a linguist in the midst of a multinational conference, with people speaking in their native tongues. Through the ability to understand different languages comes the capability to obtain cooperative effort from people who would otherwise never be able to achieve a common objective. This capability enables systems engineers to operate as leaders and troubleshooters, solving problems that no one else is capable of solving. It truly amounts to a power that gives systems engineers a central and decisive role to play in the development of a system.

It is important to note that the depth of interdisciplinary knowledge, which is required to interact effectively with specialists in a given field, is a very small fraction of the depth necessary to work effectively in that field. The number of new acronyms that one has to learn in a given technical area is nearer to a dozen of the more frequently used ones than to a hundred. It also turns out that once one gets past the differences in semantics, there are many common principles in different disciplines and many similar relationships. For instance, the equation used in communications, connecting signal, noise, antenna gain, receiver sensitivity, and other factors, is directly analogous to a similar relationship in acoustics.

These facts mean that a systems engineer does not need to spend a lifetime becoming expert in associated disciplines, but rather can accumulate a working knowledge of related fields through selected readings, and more particularly, discussion with colleagues knowledgeable in each field. The important thing is to know which principles, relationships, acronyms, and the like are important at the system level and which are details. The power of multidisciplinary knowledge is so great that, to a systems engineer, the effort required to accumulate it is well worth the learning time.

The Power of Approximate Calculation

The practice of systems engineering requires another talent besides multidisciplinary knowledge. The ability to carry out “back of the envelope” calculations to obtain a “sanity check” on the result of a complex calculation or test is of inestimable value to the systems engineer. In a few cases, this can be done intuitively on the basis of past experience, but more frequently, it is necessary to make a rough estimate to ensure that a gross omission or error has not been committed. Most successful systems engineers have the ability, using first principles, to apply basic relationships, such as the communications equation or other simple calculation, to derive an order of magnitude result

to serve as a check. This is particularly important if the results of the calculation or experiment turn out very differently from what had been originally expected.

When the sanity check does not confirm the results of a simulation or experiment, it is appropriate to go back to make a careful examination of the assumptions and conditions on which the latter were based. As a matter of general experience, more often than not, such examinations reveal an error in the conditions or assumptions under which the simulation or experiment was conducted.

The Power of Skeptical Positive Thinking

The above seemingly contradictory title is meant to capture an important characteristic of successful systems engineering. The skeptical part is important to temper the traditional optimism of the design specialist regarding the probability of success of a chosen design approach. It is the driving force for the insistence of validation of the approach selected at the earliest possible opportunity.

The other dimension of skepticism, which is directly related to the characteristic of positive thinking, refers to the reaction in the face of failure or apparent failure of a selected technique or design approach. Many design specialists who encounter an unexpected failure are plunged into despair. The systems engineer, on the other hand, cannot afford the luxury of hand wringing but must have, first of all, a healthy skepticism of the conditions under which the unexpected failure occurred. Often, it is found that these conditions did not properly test the system. When the test conditions are shown to be valid, the systems engineer must set about finding ways to circumvent the cause of failure. The conventional answer that the failure must require a new start along a different path, which in turn will lead to major delays and increases in program cost, is simply not acceptable unless heroic efforts to find an alternative solution do not succeed. This is where the power of multidisciplinary knowledge permits the systems engineer to look for alternative solutions in other parts of the system, which may take the stress off the particular component whose design proved to be faulty.

The characteristic of positive thinking is absolutely necessary in both the systems engineer and the project manager so that they are able to generate and sustain the confidence of the customer and of company management, as well as the members of the design team. Without the “can-do” attitude, the esprit de corps and productivity of the project organization is bound to suffer.

1.7 SUMMARY

What Is Systems Engineering?

The function of systems engineering is to guide the engineering of complex systems. And a system is defined as a set of interrelated components working together toward a common objective. Furthermore, a complex engineered system (as defined in this book) is (1) composed of a multiplicity of intricately interrelated diverse elements and (2) requires systems engineering to lead its development.

Systems engineering differs from traditional disciplines in that (1) it is focused on the system as a whole; (2) it is concerned with customer needs and operational environment; (3) it leads system conceptual design; and (4) it bridges traditional engineering disciplines and gaps between specialties. Moreover, systems engineering is an integral part of project management in that it plans and guides the engineering effort.

Origins of Systems Engineering

Modern systems engineering originated because advancing technology brought risks and complexity with the growth of automation; competition required expert risk taking; and specialization required bridging disciplines and interfaces.

Examples of Systems Requiring Systems Engineering

Examples of engineered complex systems include

- weather satellites,
- terminal air traffic control,
- truck location systems,
- airline navigation systems,
- clinical information systems,
- passenger aircraft,
- modern harvester combines,
- oil refineries,
- auto assembly plants, and
- electric power plants.

Systems Engineering as a Profession

Systems engineering is now recognized as a profession and has an increasing role in government and industry. In fact, numerous graduate (and some undergraduate) degree programs are now available across the country. And a formal, recognized organization exists for systems engineering professionals: the INCOSE.

Technical professionals have specific technical orientations—technical graduates tend to be highly specialized. Only a few become interested in interdisciplinary problems—it is these individuals who often become systems engineers.

Systems Engineer Career Development Model

The systems engineering profession is difficult but rewarding. A career in systems engineering typically features technical satisfaction—finding the solution of abstract and ambiguous problems—and recognition in the form of a pivotal program role. Consequently, a successful systems engineer has the following traits and attributes:

- a good problem solver and should welcome challenges;
- well grounded technically, with broad interests;
- analytical and systematic, but also creative; and
- a superior communicator, with leadership skills.

The "T" model represents the proper convergence of experience, education, mentoring, and technical depth necessary to become a successful and influential systems engineer.

The Power of Systems Engineering

Overall, systems engineering is a powerful discipline, requiring a multidisciplinary knowledge, integrating diverse system elements. Systems engineers need to possess the ability to perform approximate calculations of complex phenomena, thereby providing sanity checks. And finally, they must have skeptical positive thinking as a prerequisite to prudent risk taking.

PROBLEMS

1.1 Write a paragraph explaining what is meant by the statement "Systems engineering is focused on the system as a whole." State what characteristics of a system you think this statement implies and how they apply to systems engineering.

1.2 Discuss the difference between engineered complex systems and complex systems that are not engineered. Give three examples of the latter. Can you think of systems engineering principles that can also be applied to nonengineered complex systems?

1.3 For each of the following areas, list and explain how at least two major technological advances/breakthroughs occurring since 1990 have radically changed them. In each case, explain how the change was effected in
(a) transportation,
(b) communication,
(c) financial management,
(d) manufacturing,
(e) distribution and sales,
(f) entertainment, and
(g) medical care.

1.4 What characteristics of an airplane would you attribute to the system as a whole rather than to a collection of its parts? Explain why.

1.5 List four pros and cons (two of each) of incorporating some of the latest technology into the development of a new complex system. Give a specific example of each.

1.6 What is meant by the term "modularity?" What characteristics does a modular system possess? Give a specific example of a modular system and identify the modules.

1.7 The section Orientation of Technical Professionals uses three components to describe this characteristic: science, mathematics, and engineering. Using this model, describe what you think your orientation is in terms of *x%* science, *y%* mathematics, and *z%* engineering. Note that your "orientation" does not measure your knowledge or expertise, but rather your interest and method of thought. Consider your relative interest in discovering new truths, finding new relationships, or building new things and making them work. Also, try to remember what your orientation was when you graduated from college, and explain how and why it has changed.

1.8 Systems engineers have been described as being an advocate for the whole system. Given this statement, which stakeholders should the systems engineer advocate the most? Obviously, there are many stakeholders and the systems engineer must be concerned with most, if not all, of them. Therefore, rank your answer in priority order—which stakeholder is the most important to the systems engineer; which is second; which is third?

FURTHER READING

B. Blanchard. *Systems Engineering Management*, Third Edition. John Wiley & Sons, 2004.

B. Blanchard and W. Fabrycky. *Systems Engineering and Analysis*, Fourth Edition. Prentice Hall, 2006, Chapter 1.

W. P. Chase. *Management of System Engineering*. John Wiley, 1974, Chapter 1.

H. Chesnut. *System Engineering Methods*. John Wiley, 1967.

H. Eisner. *Essentials of Project and Systems Engineering Management*, Second Edition. Wiley, 2002, Chapter 1.

C. D. Flagle, W. H. Huggins, and R. R. Roy. *Operations Research and Systems Engineering*. Johns Hopkins Press, 1960, Part I.

A. D. A. Hall. *Methodology for Systems Engineering*. Van Nostrand, 1962, Chapters 1–3; *Systems Engineering Handbook*. International Council on Systems Engineering, *A Guide for System Life Cycle Processes and Activities*, Version 3.2, July 2010.

E. Rechtin. *Systems Architecting: Creating and Building Complex Systems*. Prentice Hall, 1991, Chapters 1 and 11.

E. Rechtin and M. W. Maier. *The Art of Systems Architecting*. CRC Press, 1997.

A. P. Sage. *Systems Engineering*. McGraw Hill, 1992, Chapter 1.

A. P. Sage and J. E. Armstrong, Jr. *Introduction to Systems Engineering*. Wiley, 2000, Chapter 1.

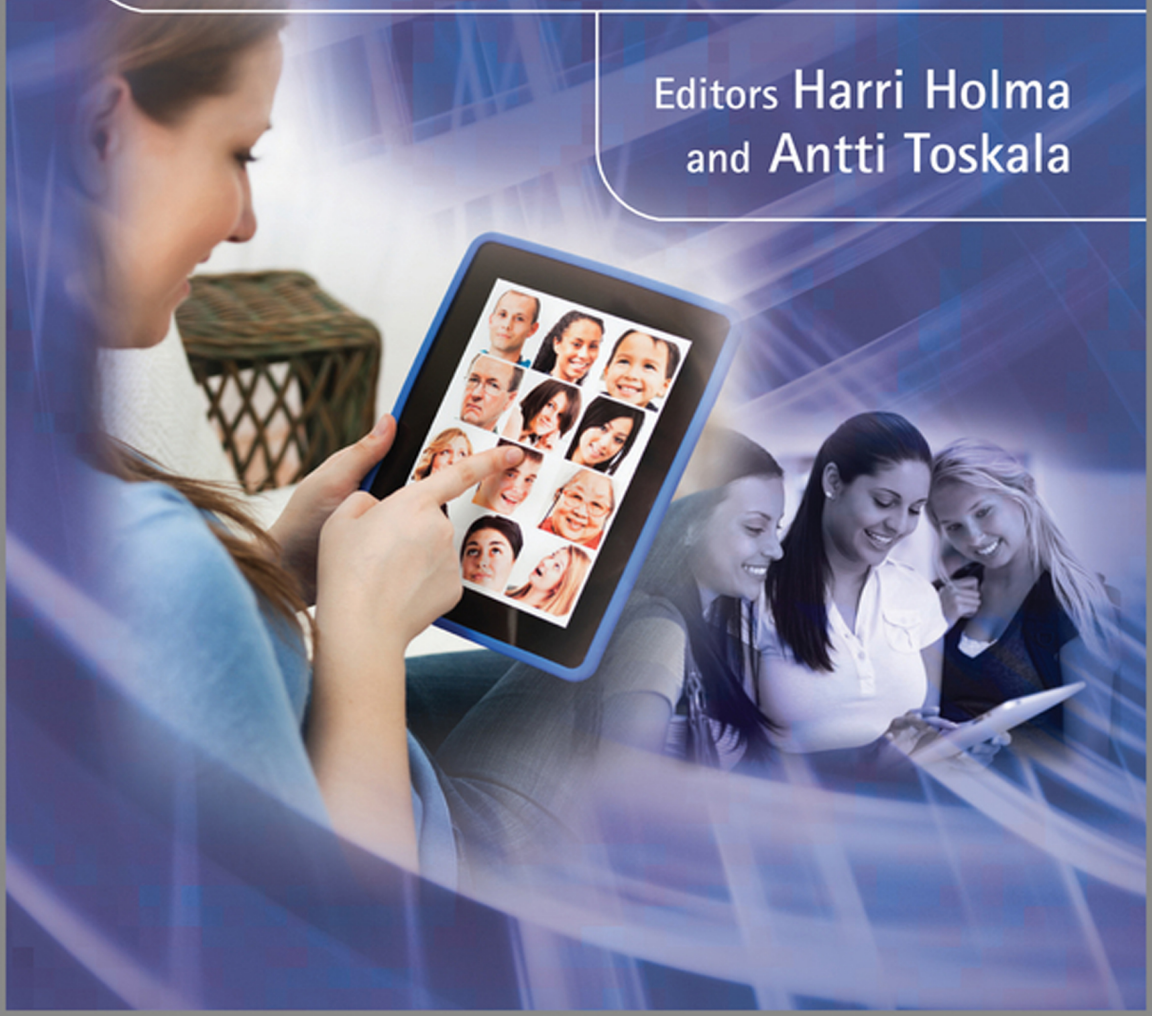
R. Stevens, P. Brook, K. Jackson, and S. Arnold. *Systems Engineering, Coping with Complexity*. Prentice Hall, 1988.

WILEY

LTE-Advanced

3GPP Solution for IMT-Advanced

Editors Harri Holma and Antti Toskala



Contents

1

Introduction

Harri Holma and Antti Toskala

1.1 Introduction

The huge popularity of smartphones and tablet computers has pushed the need for mobile broadband networks. Users find increasing value in mobile devices combined with a wireless broadband connection. Users and new applications need faster access speeds and lower latency while operators need more capacity and higher efficiency. LTE is all about fulfilling these requirements. GSM made voice go wireless, HSPA made initial set of data connections go wireless and now LTE offers massive capabilities for the mobile broadband applications.

The first set of LTE specifications were completed in 3GPP in March 2009. The first commercial LTE network opened in December 2009. There were approximately 50 commercial LTE networks by the end of 2011 and over 100 networks are expected by the end of 2012. The first LTE smartphones were introduced in 2011 and a wide selection of devices hit the market during 2012. An example LTE smartphone is shown in Figure 1.1: the Nokia 900 with 100 Mbps LTE data rate and advanced multimedia capabilities. Overall, LTE technology deployment has been a success story. LTE shows attractive performance in the field in terms of data rates and latency and the technology acceptance has been very fast. The underlying technology capabilities evolve further which allows pushing also LTE technology to even higher data rates, higher base station densities and higher efficiencies. This book describes the next step in LTE evolution, called LTE-Advanced, which is set to increase the data rate even beyond 1 Gbps.

1.2 Radio Technology Convergence Towards LTE

The history of mobile communications has seen many competing radio standards for voice and for data. LTE changes the landscape because all the existing radios converge towards LTE. LTE is the evolution of not only GSM/HSPA operators but also CDMA and WiMAX operators. Therefore, LTE can achieve the largest possible ecosystem. LTE co-exists smoothly with the current radio networks. Most GSM/HSPA operators keep their existing

LTE-Advanced: 3GPP Solution for IMT-Advanced, First Edition. Edited by Harri Holma and Antti Toskala.

Figure 1.1 An example of an LTE smartphone – Nokia Lumia 900.

GSM and HSPA radio networks running for long time together with LTE, and they also keep enhancing the existing networks with GSM and HSPA evolutions. The LTE terminals are multimode capable supporting also GSM and HSPA. The radio network solution is based on multi-radio base station which is able to run simultaneously all three radios. Many operators introduce multi-radio products to their networks together with LTE rollouts to simplify the network management and to modernize the existing networks.

The starting point for CDMA and WiMAX operators is different since there is no real evolution for those radio technologies happening. Therefore, CDMA and WiMAX operators tend to have the most aggressive plans for LTE rollouts to get quickly to the main stream 3GPP radio technology to enjoy the LTE radio performance and to get access to the world market terminals.

The high level technology evolution is illustrated in Figure 1.2.

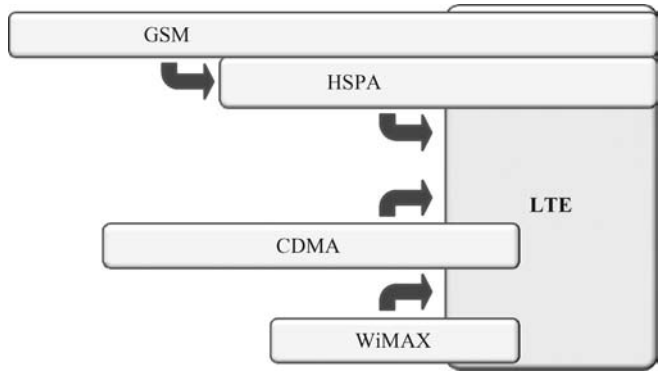


Figure 1.2 Radio technology convergence towards LTE.

1.3 LTE Capabilities

LTE Release 8 offers peak data rate of 150 Mbps in downlink by using 20 MHz of bandwidth and 2×2 MIMO. The first LTE devices support up to 100 Mbps while the network capability is up to 150 Mbps. The average data rates in the commercial networks range between 20 and 40 Mbps in downlink and 10–20 Mbps in uplink with 20 MHz bandwidth. Example drive test results are shown in Figure 1.3. Practical LTE data rates in many cases are higher than the available data rates in fixed Asymmetric Digital Subscriber Lines (ADSL). LTE has been deployed using number of different bandwidths: most networks use bandwidth from 5 to 20 MHz. If the LTE bandwidth is smaller than 20 MHz, the data rates scale down correspondingly. LTE has been rolled out both with Frequency Division Duplex (FDD) and Time Division Duplex) TDD variants. LTE has the benefit that both the FDD and TDD modes are highly harmonized in standardization.

The end user performance is also enhanced by low latency: the LTE networks can offer round trip times of 10–20 ms. The LTE connections support full mobility including seamless intra-frequency LTE handovers and inter-RAT (Radio Access Technology) mobility between LTE and legacy radio networks. The terminal power consumption is optimized by using discontinuous reception and transmission (DRX/DTX).

LTE also offers benefits for the operators in terms of simple network deployment. The flat architecture reduces the number of network elements and the interfaces. Self-Organizing Network (SON) has made the network configuration and optimization simpler enabling faster and more efficient network rollout.

LTE supports large number of different frequency bands to cater the needs of all global operators. The large number of RF bands makes it challenging to make universal LTE devices. The practical solution is to have several different device variants for the different markets. The roaming cases are handled mainly by legacy radios.

Initial LTE smartphones have a few different solutions for voice: Circuit Switched Fallback (CSFB) handover from LTE to legacy radio (GSM, HSPA, CDMA) or dual radio CDMA + LTE radio. Both options use the legacy circuit switched network for voice and

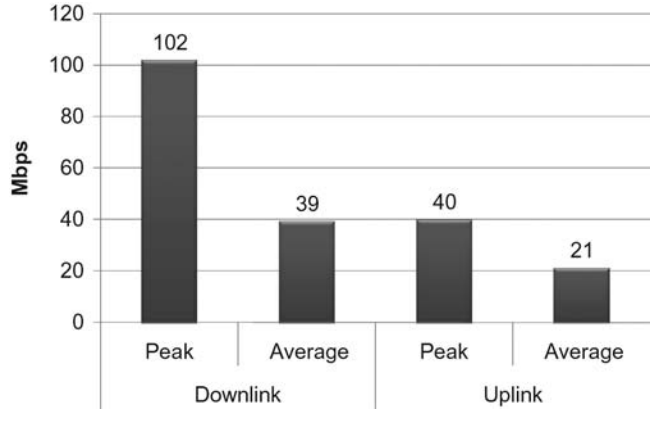


Figure 1.3 Example drive test data rates in LTE network with 20 MHz bandwidth.

LTE network for data. The Voice over LTE (VoLTE) solution with Voice over IP (VoIP) also started during 2012.

1.4 Underlying Technology Evolution

The radio technology improvements need to be supported by the evolution of the underlying technologies. The technology components – including mass storage, baseband, RF and batteries – keep evolving and help the radio improvements to materialize. The size of the mass storage is expected to have fastest growth during the next ten years which allows for storing more data on the device and which may fuel data download over the radio. The memory size can increase from tens of Gigabytes to several Terabytes. Also the digital processing has its strong evolution. The digital processing power has improved according to Moore's law for several decades. The evolution of the integration level will not be as easy as in earlier times, especially when we need to minimize the device power consumption. Still, the digital processing capabilities will improve during the 2010s, which allows for processing of higher data rates and more powerful interference cancellation techniques. Another area of improvement is the RF bandwidth which increases mainly because of innovations in digital front end processing. The terminal power consumption remains one of the challenges because the battery capacity is expected to have relatively slow evolution. Therefore, power saving features in the devices will still be needed. The technology evolution is illustrated in Figure 1.4.

LTE-Advanced devices and base stations will take benefit of the technology evolution. Higher data rates and wider bandwidth require baseband and RF evolution. The attractive LTE-Advanced devices also benefit from larger memory sizes and from improved battery capacity.

1.5 Traffic Growth

The data volumes in mobile networks have increased considerably during the last few years and the growth is expected to continue. The traffic growth since 2007 and the expected growth until 2015 are illustrated in Figure 1.5. The graph shows the total global mobile network data volume in Exabytes; that is, millions of Terabytes. The traffic is split into voice traffic and data traffic from laptops, tablets and smartphones. The data traffic exceeded the voice traffic during 2009 in terms of carried bytes. The initial data growth was driven by

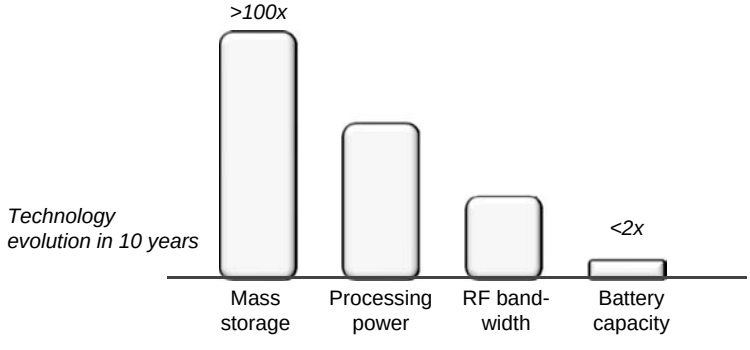


Figure 1.4 Evolution of underlying technology components.

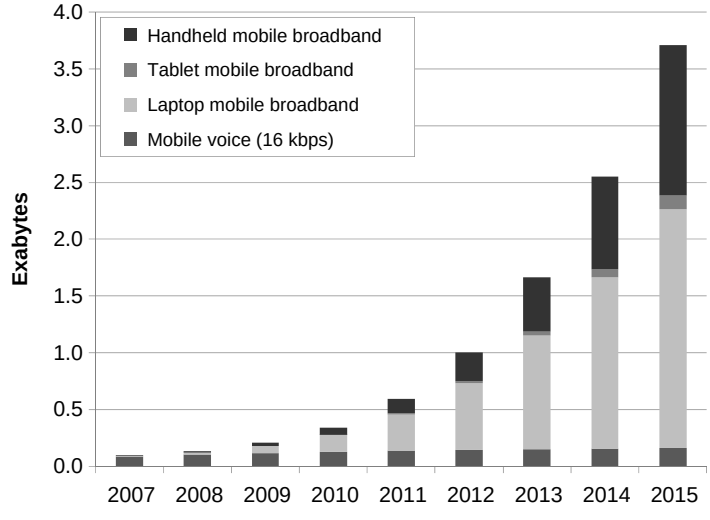


Figure 1.5 Expected traffic growth (Nokia Siemens Network estimate 2011).

the laptop modems; see an example in Figure 1.6. It is also expected that the LTE-Advanced capabilities, like higher data rates, are first introduced for the laptop modems. The relatively fastest growth from 2012 to 2015 is expected to come from smartphones. The smartphones make nearly half of the traffic by 2015. The total traffic by 2015 will be approximately 40 times more than the traffic 2007. The share of voice traffic is expected to shrink to less than 5% by 2015. Some of the advanced markets already have the total traffic 50 times more than the voice traffic; that means voice is less than 2% of total traffic.

It is not only the data volume that is growing in the networks but also the amount of signalling grows and the number of connected devices grows. The radio evolution work needs to address all these growth factors.

Figure 1.6 Example of a 100 Mbps USB modem – Nokia Siemens Networks 7210.

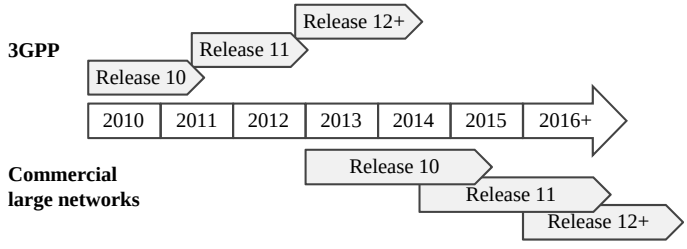


Figure 1.7 3GPP timing of LTE-Advanced.

1.6 LTE-Advanced Schedule

The first set of LTE-Advanced is specified in 3GPP Release 10. That release was completed in June 2011. The target date for Release 11 is December 2012. The typical release cycle in 3GPP has been 1.5 years – except for some smaller releases like Release 9 that was completed in a year. It tends to take another 1.5 years from the specification's completion until the first commercial networks and devices are available. Some small features can be implemented faster while some major features requiring heavy redesign may take more time. We could then expect that the first LTE-Advanced features are commercially available during 2013, and Release 11 features towards end of 2014. The LTE-Advanced schedule is shown in Figure 1.7.

1.7 LTE-Advanced Overview

The main features of LTE-Advanced are summarized in Figure 1.8.

- Downlink carrier aggregation to push the data rate initially to 300 Mbps with $20 + 20$ MHz spectrum and 2×2 MIMO, and later to even 3 Gbps by using 100 MHz bandwidth and 8×8 MIMO. More bandwidth is the handy solution to increase the data rates.
- Multiantenna MIMO evolution to 8×8 in downlink and 4×4 in uplink. The multiantenna MIMO can also be used at the base station while keeping the number of terminal antennas low. This approach offers the beamforming benefits increasing the network capacity while

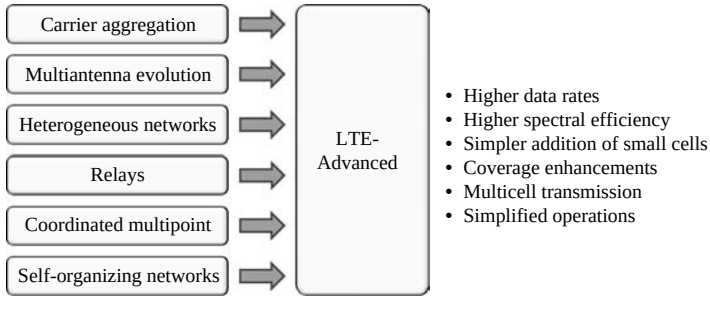


Figure 1.8 Overview of LTE-Advanced main features.

keeping the terminal complexity low. Multiantennas increase the data rates and the network capacity.
- Heterogeneous network (HetNet) for the co-channel deployment of macrocells and small cells. HetNet features enable interference coordination between the cell layers. Those features enhance the network capacity and coverage with high density of small cells while sharing the frequency with large macrocells.
- Relay nodes for backhauling the base stations via LTE radio interface. The transmission link can use inband or outband transmission. Relays are practical for increasing network coverage if the backhaul connections are not available.
- Coordinated multipoint transmission and reception allows using several cells for the data connection towards one terminal. Coordinated multipoint improves especially the cell edge data rates that are limited by inter-cell interference.
- Self-organizing network features make the network rollout faster and simpler, and improves the end user performance by providing correct configurations and optimized parameter setting.

LTE-Advanced features in Release 10 can be upgraded flexibly on top of Release 8 network on the same frequencies while still supporting all legacy Release 8 terminals. Therefore, the evolution from LTE to LTE-Advanced will be a smooth one. All these features will be described in detail in this book.

1.8 Summary

LTE Release 8 has turned out to be a successful technology in terms of practical performance and in terms of commercial network and terminal launches. At the same time the high popularity of smartphones pushes the need for further mobile broadband evolution. LTE-Advanced is designed to enhance LTE capabilities in terms of data rates, capacity, coverage and operational simplicity. The first set of LTE-Advanced specifications was completed in 3GPP during 2011 and the features are expected to be commercially available 2013. LTE-Advanced is backwards compatible with LTE and can co-exist with LTE Release 8 terminals on the same frequency.

THE DEATH OF THE INTERNET

Edited by

MARKUS JAKOBSSON

WILEY
高等教育出版社 HIGHER EDUCATION PRESS
IEEE
IEEE computer society

Contents

Chapter 1

What Could Kill the Internet? And so What?

Anything that makes the Internet either *dangerous* or *meaningless* could kill it.

The dangers may be to your machine, to proprietary information, to your financial situation, or even to *you*.

Malware can corrupt your machine. It can destroy data and software. It can even destroy hardware—for example, by rewriting your computer's EEPROM or flash memories so many times that they burn out. That takes only a few seconds per block, and if strategically chosen blocks are damaged, the hardware is rendered useless. Malware can also affect external equipment or processes as the em Stuxnet worm gave an example in 2010. It can be used to turn on the microphone of your phone, turning you into a walking eavesdropping bot—and you would not even know it! Malware is believed to commonly be used to be used to steal corporate and national secrets.

Most of the time, though, malware will only attempt to steal your money. That is the same goal as phishers have. And it is the same goal as scam artists have, attempting to convince their victims to send them money or merchandise. Often referred to as Nigerian scammers, these are certainly not all in Nigeria, although a surprisingly high number is.

The Internet—as well as wireless networks—can also be used to spy on people, to determine their location, for example. This can have direct physical consequences, whether the attack is mounted by a crazed expartner, political enemies, or common criminals. While this type of tracking is not commonly heard about today, it does not mean that it does not happen. And it certainly does not mean that it *cannot* happen. In fact, and as we describe in Section 6.1, it can be done on a grand scale without any significant investment.

Those are just a few examples of dangers that did not exist just a few years ago, and which soon may take up first-page newspaper space. There are also plenty of ways in which the Internet may become *meaningless*.

When we speak of spam, almost everybody thinks of unwanted email. A similar type of spam affects mobile communications—SMS spam. Voice spam is closely related to telemarketing. Instant messaging and online game messaging are also vulnerable to spam. But not all spam is about selling counterfeit Viagra or Rolexes. The term is also used to refer to other junk material, whether it is intended to fool search engines to rank particular pages higher than they otherwise would have. It can be used to manipulate reputations of

The Death of the Internet, First Edition. Edited by Markus Jakobsson.

sites and services—typically to make them look more attractive than they are, but sometimes used the other way to stab competitors. Spam is used to mean polluted peer-to-peer material—material that claims to be things it is not.

Spam is not the only source of pollution of information, though. Criminals can deceive news organizations to broadcast untruthful information. Given the increased competition to be first in online media, it is sometimes hard for journalists to balance the need to validate information—and to be first. Malware and spoofing can be used to make information appear to have originated with trusted sources. Criminals may benefit from the pollution of information in many ways. Politically, by sowing doubts and causing fear and confusion. Financially, by manipulating the markets.

The Internet could also become meaningless by becoming so dangerous that typical users restrain their activities and only dare to engage in a minimal manner.

But "meaningless" is in the eye of the beholder. Typical users would have one view of what could make the Internet meaningless. Service providers have a very different view. To online merchants, the Internet would be meaningless if nobody buys their products using it, or if it cannot be used to advertise products that are sold off-line. If this were to happen, advertising would plummet. Since many free services depend of advertisements, that type of development would affect them, and they would scale back or vanish. A lot of services we have come to take for granted fall into this category, starting with search engines, but also including online news services, many content distribution sites, email service providers (do you remember—we used to pay for email . . .) and other services, such as translation services, recommendation services, navigation services, consumer advice services . . . you name it, it is probably on the list.

So what happens if people do not dare to watch advertisements? Or if click-fraud runs rampant? It is the same end result. No advertisements . . . no services.

Severe attacks on the Internet will send shockwaves through society.

If your livelihood depends on the Internet—like mine does—then you are surely aware of what the impact would be to you of any severe problems with it. You know that you would not be happy if the Internet were crippled by fraud. But if that does not describe you, you might shrug, thinking that this is not such a big deal. After all, you may think, you can live just fine without reading the news online, and you can drive to the store instead of shopping online. Right? Wrong.

- "My phone will still work." Well, maybe not. *You* may not use VoIP services, but most phone calls are still routed over the Internet. If the Internet goes down, your phone goes dead. And so will the phone of your local 911 dispatcher.
- "My lights and heat will still work." Maybe. Maybe not. Our electricity infrastructure is almost as complex as the Internet. Power is routed to where it is needed. The production is ramped up and down to meet the demand. The failure of one part of the system can cause failures in other parts of the system. And since the coordination of this complex system is done using the Internet, even electricity delivery may suffer from severe attacks on the Internet.
- "I can still walk down to the grocery store and get what I need." Yes, you can. But what if their ordering system or delivery system depends on the Internet, or on companies who depend on the Internet? Will the shelves still be full? Maybe not.

- "I still have money in the bank." You may not lose your password to phishing or malware, but what if your bank clerk loses it—or accidentally leaks your mothers maiden name? It may take a while for you to get your money back. And what if the financial system is hampered by a lack of trust; by invalid trades; by general abuse?

Even if the Internet is not taken down by attacks, we may all be affected by rising levels of fraud.

You and I may have bulletproof antivirus software on our computers —and phones—and still be affected. For example, if people passing you on town have infected phones, these phones may render your phone useless simply by making phone calls or web accesses in dramatic quantities. It would be hard for you to get a connection when you want one.

If you use a Bluetooth enabled headset and let your phone be discoverable, then your phone can be tracked by infected phones in your neighborhood. In fact, it may not matter whether your phone is discoverable or not if nearby devices can eavesdrop on signals: your phone will send its Bluetooth device identifier in the clear.

But it is not all about phones. Do you use social networking? Many services will detect if you are online or not. You online/offline patterns may say a lot about you. Who you are, what you do.

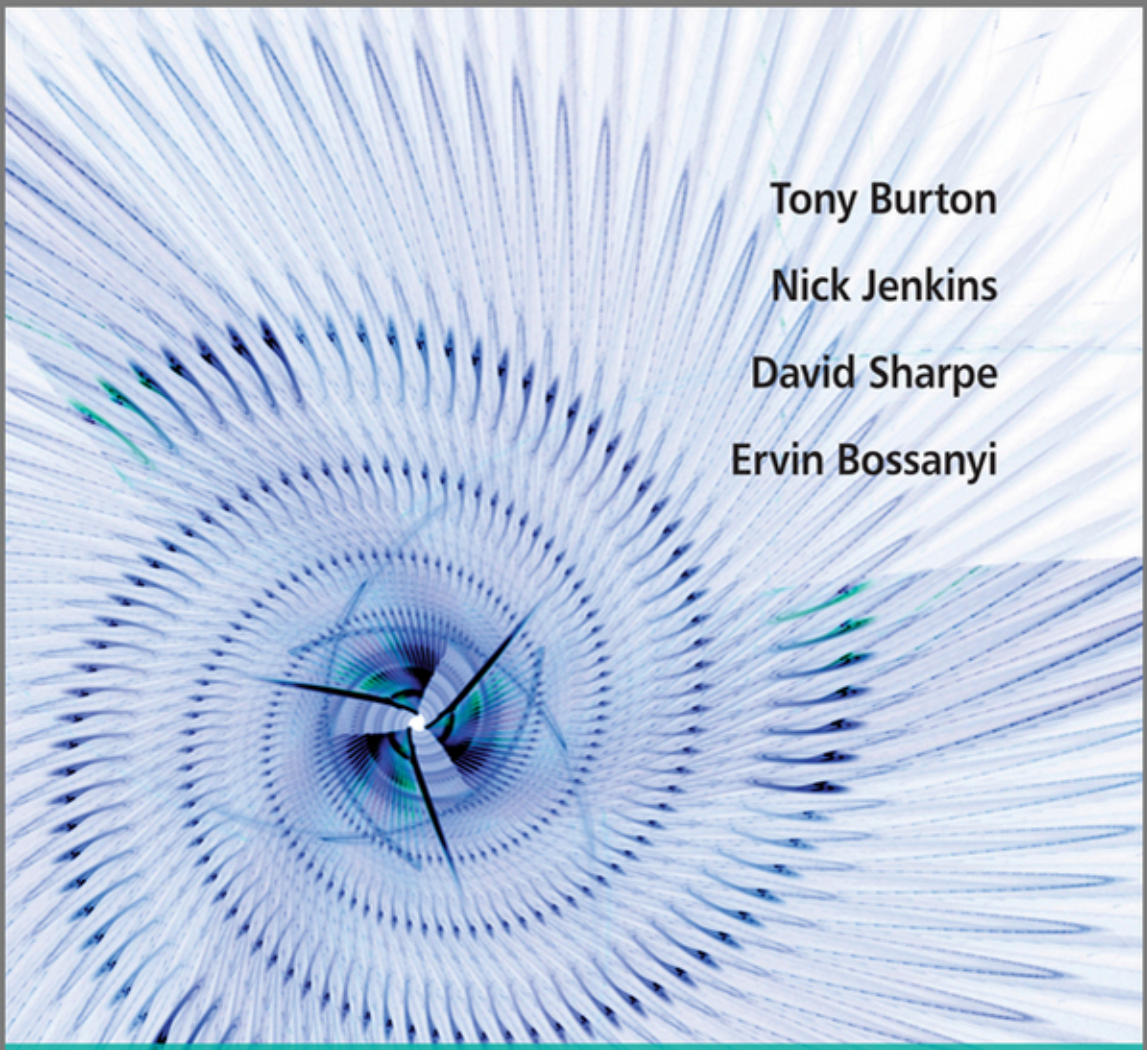
Internet terrorism is easy, and we are weak.

So far, I have argued that online attacks may result in problems in society. In lack of trust, degradation of our infrastructure, increases of costs to do business. *But the consequences of online attacks could also be what invites abuse.* If a hostile country or organization wants to hurt us, they may find that the easiest way of doing it is by attacking the Internet. Our dependence on the Internet will *invite* attacks. We have already seen instances of massive cyber attacks, such as those on Estonia in 2007. We are not safe from such attacks. If anything, we may be more vulnerable to them, as our dependence on the Internet is greater—and increasing by the day.

Since 2001, we are all aware of terrorism. What makes it terrifying is not only its arbitrariness, but its asymmetric nature. A small number of dedicated aggressors can inflict massive damage and suffering to large numbers of victims. The terrorists, of course, do not attack us because it is *fun*—they do it to further their political agendas. From their point of view, what they do is justified by the needs.

Of course, every society does what they think is justified by their needs—if they think they can get away with it, at least. Now, imagine that you belonged to an organization that needed to send a strong signal to a society or organization you disagree with. You may not have a powerful army to engage to pressure your enemy with. But you have other, and simpler, ways. You can degrade their infrastructure—with the click of a mouse. You can cause severe interruptions, degrade their economy, spread fear and confusion. Would you be tempted to click? *Of course you would.* And if you would not, then somebody else in your organization surely would.

That is also what we are up against. It is not *only* about fraudsters trying to make a profit.



Tony Burton
Nick Jenkins
David Sharpe
Ervin Bossanyi

Wind Energy Handbook

SECOND EDITION

WILEY

Contents

1

Introduction

1.1 Historical development

Windmills have been used for at least 3000 years, mainly for grinding grain or pumping water; while in sailing ships the wind has been an essential source of power for even longer. From medieval times, horizontal axis windmills were an integral part of the rural economy and only fell into disuse with the advent of cheap fossil-fuelled stationary engines and then the spread of rural electrification (Musgrove, 2010). The use of windmills (or wind turbines) to generate electricity can be traced back to the late nineteenth century with the 12 kW direct current windmill generator constructed by Charles Brush in the USA and the research undertaken by Poul la Cour in Denmark. However, for much of the twentieth century there was little interest in using wind energy for electricity generation, other than for battery charging for remote dwellings; and these low power systems were quickly removed once access to the electricity grid became available. One notable development was the 1250 kW Smith-Putnam wind turbine constructed in the USA in 1941. This remarkable machine had a steel rotor 53 m in diameter, full span pitch control and flapping blades to reduce loads. Although a blade spar failed catastrophically in 1945, it remained the largest wind turbine constructed for some 40 years (Putnam, 1948).

Golding (1955) and Shepherd and Divone in Spera (1994) provide a fascinating history of early wind turbine development. They record the 100 kW 30 m diameter Balaclava wind turbine in the then USSR in 1931 and the Andrea Enfield 100 kW 24 m diameter pneumatic design constructed in the UK in the early 1950s. In this turbine, hollow blades, open at the tip, were used to draw air up through the tower where another turbine drove the generator. In Denmark the 200 kW 24 m diameter Gedser machine was built in 1956, while Electricite de France tested a 1.1 MW 35 m diameter turbine in 1963. In Germany, Professor Ulrich Hutter constructed a number of innovative, lightweight turbines in the 1950s and 1960s. In spite of these technical advances and the enthusiasm of Golding at the Electrical Research Association in the UK, among others, there was little sustained interest in wind generation until the price of oil rose dramatically in 1973.

Wind Energy Handbook, Second Edition. Tony Burton, Nick Jenkins, David Sharpe and Ervin Bossanyi.

The sudden increase in the price of oil stimulated a number of substantial, government funded programmes of research, development and demonstration. In the USA this led to the construction of a series of prototype turbines starting with the 38 m diameter 100 kW Mod-0 in 1975 and culminating in the 97.5 m diameter 2.5 MW Mod-5B in 1987. Similar programmes were pursued in the UK, Germany and Sweden. There was considerable uncertainty as to which architecture might prove most cost-effective and several innovative concepts were investigated at full scale. In Canada, a 4 MW vertical axis Darrieus wind turbine was constructed and this concept was also investigated in the 34 m diameter Sandia Vertical Axis test facility in the USA. In the UK, an alternative, vertical axis design using straight blades to give an 'H' type rotor was proposed by Dr Peter Musgrove and a 500 kW prototype constructed (Musgrove, 2010). In 1981 an innovative horizontal axis 3 MW wind turbine was built and tested in the USA. This used hydraulic transmission and, as an alternative to a yaw drive, the entire structure was orientated into the wind. The best choice for the number of blades remained unclear for some while and large horizontal axis turbines were constructed with one, two or three blades.

Much important scientific and engineering information was gained from these government funded research programmes and the prototypes generally worked as designed. However, the problems of operating very large wind turbines, unmanned and in difficult wind climates, were often underestimated and the reliability of the prototypes was not good. At the same time as the multi-megawatt prototypes were being constructed private companies, often with considerable state support, were manufacturing much smaller, often simpler, turbines for commercial sale. In particular the financial support mechanisms in California in the mid-1980s resulted in installation of a very large number of quite small (<100 kW) wind turbines. A number of these designs also suffered from various problems but, being smaller, they were generally easier to repair and modify. The Danish wind turbine concept emerged of a three-bladed, upwind stall regulated rotor and a fixed-speed, induction generator drive train. This deceptively simple architecture proved to be remarkably successful and was implemented on turbines as large as 60 m in diameter and at ratings of up to 1.5 MW. However, at large rotor diameters and generator ratings, the architecture ceases to be effective as aerodynamic stall is increasingly difficult to predict, an induction generator no longer easily provides enough damping and torsional compliance in the drive train and the requirements of the electrical Transmission System Operators for connection to the network, the Grid Codes become difficult to meet. Hence, as the size of commercially available turbines approached or exceeded that of the large prototypes of the 1980s the concepts investigated then of variable speed operation, full-span control of the blade pitch and advanced materials were used increasingly by designers.

In 1991 the first offshore wind farm was constructed at Vindeby consisting of eleven, 450 kW wind turbines located up to 3 km offshore. Throughout the 1990s small numbers of offshore wind turbines were placed close to shore, while in 2002 the Horns Rev, 160 MW wind farm, some 20 km off the western coast of Denmark, was constructed. This was the first project to use an offshore substation that increased the power collection voltage of 30 kV to 150 kV for transmission to shore. At the time of writing (2010) a 500 MW wind offshore wind farm (Greater Gabbard) is under construction off the coast of England with 1000 MW projects under development. The wind turbines that have been installed in these offshore wind farms have been marinised versions of 3-bladed, upwind terrestrial designs. However, the possibility of higher blade tip-speeds, because of more relaxed noise constraints and a reduced emphasis on visual appearance in sites far from land, are leading to an interest in the development of very large, lower solidity rotors with two or even one blade.

The stimulus for the development of wind energy in 1973 was the increase in the price of oil and concern over limited fossil fuel resources. From around 1990, the main driver for use of wind turbines to generate electrical power was the very low CO_2 emissions (over the entire life cycle of manufacture, installation, operation and de-commissioning) and the potential of wind energy to help mitigate climate change. Then from around 2006 the very high oil price and concerns over security of energy supplies led to a further increase of interest in wind energy and a succession of policy measures were put in place in many countries to encourage its use. In 2007 the European Union declared a policy that 20% of all energy should be from renewable sources by 2020. Because of the difficulty of using renewable energy for transport and heat, this implies that in some countries 30–40% of electrical energy should come from renewables, with wind energy likely to play a major part. Energy policy continues to develop rapidly with many countries adopting ambitions to reduce greenhouse gas emissions of up to 80% by 2050 in order to mitigate climate change.

The development of wind energy in some countries has been more rapid than in others and this difference cannot be explained simply by differences in the wind speeds. Important factors include the financial support mechanisms for wind generated electricity, access to the electrical network, the process by which the local authorities give permission for the construction of wind farms and the perception of the general population, particularly with respect to visual impact. The development of offshore sites, although at considerably increased cost, is in response to these concerns over the environmental impact of wind farms.

As a relatively new generation technology, wind energy requires financial support to encourage its development and stimulate investment from private companies. Such support is provided in many countries and recognises the contribution wind generation makes to climate change mitigation and security of national energy supplies. There is presently an active debate as to the best mechanism of providing such support so that it stimulates the development of wind energy at minimum cost and without distorting the electricity market.

Feed-in-Tariffs are offered in a number of countries, most notably Germany and Spain. A fixed price is paid for each kWh generated from renewable sources with different rates for wind energy, photovoltaic solar energy and other renewable energy technologies. This support mechanism has the benefit of giving certainty of the revenue stream from a successful project and is credited by its supporters for the very rapid development of wind energy, and other renewables, in these countries. An alternative approach is a quota or Renewable Portfolio Standard system where a government places an obligation on electricity suppliers to source a certain fraction of the energy they supply from renewable sources. An example is the UK Renewable Obligation Certificate system where renewable energy generators are awarded green certificates for energy generated from renewable sources. These green certificates are traded independently from the electrical energy and electricity suppliers who fail to acquire their quota pay a buy-out price. With this support mechanism risk is transferred to the project developer and only the most commercially attractive renewable technologies are developed. Historically, Capacity Auctions have also been used, such as the earlier UK NFFO mechanism and similar examples in Ireland and France. A national government determines the volume of wind energy required and conducts an auction for capacity on price. Capacity Auctions suffered from some wind farm developers bidding low to secure agreements and then not constructing projects.

Although the form of these support mechanisms, and particularly their stability, is important, it may be argued that other factors including access to the electricity grid, speed of the planning/permitting system and public acceptability play a critical role in determining the rate

of deployment of wind energy. It is also likely that more general support measures for low carbon electricity generation, for example the European Union Emissions Trading Scheme or wider carbon taxes, will provide significant support for the development of wind energy in the future.

1.2 Modern wind turbines

The power output from a wind turbine is given by the well-known expression:

$$P = \frac{1}{2} C_p \rho A U^3 \tag{1.1}$$

where

ρ is the density of air (1.25 kg/m^3)
C_p is the power coefficient
A is the rotor swept area
U is the wind speed

The density of air is rather low, 800 times less than water which powers hydro plant, and this leads directly to the large size of a wind turbine. Depending on the design wind speed chosen, a 3 MW wind turbine may have a rotor that is more than 90 m in diameter. The power coefficient describes that fraction of the power in the wind that may be converted by the turbine into mechanical work. It has a theoretical maximum value of 0.593 (the Betz limit) and rather lower peak values are achieved in practice (see Chapter 3). The power coefficient of a rotor varies with the tip speed ratio, (the ratio of rotor tip speed to free wind speed) and is only a maximum for a unique tip speed ratio. Incremental improvements in the power coefficient are continually being sought by detailed design changes of the rotor and by operating at variable speed it is possible to maintain the maximum power coefficient over a range of wind speeds. However, these measures will give only a modest increase in the power output. Major increases in the output power can only be achieved by increasing the swept area of the rotor or by locating the wind turbines on sites with higher wind speeds.

Hence, over the last 40 years there has been a continuous increase in the rotor diameter of commercially available wind turbines from less than 30 m to more than 100 m. A tripling of the rotor diameter leads to a nine times increase in power output. The influence of the wind speed is, of course, more pronounced with a doubling of wind speed leading to an eight fold increase in power. Thus, there have been considerable efforts to ensure that wind farms are developed in areas of the highest wind speeds and the turbines optimally located within wind farms. In certain countries very high towers are being used (more than 100 m high) to take advantage of the increase of wind speed with height.

In the past a number of studies were undertaken to determine the 'optimum' size of a wind turbine by balancing the complete costs of manufacture, installation and operation of various sizes of wind turbines against the revenue generated (Molly *et al*., 1993). The results indicated a minimum cost of energy would be obtained with wind turbine diameters in the range of 35–60 m, depending on the assumptions made. However, these estimates would now appear to be too low and there is no obvious point at which rotor diameters and, hence, output

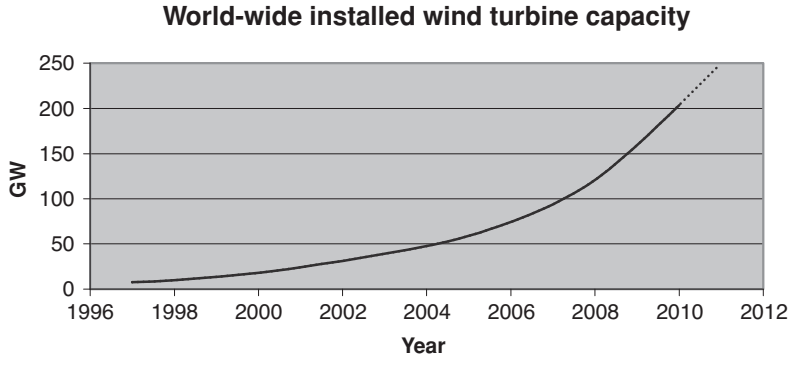


Figure 1.1 Wind power capacity world-wide (World Wind Energy Association, 2009)

power will be limited particularly for offshore wind turbines where the very large components can be transported by ship directly from the factory to site.

All modern electricity generating wind turbines use the lift force derived from the blades to drive the rotor. A high rotational speed of the rotor is desirable in order to reduce the gearbox ratio required and this leads to low solidity rotors (the ratio of blade area/rotor swept area). The low solidity rotor acts as an effective energy concentrator and as a result the energy generated over a wind turbine's life is much less than that used for its manufacture and installation. An energy balance analysis of a 3 MW wind turbine showed that the expected average time to generate a similar quantity of energy to that used for its manufacture, operation, transport, dismantling and disposal was 6–7 months (European Wind Energy Association (EWEA), 2009). A similar time was calculated for installation both onshore and offshore.

Until around 2000, the installed wind turbine generating capacity was so low that its output was viewed by electricity Transmission System Operators simply as negative load that supplied energy but played no part in supporting the operation of the power system and maintaining its stability. Since then, with the very much increased capacity of wind generation (Figure 1.1), turbines are required to contribute to the operation of the power system. The requirements for their performance are defined through the Grid Codes, issued by the Transmission System Operators. Compliance is mandatory before connection to the network is allowed. The Grid Codes specify operational requirements so that in addition to contributing energy, or real power,[1] the wind turbines provide ancillary services particularly for voltage and frequency control. At present the Grid Codes differ in detail from country to country but generally specify that wind turbines must remain stable and connected to the network in case of electrical faults on the network, define their performance in terms of reactive power for voltage control and their ability to vary real power for frequency support. As wind turbines become an ever increasing fraction of electricity generating capacity, it is likely that further requirements will be placed on them to replicate the ancillary services previously provided by conventional synchronous generators. Compliance with the Grid Code requirements is difficult to achieve

[1] In a large interconnected electric power transmission system, real power controls system frequency while reactive power determines the voltages of the network.

with simple fixed speed induction generators using the Danish concept and these regulations are a major driver for the use of variable speed generators.

1.3 Scope of the book

The use of wind energy to generate electricity is now well accepted with a large industry manufacturing and installing tens of GWs of new capacity each year. Although there are exciting new developments, particularly in very large wind turbines, and many challenges remain, there is a considerable body of established knowledge concerning the science and technology of wind turbines. This book records some of this knowledge and presents it in a form suitable for use by students (at final year undergraduate or post-graduate level) and by those involved in the design, manufacture or operation of wind turbines. The overwhelming majority of wind turbines presently in use are horizontal axis connected to a large electricity network. These turbines are the subject of this book.

Chapter 2 discusses the wind resource. Particular reference is made to wind turbulence due to its importance in wind turbine design. Chapter 3 sets out the basis of the aerodynamics of horizontal axis wind turbines, while Chapter 4 discusses aspects of their performance. Any wind turbine design starts with establishing the design loads and these are discussed in Chapter 5. Chapter 6 sets out the various design options for horizontal axis wind turbines with approaches to the design of some of the important components examined in Chapter 7. The functions of the wind turbine controller and some of the possible analysis techniques described are discussed in Chapter 8. In Chapter 9 wind farms and the development of wind energy projects are reviewed with particular emphasis on environmental impact. Chapter 10 considers how wind turbines interact with the electrical power system while Chapter 11 deals with the important topic of offshore wind energy.

The book attempts to record well-established knowledge that is relevant to wind turbines which are currently commercially significant. Thus, it does not discuss a number of interesting research topics or areas where wind turbine technology is still evolving. Although they were investigated in considerable detail in the 1980s, large vertical axis wind turbines have not proved to be commercially competitive and are not currently manufactured in significant numbers. Hence, the particular issues of vertical axis turbines are not dealt with in this text.

There are presently some 2 billion people in the world without access to mains electricity and, in conjunction with other generators (e.g. diesel engines), wind turbines may in the future be an effective means of providing some of them with power. However, autonomous power systems are extremely difficult to design and operate reliably, particularly in remote areas of the world and with limited budgets. A small autonomous AC power system has all the technical challenges of a large national electricity system but, due to the low inertia of the generators, requires a very fast, sophisticated control system to maintain stable operation. Over the last 30 years there have been a number of attempts to operate autonomous wind-diesel systems on islands or for other remote communities throughout the world, but with only limited success. This class of installation has its own particular problems and again, given the very limited size of the market at present, this specialist area is not dealt with in this book.

Installations of offshore wind turbines are now commencing in significant numbers (Figure 1.2). The first offshore wind farms were installed in rather shallow waters and used marinised terrestrial designs. A number of larger offshore wind farms used offshore substations to increase the transmission voltage, while prototype floating wind turbines for deeper waters

(a) (b)

(c)

Figure 1.2 (a) Rhyl Flats Offshore Wind Farm. See Plate 1 for the colour figure. (b) 3.6 MW nacelle prior to lifting. (c) Assembly of 3.6 MW wind turbine. Rhyl Flats Offshore Wind Farm consists of 25 × 3.6 MW Siemens wind turbines. Hub height – 80 m above mean sea level (MSL). Height to blade tip – 134 m above MSL. Rhyl Flats Offshore Wind Farm was built and is operated by RWE npower renewables. Photographer: Guy Woodland. Photos reproduced courtesy of RWE npower renewables. See Plate 2 for the colour figure

have been deployed. Very large wind farms with multi-megawatt turbines many kilometres offshore are now being planned and will be constructed over the coming years.

References

European Wind Energy Association (EWEA) (2009) *Wind energy the facts*. Earthscan, London.

Golding, E.W. (1955) '*The generation of electricity from wind power*', E & F.N. Spon, London (reprinted R.I. Harris, London 1976).

Molly, J.P., Keuper, A. and Veltrup, M. (1993) '*Statistical WEC Design and Cost Trends*', Proceedings of the European Wind Energy Conference, Travemunde 8–12 March 1993, pp. 57–59.

Musgrove, P. (2010) '*Wind Power*', Cambridge University Press, Cambridge.

Putnam, G.C. (1948) '*Power from the Wind*', Van Nostrand Rheinhold, New York.

Spera, D.A. (1994) '*Wind Turbine Technology, Fundamental Concepts of Wind Turbine Engineering*', ASME Press, New York.

World Wind Energy Association (2009), '*World wind energy report 2009*'. www.wwindea.org.

Further Reading

Eggleston, D.M. and Stoddard, F.S. (1987) '*Wind Turbine Engineering Design*', Van Nostrand Rheinhold, New York.

Freris, L.L. (ed.) (1990) '*Wind Energy Conversion Systems*', Prentice-Hall, New York.

Gipe, P. (1995) '*Wind Energy Comes of Age*', John Wiley & Sons Inc., New York.

Harrison, R., Hau, E., and Snel, H. (2000) '*Large Wind Turbines, Design and Economics*', John Wiley & Sons Ltd., Chichester.

Hau, E. (2006) '*Wind Turbines: Fundamentals, Technologies, Application, Economics*, 2nd edition', Springer, Heidelberg.

Johnson, L. (1985) '*Wind Energy Systems*', Prentice-Hall, New York.

Le Gourieres, D. (1992) '*Wind Power Plants Theory and Design*', Pergamon Press, Oxford.

Manwell, J.W., McGowan, J.G. and Rogers, A.L. (2009) '*Wind Energy Explained, Theory, Design and Application*, 2nd edition', Wiley-Blackwell, Oxford.

Twiddell, J.W. and Weir, A.D. (1986) '*Renewable Energy Resources*', E & F.N. Spon, London.

Twidell, J.W. and Gaudiosi, G. (eds) (2009) '*Offshore Wind Power*', Multi-science Publishing Co., Brentwood.