Analysis, Synthesis, and Implementation of Networks of Multiple-Input Translinear Elements

Thesis by

Bradley A. Minch

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1997

(Submitted May 20, 1997)

ACKNOWLEDGMENTS

I want to express my gratitude to my advisor, Carver Mead, who has done much more than merely advise me—he has been a mentor, he has shared his insights, and he has created a wonderfully rich environment in which to do research. I am particularly grateful that he allowed me the intellectual freedom to develop a tangent until it became my thesis.

I want to thank Barrie Gilbert for being an inspiration to me, for documenting the development of his ideas, and for sharing his keen insights into whence little circuits come.

I want to thank Ron Benson, Kwabena Boahen, Tobi Delbrück, Jeff Dickson, Doug Kerns, Shih-Chii Liu, Dick Lyon, Lena Peterson, Rahul Sarpeshkar, and Lloyd Watts for many enlightening conversations and for much help over the years. I am grateful to Rahul Sarpeshkar for pointing me toward Samuel Mason's work on signal-flow graphs. I am especially grateful to Paul Hasler and Chris Diorio for many helpful conversations and close collaborations over the last five years—they have been more than mere colleagues, but good friends as well. All for one, and one for all!

I want to thank Yaser Abu-Mostafa, Al Barr, Jehoshua Bruck, and R. David Middlebrook for serving on my thesis defense committee.

I want to thank Jim Campbell, Donna Fox, Calvin Jackson, and Candace Schmidt for keeping things running smoothly in Carverland.

I want to thank Lyn Dupré for editing my writing and for improving my ear. Having extensively worked with Lyn, the active voice has become much more prominent in my writing.

I want to thank my parents, Roland and Laurel Minch, for their continual loving support, guidance, and encouragement. In particular, I want to thank my Dad for taking me to his office when I was a little boy and letting me play with the stuff in his middle desk drawer—without that experience, I doubt that I would ever have seen a slide rule.

I want to express my profound gratitude to my wife, Sherri Minch, whose love, friendship, support, and encouragement have made these last five years the best of my life.

Now, to Him who upholds all things by His word of power, and in whom all things cohere be glory and thanks. Because He created all things and by His will they have their being, He is worthy to receive praise and honor both now and forever!

ABSTRACT

At the time of its invention in the seventeenth century, the logarithmic slide rule literally revolutionized the way calculation was done. From then until the advent of the pocket calculator, this analog computational device was widely used to perform multiplications and divisions, to raise numbers to fixed powers and extract fixed roots of numbers. Today, the slide rule may be gone, but it is not forgotten. In this thesis, I present a class of simple translinear network circuits which essentially function as electronic slide rules, accurately computing products, quotients, powers, and roots. I describe two different analysis procedures that allow us to determine the steady-state relationship between input and output currents. I also describe systematic techniques for synthesizing such circuits whereby we can produce a circuit whose steady-state transfer characteristics embody some desired product-of-power-law relationship between input and output currents. These circuits are made from multiple-input translinear elements; such elements produce output currents that are proportional to the exponential of a weighted sum of their input voltages. We can implement the weighted voltage summations with either resistive or capacitive voltage dividers. We can obtain the required exponential voltage-to-current transformations from either bipolar transistors or subthreshold MOS transistors. The subthreshold floating-gate MOS transistor naturally implements the exponential-of-a-weighted-sum operation in a single device. I will present experimental results from several of these translinear network circuits breadboarded from subthreshold floating-gate MOS transistors. I will also describe and present experimental data from a variety of other implementations of the multiple-input translinear element.

CONTENTS

PROLOG: WHY ANALOG? viii

CHAPTER 1: THE TRANSLINEAR PRINCIPLE 1

1.1. The Emergence of Translinear Circuits 1
1.2. The Translinear Principle 3
1.3. References 7
Figures 14

CHAPTER 2: TRANSLINEAR CIRCUITS COMPRISING MULTIPLE-INPUT TRANSLINEAR ELEMENTS 17

2.1. The Multiple-Input Translinear Element 17
2.2. Three Configurations of the Multiple-Input Translinear Element 19
2.3. Basic Current-In, Current-Out Circuits 20
2.4. Prior Work 23
2.5. Matrix Analysis of Multiple-Input Translinear Element Networks 27
2.6. Stability of Multiple-Input Translinear Element Networks 30
2.7. Relationship to Translinear Loop Circuits 34
2.8. Appendix 2.A 37
2.9. Appendix 2.B 38
2.10. Appendix 2.C 40
2.11. References 46
Figures 47

CHAPTER 3: SIGNAL-FLOW–GRAPH ANALYSIS OF MULTIPLE-INPUT TRANSLINEAR ELEMENT NETWORKS 61

3.1. Signal-Flow Graphs 61
3.2. Signal-Flow–Graph Analysis of Multiple-Input Translinear Element Networks 66
3.3. Reduced Signal-Flow Graph Construction 72
3.4. Examples of Construction and Analysis of Reduced Signal-Flow Graphs 72
3.4.1. Single-Branch Paths 73
3.4.2. Two-Branch Paths 73
3.4.3. Parallel Paths 74
3.4.4. Reciprocal Connections 75
3.5. References 76
Figures 77

CHAPTER 4: SYNTHESIS OF MULTIPLE-INPUT TRANSLINEAR ELEMENT NETWORKS 87

4.1. The Scope of the Synthesis Problem 87
4.2. The Myriad Solutions to the Synthesis Problem 88
4.3. Synthesis of Multiple-Input Translinear Element Networks 98
4.3.1. Construction of a Two-Layer Network 98
4.3.2. Construction of a Cascade Network 100
4.3.3. An Illustrative Example 101
4.4. The Synthesis Problem as a Integer Linear Program 103
4.5. Appendix 4.A 107
4.6. Appendix 4.B 110
4.7. Appendix 4.C 115
4.8. References 119
Figures 120

CHAPTER 5: THE SUBTHRESHOLD FLOATING-GATE MOS TRANSISTOR 127

5.1. Floating-Gate MOS Transistors: They're Not Just for Storing Information Anymore 127
5.2. Subthreshold Floating-Gate MOS Transistor Model 129
5.3. Measurements from a Subthreshold Floating-Gate MOS Transistor 132
5.4. The Matching of Small Capacitors for Analog VLSI 136
5.4.1. Experimental Method to Assess Capacitor Mismatch 137
5.4.2. Experimental Results and Discussion 138
5.5. Appendix 5.A 140
5.6. References 142
Figures 146

CHAPTER 6: THE SUBTHRESHOLD FLOATING-GATE MOS TRANSISTOR AS A MULTIPLE-INPUT TRANSLINEAR ELEMENT 159

6.1. The Subthreshold Floating-Gate MOS Transistor as a Multiple-Input Translinear Element 159
6.2. Experimental Results 163
6.2.1. Two-Input Geometric-Mean Circuit 164
6.2.2. Squaring-Reciprocal Circuit 166
6.2.3. Power-Law Circuits 168
6.2.4. Product-Reciprocal Circuit 171
6.3. Appendix 6.A 173
6.4. References 175
Figures 176

CHAPTER 7: CONCLUSIONS AND FUTURE RESEARCH 195

7.1. Summary of Primary Contributions 195
7.2. Directions for Future Research 196
7.2.1. Other Multiple-Input Translinear Element Implementations 197
7.2.2. Continuous Floating-Gate Charge Adaptation 200
7.3. Appendix 7.A 201
7.4. References 202
Figures 204

PROLOG
WHY ANALOG?

Today, nearly all computation is done digitally. Because my research involves analog computation, people frequently ask me questions such as: Why bother with analog computation? Can't any given information-processing task be performed both faster and more precisely by a digital computer than by an analog one? Given no qualifications or constraints, it is likely that the answer to the latter question is yes. However, speed and precision have real costs in terms of both power consumption and circuit complexity. In many situations, it makes no sense for us to pay these costs; in some cases, we simply cannot afford to pay them.

Faster isn't necessarily better. For example, if a portable computer user doesn't notice any difference between running a word processor at 40 MHz and running it at 160 MHz, but the computer's battery will last four times longer at 40 MHz, then it makes sense to run the computer at 40 MHz, even though 160 MHz is possible. In many situations, we are forced by a limited power budget to ask ourselves: How much power can we afford to spend? Examples of such applications include portable computers, mobile communications devices, autonomous mobile robots, and prosthetic devices. In such cases, low-power, analog information-processing subsystems can be used to great advantage.

More precise isn't necessarily better either. The familiar computer-science adage, *garbage in, garbage out,* is true—it makes little sense to digitize signals that contain only a few bits of "meaningful" information to 16 bits of precision, and then to perform costly 32-bit operations on them. Rather, it pays to ask two questions: How much precision is needed for this application? How much precision is warranted by the incoming data? In both analog and digital information-processing systems, an increase in precision is purchased with an increase in power consumption, in silicon area, or in both. If a computation is executed with more precision than is necessary, resources are wasted—resources that could be either spent on another computation or saved.

At low to moderate precision, analog implementations of certain computations can be vastly more efficient—in terms of power dissipation, circuit area, or both—than equivalent digital implementations. In part, this efficiency is a direct result of working with

the medium, instead of against it—that is, of letting the silicon impose a framework on us, rather than imposing our preconceived framework on the silicon. In the design of digital information-processing systems, we usually decide *a priori* what the representation of information and what the primitive operations shall be. We don't care much about the detailed behavior of the transistors—we are content as long as we can use these devices as switches. From these switches, we construct primitive logic gates, such as AND, OR, and NOT gates. By combining these gates, we can compute addition, multiplication, and more complex operations. Building in this fashion, we must use many devices to implement complex operations. On the other hand, for an analog implementation, we examine the medium in which we are working, searching for physical quantities that efficiently represent information and device characteristics that are useful for information-processing tasks. From conservation laws, we obtain addition and subtraction. From capacitors, we obtain integration and differentiation. From capacitive and resistive dividers, we obtain weighted summation. From energy barriers and the Boltzmann distribution, we obtain exponentials and logarithms. From negative feedback around a high-gain amplifier, we obtain function inversion. By combining these primitives cleverly, we can often implement complex operations with just a few devices.

The class of circuits that is described in this thesis illustrates the efficiency of implementation that is possible if we utilize the physics of readily available devices to perform computations. Capacitors and MOS transistors are ubiquitous in today's technology; from these devices, we can make floating-gate MOS (FGMOS) transistors. People usually think of FGMOS transistors in the context of nonvolatile information storage; nonetheless, we can use these devices to process information as well. If multiple control gates capacitively couple into the floating gate of such a transistor, the floating-gate voltage is established as a weighted summation of the control-gate voltages via a capacitive divider. Because the channel current of a subthreshold MOS transistor is an exponential function of the gate voltage, the channel current of a subthreshold FGMOS transistor is an exponential function of a weighted sum of the control-gate voltages. Now, if we represent input signals by subthreshold currents, we can generate, with diode-connected FGMOS transistors, voltages that are logarithmic in the input currents. If we then apply these voltages to the control gates of other FGMOS transistors, we obtain output currents that are exponential functions of the weighted sum of logarithms of the input currents. Because of the familiar identities involving logarithms and exponentials, $x \log y = \log y^x$, $\log x + \log y = \log xy$, and $\exp(\log x) = x$, the output currents are thus products of powers of the input currents. Because these powers are set by capacitor ratios, we can obtain

accurate power-law relationships with these circuits. By capitalizing on the exponential current–voltage relationship of a subthreshold transistor and the weighted summation of a capacitive divider, we can implement many useful information-processing circuits using only a few capacitors and a single transistor for each input and output.

There is still a great deal of information-processing potential latent in the physics of silicon semiconductor devices. Rather than pitting analog against digital or using one to the exclusion of the other, we should determine under what circumstances analog implementations are more effective than digital ones, and vice versa. Equipped with this knowledge, we should use analog in those situations in which analog excels, and digital in those in which digital excels.

CHAPTER 1
THE TRANSLINEAR PRINCIPLE

In this chapter, I give a brief historical overview of the emergence of the class of translinear loop circuits, I discuss the two-fold meaning of the word *translinear*, and I provide a simple derivation of the translinear principle for a single loop comprising N idealized translinear elements.

1.1. The Emergence of Translinear Circuits

In the years immediately following the nearly simultaneous invention of the integrated circuit by Robert Noyce and Jack Kilby in the late 1950s [1], several of the ground rules by which the game of analog-circuit design was played began to change. For analog circuits built from discrete components, passive devices are typically more plentiful than are active components in circuit designs. Contrariwise, for analog integrated-circuit designs, active devices are ubiquitous, whereas passive components with suitable values are usually unavailable. Discrete components are in poor thermal contact with one another, whereas miniaturized devices that are integrated on the same chip are in intimate thermal contact with one another. Precisely matched discrete components are like precious jewels—costly and rare—whereas monolithic devices with reasonably well-matched characteristics are like common coins—minted together by the thousands under nearly identical conditions.

At that time, the most commonly held view of the bipolar transistor was that of a current-controlled current source. From this standpoint, the bipolar transistor is seen as a linear current amplifier whose most important property is the forward current gain, β. The exquisitely temperature-sensitive exponential current–voltage characteristic of these devices was shunned by analog-circuit designers as incidental; as of secondary importance; and as the source of a mildly irritating, temperature-dependent diode drop of approximately 700 mV, which, at times, appeared in the most inconvenient locations in their circuits.

In the mid-1960s, a complementary view of the bipolar transistor began to emerge. From this point of view, the bipolar transistor is seen as a voltage-controlled current source, whose most important property is that its transconductance is proportional to the current flowing through it [2–4], which is a direct consequence of its exponential

current–voltage relationship. From this standpoint, it is the bipolar's forward current gain that is of secondary importance, and the existence of a finite base current is often a troublesome source of errors in analog-circuit designs. The bipolar transistor's exponential current–voltage relationship is indeed highly temperature dependent, but if the base-emitter voltage of a bipolar transistor is set by another such device that is operating at the same temperature—as, for example, is the case in a simple current mirror—then overall circuit function is highly reliable and is temperature invariant.

From the mid-1960s to the early 1970s, a number of specific circuits and circuit techniques emerged—each of which was based on this view of the bipolar transistor—including amplifiers/multipliers [5, 7–15, 18], various components of operational amplifiers [6, 16], and various product-of-power-law circuits [8, 17]. Of particular note are the significant contributions made by Barrie Gilbert [7–9, 18]. Such circuits went by a variety of different names [19], including nonlinear circuits [20], function circuits [21], shaping circuits, processing circuits, and log–antilog circuits. However, none of these names (except perhaps for log–antilog) captured the essence of the principle underlying this important class of nonlinear circuits.

In 1975, Gilbert [19] coined the word **translinear** to describe these circuits, and succinctly enunciated a general circuit principle, the **translinear principle**, by which the steady-state characteristics of such circuits can be analyzed quickly. The word *translinear* derives from a contraction of one way of stating the exponential current–voltage property of the bipolar transistor that is central to the functioning of these circuits—that is, the bipolar transistor's *trans*conductance is *linear* in the current flowing through the transistor. The word was also meant to convey the notion of analysis and design techniques (e.g., the translinear principle) that bridge the gap between the well-established domain of linear-circuit design and the domain of nonlinear-circuit design, for which precious little can be said in general [2–4]. As I show in Section 1.2, the *translinear* principle is essentially a *trans*lation of a *linear* algebraic constraint on the voltages in a circuit into a product-of-power-law constraint on the currents flowing in the circuit.

Since the mid-1970s, the translinear principle has been the basis of a whole host of useful nonlinear circuits [22–61], including wideband analog multipliers with state-of-the-art precision [40], translinear current conveyors [46, 47], translinear frequency multipliers [28, 34, 35, 55], and operational current amplifiers [49, 51, 53, 57]. In 1979, Hart [62] extended the translinear principle to include voltage sources in translinear loops. Designers of translinear circuits sometimes use such voltage sources to compensate for scale-factor errors resulting from device mismatch [3, 4]. In the 1980s, Everet Seevinck [63, 64] made

significant contributions to the art of translinear-circuit design by developing systematic procedures for the analysis and synthesis of such circuits.

In his dissertation, Seevinck [63] noted that, in principle, we can build translinear circuits using subthreshold MOS transistors [65, 66] instead of bipolar transistors. Compared to bipolar transistors, subthreshold MOS transistors have limitations from the standpoint of translinear-circuit implementation [65, 68]. These limitations include a smaller range of exponential current–voltage behavior (i.e., four decades for a subthreshold MOS, versus eight or 10 decades for a bipolar), lower current levels (i.e., at most a few microamperes for a subthreshold MOS, versus milliamperes for a bipolar), lower achievable bandwidths (i.e., from 100 kHz to perhaps a few MHz for subthreshold MOS implementations, versus hundreds of MHz for bipolar implementations), and larger device mismatch (i.e., a few percent for subthreshold MOS transistors, versus one percent or less for bipolars).

Nonetheless, subthreshold MOS transistors have unique properties that allow us to implement certain translinear circuits with fewer transistors than we could using bipolars. Because the MOS transistor is symmetric with respect to its source and drain terminals, and because conduction through the channel is lossless, we can decompose the channel current of a subthreshold MOS transistor into a forward component and a reverse component, and can show that a subthreshold MOS transistor operating in the ohmic regime is, in a sense, two translinear devices instead of one [67, 68]. Because the MOS transistor is a four-terminal device, we can use the bulk connection (i.e., the back gate) to build novel intersecting translinear-loop structures [68–71]. Because the gate of the MOS transistor is completely insulated, we can make these transistors into floating-gate MOS (FGMOS) transistors. By capacitively coupling multiple inputs into the floating gates of these devices, we can, in a sense, generalize the use of the back gate, and build low-voltage translinear circuits comprising intersecting translinear loops [72]. In fact, exploring the possibility of using subthreshold FGMOS transistors to build translinear circuits is what led me to the material that I describe in Chapters 2 through 7 of this thesis.

1.2. The Translinear Principle

In this section, I provide a simple derivation of the translinear principle for a single loop of ideal **translinear elements** (**TEs**). A circuit symbol for an ideal TE is shown in Figure 1.1. Such a device has a current–voltage relationship given by

$$I = \lambda \mathrm{I_s} \exp\left[\frac{V}{\mathrm{U_T}}\right], \tag{1.1}$$

where $\mathrm{I_s}$ is a pre-exponential scaling current, which could be temperature dependent; λ is a dimensionless quantity that proportionally scales $\mathrm{I_s}$; and $\mathrm{U_T}$ is the thermal voltage, $\frac{\mathrm{kT}}{\mathrm{q}}$. We can think of this idealized device as a bipolar transistor with an infinite forward current gain. An ideal TE is well-approximated either by a bipolar transistor with a large forward current gain or by a saturated subthreshold MOS transistor with its source connected to its local substrate. For a bipolar transistor, the quantity λ corresponds to an emitter-area ratio. For a subthreshold MOS transistor, λ corresponds to a $\frac{W}{L}$ ratio.

To show that the ideal TE is *translinear* in the first sense of the word that I described in Section 1.1, I differentiate Equation 1.1 with respect to V as follows:

$$\begin{aligned} g_\mathrm{m} &= \frac{\partial I}{\partial V} \\ &= \lambda \mathrm{I_s} \exp\left[\frac{V}{\mathrm{U_T}}\right] \frac{1}{\mathrm{U_T}} \\ &= \frac{I}{\mathrm{U_T}}. \end{aligned}$$

Thus, the *trans*conductance of an ideal TE is *linear* in the current flowing through it.

Now, consider the closed loop of N ideal TEs that is shown in Figure 1.2. The large arrow shows the clockwise direction around the loop. If the arrow in the symbol of a TE points in the clockwise direction, we classify the TE as a **clockwise element**. Contrariwise, if the arrow in the symbol of a TE points in the counterclockwise direction, we classify the TE as a **counterclockwise element**. I denote by CW the set of clockwise element indices, and I denote by CCW the set of counterclockwise element indices.

Note that the voltage, V, across a counterclockwise element corresponds to a voltage increase as we proceed around the loop in the clockwise direction, and that the voltage across a clockwise element corresponds to a voltage drop as we proceed around the loop in the clockwise direction. One way of stating Kirchhoff's voltage law is that the sum of the voltage increases around a closed loop is equal to the sum of the voltage drops around the loop. Consequently, by applying Kirchhoff's voltage law around the loop of Figure 1.2, I have that

$$\sum_{n \in CCW} V_n = \sum_{n \in CW} V_n. \tag{1.2}$$

By solving Equation 1.1 for V in terms of I and substituting the resulting expression for

each V_n in Equation 1.2, I obtain

$$\sum_{n\in CCW} \mathrm{U_T} \log \frac{I_n}{\lambda_n \mathrm{I_s}} = \sum_{n\in CW} \mathrm{U_T} \log \frac{I_n}{\lambda_n \mathrm{I_s}}. \tag{1.3}$$

Assuming that all the TEs are operating at the same temperature, I can cancel the common factor of $\mathrm{U_T}$ in all the terms in Equation 1.3 to obtain

$$\sum_{n\in CCW} \log \frac{I_n}{\lambda_n \mathrm{I_s}} = \sum_{n\in CW} \log \frac{I_n}{\lambda_n \mathrm{I_s}}. \tag{1.4}$$

Because $\log x + \log y = \log xy$, I can rewrite Equation 1.4 as

$$\log \prod_{n\in CCW} \frac{I_n}{\lambda_n \mathrm{I_s}} = \log \prod_{n\in CW} \frac{I_n}{\lambda_n \mathrm{I_s}}. \tag{1.5}$$

By exponentiating both sides of Equation 1.5, I get

$$\prod_{n\in CCW} \frac{I_n}{\lambda_n \mathrm{I_s}} = \prod_{n\in CW} \frac{I_n}{\lambda_n \mathrm{I_s}},$$

which I can rearrange to obtain

$$\prod_{n\in CCW} \frac{I_n}{\lambda_n} = \mathrm{I_s}^{N_{\mathrm{CCW}} - N_{\mathrm{CW}}} \prod_{n\in CW} \frac{I_n}{\lambda_n}, \tag{1.6}$$

where N_{CCW} denotes the number of counterclockwise elements, and N_{CW} denotes the number of clockwise elements. Now, it is easy to see that, if $N_{\mathrm{CW}} = N_{\mathrm{CCW}}$, then Equation 1.6 reduces to

$$\prod_{n\in CCW} \frac{I_n}{\lambda_n} = \prod_{n\in CW} \frac{I_n}{\lambda_n}, \tag{1.7}$$

which has no remaining temperature dependence. Equation 1.7 is the **translinear principle**, which can be stated as follows.

> In a closed loop of ideal TEs comprising an equal number of clockwise elements and counterclockwise elements, the product of the current densities flowing through the counterclockwise elements is equal to the product of the current densities flowing through the clockwise elements.

If all TEs in the loop have identical geometry (i.e., each TE has the same value of λ), then Equation 1.7 becomes

$$\prod_{n\in CCW} I_n = \lambda^{N_{\mathrm{CCW}} - N_{\mathrm{CW}}} \prod_{n\in CW} I_n,$$

which, if $N_{\mathrm{CW}} = N_{\mathrm{CCW}}$, further reduces to

$$\prod_{n\in CCW} I_n = \prod_{n\in CW} I_n. \tag{1.8}$$

Equation 1.8 is an important special case of the translinear principle that can be stated as follows.

> In a closed loop of identical ideal TEs comprising an equal number of clockwise elements and counterclockwise elements, the product of the currents flowing through the counterclockwise elements is equal to the product of the currents flowing through the clockwise elements.

Note that the derivation of the translinear principle just described can be characterized as a *trans*lation of a *linear* algebraic constraint on the voltages in the circuit (i.e., Kirchhoff's voltage law applied around the loop of Figure 1.2) into a product-of-power-law constraint on the currents flowing in the circuit. This characterization of the translinear principle is one way to state the second connotation of the word *translinear* originally intended by Gilbert [2–4, 19].

To illustrate the use of the translinear principle, I now analyze the simple translinear circuit shown in Figure 1.3. This circuit has a single translinear loop comprising four identical TEs, two of which face in the counterclockwise direction and two of which face in the clockwise direction. Input current I_1 passes through both counterclockwise elements. Input current I_2 passes through one of the clockwise elements. Output current I_3 passes through the other clockwise element. Consequently, to analyze this circuit, I apply the translinear principle, as stated in Equation 1.8, and write that

$$I_1^2 = I_2 I_3,$$

which I rearrange to obtain

$$I_3 = \frac{I_1^2}{I_2}.$$

Thus, the circuit shown in Figure 1.3 is a squaring-reciprocal circuit.

1.3. References

1. E. Braun and S. Macdonald, *Revolution in Miniature: The History and Impact of Semiconductor Electronics*, 2nd ed., Cambridge, England: Cambridge University Press, 1992.

2. B. Gilbert, "Translinear Circuits—25 Years On, Part I: The Foundations," *Electronic Engineering*, vol. 65, no. 800, pp. 21–24, 1993.

3. B. Gilbert, "Current-Mode Circuits from a Translinear Viewpoint: A Tutorial," in C. Toumazou, F. J. Lidgey, and D. G. Haigh, eds., *Analogue IC Design: The Current-Mode Approach*, London: Peter Peregrinus, pp. 11–91, 1990.

4. B. Gilbert, "Translinear Circuits: An Historical Review," *Analog Integrated Circuits and Signal Processing*, vol. 9, no. 2, pp. 95–118, 1996.

5. W. L. Paterson, "Multiplication and Logarithmic Conversion by Operational Amplifier–Transistor Circuits," *Review of Scientific Instruments*, vol. 34, no. 12, pp. 1311–1316, 1963.

6. G. R. Wilson, "A Monolithic Junction FET–*n*-*p*-*n* Operational Amplifier," *IEEE Journal of Solid-State Circuits*, vol. SC-3, no. 4, pp. 341–348, 1968.

7. B. Gilbert, "A DC–500 MHz Amplifier/Multiplier Principle," in *Digest of Technical Papers of the 1968 International Solid-State Circuits Conference*, Philadelphia, PA, pp. 114–115, February 1968.

8. B. Gilbert, "A New Wide-Band Amplifier Technique," *IEEE Journal of Solid-State Circuits*, vol. SC-3, no. 4, pp. 353–365, 1968.

9. B. Gilbert, "A Precise Four-Quadrant Multiplier with Subnanosecond Response," *IEEE Journal of Solid-State Circuits*, vol. SC-3, no. 4, pp. 365–373, 1968.

10. H. Brüggemann, "Feedback Stabilized Four-Quadrant Analog Multiplier," *IEEE Journal of Solid-State Circuits*, vol. SC-5, no. 4, pp. 150–159, 1970.

11. E. W. Scratchley, “Single-Ended-Input Single-Ended-Output Four-Quadrant Analog Multiplier,” *IEEE Journal of Solid-State Circuits*, vol. SC-6, no. 6, pp. 394–395, 1971.

12. B. Gilbert, “Comment on ‘Single-Ended-Input Single-Ended-Output Four-Quadrant Analog Multiplier,’” *IEEE Journal of Solid-State Circuits*, vol. SC-7, no. 5, p. 434, 1972.

13. B. J. M. Overgoor, “Error Sources in Analog Multipliers,” *Electronics Applications Bulletin*, vol. 31, no. 3, pp. 187–204, 1972.

14. K. G. Schlotzhauer and T. R. Viswanathan, “New Bipolar Analogue Multiplier,” *Electronics Letters*, vol. 8, no. 16, pp. 425–427, 1972.

15. W. M. C. Sansen and R. G. Meyer, “Distortion in Bipolar Transistor Variable-Gain Amplifiers,” *IEEE Journal of Solid-State Circuits*, vol. SC-8, no. 4, pp. 275-282, 1973.

16. R. G. van Vliet, “Integrated Class-B End Stage,” *Electronics Letters*, vol. 10, no. 15, pp. 317–319, 1974.

17. R. W. J. Barker and B. L. Hart, “Root-Law Circuit Using Monolithic Bipolar-Transistor Arrays,” *Electronics Letters*, vol. 10, no. 21, pp. 439–440, 1974.

18. B. Gilbert, “A High-Performance Monolithic Multiplier Using Active Feedback,” *IEEE Journal of Solid-State Circuits*, vol. SC-9, no. 6, pp. 364–373, 1974.

19. B. Gilbert, “Translinear Circuits: A Proposed Classification,” *Electronics Letters*, vol. 11, no. 1, pp. 14–16, 1975; and errata, vol. 11, no. 6, p. 136, 1975.

20. D. H. Sheingold, ed., *Nonlinear Circuits Handbook: Designing with Analog Functional Modules and ICs*, Norwood, MA: Analog Devices, 1976.

21. Y. J. Wong and W. E. Ott, *Function Circuits: Design and Applications*, New York: McGraw-Hill, 1976.

22. B. Gilbert, “Wideband Negative-Current Mirror,” *Electronics Letters*, vol. 11, no. 6, pp. 126–127, 1975.

23. E. Traa, "An Integrated Function Generator with Two-Dimensional Electronic Programming Capability," *IEEE Journal of Solid-State Circuits*, vol. SC-10, no. 6, pp. 458–463, 1975.

24. B. Gilbert and L. W. Counts, "A Monolithic RMS-DC Convertor with Crest-Factor Compensation," in *Digest of Technical Papers of the 1976 International Solid-State Circuits Conference*, Philadelphia, PA, pp. 110–111, February 1976.

25. J. H. Huijsing and F. Tol, "Monolithic Operational Amplifier Design with Improved HF Behavior," *IEEE Journal of Solid-State Circuits*, vol. SC-11, no. 2, pp. 323–328, 1976.

26. S. Ashok, "Translinear Root-Difference-of-Squares Circuit," *Electronics Letters*, vol. 12, no. 8, pp. 194–195, 1976.

27. B. Gilbert, "High-Accuracy Vector-Difference and Vector-Sum Circuits," *Electronics Letters*, vol. 12, no. 11, pp. 293–294, 1976.

28. S. Ashok, "Integrable Sinusoidal Frequency Doubler," *IEEE Journal of Solid-State Circuits*, vol. SC-11, no. 2, pp. 341–343, 1976.

29. F. Doorenbosch and Y. Goinga, "Integrable, Wideband, Automatic Volume Control (A.V.C.) Using Pythagoras's Law for Amplitude Detection," *Electronics Letters*, vol. 12, no. 16, pp. 418-420, 1976.

30. P. A. McGovern, "Linearisation of Microwave Point-Contact-Detector Characteristics," *Electronics Letters*, vol. 12, no. 22, p. 585, 1976.

31. P. A. McGovern, "Tangent Sweep Circuit," *Electronics Letters*, vol. 12, no. 23, pp. 613–614, 1976.

32. W. J. Barker and B. L. Hart, "Negative Current-Mirror Using *n-p-n* Transistors," *Electronics Letters*, vol. 13, no. 11, pp. 311–312, 1977.

33. D. J. Hamilton and K. B. Finch, "A Single-Ended Current Gain Cell with AGC, Low Offset Voltage, and Large Dynamic Range," *IEEE Journal of Solid-State Circuits*, vol. SC-12, no. 3, pp. 322–323, 1977.

34. R. Genin and R. Konn, "Sinusoidal Frequency Doubler," *Electronics Letters*, vol. 15, no. 2, pp. 47–48, 1979.

35. R. Konn and R. Genin, "High-Performance Aperiodic Frequency Multiplying," *Electronics Letters*, vol. 15, no. 6, pp. 187–189, 1979.

36. B. Gilbert and P. Holloway, "A Wideband Two-Quadrant Analog Multiplier," in *Digest of Technical Papers of the 1980 International Solid-State Circuits Conference*, San Francisco, CA, pp. 200–201, February 1980.

37. E. Seevinck and G. H. Renkema, "A 4-Quadrant Cosine Synthesis Circuit," in *Digest of Technical Papers of the 1982 International Solid-State Circuits Conference*, San Francisco, CA, pp. 40–41, February 1982.

38. J. H. Huijsing, P. Lucas, and B. de Bruin, "Monolithic Analog Multiplier-Divider," *IEEE Journal of Solid-State Circuits*, vol. SC-17, no. 1, pp. 9–15, 1982.

39. C. K. Wong, R. F. Wassenaar, and E. Seevinck, "A Wideband Accurate Vector-Sum Circuit," in *Proceedings of the 1983 European Solid-State Circuits Conference*, pp. 135–138, 1983.

40. B. Gilbert, "A Four-Quadrant Analog Divider/Multiplier with 0.01% Distortion," in *Digest of Technical Papers of the 1983 International Solid-State Circuits Conference*, Philadelphia, PA, pp. 248–249, February 1983.

41. B. Gilbert, "An Analog Array Processor," in *Digest of Technical Papers of the 1984 International Solid-State Circuits Conference*, San Francisco, CA, pp. 286–287, February 1984.

42. E. Seevinck, R. F. Wassenaar, and C. K. Wong, "A Wide-Band Technique for Vector Summation and RMS–DC Conversion," *IEEE Journal of Solid-State Circuits*, vol. SC-19, no. 3, pp. 311–318, 1984.

43. B. Gilbert, "A Monolithic 16-Channel Analog Array Normalizer," *IEEE Journal of Solid-State Circuits*, vol. SC-19, no. 6, pp. 956–963, 1984.

44. A. Fabre and P. Rochegude, "Ultra-Low-Distortion Current-Conversion Technique," *Electronics Letters*, vol. 20, no. 17, pp. 675–676, 1984.

45. A. Fabre, "An Integrable Multiple Output Translinear Current Convertor," *International Journal of Electronics*, vol. 57, no. 5, pp. 713–717, 1984.

46. A. Fabre, "Translinear Current Conveyor Implementation," *International Journal of Electronics*, vol. 59, no. 5, pp. 619–623, 1985.

47. G. Normand, "Translinear Current Conveyors," *International Journal of Electronics*, vol. 59, no. 6, pp. 771–777, 1985.

48. W. Surakampontorn, "Integrable Wide-Dynamic-Range Negative Resistance Circuits," *Electronics Letters*, vol. 21, no. 11, pp. 506–508, 1985.

49. A. Fabre, "The Translinear Operational Current Amplifier: A New Building Block," *International Journal of Electronics*, vol. 60, no. 2, pp. 275–279, 1986.

50. G. Normand, "Current-Controlled Translinear IPIC," *International Journal of Electronics*, vol. 61, no. 2, pp. 225–231, 1986.

51. A. Fabre and P. Rochegude, "Current Processing Circuits with Translinear Operational Current Amplifiers," *International Journal of Electronics*, vol. 63, no. 1, pp. 9–28, 1987.

52. E. Seevinck, W. de Jager, and P. Buitendijk, "A Low-Distortion Output Stage with Improved Stability for Monolithic Power Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 3, pp. 794–801, 1988.

53. A. Fabre, "Translinear Current-Controlled Current-Amplifier," *Electronics Letters*, vol. 24, no. 9, pp. 548–549, 1988.

54. A. Fabre and J. P. Longuemard, "A Translinear Floating Current-Source with Current-Control," *International Journal of Electronics*, vol. 65, no. 6, pp. 1137–1142, 1988.

55. W. Surakampontorn, S. Jutaviriya, and T. Apajinda, "Dual Translinear Sinusoidal Frequency Doubler and Full-Wave Rectifier," *International Journal of Electronics*, vol. 65, no. 6, pp. 1203–1208, 1988.

56. R. F. Wassenaar, E. Seevinck, M. G. van Leeuwen, C. J. Speelman, and E. Holle, "New Techniques for High-Frequency RMS-to-DC Conversion Based on a Multifunctional *V*-to-*I* Convertor," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 3, pp. 802–815, 1988.

57. C. Toumazou, F. J. Lidgey, and M. Yang, "Translinear Class-AB Current-Amplifier," *Electronics Letters*, vol. 25, no. 13, pp. 873–874, 1989.

58. A. Fabre, P. Siarry, and M. Lameche, "Current-Controlled Translinear Impedance Converter," *International Journal of Electronics*, vol. 70, no. 4, pp. 795–801, 1991.

59. W. Surakampontorn, Y. Chonbodeechalermroong, and S. Bunjongjit, "An Analog Sinusoidal Frequency-to-Voltage Converter," *IEEE Transactions on Instrumentation and Measurement*, vol. 40, no. 6, pp. 925–929, 1991.

60. A. Fabre, "Bidirectional Current-Controlled PTAT Current Source," *IEEE Transactions on Circuits and Systems I*, vol. 41, no. 12, pp. 922–925, 1994.

61. O. Saaid and A. Fabre, "Class AB Current-Controlled Resistor for High-Performance Current-Mode Applications," *Electronics Letters*, vol. 32, no. 1, pp. 4–5, 1996.

62. B. L. Hart, "Translinear Circuit Principle: A Reformulation," *Electronics Letters*, vol. 15, no. 24, pp. 801–803, 1979.

63. E. Seevinck, *Analysis and Synthesis of Translinear Integrated Circuits*, D.Sc. Dissertation, University of Pretoria, Pretoria, South Africa, May 1981.

64. E. Seevinck, *Analysis and Synthesis of Translinear Integrated Circuits*, Amsterdam: Elsevier, 1988.

65. E. Vittoz and J. Fellrath, "CMOS Analog Integrated Circuits Based on Weak-Inversion Operation," *IEEE Journal of Solid-State Circuits*, vol. SC-12, no. 3, pp. 224–231, 1977.

66. C. Mead, *Analog VLSI and Neural Systems*, Reading, MA: Addison-Wesley, 1989.

67. A. G. Andreou and K. A. Boahen, "Neural Information Processing II," in M. Ismail and T. Fiez, eds., *Analog VLSI Signal and Information Processing*, New York: McGraw-Hill, pp. 358–413, 1994.

68. A. G. Andreou and K. A. Boahen, "Translinear Circuits in Subthreshold MOS," *Analog Integrated Circuits and Signal Processing*, vol. 9, no. 2, pp. 141–166, 1996.

69. M. van der Gevel and J. C. Kuenen, "$\sqrt{X}$ Circuit Based on a Novel, Back-Gate-Using Multiplier," *Electronics Letters*, vol. 30, no. 3, pp. 183–184, 1994.

70. J. Mulder, A. C. van der Woerd, W. A. Serdijn, and A. H. M. van Roermund, "Application of the Back Gate in MOS Weak Inversion Translinear Circuits," *IEEE Transactions on Circuits and Systems I*, vol. 42, no. 11, pp. 958–962, 1995.

71. R. Fried and C. C. Enz, "CMOS Parametric Current Amplifier," *Electronics Letters*, vol. 32, no. 14, pp. 1249–1250, 1996.

72. B. A. Minch, C. Diorio, P. Hasler, and C. A. Mead, "Translinear Circuits Using Subthreshold Floating-Gate MOS Transistors," *Analog Integrated Circuits and Signal Processing*, vol. 9, no. 2, pp. 167–180, 1996.

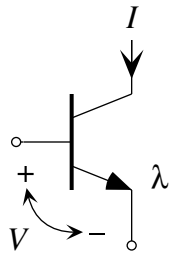


Figure 1.1. Circuit symbol for an ideal translinear element (TE). The current, I, flowing through the element is exponential in the voltage V. The quantity, λ, shown in the symbol next to the emitter, is a dimensionless number that scales the current, I, proportionally. A bipolar transistor with a very large forward current gain is one possible implementation of an ideal TE; in this case, λ corresponds to an emitter-area ratio. A subthreshold MOS transistor with its local substrate connected to its source is another possible implementation of the ideal TE; in this case, λ corresponds to a $\frac{W}{L}$ ratio.

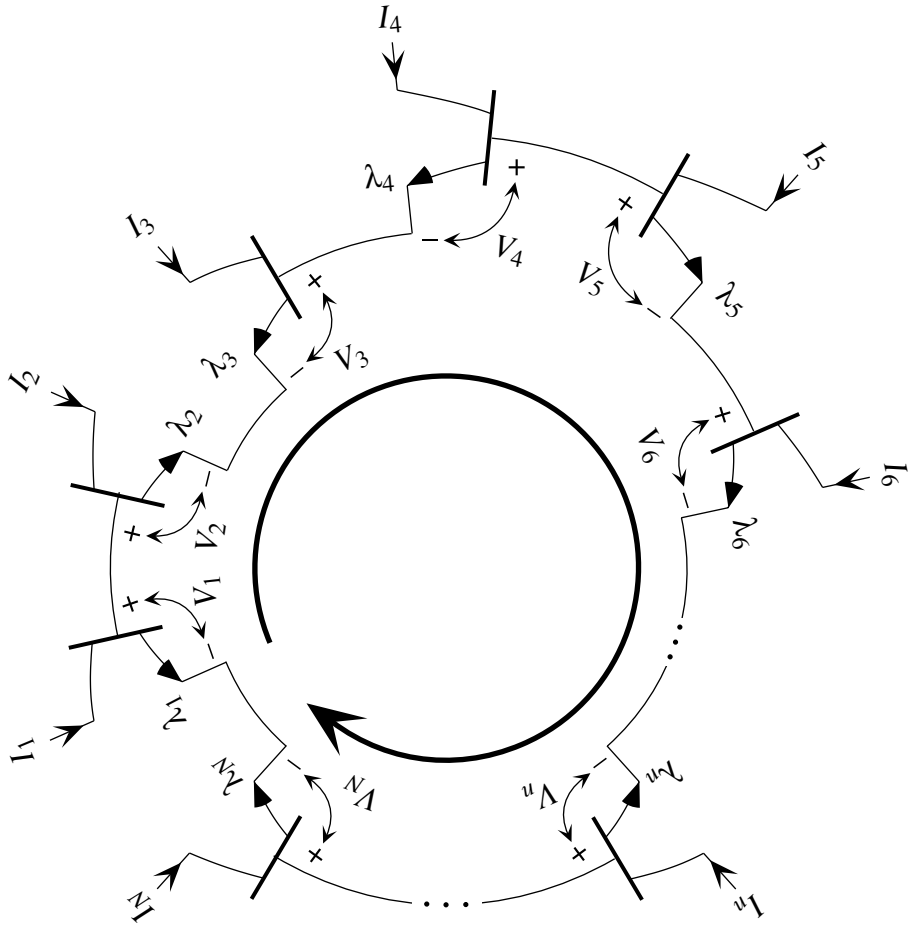


Figure 1.2. A conceptual translinear loop comprising *N* ideal TEs. The large arrow shows the direction around the loop according to which we judge whether each element is a *clockwise element* or a *counterclockwise element*. If a TE symbol's arrow points in the direction opposite to that of the large arrow, we consider the element a counterclockwise element. If a TE symbol's arrow points in the same direction as the large arrow, then the element is a clockwise element. The translinear principle states that, if the number of counterclockwise elements is equal to the number of clockwise elements, then the product of the currents flowing through the counterclockwise elements is equal to the product of the currents flowing through the clockwise elements.

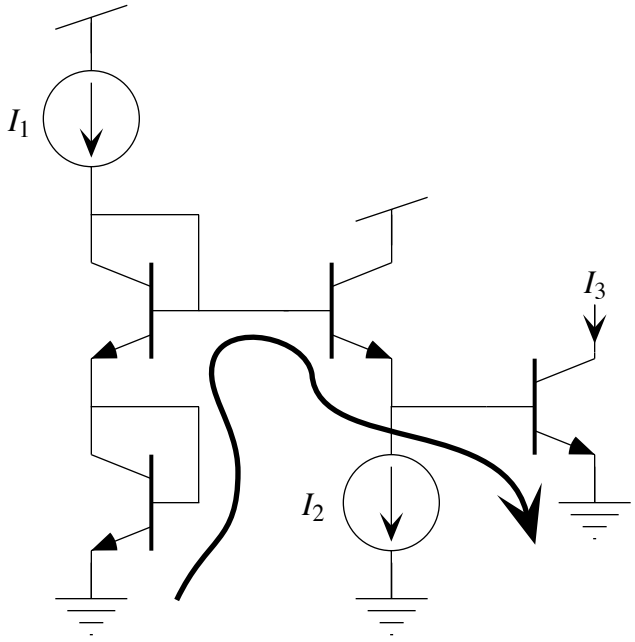


Figure 1.3. A simple translinear-loop circuit comprising four identical TEs, two of which face in the counterclockwise direction and two of which face in the clockwise direction.

CHAPTER 2
TRANSLINEAR CIRCUITS COMPRISING MULTIPLE-INPUT TRANSLINEAR ELEMENTS

In this chapter, I describe a class of translinear circuits made from **multiple-input translinear elements** (**MITEs**). The output currents of circuits from this class are given by products of powers of the input currents; I refer to these circuits both as **MITE translinear circuits** and as **MITE networks**. A MITE is an element that has K different *trans*conductances, g_1 through g_K, each of which is *linear* in the current flowing through the element; hence, the MITE is *translinear* in the first sense of the word that I discussed in Chapter 1. After describing the MITE in Section 2.1, I develop an intuition for the functioning of MITE translinear circuits in Sections 2.2 and 2.3 by considering various combinations of three basic circuit stages made from these MITEs. Then, in Section 2.5, I analyze the steady-state behavior of the general case of such a circuit systematically using matrices. In Section 2.6, I consider the stability of the general case of a MITE translinear circuit. Finally, in Section 2.7, I discuss the relationship of the class of MITE networks to the class of translinear loop circuits. In Chapter 3, I return to the intuitive analysis with which I begin this chapter, and develop a rigorous by-inspection analysis procedure based on the theory of linear signal-flow graphs.

2.1. The Multiple-Input Translinear Element

A circuit symbol for an ideal MITE is depicted in Figure 2.1; such an element sums K input voltages, V_1 through V_K, scaled, respectively, by dimensionless positive coefficients, w_1 through w_K. It then generates a current that is exponential in this weighted sum. I assume that we, as designers, have the ability to control the values of the weighting coefficients proportionally, so we can make accurate ratios of weighting coefficients.

The bipolar transistor shown in Figure 2.1 represents an ideal device with an exponential current–voltage characteristic such that the exponential current, I, flows into a terminal different from the one to which the controlling voltage, U, is applied. This device does not need to be a bipolar transistor. A diode would not be appropriate in this context,

but a subthreshold MOSFET would be a suitable replacement for the bipolar. We can implement the weighted summation operation in a purely passive manner using either a resistive or a capacitive voltage divider. The triangular amplifier symbols, which represent the weighting operation in Figure 2.1, convey two different notions. First, they denote that an input voltage, V_k, is scaled by a constant gain whose value is given by a nearby w_k. Second, they suggest that the input terminals should draw a negligible amount of DC current; hence, if we used a resistive divider to implement the weighted summation, then we would need to buffer the input voltages into the resistive network. In practice, we obtain the most accurate ratios of weighing coefficients by connecting an integral number of unit cells, each with weight *w*, in parallel. In such cases, we are interested primarily in the number of cells rather than the actual weight value involved; consequently, to prevent unnecessary clutter in circuit schematics, I omit the *w* associated with each of these amplifier symbols.

Without loss of generality, I assume that the weighted summation of the input voltages, *U*, is of the form

$$U = \sum_{k=1}^{K} w_k V_k,$$

where V_k is the *k*th input voltage, and w_k is a dimensionless positive weighting coefficient that scales V_k. Further, I assume that the output current, *I*, is of the form

$$I = \lambda \mathrm{I_s} \exp\left[\sum_{k=1}^{K} \frac{w_k V_k}{\mathrm{U_T}}\right], \tag{2.1}$$

where $\mathrm{I_s}$ is a pre-exponential scaling current, which could be temperature dependent, λ is a dimensionless quantity that proportionally scales $\mathrm{I_s}$, and $\mathrm{U_T}$ is the thermal voltage, $\frac{\mathrm{kT}}{\mathrm{q}}$. Note that, if the weighted summation, *U*, included a voltage offset term, U_{offset}, the form of the output current would remain unchanged; there would simply be a new scaling current, $\mathrm{I_s'}$, given by

$$\mathrm{I_s'} = \mathrm{I_s} \exp\left[\frac{U_{\mathrm{offset}}}{\mathrm{U_T}}\right].$$

This MITE does indeed have *K* different transconductances, g_1 through g_K, each of which is linear in the output current, *I*. To demonstrate this property, I simply differentiate Equation 2.1 with respect to V_k as follows:

$$\begin{aligned} g_k &= \frac{\partial I}{\partial V_k} \\ &= \lambda \mathrm{I_s} \exp\left[\sum_{k=1}^{K} \frac{w_k V_k}{\mathrm{U_T}}\right] \frac{w_k}{\mathrm{U_T}} \end{aligned}$$

$$= \frac{w_k I}{\mathrm{U_T}}.$$

2.2. Three Configurations of the Multiple-Input Translinear Element

In this section, I describe the three basic circuit stages comprising a single MITE that are shown in Figure 2.2. These three MITE stages are composed in various basic ways to make up all the MITE translinear circuits for which we would like to gain intuition.

The first of the three basic MITE stages is a **voltage-in, current-out** (**VICO**) stage, shown in Figure 2.2a. Here, I apply input voltages, V_i and V_k, to two different input terminals of the MITE, Q_n, which, in response, generates an output current, I_n. To see how I_n depends on V_i and V_k, using Equation 2.1, I write

$$I_n \propto \exp\left[\frac{w_{ni}V_i + w_{nk}V_k + \ldots}{\mathrm{U_T}}\right].$$

By breaking out the first two terms of the weighted summation and using the fact that $\exp(x+y) = \exp(x)\exp(y)$, I rewrite the preceding expression as

$$I_n \propto \exp\left[\frac{w_{ni}V_i}{\mathrm{U_T}}\right]\exp\left[\frac{w_{nk}V_k}{\mathrm{U_T}}\right]. \tag{2.2}$$

The second of the three basic MITE stages is effectively the inverse of the first: a **current-in, voltage-out** (**CIVO**) stage (Figure 2.2b). Here, I source a current, I_i, into the output of MITE, Q_i, and feed the output voltage, V_i, back through the self-coupling coefficient, w_{ii}. This feedback configuration adjusts V_i, so that the current sunk by Q_i just balances the input current, I_i. This feedback configuration is analogous to a diode connection of either a bipolar transistor or an MOS transistor, so I say that a MITE in the configuration of Figure 2.2b is **diode connected** through w_{ii}. To determine how the output voltage, V_i, depends on the input current, I_i, I begin with Equation 2.1, and solve for V_i in terms of I_i. So, I write

$$I_i \propto \exp\left[\frac{w_{ii}V_i + \ldots}{\mathrm{U_T}}\right],$$

which I rearrange to find that

$$V_i = \frac{\mathrm{U_T}}{w_{ii}}\log I_i - \ldots. \tag{2.3}$$

Thus, the output voltage, V_i, is equal to the logarithm of the input current, I_i, inversely scaled by the self-coupling weight, w_{ii}, minus a number of other terms, which depend neither on V_i nor on I_i.

The third basic MITE stage is a **voltage-in, voltage-out** (**VIVO**) configuration, shown in Figure 2.2c. This configuration is identical to the CIVO stage of Figure 2.2b, except that I now hold the current, I_i, fixed, and I am instead concerned with how the output voltage, V_i, depends on an input voltage, V_j, which I apply to another of the input terminals of the MITE Q_i. Beginning with Equation 2.1, I write that

$$I_i \propto \exp\left[\frac{w_{ii}V_i + w_{ij}V_j + ...}{U_T}\right],$$

which I rearrange to solve for V_i in terms of V_j as follows:

$$V_i = -\frac{w_{ij}}{w_{ii}}V_j - .... \tag{2.4}$$

So, the output voltage is an inverted version of the input voltage scaled by a gain that depends on the ratio of the feedforward weighting coefficient, w_{ij}, to the feedback weighting coefficient, w_{ii}. This result is not particularly surprising, because the current source, I_i, and the transistor, Q_i, together can be thought of as a high-gain inverting voltage amplifier around which I am placing negative feedback through the weighing coefficients, w_{ij} and w_{ii}. The closed-loop gain of this well-known inverting-amplifier configuration is $-\frac{w_{ij}}{w_{ii}}$, which is precisely what I obtained in Equation 2.4.

Occasionally, I will be interested in how the output voltage of the configuration of Figure 2.2c depends both on the input voltage, V_j, and on the input current, I_i. In this case, it is easy to see that I have a combination of Equations 2.3 and 2.4, as follows:

$$V_i = \frac{U_T}{w_{ii}}\log I_i - \frac{w_{ij}}{w_{ii}}V_j - .... \tag{2.5}$$

2.3. Basic Current-In, Current-Out Circuits

In this section, I describe three simple **current-in, current-out** (**CICO**) circuits made from combinations of the three basic MITE stages. The simplest CICO circuit that I can make from the three stages of Figure 2.2 consists of a single CIVO stage connected to one input of a VICO stage as shown in Figure 2.3. I would like to determine how the output current, I_n, depends on the input current, I_i. Beginning with Equation 2.1 applied to the output stage, I write that

$$I_n \propto \exp\left[\frac{w_{ni}V_i + ...}{U_T}\right]. \tag{2.6}$$

Substituting Equation 2.3 into Equation 2.6, I find that

$$I_n \propto \exp\left[\frac{w_{ni}}{U_T}\left(\frac{U_T}{w_{ii}}\log I_i - ...\right) + ...\right].$$

When I break out the first term from the extended summation and regroup, I get

$$I_n \propto \exp\left[\frac{\mathrm{U_T}}{\mathrm{U_T}}\frac{w_{ni}}{w_{ii}}\log I_i\right]. \tag{2.7}$$

Note that, if the transistors Q_i and Q_n are operating at the same temperature, then the primary temperature dependence of the relationship between I_i and I_n disappears from Equation 2.7. In this analysis, I have not kept track of the scaling currents, I_s, which can be strongly temperature dependent, but, as I show in Section 2.5, if the product of the input currents raised to their respective powers has units of amperes (i.e., as opposed to amperes raised to a power other than unity), then the relationship between the output current and the input currents is generally insensitive to isothermal variations. Now, because $x\log y = \log y^x$ and $\exp(\log x) = x$, I can rewrite Equation 2.7 as

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}}.$$

Thus, the output current is proportional to the input current raised to a power that is set by a ratio of weighting coefficients, which I assumed that we, as designers, could control accurately.

Now, I can add a second CIVO stage to the circuit of Figure 2.3, and connect its output voltage to the output VICO stage through a second input terminal to arrive at the circuit of Figure 2.4. I would like to determine how the output current, I_n, depends on each of the input currents, I_i and I_k. I begin with Equation 2.2 applied to the output stage as follows:

$$I_n \propto \exp\left[\frac{w_{ni}V_i}{\mathrm{U_T}}\right]\exp\left[\frac{w_{nk}V_k}{\mathrm{U_T}}\right]. \tag{2.8}$$

Substituting Equation 2.3 into Equation 2.8 for each of V_i and V_k, I obtain

$$I_n \propto \exp\left[\frac{w_{ni}}{\mathrm{U_T}}\left(\frac{\mathrm{U_T}}{w_{ii}}\log I_i - ...\right)\right]\exp\left[\frac{w_{nk}}{\mathrm{U_T}}\left(\frac{\mathrm{U_T}}{w_{kk}}\log I_k - ...\right)\right].$$

When I break out the first term in each of the two summations and regroup, this equation becomes

$$I_n \propto \exp\left[\frac{\mathrm{U_T}}{\mathrm{U_T}}\frac{w_{ni}}{w_{ii}}\log I_i\right]\exp\left[\frac{\mathrm{U_T}}{\mathrm{U_T}}\frac{w_{nk}}{w_{kk}}\log I_k\right]. \tag{2.9}$$

Again, because $x\log y = \log y^x$ and $\exp(\log x) = x$, I can rewrite Equation 2.9 as

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}} \times I_k^{\frac{w_{nk}}{w_{kk}}}.$$

Thus, the output current is proportional to the product of the two input currents, each raised to a power that is set by a ratio of weighting coefficients and is completely independent of

the other power. This product-of-powers relationship is also insensitive to isothermal variations.

There is a third basic way in which I can combine the stages of Figure 2.2. Instead of connecting the output of the second CIVO stage directly to a second input of the output VICO stage, as I did in the circuit of Figure 2.4, I connect the output of the second CIVO stage to the output stage through the first CIVO stage, as shown in Figure 2.5. This first CIVO stage both generates a voltage that is logarithmic in the input current, I_i, and serves as a VIVO stage for the second CIVO stage. This connection will allow us, as designers, to obtain negative powers. To show that it will, I again begin with Equation 2.1 applied to the output stage as follows:

$$I_n \propto \exp\left[\frac{w_{ni}V_i + ...}{\mathrm{U_T}}\right]. \tag{2.10}$$

Substituting Equation 2.5 into Equation 2.10, I get

$$I_n \propto \exp\left[\frac{w_{ni}}{\mathrm{U_T}}\left(\frac{\mathrm{U_T}}{w_{ii}}\log I_i - \frac{w_{ij}}{w_{ii}}V_j - ...\right)\right],$$

into which I substitute Equation 2.3 for V_j, and thus obtain

$$I_n \propto \exp\left[\frac{w_{ni}}{\mathrm{U_T}}\left(\frac{\mathrm{U_T}}{w_{ii}}\log I_i - \frac{w_{ij}}{w_{ii}}\left(\frac{\mathrm{U_T}}{w_{jj}}\log I_j - ...\right) - ...\right)\right].$$

Now, if I break out the first two terms of the summation and regroup, I find that

$$I_n \propto \exp\left[\frac{\mathrm{U_T}}{\mathrm{U_T}}\frac{w_{ni}}{w_{ii}}\log I_i\right]\exp\left[-\frac{\mathrm{U_T}}{\mathrm{U_T}}\frac{w_{ni}}{w_{ii}}\frac{w_{ij}}{w_{jj}}\log I_j\right]. \tag{2.11}$$

Again, because $x\log y = \log y^x$ and $\exp(\log x) = x$, I can express Equation 2.11 as

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}} \times I_j^{-\frac{w_{ni}}{w_{ii}}\frac{w_{ij}}{w_{jj}}}, \tag{2.12}$$

which, in turn, I can write as

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}} \div I_j^{\frac{w_{ni}}{w_{ii}}\frac{w_{ij}}{w_{jj}}}.$$

Thus, the output current is proportional to the quotient of the two input currents, each raised to a power that is set by ratios of weighting coefficients. Here, the powers are not completely independent of each other; for any value of $\frac{w_{ni}}{w_{ii}}$, however, as designers, we can adjust the value of $\frac{w_{ij}}{w_{jj}}$ to set the power of I_j as desired. This quotient-of-powers relationship is also insensitive to isothermal variations.

These three basic CICO circuits capture the intuition behind all MITE translinear circuits. We generate voltages that are logarithmic in the input currents using diode-connected MITEs. We set power laws through ratios of weighting coefficients. We obtain

negative powers by using voltage-inversion stages. We get products by summing two or more logarithmic voltages on MITEs. I formalize this intuition in Chapter 3 through the theory of linear signal-flow graphs; I thereby obtain a by-inspection analysis procedure that we can apply to determine with what power any input current factors into any output current.

2.4. Prior Work

The basic principle on which MITE translinear circuits operate is not a new one; it dates back at least to the invention of the logarithmic slide rule in the mid-seventeenth century. For nearly three centuries, scientists and engineers used the slide rule to multiply numbers, to divide numbers, to raise numbers to various powers, and to extract various roots of numbers. With these devices, scientists and engineers could perform these relatively complex operations much more rapidly than ever before. On a typical slide rule, there are two fixed logarithmic scales, A and D, with scale units in a ratio of 1 to 2. There are also two sliding logarithmic scales, B and C, with the same scale units, respectively, as the A and D scales. As illustrated in Figure 2.6, by moving the sliding scales relative to the fixed scales, thereby adding or subtracting logarithmic distances, and by employing scales with scale units in a ratio of 1 to 2, we can multiply, divide, square, square root, and perform a variety of compound operations. In Figure 2.6, the quantity d_s is the scale unit of the A and B scales; if natural logarithms are used, then d_s is the distance from the origin to the position of *e* on the A and B scales.

We can multiply two numbers, *x* and *y*, on a slide rule as shown in Figure 2.6a. We make use of two scales, A and B, with the same scale units, d_s, that can be moved relative to each other. We begin on the A scale by locating where *x* is marked; this point will be located a distance $d_s \log x$ from the origin of the A scale. We then align the origin of the B scale with the position of *x* on the A scale. Next, we find the location of *y* on the B scale, which will be a distance $d_s \log y$ from the origin of the B scale. At this point, we have moved a distance on the A scale equal to $d_s \log x + d_s \log y$, which, because $\log x + \log y = \log xy$, is equal to $d_s \log xy$. Thus, the reading on the A scale at the location corresponding to *y* on the B scale will be the product *xy*. We can obtain the inverse operation of division by performing these steps in reverse and thus subtracting two distances that are logarithmic in the readings of interest on the A and B scales.

We can square a number, *y*, on a slide rule as illustrated in Figure 2.6b. In this case, we make use of two scales, A and D, with scale units in a ratio of 1 to 2. We begin

on the D scale by finding the location corresponding to the value of *y*; this point will be a distance of $2d_s \log y$ from the origin of the D scale. We then read off the value at the corresponding location on the A scale, a distance $d_s \log x$ from the origin of the A scale. Because $x \log y = \log y^x$, the value of *x* there will be equal to y^2. Again, we can obtain the inverse operation of square rooting by performing these steps in reverse, starting with the A scale and moving to the D scale.

We can also perform compound operations, such as $x\sqrt{y}$, on a slide rule, as shown in Figure 2.6c. In this case we make use of two scales, B and D, with scale units in a ratio of 1 to 2, that can be moved relative to each other. We begin on the D scale by locating the point where *x* is marked, a distance of $2d_s \log x$ from the origin of the D scale. We then align the origin of the B scale with the position of *x* on the D scale. Next, we find the location of *y* on the B scale. At this point, we have moved a distance on the D scale equal to $2d_s \log x + d_s \log y$ which is equal to $2d_s \log x + 2d_s \log \sqrt{y}$, which, in turn, is equal to $2d_s \log x\sqrt{y}$. Thus, the reading on the D scale at the location corresponding to the location of *y* on the B scale will be the value of $x\sqrt{y}$. Depending on with which scale we begin and whether we add or subtract the distances involved, we can obtain a number of similar compound operations, including

$$\frac{x}{\sqrt{y}}, \quad \frac{\sqrt{x}}{y}, \quad xy^2, \quad \frac{x}{y^2}, \text{ and } \quad \frac{x^2}{y}.$$

Thus, there is a close analogy between the way that slide rules work and the way that MITE translinear circuits work. Voltages in a MITE translinear circuit are analogous to distances in a slide rule. Currents in a MITE translinear circuit are analogous to readings on the various scales of a slide rule. In both cases, we obtain products by adding logarithmic quantities, and we set powers through ratios of scale units. In fact, we could aptly describe MITE translinear circuits as microelectronic slide rules.

The idea of building electronic circuits that function along the same lines as the slide rule is not a new one either. In the *Nonlinear Circuits Handbook* from Analog Devices, which was published in the mid-1970s, we find the following clear description of the general principle involved:

> When compound multiplications, involving roots and powers are performed (e.g., $x_1^{\alpha} \cdot x_2^{\beta} \cdot x_3^{\gamma} \cdot x_4^{\delta} \cdot ...$), each input is "logged," multiplied by a constant (or variable) exponent of appropriate magnitude and polarity, the terms are summed and/or differenced, then the antilog is taken to convert the result back into the "world of phenomena" [1, p. 469].

A few power-law circuits that function according to this general principle have been described in the literature since the early 1980s [2–5]. Vittoz [5] cites Arreguit and his associates [3], and indicates that such circuits are based on a "generalization of the translinear principle" [5, p. 37]. Arreguit and his associates, in turn, cite the *Nonlinear Circuits Handbook* [1], and mention that, in analyzing such circuits, they can "apply the generalized translinear principle that translates the sum of voltages into a product of currents and their multiplication by a constant *k* into the elevation of the currents to the power *k*" [3, p. 443]. It seems likely that Arreguit and his colleagues are referring to the lines just quoted from the *Nonlinear Circuits Handbook* [1]. Despite these claims, these power-law circuits seem to have been conceived as a collection of special forms: one for powers between zero and unity, one for powers greater than unity, and one for negative powers. From all appearances, there has not heretofore been a systematic study of this important class of circuits. In Section 2.6 and the chapters to follow, I provide two different systematic analysis procedures for obtaining the steady-state behavior of such circuits, I discuss procedures for synthesizing such circuits systematically, and I both describe and experimentally demonstrate several novel implementations of the MITE based on floating-gate MOS transistors.

Nauta [2], Arreguit [3, 4], and Vittoz [5] have realized MITEs with resistive voltage dividers implementing the weighted voltage summation and with bipolar transistors implementing the exponential voltage-to-current transduction. Such an implementation of the MITE is valid as long as the collector current levels are such that the base resistance of the bipolars is much higher than the level of resistance used in the resistive voltage divider. It is easy to show that this condition corresponds to the condition

$$I_c << \frac{\beta U_T}{R},$$

where β is the forward current gain of the bipolar transistors employed, U_T is the thermal voltage, $\frac{kT}{q}$, and R is the resistance level used in the resistive voltage divider. At these current levels, the base-emitter junctions of the bipolar transistors begin to clamp the base voltages so that further increases in the collector voltages only linearly increase the collector current, instead of doing so exponentially. Also, for this implementation of the MITE, the collector voltages must be buffered into the resistive network, and these buffers must be able to drive the impedance level of the resistive network.

As an example, I now consider the circuit depicted in Figure 2.7, which is of the same form as that shown in Figure 2.5. In this circuit, I use the bipolar–resistive-divider

implementation of the MITE. The collector voltages, V_1 and V_2, are buffered into the resistive network through unity-gain followers. For this circuit, w_{22}=w_{31}=1 and w_{11}=w_{12}= $\frac{R}{2R} = \frac{1}{2}$. From Equation 2.12, I expect I_1 to factor into I_3 raised to the power $\frac{w_{31}}{w_{11}} = \frac{1}{1/2} = 2$, and I expect I_2 to factor into I_3 raised to the power $-\frac{w_{31}}{w_{11}}\frac{w_{12}}{w_{22}} = -\frac{1}{1/2}\frac{1/2}{1} = -1$. Combining these results, I have that

$$I_3 = \frac{I_1^2}{I_2};$$

thus, this circuit is a squaring-reciprocal circuit.

Figure 2.8 shows experimental data from a breadboarded version of the circuit of Figure 2.7. For this implementation, I used three bipolar transistors from a quad TPQ3904 NPN transistor array, I used resistors from an integrated 10 k resistor array, and I made the voltage buffers from two rail-to-rail CMOS opamps connected as unity-gain voltage followers. Figure 2.8a shows I_3 plotted as a function of I_1 on a log-log plot for various values of I_2 ranging from 37.0 pA to 31.6 μA. Figure 2.8b shows I_3 plotted as a function of I_2 on a log-log plot for various values of I_1 again ranging from 37.0 pA to 31.6 μA. In both cases, circles represent actual measured values of I_3, and solid lines show values of the ideal theoretical expression

$$I_3 = \frac{I_1^2}{I_2}$$

calculated from the values of I_1 and I_2 at each point. Apart from a slight scale-factor error on the order of a few percent, the data and ideal curves agree nicely over a seven-decade range of currents. In particular, the slopes of the lines, which represent the power laws, agree remarkably well. For the bipolars used, the current gain, β, is on the order of 100. Given this value of β and the values of 10 k used in the resistive divider, we would expect to start seeing deviation from the desired behavior at current levels approaching

$$\frac{\beta U_T}{R} = \frac{100 \times 25\text{mV}}{10\text{k}\Omega} = 250\mu\text{A},$$

and we do begin to see such deviations along the upper and right edge of Figure 2.8b.

Recently, there has been interest in the literature in using the back gate of a subthreshold MOSFET, in addition to the front gate, to build translinear circuits [6–9]. Of particular note is a one-quadrant product-reciprocal circuit, which was first described by van der Gevel and Kuenen [6]. The four-terminal subthreshold MOSFET can be viewed from the standpoint that I have developed in this chapter as a two-input translinear element with weighting coefficients of κ and 1-κ for the front and back gates, respectively. Hence, the analysis procedures that I develop in Section 2.5 and Chapter 3 are applicable to many

of these circuits, including the product-reciprocal circuit just mentioned. The weighting coefficients, however, are not controllable by the designer, and they change as a function of gate-to-bulk voltage. We can build circuits that compute products of currents raised to powers of $\pm\frac{\kappa}{1-\kappa}$, $\pm\frac{1-\kappa}{\kappa}$, and $\pm\frac{1-\kappa}{1-\kappa}=\pm\frac{\kappa}{\kappa}=\pm 1$ using these devices. However, we must exercise care with such circuits to ensure that the source-to-bulk junctions, which also form the base-emitter junctions of parasitic vertical bipolar transistors, do not become appreciably forward biased.

2.5. Matrix Analysis of Multiple-Input Translinear Element Networks

Now, I consider the circuit shown in Figure 2.9. There are N input MITEs, labeled Q_1 through Q_N, and M output MITEs, labeled Q_{N+1} through Q_{N+M}. The output voltage of Q_k couples into the weighted voltage summation of Q_n through weighting coefficient w_{nk}; here, k can range from 1 to N, and n can range from 1 to $N+M$. If the output voltage of Q_k does not enter into the weighted voltage summation of Q_n, then the value of w_{nk} is zero. Together, these weighting coefficients constitute an $(N+M)\times N$ connectivity matrix, $\mathbf{W}$. I partition $\mathbf{W}$ into an input connectivity matrix, $\mathbf{W}_{\text{in}}$, and an output connectivity matrix $\mathbf{W}_{\text{out}}$; $\mathbf{W}_{\text{in}}$ comprises the first N rows of $\mathbf{W}$, whereas $\mathbf{W}_{\text{out}}$ comprises the last M rows of $\mathbf{W}$.

The N input currents, I_1 through I_N, are sourced into the outputs of MITEs Q_1 through Q_N, respectively. As a result, N voltages, V_1 through V_N, develop that are each a linear combination of logarithms of the N input currents. The particular coefficients appearing in these linear combinations depend on the input connectivity matrix, $\mathbf{W}_{\text{in}}$. The circuit then forms M output currents, I_{N+1} through I_{N+M}, in output MITEs Q_{N+1} through Q_{N+M}, respectively, by linearly combining the voltages V_1 through V_N according to the output connectivity matrix, $\mathbf{W}_{\text{out}}$, and exponentiating the resulting weighted sums. In this section, I show that the mth output current, I_{N+m}, is given by

$$I_{N+m} = I_s^{1-\sum_{n=1}^{N}\Lambda_{mn}} K_m \prod_{n=1}^{N} I_n^{\Lambda_{mn}}, \tag{2.13}$$

where

$$K_m \equiv \lambda_{N+m}\prod_{n=1}^{N}\lambda_n^{-\Lambda_{mn}}, \tag{2.14}$$

and the values of Λ_{mn} are given by the matrix product

$$\Lambda \equiv \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}. \tag{2.15}$$

In other words, the mth output current is a product of the N input currents; I_n factors into the product raised to the power Λ_{mn}, which, in general, will be equal to a sum of products

of ratios of weighting coefficients. Now, it is easy to see that, if the powers contained in Λ are such that, for each *m*, $\sum_{n=1}^{N}\Lambda_{mn}=1$, then Equation 2.13 reduces to

$$I_{N+m}=K_m\prod_{n=1}^{N}I_n^{\Lambda_{mn}},$$

which is both independent of process parameters and insensitive to isothermal variations.

I also show that, if the value of λ_n for each MITE is the same (i.e., $\lambda_1=...=\lambda_{N+M}=\lambda$), then K_m=1 for all *m*, and Equation 2.13 further reduces to

$$I_{N+m}=\prod_{n=1}^{N}I_n^{\Lambda_{mn}}.$$

Finally, I show that, if the circuit of Figure 2.9 is made from MITEs that each have an identical set of weighting coefficients, and if all the MITE inputs are connected to one of the V_n, then the powers in Λ are such that, for each *m* , $\sum_{n=1}^{N}\Lambda_{mn}=1$, so the I_s dependence of Equation 2.13 disappears.

In the analysis that follows, I assume that all the MITEs are operating at the same temperature, and that they have well-matched values of I_s. I also assume that the input connectivity matrix, $\mathbf{W}_{in}$, has an inverse, $\mathbf{W}_{in}^{-1}$, so that Λ is well defined. I begin by noting that we assumed that the input terminals of the MITE draw negligible DC current, so that **Kirchhoff's current law** (**KCL**) implies that, at equilibrium, the *n*th input current, I_n, will just balance the current sunk by the MITE Q_n. Thus, I can apply Equation 2.1 to each of the input elements, and I write that

$$I_n=\lambda_n I_s\exp\left[\sum_{k=1}^{N}\frac{w_{nk}V_k}{U_T}\right].$$

After rearranging, taking logarithms, and solving for V_k, I obtain

$$V_k=U_T\sum_{n=1}^{N}\left(\mathbf{W}_{in}^{-1}\right)_{kn}\log\frac{I_n}{\lambda_n I_s}. \tag{2.16}$$

From Equation 2.1, the *m*th output current, I_{N+m}, is given by

$$I_{N+m}=\lambda_{N+m}I_s\exp\left[\sum_{k=1}^{N}\frac{w_{(N+m)k}V_k}{U_T}\right]. \tag{2.17}$$

Substituting Equation 2.16 into Equation 2.17, I obtain

$$\begin{aligned}I_{N+m} &= \lambda_{N+m}I_s\exp\left[\frac{U_T}{U_T}\sum_{n=1}^{N}\sum_{k=1}^{N}w_{(N+m)k}\left(\mathbf{W}_{in}^{-1}\right)_{kn}\log\frac{I_n}{\lambda_n I_s}\right]\\ &= \lambda_{N+m}I_s\exp\left[\sum_{n=1}^{N}\sum_{k=1}^{N}\left(\mathbf{W}_{out}\right)_{mk}\left(\mathbf{W}_{in}^{-1}\right)_{kn}\log\frac{I_n}{\lambda_n I_s}\right].\end{aligned} \tag{2.18}$$

If I apply the definition of Λ from Equation 2.15, then Equation 2.18 becomes

$$\begin{aligned} I_{N+m} &= \lambda_{N+m} \mathrm{I_s} \exp\left[\sum_{n=1}^{N} \Lambda_{mn} \log \frac{I_n}{\lambda_n \mathrm{I_s}}\right] \\ &= \lambda_{N+m} \mathrm{I_s} \prod_{n=1}^{N} \exp\left[\log\left(\frac{I_n}{\lambda_n \mathrm{I_s}}\right)^{\Lambda_{mn}}\right] \\ &= \lambda_{N+m} \mathrm{I_s} \prod_{n=1}^{N} \left(\frac{I_n}{\lambda_n \mathrm{I_s}}\right)^{\Lambda_{mn}} \\ &= \mathrm{I_s}^{1-\sum_{n=1}^{N} \Lambda_{mn}} \lambda_{N+m} \prod_{n=1}^{N} \lambda_n^{-\Lambda_{mn}} \prod_{n=1}^{N} I_n^{\Lambda_{mn}} . \end{aligned} \tag{2.19}$$

Applying the definition of K_m from Equation 2.14, I have that

$$I_{N+m} = \mathrm{I_s}^{1-\sum_{n=1}^{N} \Lambda_{mn}} K_m \prod_{n=1}^{N} I_n^{\Lambda_{mn}} . \tag{2.20}$$

If the powers contained in Λ are such that, for each m, $\sum_{n=1}^{N} \Lambda_{mn} = 1$, then it is easy to see that Equation 2.20 reduces to

$$I_{N+m} = K_m \prod_{n=1}^{N} I_n^{\Lambda_{mn}} , \tag{2.21}$$

which is both independent of process parameters and insensitive to isothermal variations. Moreover, if the value of λ_n for each MITE is the same (i.e., $\lambda_1 = ... = \lambda_{N+M} = \lambda$), then I have that

$$\begin{aligned} K_m &= \lambda_{N+m} \prod_{n=1}^{N} \lambda_n^{-\Lambda_{mn}} \\ &= \lambda^{1-\sum_{n=1}^{N} \Lambda_{mn}} \\ &= 1. \end{aligned}$$

Thus, under these conditions, Equation 2.20 further reduces to

$$I_{N+m} = \prod_{n=1}^{N} I_n^{\Lambda_{mn}} .$$

These results are just what I set out to show.

Now, if each of the $M + N$ MITEs in the circuit of Figure 2.9 has an identical set of K weighting coefficients, w_1 through w_K, such that $\sum_{k=1}^{K} w_k = \mathrm{w}$, where w is a constant, and if each MITE input is connected to one of the N input node voltages, V_1 through V_N, then it is easy to see that these conditions imply that the sum of each of the rows of the connectivity matrix, **W**, sums to the constant w (i.e., for each n between 1 and $N+M$, $\sum_{k=1}^{N} w_{nk} = \mathrm{w}$). In Appendix 2.A, I show that this condition on **W** is sufficient to guarantee that Λ is such that, for each m between 1 and M, $\sum_{n=1}^{N} \Lambda_{mn} = 1$, which, in turn, implies that Equation 2.20 reduces to Equation 2.21. Note that this condition on **W** is *not*

a necessary one—each of the rows of Λ may sum to unity even though the rows of $\mathbf{W}$ do not sum to the same number. In Chapter 4, I show examples of translinear circuits that comprise MITEs with identical sets of weighting coefficients, but that do not have all MITE inputs connected to one of the input-node voltages. In such cases, however, we, as designers, are wasting some MITE transconductance; consequently, some of the input-node voltage swings may be larger than they need to be.

2.6. Stability of Multiple-Input Translinear Element Networks

In this section, I consider the stability of the general MITE translinear circuit of Figure 2.9 in response to a fixed set of input currents. I would like to find sufficient conditions on the input connectivity matrix $\mathbf{W}_{\text{in}}$ that will ensure the stability of this circuit. To do so, I add capacitors, C_1 through C_N, to the circuit of Figure 2.9 to arrive at the circuit shown in Figure 2.10, and I consider the response of this circuit to a set of fixed input currents, I_1 through I_N. Capacitor C_n models the interconnect capacitance associated with node V_n and the input capacitance of each MITE to which node V_n is connected.

All the assumptions that I made and all the analysis that I developed in Section 2.5 apply to this situation, except that, instead of assuming that the current sunk by MITE Q_n balances the nth input current instantaneously, I assume that these currents are transiently different from each other. I begin by applying KCL to the kth node, labeled V_k, and I write that

$$C_k\dot{V}_k = I_k - J_k, \tag{2.22}$$

where

$$J_n = \lambda_n \mathrm{I_s} \exp\left[\sum_{k=1}^{N} \frac{w_{nk}V_k}{\mathrm{U_T}}\right]. \tag{2.23}$$

I would like to express Equation 2.22 completely in terms of the currents, J_1 through J_N. To do so, I differentiate Equation 2.23 with respect to time as follows:

$$\begin{aligned} \dot{J}_n &= \lambda_n \mathrm{I_s} \exp\left[\sum_{k=1}^{N} \frac{w_{nk}V_k}{\mathrm{U_T}}\right] \sum_{k=1}^{N} \frac{w_{nk}\dot{V}_k}{\mathrm{U_T}} \\ &= J_n \sum_{k=1}^{N} \frac{w_{nk}\dot{V}_k}{\mathrm{U_T}}. \end{aligned} \tag{2.24}$$

Next, by rearranging and premultiplying both sides of Equation 2.24 by $\mathbf{W}_{\text{in}}^{-1}$, I solve for $\dot{V}_k$ and get

$$\dot{V}_k = \mathrm{U_T} \sum_{n=1}^{N} \left(\mathbf{W}_{\text{in}}^{-1}\right)_{kn} \frac{\dot{J}_n}{J_n}. \tag{2.25}$$

Substituting Equation 2.25 into Equation 2.22, I find that

$$C_k \mathrm{U_T} \sum_{n=1}^{N} \left(\mathbf{W}_{\mathrm{in}}^{-1} \right)_{kn} \frac{\dot{J}_n}{J_n} = I_k - J_k. \tag{2.26}$$

By rearranging and premultiplying both sides of Equation 2.26 by $\mathbf{W}_{\mathrm{in}}$, I obtain

$$\frac{\dot{J}_n}{J_n} = \sum_{k=1}^{N} w_{nk} \left(\frac{I_k - J_k}{C_k \mathrm{U_T}} \right),$$

which I rewrite as

$$\dot{J}_n = J_n \sum_{k=1}^{N} w_{nk} \left(\frac{I_k - J_k}{C_k \mathrm{U_T}} \right). \tag{2.27}$$

Now, the system described by Equation 2.27 appears to have many possible fixed points. One occurs when I_k is equal to J_k for each value of k between 1 and N. Many other possibilities result when some of the $J_n = 0$ and, for the rest,

$$\sum_{k=1}^{N} w_{nk} \left(\frac{I_k - J_k}{C_k \mathrm{U_T}} \right) = 0.$$

These last possibilities, however, cannot be physical because, from Equation 2.23, the values of J_k are strictly positive. Consequently, the circuit of Figure 2.10 has a single fixed point that occurs when the current sunk by each input MITE just balances its respective input current. It will prove convenient to translate this fixed point to the origin of the system via the following change of variables:

$$J_n^* \equiv J_n - I_n. \tag{2.28}$$

Physically, this current, J_n^*, is the current flowing out of capacitor C_n. By substituting Equation 2.28 into Equation 2.27, I find that

$$\begin{aligned} \dot{J}_n^* &= (I_n + J_n^*) \sum_{k=1}^{N} \frac{w_{nk}(-J_k^*)}{C_k \mathrm{U_T}} \\ &= -\frac{I_n}{C_n \mathrm{U_T}} \sum_{k=1}^{N} w_{nk} \frac{C_n}{C_k} J_k^* - \frac{I_n}{C_n \mathrm{U_T}} \sum_{k=1}^{N} w_{nk} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n}. \end{aligned} \tag{2.29}$$

Next, I define

$$\tau_n \equiv \frac{C_n \mathrm{U_T}}{I_n}, \tag{2.30}$$

so Equation 2.29 becomes

$$\tau_n \dot{J}_n^* = -\sum_{k=1}^{N} w_{nk} \frac{C_n}{C_k} J_k^* - \sum_{k=1}^{N} w_{nk} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n}. \tag{2.31}$$

From the analysis of Section 2.5, the mth output current, I_{N+m}, is given by

$$I_{N+m} = \mathrm{I_s}^{1 - \sum_{n=1}^{N} \Lambda_{mn}} K_m \prod_{n=1}^{N} J_n^{\Lambda_{mn}},$$

which, when expressed in terms of J_n^*, becomes

$$I_{N+m} = \mathrm{I}_\mathrm{s}^{1-\sum_{n=1}^{N}\Lambda_{mn}} K_m \prod_{n=1}^{N}\left(I_n + J_n^*\right)^{\Lambda_{mn}}. \tag{2.32}$$

Thus, the response of the circuit of Figure 2.10 to a set of constant input currents is described by an autonomous set of N coupled nonlinear ordinary differential equations for the capacitor currents J_1^* through J_N^* as follows:

$$\begin{cases} \tau_1 \dot{J}_1^* = -\sum_{k=1}^{N} w_{1k} \dfrac{C_1}{C_k} J_k^* - \sum_{k=1}^{N} w_{1k} \dfrac{C_1}{C_k} \dfrac{J_k^* J_1^*}{I_1}, \\ \quad \vdots \\ \tau_N \dot{J}_N^* = -\sum_{k=1}^{N} w_{Nk} \dfrac{C_N}{C_k} J_k^* - \sum_{k=1}^{N} w_{Nk} \dfrac{C_N}{C_k} \dfrac{J_k^* J_N^*}{I_N}. \end{cases} \tag{2.33}$$

The output currents, I_{N+1} through I_{N+M}, depend on the input currents, I_1 through I_N, and the capacitor currents, J_1^* through J_N^*, through a set of M nonlinear equations as follows:

$$\begin{cases} I_{N+1} = \mathrm{I}_\mathrm{s}^{1-\sum_{n=1}^{N}\Lambda_{1n}} K_1 \prod_{n=1}^{N}\left(I_n + J_n^*\right)^{\Lambda_{1n}}, \\ \quad \vdots \\ I_{N+M} = \mathrm{I}_\mathrm{s}^{1-\sum_{n=1}^{N}\Lambda_{Mn}} K_M \prod_{n=1}^{N}\left(I_n + J_n^*\right)^{\Lambda_{Mn}}. \end{cases} \tag{2.34}$$

Now, I would like to determine the stability of the fixed point, $\mathbf{J}^* = \mathbf{0}$, of the system given by Equations 2.33 and 2.34. The following theorem will prove useful in this endeavor.

Theorem 2.1: Consider an autonomous system of ordinary differential equations of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{f}(\mathbf{x}), \tag{2.35}$$

with a fixed point at the origin, $\mathbf{x}=\mathbf{0}$, where $\mathbf{f}(\mathbf{x})$ is continuous in a neighborhood of the origin and satisfies

$$\lim_{|\mathbf{x}|\to 0} \frac{|\mathbf{f}(\mathbf{x})|}{|\mathbf{x}|} = 0,$$

and where $\mathbf{A}$ is a constant matrix whose eigenvalues are distinct. If the eigenvalues of $\mathbf{A}$ all have negative real parts, then the system is asymptotically stable; if they all have positive real parts, then the system is unstable.

Proof: See Lefschetz [10, pp. 87–90 and pp. 117–118]. ■

I can put Equation 2.33 in the form of Equation 2.35 by making the following

identifications:

$$(\mathbf{A})_{nk} = -\frac{w_{nk}}{\tau_n}\frac{C_n}{C_k} \tag{2.36}$$

and

$$f_n(\mathbf{J}^*) = -\sum_{k=1}^{N}\frac{w_{nk}}{\tau_n}\frac{C_n}{C_k}\frac{J_k^* J_n^*}{I_n}. \tag{2.37}$$

Clearly, $\mathbf{f}(\mathbf{J}^*)$ is continuous near the origin. In Appendix 2.B, I show that $\mathbf{f}(\mathbf{J}^*)$ also satisfies the condition

$$\lim_{|\mathbf{J}^*|\to 0}\frac{|\mathbf{f}(\mathbf{J}^*)|}{|\mathbf{J}^*|} = 0.$$

Consequently, I can apply Theorem 2.1 to the system given by Equation 2.33, and I can conclude that, if the eigenvalues of $\mathbf{A}$ are distinct and all have negative real parts, then the origin of the system given by Equation 2.33 is asymptotically stable.

For convenience, I define two diagonal matrices, $\mathbf{T}$ and $\mathbf{C}$, as follows:

$$\mathbf{T} \equiv \begin{bmatrix} \tau_1 & & 0 \\ & \ddots & \\ 0 & & \tau_N \end{bmatrix} \text{ and } \mathbf{C} \equiv \begin{bmatrix} C_1 & & 0 \\ & \ddots & \\ 0 & & C_N \end{bmatrix}.$$

Because $\mathbf{T}$ and $\mathbf{C}$ are diagonal, it follows that

$$\mathbf{T}^{-1} = \begin{bmatrix} \frac{1}{\tau_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\tau_N} \end{bmatrix} \text{ and } \mathbf{C}^{-1} = \begin{bmatrix} \frac{1}{C_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{C_N} \end{bmatrix}.$$

Now, I can express the matrix $\mathbf{A}$ from Equation 2.36 in terms of $\mathbf{T}$, $\mathbf{C}$, and $\mathbf{W}_{\text{in}}$ as

$$\mathbf{A} = -\mathbf{C}\mathbf{T}^{-1}\mathbf{W}_{\text{in}}\mathbf{C}^{-1}.$$

I want to find conditions on the input connectivity matrix, $\mathbf{W}_{\text{in}}$, that are sufficient to guarantee that the eigenvalues of $\mathbf{A}$ all have negative real parts. These conditions on the eigenvalues of $\mathbf{A}$, in turn, ensures the stability of the circuit of Figure 2.10. Now, the eigenvalues of $\mathbf{A}$ are given by solutions to the characteristic equation

$$|\lambda\mathbf{I} - \mathbf{A}| = 0.$$

Consider the quantity

$$\begin{aligned} \lambda\mathbf{I} - \mathbf{A} &= \lambda\mathbf{I} + \mathbf{C}\mathbf{T}^{-1}\mathbf{W}_{\text{in}}\mathbf{C}^{-1} \\ &= \lambda\mathbf{C}\mathbf{C}^{-1} + \mathbf{C}\mathbf{T}^{-1}\mathbf{W}_{\text{in}}\mathbf{C}^{-1} \\ &= \mathbf{C}\left(\lambda\mathbf{I} + \mathbf{T}^{-1}\mathbf{W}_{\text{in}}\right)\mathbf{C}^{-1}. \end{aligned}$$

Taking the determinant of both sides of the preceding equation, I have that

$$\begin{aligned} |\lambda\mathbf{I} - \mathbf{A}| &= \left|\mathbf{C}\left(\lambda\mathbf{I} + \mathbf{T}^{-1}\mathbf{W}_{\text{in}}\right)\mathbf{C}^{-1}\right| \\ &= \left|\mathbf{C}\right|\left|\lambda\mathbf{I} + \mathbf{T}^{-1}\mathbf{W}_{\text{in}}\right|\left|\mathbf{C}^{-1}\right| \end{aligned}$$

$$= \left(\prod_{n=1}^{N} C_n\right)\left|\lambda\mathbf{I} + \mathbf{T}^{-1}\mathbf{W}_{\text{in}}\right|\left(\prod_{n=1}^{N}\frac{1}{C_n}\right)$$
$$= \left|\lambda\mathbf{I} + \mathbf{T}^{-1}\mathbf{W}_{\text{in}}\right|.$$

Thus, I have established that $|\lambda\mathbf{I} - \mathbf{A}| = 0$ if and only if $\left|\lambda\mathbf{I} + \mathbf{T}^{-1}\mathbf{W}_{\text{in}}\right| = 0$. Consequently, the condition that the eigenvalues of $\mathbf{A}$ all have negative real parts is equivalent to the condition that the eigenvalues of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ all have positive real parts. In Appendix 2.C, I show that, if, by renumbering the inputs, I can put $\mathbf{W}_{\text{in}}$ into the form

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} \mathbf{W}_{\text{DD}} & * \\ 0 & \mathbf{W}_{\text{UT}} \end{bmatrix},$$

where $\mathbf{W}_{\text{DD}}$ is a $J \times J$ strictly diagonally dominant submatrix (i.e., for each n and k between 1 and J, $w_{nn} > \sum_{k \neq n} w_{nk}$), $\mathbf{W}_{\text{UT}}$ is an $(N-J)\times(N-J)$ upper triangular submatrix (i.e., for each n between $J+1$ and N, if $k<n$, then $w_{nk} = 0$), and * denotes any entry, then the eigenvalues of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ all have positive real parts. This condition on $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$, in turn, implies that the eigenvalues of $\mathbf{A}$ all have negative real parts. Hence, the circuit of Figure 2.10 is asymptotically stable. The synthesis procedures that I develop in Chapter 4 are guaranteed to produce MITE translinear circuits that satisfy these conditions on the input connectivity matrix. Hence, the circuits will be guaranteed asymptotically stable by construction.

2.7. Relationship to Translinear Loop Circuits

In this section, I show that any expression that can be implemented with a single-output MITE translinear network also can be implemented with a single translinear loop. In Chapter 4, when I develop a systematic synthesis procedure for MITE networks, I will be in a position to show the converse—that is, any expression that can be implemented with a single translinear loop can also be implemented with a single-output MITE network. At that point, I will have shown, in a constructive manner, that the two classes of circuits are, in a sense, equivalent.

I would like to show that any expression of the form

$$I_{N+1} = \prod_{n=1}^{N} I_n^{\Lambda_n} \tag{2.38}$$

can be implemented with a single translinear loop. I restrict the values of Λ_n to be rational numbers, either positive or negative. So, I have that

$$|\Lambda_n| = \frac{p_n}{q_n},$$

where p_n and q_n are positive integers for each n. In the case of translinear loop circuits, the reason for this restriction is obvious: We raise a current to a certain power in a translinear loop circuit by passing that current through an integral number of TEs facing in the same direction. We cannot pass a current through a fractional number of TEs. In principle, it may seem that there is no reason for such a restriction with MITE translinear circuits; in practice, however, as designers, we obtain the most accurate ratios of weighting coefficients by connecting an integral number of identical unit cells in parallel with one another. Even if we choose not to follow this practice, we draw capacitors and resistors on a finite grid, which implies that we are ultimately restricted to the rationals with MITE translinear circuits as well.

Without loss of generality, I assume that Λ_1 through Λ_J are positive and Λ_{J+1} through Λ_N are negative, so I_{N+1} is of the form

$$I_{N+1} = \frac{\prod_{n=1}^{J} I_n^{\Lambda_n}}{\prod_{n=J+1}^{N} I_n^{|\Lambda_n|}},$$

which I can equivalently express as

$$\prod_{n=1}^{J} I_n^{\Lambda_n} = I_{N+1} \prod_{n=J+1}^{N} I_n^{|\Lambda_n|}. \tag{2.39}$$

Now, let k_{N+1} denote the least common multiple of the q_n. I then define a new set of integers, k_n, as follows:

$$k_n \equiv k_{N+1}|\Lambda_n| = \frac{p_n k_{N+1}}{q_n}.$$

Thus, by raising both sides of Equation 2.39 to the power k_{N+1}, I have that

$$\prod_{n=1}^{J} I_n^{k_n} = I_{N+1}^{k_{N+1}} \prod_{n=J+1}^{N} I_n^{k_n},$$

which I rewrite as

$$\prod_{n=1}^{J} I_n^{k_n} = \prod_{n=J+1}^{N+1} I_n^{k_n}. \tag{2.40}$$

This form of Equation 2.38 is suitable for direct realization with a single translinear loop. I take the currents on the left-hand side of Equation 2.40 to be the counterclockwise currents, and take those on the right-hand side of Equation 2.40 to be the clockwise currents. A translinear loop with a stacked topology implementing the general case of

Equation 2.40 is shown in Figure 2.11. There are two special cases that have loop structures slightly different from that shown in Figure 2.11. First, when the values of Λ_n are all positive and form a partition of unity, the translinear loop of Figure 2.11 reduces to that shown in Figure 2.12a. In this case, J is equal to N. Second, when the Λ_n are all negative except for one, the translinear loop of Figure 2.11 reduces to that shown in Figure 2.12b. In this case, J is equal to unity.

As designers, we can implement any expression of the form of Equation 2.38 with either a translinear loop or a MITE network; which should we use? Do we gain anything by using a MITE translinear circuit over a translinear loop circuit? In general, the answer to these questions depends on the desired relationship between the output current and the input currents. I can, however, make two general comparisons between these two alternatives. First, in a MITE translinear circuit, the number of MITEs required is given by the number of input and output currents. In Chapters 6 and 7, when I describe specific MITE implementations, this required number of MITEs translates into between one and three active devices per input or output current. In a translinear loop implementation, the number of active devices required is given by the sum of all the powers in Equation 2.40 (i.e., the value of $\sum_{n=1}^{N+1} k_n$). Second, in a MITE network, we get a power-law relationship of $\frac{m}{n}$ by passing the input current through a MITE with scale units proportional to n, and by passing the output current through a MITE with scale units proportional to m, so that the ratio of these scale units is $\frac{m}{n}$. In a translinear loop circuit, we obtain a power-law relationship of $\frac{m}{n}$ by passing the input current through a stack of m TEs connected in one direction around the loop, and by passing the output current through a stack of n TEs connected in the other direction. Together, these differences can imply—depending on the power laws in the relationship that we need to implement—that a MITE network implementation is more compact and operates with lower voltage than a translinear loop implementation.

To illustrate these comparisons, I now describe a specific example. Suppose that I am interested in implementing the following function:

$$I_3 = \frac{I_1^{k_1/k_3}}{I_2^{k_2/k_3}}, \tag{2.41}$$

where k_1, k_2, and k_3 are all positive integers. For the units of Equation 2.41 to balance properly, I must have that $\frac{k_1}{k_3} - \frac{k_2}{k_3} = 1$, or equivalently that $k_1 = k_2 + k_3$.

A translinear loop implementation of Equation 2.41 is shown in Figure 2.13a. I source current I_1 into a stack of k_1 diode-connected transistors, generating voltage V_1. I sink current I_2 from a stack of k_2 transistors that comprises one emitter follower and

$k_2 - 1$ diode-connected transistors, generating voltage V_2. The circuit generates current I_3 from V_2 in a stack of k_3 transistors that comprises one emitter follower and $k_3 - 1$ diode-connected transistors. In all, I need $k_1 + k_2 + k_3$ transistors. If I e-fold I_1, V_1 will increase by $k_1 \mathrm{U_T}$. If I e-fold I_2, V_2 will decrease by $k_2 \mathrm{U_T}$.

A MITE network implementation of Equation 2.41 is shown in Figure 2.13b. In this case, all the weighting coefficients have the same value, w, and I obtain accurate ratios of weighing coefficients by connecting integral numbers of unit cells in parallel. I source I_1 into MITE $\mathrm{Q_1}$, which I diode connect through a weight of $k_3 w$, resulting in voltage V_1. I source I_2 into MITE $\mathrm{Q_2}$, which I diode connect through a weight of $k_1 w$, resulting in voltage V_2. Here, V_2 couples into MITE $\mathrm{Q_1}$ through a weight of $k_2 w$. The circuit generates current I_3 in MITE $\mathrm{Q_3}$ from V_1 through a weight of $k_1 w$. For this implementation, I need between three and nine transistors, depending on the particular MITE implementation that I use. If I e-fold I_1, V_1 will increase by approximately $\frac{k_1 + k_2}{k_1} \mathrm{U_T}$. If I e-fold I_2, V_2 will only increase by approximately $\mathrm{U_T}$.

If the values of k_1, k_2, and k_3 are each between 1 and 3, the two implementations shown in Figure 2.13 are comparable in terms of both number of active devices (which affects the circuit size) and voltage swings (which dictate the power-supply voltage). If the values of k_1, k_2, and k_3 get much larger than three, then the MITE implementation will have fewer devices (and thus a more compact circuit) and smaller voltage swings (and thus a lower power-supply voltage) than the translinear loop implementation.

2.8. Appendix 2.A

In this appendix, I show that, if the rows of **W** sum to the same constant, w (i.e., for each n between 1 and $N + M$, $\sum_{k=1}^{N} w_{nk} = \mathrm{w}$), and if the input connectivity matrix, $\mathbf{W}_{\mathrm{in}}$, is invertible, then the powers contained in Λ will be such that, for each m, $\sum_{n=1}^{N} \Lambda_{mn} = 1$. The following lemma will prove useful for this endeavor.

Lemma 2.1: If each of the rows of an invertible $N \times N$ matrix, **A**, sums to some constant, C, then each of the rows of $\mathbf{A}^{-1}$ sums to the constant $\frac{1}{\mathrm{C}}$.

Proof: The condition that the sum of each of the rows of **A** sums to some constant, C, can be written in matrix notation as follows:

$$\mathbf{A}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = \mathrm{C}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}. \tag{2.42}$$

Now, I premultiply each side of Equation 2.42 by $\mathbf{A}^{-1}$ to obtain

$$\mathbf{A}^{-1}\mathbf{A}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = \mathrm{C}\mathbf{A}^{-1}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix},$$

which implies that

$$\mathbf{I}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = \mathrm{C}\mathbf{A}^{-1}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}, \tag{2.43}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. I rewrite Equation 2.43 as

$$\mathbf{A}^{-1}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = \frac{1}{\mathrm{C}}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix},$$

which is just what I set out to show, written in matrix notation. ■

Now, I consider the quantity

$$\begin{aligned}\sum_{n=1}^{N}\Lambda_{mn} &= \sum_{n=1}^{N}\sum_{k=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{mk}\left(\mathbf{W}_{\text{in}}^{-1}\right)_{kn}\\ &= \sum_{k=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{mk}\sum_{n=1}^{N}\left(\mathbf{W}_{\text{in}}^{-1}\right)_{kn},\end{aligned}$$

which, because of Lemma 2.1

$$\begin{aligned}&= \sum_{k=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{mk}\left(\frac{1}{\mathrm{w}}\right)\\ &= \frac{1}{\mathrm{w}}\sum_{k=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{mk}\\ &= \frac{\mathrm{w}}{\mathrm{w}}\\ &= 1,\end{aligned}$$

which is what I set out to show.

2.9. Appendix 2.B

In this appendix, I show that

$$\lim_{|\mathbf{J}^*|\to 0}\frac{\left|\mathbf{f}(\mathbf{J}^*)\right|}{\left|\mathbf{J}^*\right|} = 0,$$

where the *n*th component of $\mathbf{f}(\mathbf{J}^*)$ is given by

$$f_n(\mathbf{J}^*) = -\sum_{k=1}^{N} \frac{w_{nk}}{\tau_n} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n}.$$

The following lemma will prove useful for this endeavor.

Lemma 2.2: If *x* and *y* are any two real numbers, then $|xy| \le x^2 + y^2$.

Proof: If $|x| \ge |y|$, then, by multiplying both sides of this inequality by $|x|$, I find that

$$x^2 \ge |xy|, \tag{2.44}$$

with equality if and only if $|x| = |y|$. Because $y^2 \ge 0$, I can add y^2 to the left-hand side of Equation 2.44 while preserving the sign of the inequality, to obtain

$$|xy| \le x^2 + y^2.$$

Instead, if $|y| \ge |x|$, then by multiplying both sides of this inequality by $|y|$, I find that

$$y^2 \ge |xy|, \tag{2.45}$$

with equality if and only if $|x| = |y|$. Because $x^2 \ge 0$, I can add x^2 to the left-hand side of Equation 2.45 while preserving the sign of the inequality, to obtain

$$|xy| \le x^2 + y^2.$$

Thus, regardless of which of $|x|$ and $|y|$ is larger, $|xy| \le x^2 + y^2$. ■

First, I define

$$w_{\max} \equiv \max_{n,k}\left\{w_{nk}\frac{C_n}{C_k}\right\}, \quad \tau_{\min} \equiv \min_n\{\tau_n\}, \text{ and } \quad I_{\min} \equiv \min_n\{I_n\}.$$

Now, I consider the quantity

$$\begin{aligned}
\left|\sum_{k=1}^{N} \frac{w_{nk}}{\tau_n} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n}\right| &\le \sum_{k=1}^{N} \frac{w_{nk}}{\tau_n} \frac{C_n}{C_k} \frac{\left|J_k^* J_n^*\right|}{I_n} && \text{(by the triangle inequality)} \\
&\le \sum_{k=1}^{N} \frac{w_{\max}}{\tau_{\min} I_{\min}} \left|J_k^* J_n^*\right| \\
&\le \sum_{k=1}^{N} \frac{w_{\max}}{\tau_{\min} I_{\min}} \left(J_k^{*2} + J_n^{*2}\right) && \text{(by Lemma 2.2)} \\
&= \frac{w_{\max}}{\tau_{\min} I_{\min}} \left(N J_n^{*2} + \sum_{k=1}^{N} J_k^{*2}\right) \\
&\le \frac{w_{\max}}{\tau_{\min} I_{\min}} \left(N \sum_{k=1}^{N} J_k^{*2} + \sum_{k=1}^{N} J_k^{*2}\right) \\
&= \frac{w_{\max}}{\tau_{\min} I_{\min}} (N+1) \sum_{k=1}^{N} J_k^{*2}
\end{aligned}$$

$$= \frac{w_{\max}}{\tau_{\min} I_{\min}} (N+1) \left| \mathbf{J}^* \right|^2 .$$

Next, I consider the quantity

$$\sqrt{\sum_{n=1}^{N} \left(\sum_{k=1}^{N} \frac{w_{nk}}{\tau_n} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n} \right)^2} \qquad \sqrt{\sum_{n=1}^{N} \left(\frac{w_{\max}}{\tau_{\min} I_{\min}} (N+1) \left| \mathbf{J}^* \right|^2 \right)^2}$$

$$= \sqrt{N \left(\frac{w_{\max}}{\tau_{\min} I_{\min}} (N+1) \left| \mathbf{J}^* \right|^2 \right)^2}$$

$$= \frac{w_{\max}}{\tau_{\min} I_{\min}} \sqrt{N} (N+1) \left| \mathbf{J}^* \right|^2 .$$

Thus, I have that

$$\lim_{|\mathbf{J}^*| \to 0} \frac{\left| \mathbf{f}(\mathbf{J}^*) \right|}{\left| \mathbf{J}^* \right|} = \lim_{|\mathbf{J}^*| \to 0} \frac{1}{\left| \mathbf{J}^* \right|} \sqrt{\sum_{n=1}^{N} \left(\sum_{k=1}^{N} \frac{w_{nk}}{\tau_n} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n} \right)^2}$$

$$\lim_{|\mathbf{J}^*| \to 0} \frac{w_{\max}}{\tau_{\min} I_{\min}} \sqrt{N} (N+1) \frac{\left| \mathbf{J}^* \right|^2}{\left| \mathbf{J}^* \right|}$$

$$= \lim_{|\mathbf{J}^*| \to 0} \frac{w_{\max}}{\tau_{\min} I_{\min}} \sqrt{N} (N+1) \left| \mathbf{J}^* \right|$$

$$= 0,$$

which is just what I set out to show.

2.10. Appendix 2.C

In this appendix, I show that, if, by renumbering the inputs, I can put $\mathbf{W}_{\text{in}}$ into the form

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} \mathbf{W}_{\text{DD}} & * \\ 0 & \mathbf{W}_{\text{UT}} \end{bmatrix},$$

where $\mathbf{W}_{\text{DD}}$ is a $J \times J$ strictly diagonally dominant submatrix (i.e., for each n and k between 1 and J, $w_{nn} > \sum_{k \neq n} w_{nk}$), $\mathbf{W}_{\text{UT}}$ is an $(N-J) \times (N-J)$ upper triangular submatrix (i.e., for each n between J+1 and N, if $k<n$, then $w_{nk} = 0$), and * denotes any entry, then the eigenvalues of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ all have positive real parts. The following definitions and results will be useful in this endeavor.

An $N \times N$ matrix $\mathbf{A}$ is said to be **strictly diagonally dominant** if, for each n between 1 and N,

$$|a_{nn}| > \sum_{k \neq n} |a_{nk}|.$$

Lemma 2.3: Let **A** be an $N \times N$ strictly diagonally dominant matrix with real components. If the diagonal elements of **A** are all positive, then the eigenvalues of **A** all have positive real parts.

Proof: Let λ be an eigenvalue of **A**, and let **x** (**x** **0**) be the eigenvector associated with λ. Suppose that the mth component of **x**, x_m, has the largest absolute value (i.e., $|x_m| \geq |x_n|$, for each n between 1 and N). By hypothesis, I have that $\mathbf{Ax} = \lambda\mathbf{x}$, which implies that

$$\lambda x_m = (\lambda \mathbf{x})_m = (\mathbf{Ax})_m = \sum_{k=1}^{N} a_{mk}.$$

By subtracting a_{mm} from both sides of the preceding equation, I have that

$$x_m(\lambda - a_{mm}) = \sum_{k \neq m} a_{mk}.$$

By taking the absolute value of both sides of the preceding equation, I write

$$\begin{aligned} |x_m||\lambda - a_{mm}| &= \left|\sum_{k \neq m} a_{mk} x_k\right| \\ &\leq \sum_{k \neq m} |a_{mk} x_k| \qquad \text{(by the triangle inequality)} \\ &= \sum_{k \neq m} |a_{mk}||x_k| \\ &\leq |x_m| \sum_{k \neq m} |a_{mk}|. \end{aligned} \tag{2.47}$$

Because **x** **0**, we have that $|x_m| \neq 0$; consequently, I can conclude from Equation 2.47 that

$$|\lambda - a_{mm}| \leq \sum_{k \neq m} |a_{mk}|. \tag{2.48}$$

Now, I consider the quantity

$$\begin{aligned} |\lambda - a_{mm}| &= \sqrt{\left(\mathrm{Re}\{\lambda\} - a_{mm}\right)^2 + \left(\mathrm{Im}\{\lambda\}\right)^2} \\ &\geq \sqrt{\left(\mathrm{Re}\{\lambda\} - a_{mm}\right)^2} \qquad \text{(because } \left(\mathrm{Im}\{\lambda\}\right)^2 \geq 0\text{)} \\ &= |\mathrm{Re}\{\lambda\} - a_{mm}|. \end{aligned} \tag{2.49}$$

By combining inequalities 2.48 and 2.49, I have that

$$|\mathrm{Re}\{\lambda\} - a_{mm}| \leq \sum_{k \neq m} |a_{mk}|,$$

which implies that

$$-\sum_{k \neq m} |a_{mk}| \leq \mathrm{Re}\{\lambda\} - a_{mm} \leq \sum_{k \neq m} |a_{mk}|.$$

By adding a_{mm} to both sides of each of the preceding inequalities, I have that

$$a_{mm} - \sum_{k \neq m} |a_{mk}| \leq \mathrm{Re}\{\lambda\} \leq a_{mm} + \sum_{k \neq m} |a_{mk}|. \tag{2.50}$$

By hypothesis, **A** is strictly diagonally dominant, so regardless of the value of *m*, I have that

$$a_{mm} > \sum_{k \neq m} a_{mk},$$

which implies that

$$a_{mm} - \sum_{k \neq m} a_{mk} > 0.$$

Consequently, by combining the preceding inequality with the left inequality of inequalities 2.50, I have that

$$\mathrm{Re}\{\lambda\} \geq a_{mm} - \sum_{k \neq m} a_{mk} > 0,$$

which is what I set out to prove. ■

Corollary 2.1: Let **A** be an $N \times N$ strictly diagonally dominant matrix with real components and with positive diagonal elements. If **B** is an $N \times N$ diagonal matrix with positive real diagonal elements, then the eigenvalues of the matrix product **BA** all have positive real parts.

Proof: Premultiplying a matrix, **A**, by a diagonal matrix, **B**, has the effect of multiplying each row of **A** by the diagonal element in the corresponding row of **B** (i.e., for each *n* and *k* between 1 and *N*, $(\mathbf{BA})_{nk} = b_{nn} a_{nk}$). Consequently, if **A** is strictly diagonally dominant, then so is **BA**. Further, if the diagonal elements of **A** and **B** are all positive real, the diagonal elements of **BA** will all be positive real as well. Thus, I can apply Lemma 2.3 to **BA** and conclude that the eigenvalues of **BA** all have positive real parts. ■

An $N \times N$ matrix A is said to be **upper triangular** if $a_{nk} = 0$ when $k < n$ for each *n* and *k* between 1 and *N*.

Lemma 2.4: Let **A** be an $N \times N$ upper triangular matrix with real components. If the diagonal elements of **A** are all positive, then the eigenvalues of **A** all are positive real.

Proof: The eigenvalues of **A** are given by solutions of the characteristic equation

$$|\lambda \mathbf{I} - \mathbf{A}| = 0.$$

Now, if **A** is upper triangular, then so is $\lambda \mathbf{I} - \mathbf{A}$. The determinant of an upper triangular matrix is given by the product of the diagonal elements; consequently, I have that

$$|\lambda \mathbf{I} - \mathbf{A}| = \prod_{k=1}^{N} (\lambda - a_{kk}) = 0.$$

So, $|\lambda\mathbf{I}-\mathbf{A}|=0$ precisely when $\lambda=a_{kk}$ for some value of k between 1 and N; thus, the eigenvalues of **A** are simply the diagonal elements of **A**. By hypothesis, the diagonal elements of **A** are all positive real numbers, so the eigenvalues of **A** are all positive real. ■

Corollary 2.2: Let **A** be an $N\times N$ upper triangular matrix matrix with positive real diagonal elements. If **B** is an $N\times N$ diagonal matrix with positive real diagonal elements, then the eigenvalues of the matrix product **BA** are all positive real numbers.

Proof: Premultiplying a matrix, **A**, by a diagonal matrix, **B**, has the effect of multiplying each row of **A** by the diagonal element in the corresponding row of **B** (i.e., for each n and k between 1 and N, $(\mathbf{BA})_{nk}=b_{nn}a_{nk}$). Consequently, if **A** is upper triangular, then so is **BA**. Further, if the diagonal elements of **A** and **B** are all positive real, the diagonal elements of **BA** will all be positive real as well. Thus, I can apply Lemma 2.4 to **BA** and conclude that the eigenvalues of **BA** are all positive real. ■

An $N\times N$ matrix **A** is said to be **block upper triangular** if it is of the form

$$\mathbf{A}=\begin{bmatrix}\mathbf{A}_{11} & & & * \\ & \mathbf{A}_{22} & & \\ & & \ddots & \\ 0 & & & \mathbf{A}_{KK}\end{bmatrix},$$

where, for each k between 1 and K, $\mathbf{A}_{kk}$ is an $N_k\times N_k$ matrix and $\sum_{k=1}^{K}N_k=N$, and * denotes any entry. The matrices $\mathbf{A}_{kk}$, for k between 1 and K, are called **diagonal block matrices**.

Lemma 2.5: The eigenvalues of a block upper triangular matrix, **A**, are given by the eigenvalues of the individual block diagonal matrices.

Proof: The eigenvalues of **A** are given by solutions to the characteristic equation

$$|\lambda\mathbf{I}-\mathbf{A}|=0.$$

Now, if **A** is block upper triangular, then so is $\lambda\mathbf{I}-\mathbf{A}$. The determinant of a block upper triangular matrix is given by the product of the determinants of the diagonal block matrices [11, p. 25]. Consequently, I have that

$$|\lambda\mathbf{I}-\mathbf{A}|=\prod_{k=1}^{K}\left|\lambda\mathbf{I}_{N_k}-\mathbf{A}_{kk}\right|=0,$$

where $\mathbf{I}_{N_k}$ is the $N_k \times N_k$ identity matrix. So, $|\lambda\mathbf{I} - \mathbf{A}| = 0$ precisely when $\left|\lambda\mathbf{I}_{N_k} - \mathbf{A}_{kk}\right| = 0$ for some k between 1 and K. Therefore, λ is an eigenvalue of $\mathbf{A}$ if and only if it is an eigenvalue of one of the diagonal block matrices. ■

Now, without loss of generality, we suppose that $\mathbf{W}_{\text{in}}$ is of the form

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} \mathbf{W}_{\text{DD}} & * \\ 0 & \mathbf{W}_{\text{UT}} \end{bmatrix},$$

where $\mathbf{W}_{\text{DD}}$ is a $J \times J$ strictly diagonally dominant submatrix, $\mathbf{W}_{\text{UT}}$ is an $(N-J)\times(N-J)$ upper triangular submatrix, and * denotes any entry. If $\mathbf{W}_{\text{in}}$ is not in this form, but I can put it into that form by renumbering the inputs of the circuit of Figure 2.10, then all my conclusions will still hold. The reason is that I can renumber the inputs though a sequence of row and column exchanges; that is, renumbering input j as k and vice versa corresponds to exchanging rows j and k and columns j and k of the matrices $\mathbf{T}$, $\mathbf{W}_{\text{in}}$, and $\mathbf{C}$. Exchanging the two rows of a matrix changes the sign of the determinant of that matrix, and exchanging two columns of a matrix also changes the sign of the determinant of that matrix. Thus, exchanging rows j and k and columns j and k of a matrix will leave the determinant of that matrix unchanged. Consequently, if I renumber the inputs of the circuit of Figure 2.10, I will not change the eigenvalues of $\mathbf{W}_{\text{in}}$ or those of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$.

Now, I can express the diagonal matrix $\mathbf{T}^{-1}$ in the form

$$\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{T}_1 & 0 \\ 0 & \mathbf{T}_2 \end{bmatrix},$$

where $\mathbf{T}_1$ is the $J \times J$ diagonal matrix

$$\mathbf{T}_1 = \begin{bmatrix} \frac{1}{\tau_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\tau_J} \end{bmatrix},$$

and $\mathbf{T}_2$ is the $(N-J)\times(N-J)$ diagonal matrix

$$\mathbf{T}_2 = \begin{bmatrix} \frac{1}{\tau_{J+1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\tau_N} \end{bmatrix}.$$

The matrix product $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ will then have the form

$$\mathbf{T}^{-1}\mathbf{W}_{\text{in}} = \begin{bmatrix} \mathbf{T}_1\mathbf{W}_{\text{DD}} & * \\ 0 & \mathbf{T}_2\mathbf{W}_{\text{UT}} \end{bmatrix},$$

where $\mathbf{T}_1\mathbf{W}_{\text{DD}}$ is the product of a diagonal matrix and a strictly diagonally dominant matrix; $\mathbf{T}_2\mathbf{W}_{\text{UT}}$ is the product of a diagonal matrix and an upper triangular matrix; and * denotes an arbitrary entry. By Lemma 2.5, the eigenvalues of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ are the eigenvalues of $\mathbf{T}_1\mathbf{W}_{\text{DD}}$ and of $\mathbf{T}_2\mathbf{W}_{\text{UT}}$. By Corollary 2.1, the eigenvalues of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ all have positive real parts. By Corollary 2.2, the eigenvalues of $\mathbf{T}_2\mathbf{W}_{\text{UT}}$ are all positive reals. Therefore, the eigenvalues of $\mathbf{T}^{-1}\mathbf{W}_{\text{in}}$ all have positive real parts, which is what I set out to show.

2.11. References

1. D. H. Sheingold, ed., *Nonlinear Circuits Handbook: Designing with Analog Functional Modules and ICs*, Norwood, MA: Analog Devices, 1976.

2. H. C. Nauta, "An Integrated Gamma Corrector," *IEEE Journal of Solid-State Circuits*, vol. SC-16, no. 3, pp. 238–241, 1981.

3. X. Arreguit, E. A. Vittoz, and M. Merz, "Precision Compressor Gain Controller in CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. SC-22, no. 3, pp. 442–445, 1987.

4. X. Arreguit, *Compatible Lateral Bipolar Transistors in CMOS Technology: Model and Applications*, Ph.D. Thesis, École Polytechnique Federale de Lausanne, Lausanne, Switzerland, pp. 114–117, 1989.

5. E. A. Vittoz, "Analog VLSI Signal Processing: Why, Where, and How?" *Analog Integrated Circuits and Signal Processing*, vol. 6, no. 1, pp. 27–44, 1994.

6. M. van der Gevel and J. C. Kuenen, "$\sqrt{X}$ Circuit Based on a Novel, Back-Gate-Using Multiplier," *Electronics Letters*, vol. 30, no. 3, pp. 183–184, 1994.

7. J. Mulder, A. C. van der Woerd, W. A. Serdijn, and A. H. M. van Roermund, "Application of the Back Gate in MOS Weak Inversion Translinear Circuits," *IEEE Transactions on Circuits and Systems I*, vol. 42, no. 11, pp. 958–962, 1995.

8. A. G. Andreou and K. A. Boahen, "Translinear Circuits in Subthreshold MOS," *Analog Integrated Circuits and Signal Processing*, vol. 9, no 2, pp. 141–166, 1996.

9. R. Fried and C. C. Enz, "CMOS Parametric Current Amplifier," *Electronics Letters*, vol. 32, no. 14, pp. 1249–1250, 1996.

10. S. Lefschetz, *Differential Equations: Geometric Theory*, 2nd ed., New York: Wiley, 1963.

11. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, England: Cambridge University Press, 1987.

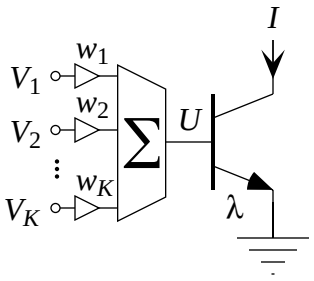
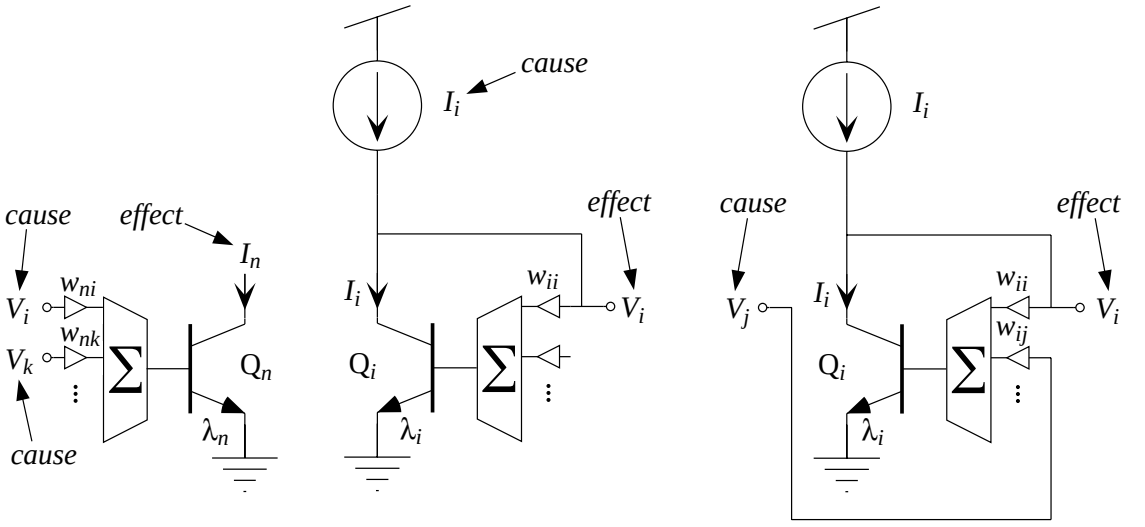


Figure 2.1. Circuit symbol for an ideal multiple-input translinear element (MITE). This element sums its K input voltages, V_k, each of which is weighted by a dimensionless positive coefficient, w_k. The element takes this weighted sum, U, and generates a current, I, that is exponential in U. The weighted summation may be implemented in a purely passive manner via either a resistive or a capacitive voltage divider. I require that the input terminals draw negligible DC current; consequently, if a resistive voltage divider is used, the input voltages must be buffered into the resistive network. The output device is shown as a bipolar transistor, but any device with an exponential current–voltage characteristic would do just as well, as long as the exponential current appears at a terminal different from the one to which the controlling voltage, U, is applied. A diode, for example, would not be appropriate in this context, but a subthreshold MOS transistor would be a suitable alternative to the bipolar transistor. The value of λ, shown under the emitter of the bipolar transistor, represents a dimensionless coefficient that proportionally scales the output current, I, such as an emitter area ratio in the case of a bipolar transistor or a $\frac{W}{L}$ ratio in the case of a subthreshold MOS transistor.



(a) (b) (c)

Figure 2.2. Three basic ways of using the MITE of Figure 2.1. (a) The MITE in a voltage-in, current-out (VICO) configuration. In this case, I apply voltages V_i and V_k, respectively, to generate an output current, I_n, which is proportional to the exponential of the weighted sum, $w_{ni}V_i+w_{nk}V_k$. (b) The MITE in a current-in, voltage-out (CIVO) configuration. Here, I force a current into the output of Q_i. The output voltage, V_i, adjusts itself through the self coupling coefficient, w_{ii}, until the current sunk by Q_i just balances the input current, I_i. The resulting output voltage, V_i, is logarithmic in the input current, I_i. (c) The MITE in a voltage-in, voltage-out (VIVO) configuration. In this case, I consider the current, I_i, to be fixed. The transistor Q_i and the current source together constitute a high-gain, inverting voltage amplifier with negative feedback around it through the weighting coefficients, w_{ij} and w_{ii}. The closed-loop gain of this well-known inverting amplifier configuration is simply $-\frac{w_{ij}}{w_{ii}}$.

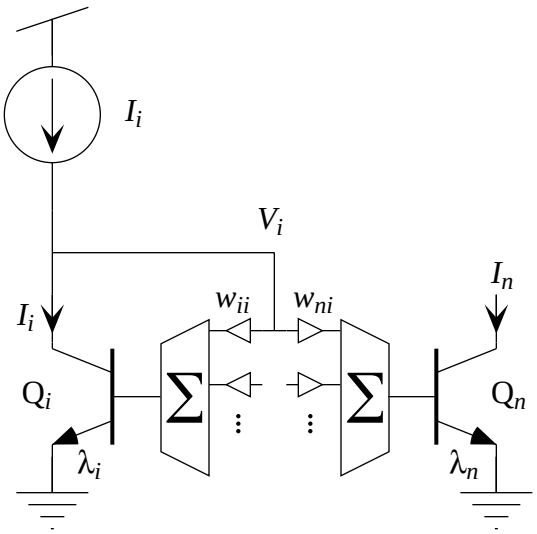


Figure 2.3. A current-in, current-out (CICO) circuit comprising a single CIVO stage connected to a single VICO stage. In this case, the output current, I_n, depends on the input current, I_i, as follows:

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}}.$$

This power-law relationship is insensitive to isothermal variations.

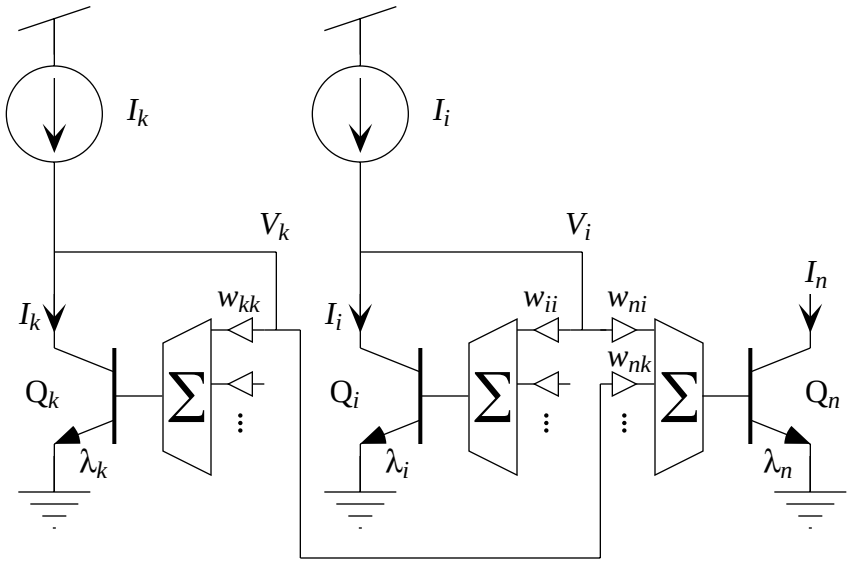


Figure 2.4. A CICO circuit comprising two CIVO stages connected to two different inputs of a single VICO stage. In this case, the output current, I_n, depends on the two input currents, I_i and I_k, as follows:

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}} \times I_k^{\frac{w_{nk}}{w_{kk}}} .$$

This product-of-powers relationship is insensitive to isothermal variations.

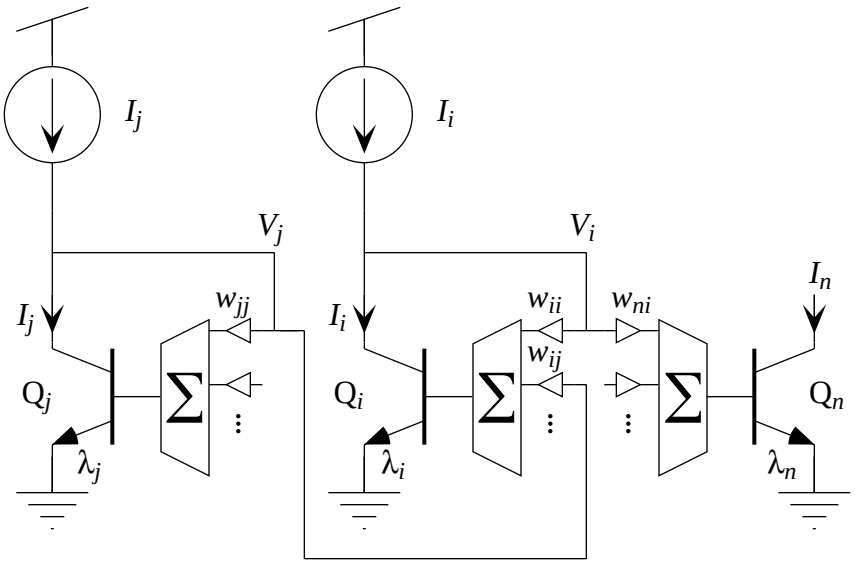


Figure 2.5. A CICO circuit comprising two CIVO stages and a single VICO stage. Unlike that of the circuit shown in Figure 2.4, the output voltage of the second CIVO stage, V_j, connects to the VICO stage through the first CIVO stage, which serves as a voltage-inversion stage for V_j. In this case, the output current, I_n, depends on the input currents, I_i and I_j, as follows:

$$I_n \propto I_i^{\frac{w_{ni}}{w_{ii}}} \div I_j^{\frac{w_{ni}}{w_{ii}}\frac{w_{ij}}{w_{jj}}}.$$

Here, the powers of I_i and I_j in the relationship are not completely independent of each other; however, for any value of $\frac{w_{ni}}{w_{ii}}$, we can adjust the value of $\frac{w_{ij}}{w_{jj}}$ to set the power of I_j as desired. This quotient-of-powers relationship is insensitive to isothermal variations.

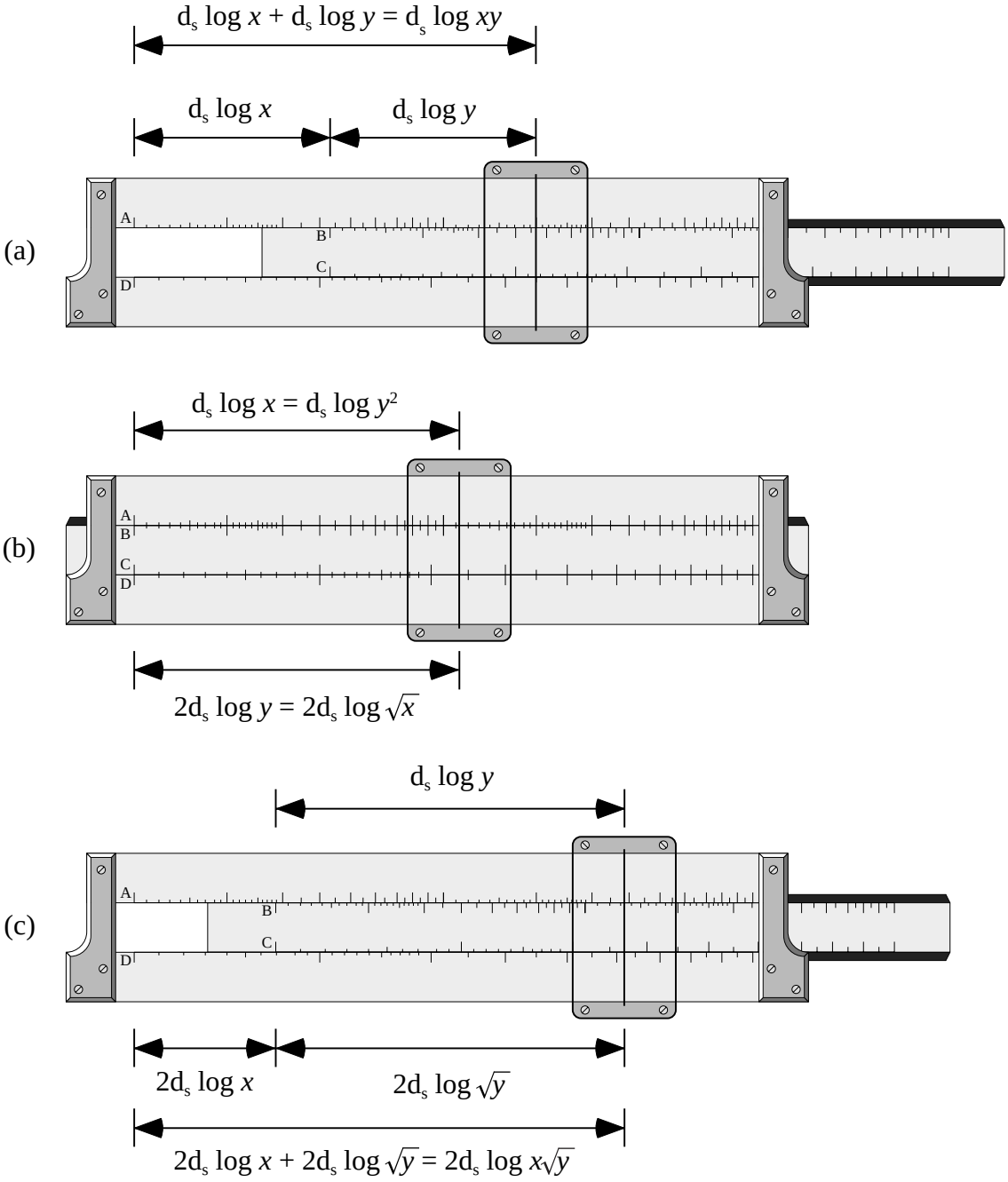


Figure 2.6. Examples of operations that we can perform on a logarithmic slide rule. Here, d_s is the scale unit of the A and B scales; if natural logarithms are used, d_s is the distance from the origin to *e* on the A and B scales. (a) We can multiply two numbers, *x* and *y*, by adding distances on two scales, A and B, with the same scale units by moving one scale relative the other. (b) We can square or square root a number by marking off equal distances on two scales, A and D, with scale units in a ratio of 1 to 2; here, d_s and $2d_s$. (c) We can perform operations such as $x\sqrt{y}$ by adding distances on two scales, B and D, with scale units in a ratio of 1 to 2, by moving one scale relative to the other.

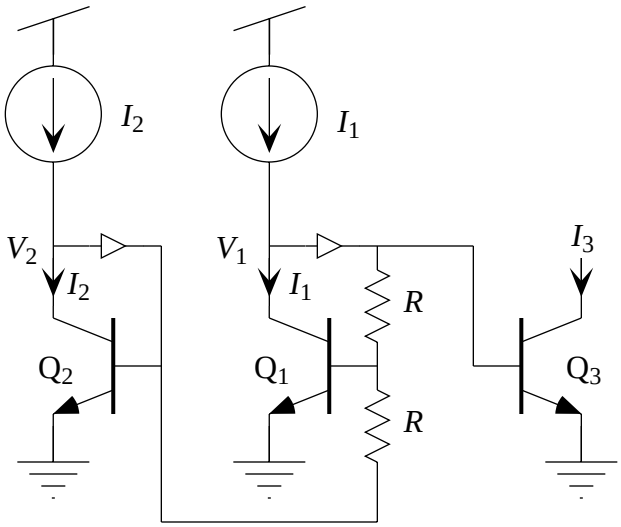


Figure 2.7. Schematic of a circuit of the same form as that of Figure 2.5. Here, the MITEs are implemented with resistive voltage dividers to generate the weighted summation, and with bipolar transistors to generate the exponential currents. The collector voltages, V_1 and V_2, are buffered into the resistive network through unity-gain voltage followers. This particular implementation of the MITE is valid at current levels such that the base resistance of the bipolars is much greater than R. For this circuit, w_{22}=w_{31}=1 and w_{11}=w_{12}=$\frac{R}{2R}=\frac{1}{2}$. From Equation 2.12, I expect I_1 to factor into I_3 raised to the power $\frac{w_{31}}{w_{11}}=\frac{1}{1/2}=2$. Moreover, I expect I_2 to factor into I_3 raised to the power $-\frac{w_{31}}{w_{11}}\frac{w_{12}}{w_{22}}=-\frac{1}{1/2}\frac{1/2}{1}=-1$. Combining these results, I have that

$$I_3=\frac{I_1^2}{I_2}.$$

Thus, this circuit is a squaring-reciprocal circuit.

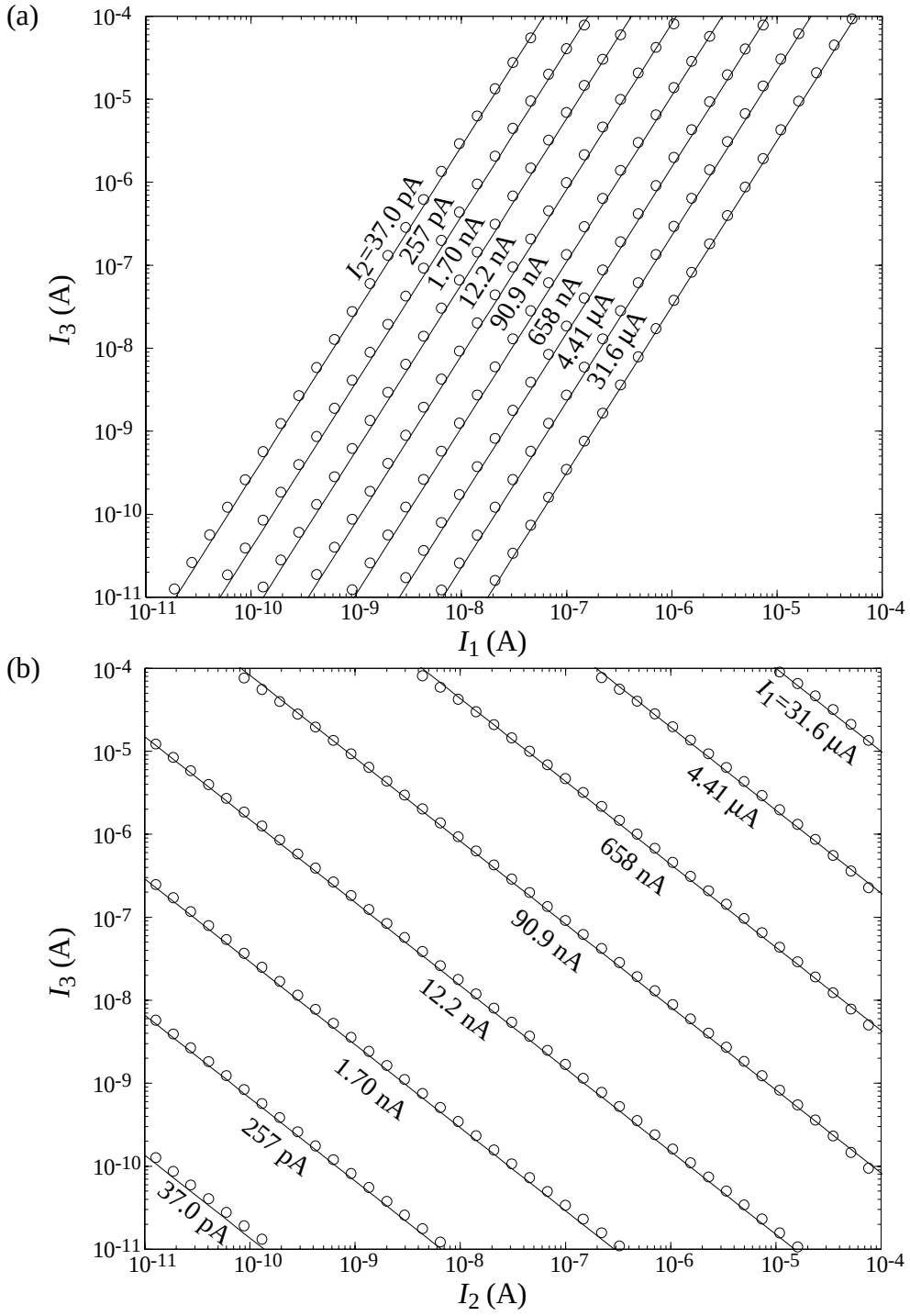


Figure 2.8. Measured data from the circuit of Figure 2.7. Circles are measured values of I_3 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1. Solid lines show the ideal expression, $I_3 = I_1^2 \div I_2$, calculated for the values of I_1 and I_2 at each point.

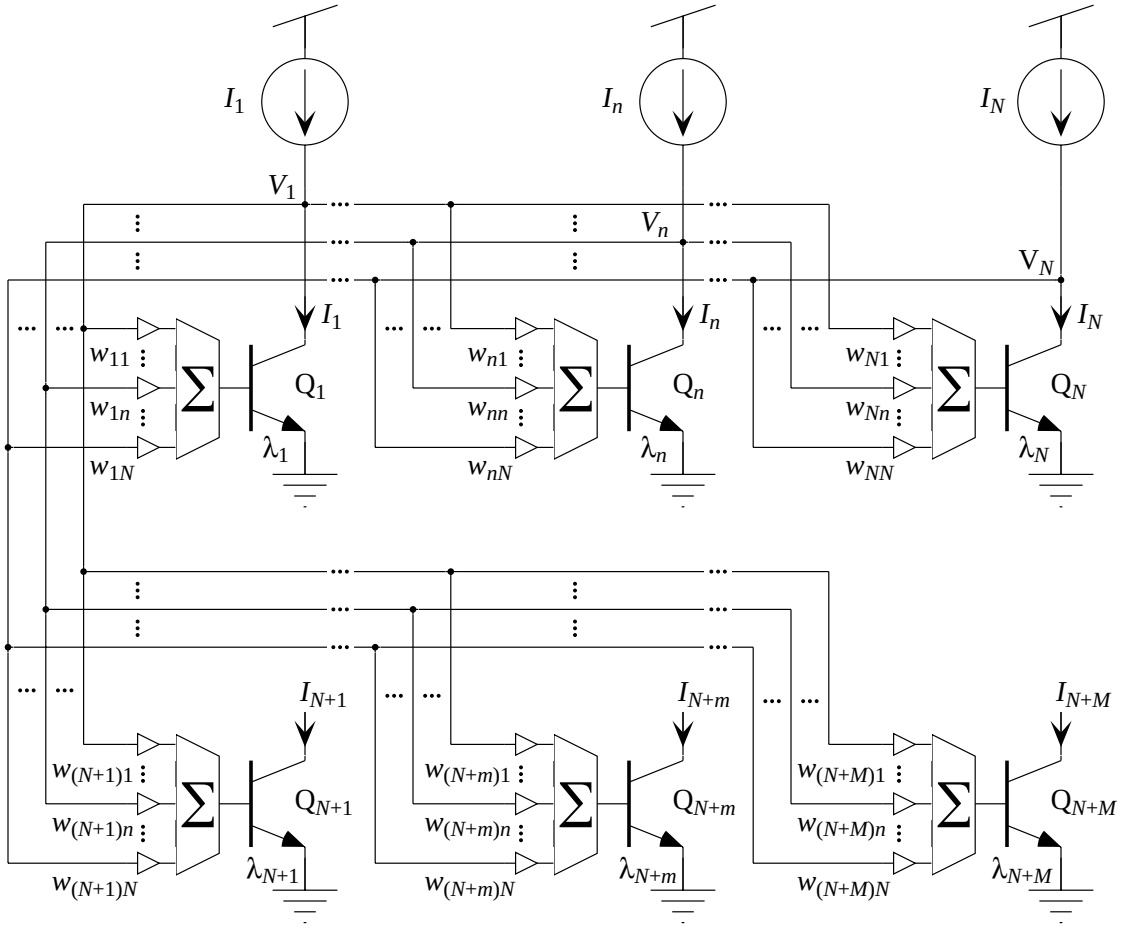


Figure 2.9. Schematic of a generalized translinear circuit comprising N input MITEs, labeled Q_1 through Q_N, and M output MITEs, labeled Q_{N+1} through Q_{N+M}. Input currents, I_1 through I_N, are sourced into the outputs of Q_1 through Q_N, respectively, causing voltages, V_1 through V_N, to develop that will depend on the $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$. $\mathbf{W}_{\text{in}}$ comprises the coupling strengths w_{nk}, where both n and k can take on integer values from 1 to N; the value of w_{nk} is a measure of the coupling strength between the output of Q_k and the weighted sum of Q_n. The circuit forms M output currents, I_{N+1} through I_{N+M}, by linearly combining the N voltages, V_1 through V_N, according to the $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, and exponentiating. $\mathbf{W}_{\text{out}}$ comprises the coupling strengths $w_{(N+m)k}$, where m can take on integer values from 1 to M, and k can take on integer values from 1 to N; the value of $w_{(N+m)k}$ is a measure of the coupling strength between the output of Q_k and the weighted sum of Q_{N+m}.

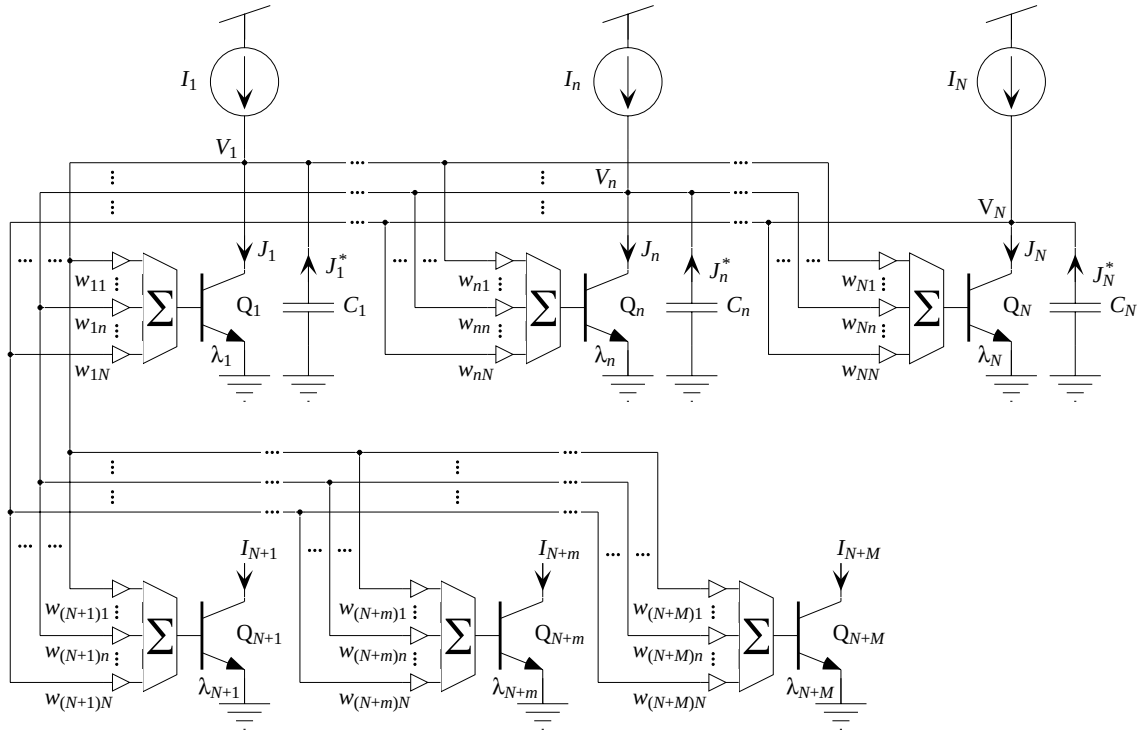


Figure 2.10. Circuit of Figure 2.9 with capacitances C_1 through C_N added to nodes labeled V_1 through V_N, respectively. Capacitance C_n models the interconnect capacitance of node V_n as well as the input capacitances of all the MITEs to which V_n connects. If the input connectivity matrix, $\mathbf{W}_{\text{in}}$, is nonsingular, the circuit has a single fixed point when, for each n between 1 and N, the current sunk by MITE Q_n balances input current I_n. I would like to asses the stability of this fixed point for a constant set of input currents, I_1 through I_N. Expressed in terms of the capacitor currents J_1^* through J_N^*, for which the fixed point is given by $J_1^* = ... = J_N^* = 0$, the circuit is governed by the following system of equations:

$$\begin{cases} \tau_n \dot{J}_n^* = -\sum_{k=1}^{N} w_{nk} \frac{C_n}{C_k} J_k^* - \sum_{k=1}^{N} w_{nk} \frac{C_n}{C_k} \frac{J_k^* J_n^*}{I_n}, & 1 \le n \le N, \\ I_{N+m} = \mathrm{I_s}^{1-\sum_{n=1}^{N} \Lambda_{mn}} K_m \prod_{n=1}^{N} \left(I_n + J_n^* \right)^{\Lambda_{mn}}, & 1 \le m \le M. \end{cases}$$

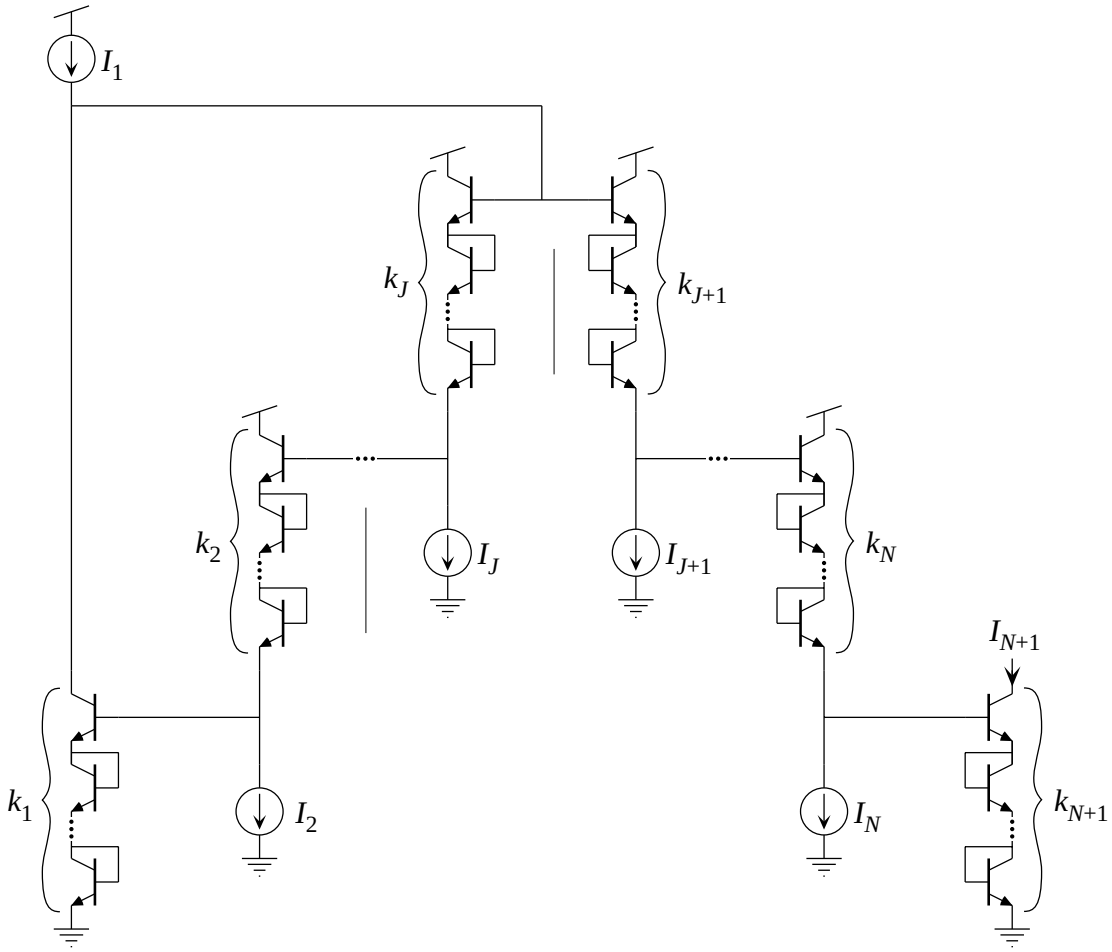
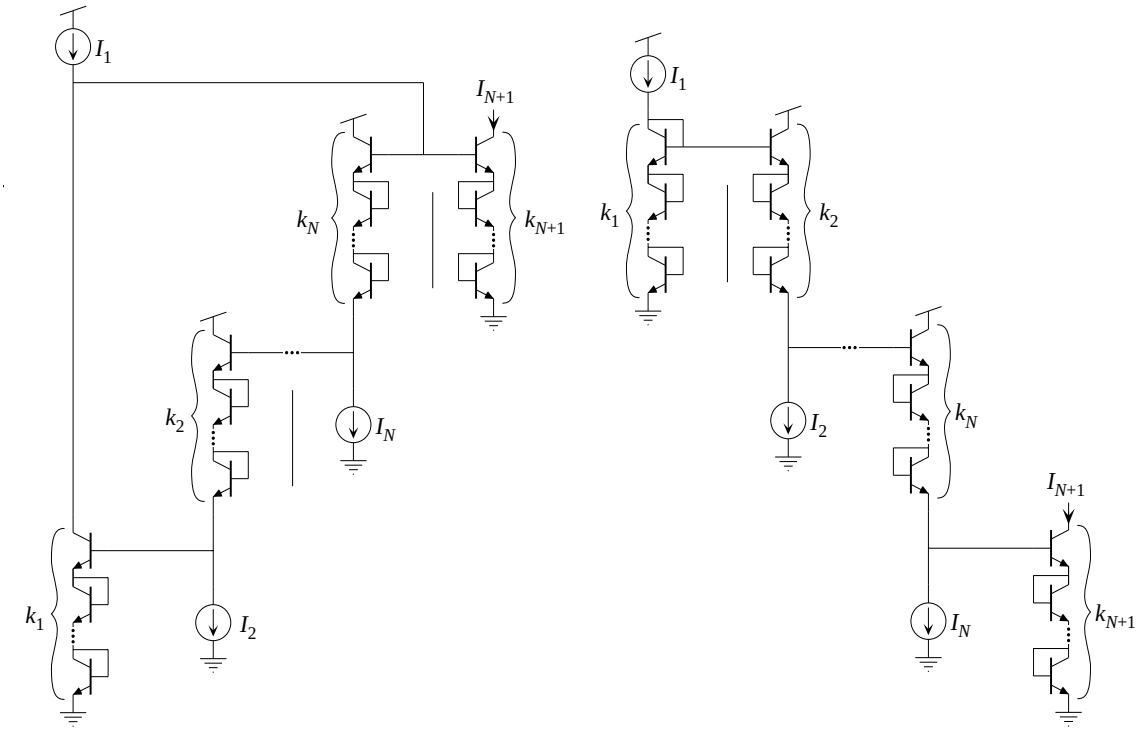


Figure 2.11. A single translinear loop with a stacked topology that implements a relationship of the form

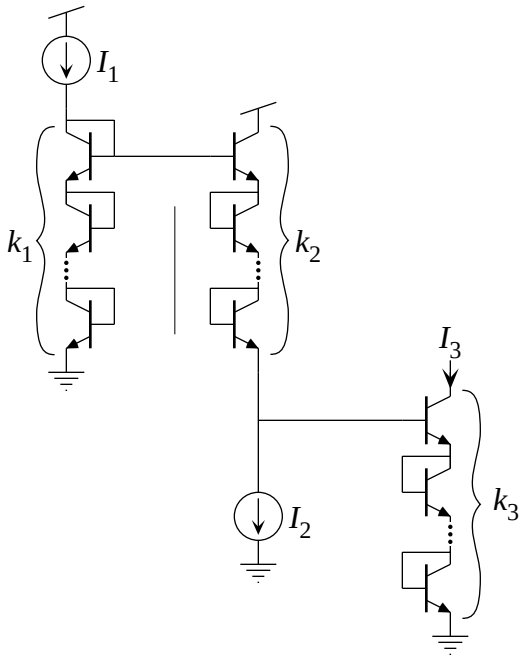
$$I_{N+1} = \prod_{n=1}^{N} I_n^{\Lambda_n},$$

where the values of Λ_n are rational numbers such that Λ_1 through Λ_J are positive and Λ_{J+1} through Λ_N are negative.

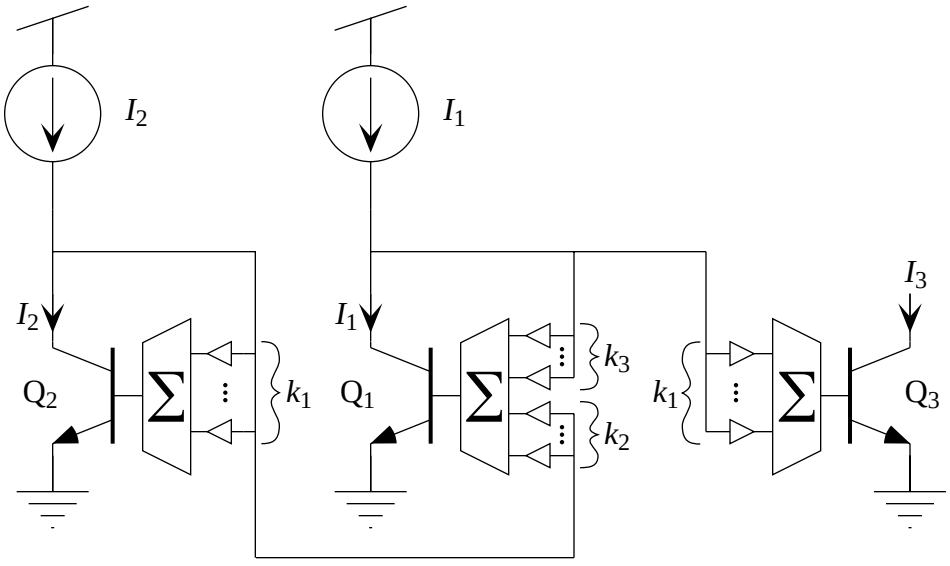


(a) (b)

Figure 2.12. Special cases of the translinear loop shown in Figure 2.11. (a) Translinear loop corresponding to the case when the values of Λ_n are all positive and form a partition of unity. In this case, $J=N$. (b) Translinear loop corresponding to the case when the values of Λ_n are all negative except for one. In this case, $J=1$.



(a)



(b)

Figure 2.13. Implementation of the relationship

$$I_3 = \frac{I_1^{k_1/k_3}}{I_2^{k_2/k_3}}$$

with (a) a translinear loop and (b) a MITE network. In the MITE network, all weighting coefficients have identical values.

(This page has been left blank intentionally.)

CHAPTER 3
SIGNAL-FLOW–GRAPH ANALYSIS OF MULTIPLE-INPUT TRANSLINEAR ELEMENT NETWORKS

In this chapter, I develop a procedure that we can use to analyze a MITE network directly by inspection of the circuit schematic. This analysis technique is based on the theory of linear signal-flow graphs. We can use a linear-analysis technique to determine the steady-state behavior of a class of highly nonlinear circuits because linear relationships hold among the logarithms of the input and output currents in these circuits. Because $x \log y = \log y^x$ and $\log x + \log y = \log xy$, these *linear* relationships that hold among the logarithms of the input and output currents *trans*late into product-of-power-law relationships among the currents themselves. Hence, MITE networks are *translinear* in the second sense of the word that I discussed in Chapter 1. The familiar principle of linear superposition manifests itself in the translinear domain as the relation *is proportional to*. Superposition allows us to consider separately how each input current affects a given output current. In the event that an input current can affect an output current in more than one way, superposition allows us to analyze each chain of cause-and-effect individually; at the end, we simply add the results.

In Section 3.1, I provide necessary background material from the theory of signal-flow graphs. Then, in Section 3.2, I set up the problem of analyzing the steady-state behavior of a MITE translinear circuit using linear signal-flow graphs; I observe several redundancies that allow us to simplify the analysis procedure and to work with a reduced signal-flow graph. In Section 3.3, I describe the construction, directly from the circuit schematic, of the reduced signal-flow graph associated with a MITE translinear circuit. Finally, in Section 3.4, I present several simple examples illustrating the signal-flow–graph analysis technique.

3.1. Signal-Flow Graphs

The notion of a signal-flow graph was first delineated by Samuel Mason in 1953 [1], and was developed subsequently by Mason [2, 3] and other researchers [4–8]. A **signal-flow**

graph is a directed graph whose nodes represent quantities of interest, which are called **node signals**. Branch *ji*, which goes to node *j* from node *i*, implies the existence of a functional dependance of the signal associated with node *j* on the signal associated with node *i*; branch *ji* is indicated graphically by an arc joining nodes *i* and *j* with an arrowhead indicating the direction of signal flow. A signal-flow graph is a graphical representation that allows us to visualize the chains of functional dependencies among various quantities that are of interest to us. We can think of a signal-flow graph as a collection of simple interconnected processing elements, each of which receives information along various incoming branches, combines this information in some way, and transmits the result on each outgoing branch.

Consider the signal-flow graph shown in Figure 3.1a. This signal-flow graph has five nodes, which are numbered 1 though 5, and eight branches, which are indexed by 21, 23, 32, 34, 42, 43, 53, and 54. If I denote by x_i the signal associated with node *i*, then the signal-flow graph of Figure 3.1a implies the following set of functional dependencies among node signals x_1 through x_5:

$$\begin{cases} x_1 = x_1, \\ x_2 = f_2(x_1, x_3), \\ x_3 = f_3(x_2, x_4), \\ x_4 = f_4(x_2, x_3), \\ x_5 = f_5(x_3, x_4). \end{cases} \tag{3.1}$$

In Figures 3.1b through 3.1e, I show subgraphs of the signal-flow graph of Figure 3.1a that correspond, respectively, to the equations for node signals x_2 through x_5. For example, the signal-flow graph of Figure 3.1b, which corresponds to the equation for x_2, asserts that x_2 depends both on x_1 and on x_3, as indicated by the presence of branches 21 and 23 in this graph. I do not show the signal-flow graph corresponding to the equation for x_1 because it comprises a single, isolated node.

Following Mason [1, 2], I now define certain terms and illustrate them with the signal-flow graph of Figure 3.1a. A **source** is a node that has only outgoing branches (e.g., node 1 in Figure 3.1a). A **sink** is a node that has only incoming branches (e.g., node 5 in Figure 3.1a). The node signals of sources correspond to independent variables or input quantities, whereas the node signals of sinks are typically output quantities. A **path** between node *i* and node *j* is any sequence of branches, which, when traversed in the indicated directions, joins nodes *i* and *j*. A **forward path** is a path from a source to a sink along which no node is encountered more than once (e.g., in Figure 3.1a,

1→2→3→4→5, 1→2→4→5, 1→2→3→5, and 1→2→4→3→5). A **feedback loop** is a path that forms a closed loop along which no node is encountered more than once per cycle (e.g., 2→3→2, 3→4→3, and 2→4→3→2, but *not* 2→3→4→3→2, in the graph of Figure 3.1a). Paths that have no nodes in common are said to be **nontouching** (e.g., path 4→5 does not touch feedback loop 2→3→2 in the graph Figure 3.1a).

If I restrict the relationships among node signals of a signal-flow graph to be linear, then I obtain a **linear signal-flow graph**. For example, a linear signal-flow graph with the same topology as that of Figure 3.1a is shown in Figure 3.2. Now, I associate a number with each branch, called the **branch gain**; the gain of branch *ji* indicates by how much the signal at node *j* changes in response to a unit change in the signal at node *i*. In Figure 3.2, I denote the gain of branch *ji* by b_{ji}. The equations relating node signals x_1 through x_5, corresponding to the signal-flow–graph shown in Figure 3.2, are as follows:

$$\begin{cases} x_1 = x_1, \\ x_2 = b_{21}x_1 + b_{23}x_3, \\ x_3 = b_{32}x_2 + b_{34}x_4, \\ x_4 = b_{42}x_2 + b_{43}x_3, \\ x_5 = b_{53}x_3 + b_{54}x_4. \end{cases} \tag{3.2}$$

The specialization to linear signal-flow graphs gives us an elegant formula, called **Mason's gain formula**, for calculating various gains in linear systems and for solving systems of linear equations directly by inspecting the associated signal-flow graph. Before I can state Mason's gain formula, however, I need to define several more terms; I illustrate these new terms with the linear signal-flow graphs of Figure 3.2. A **path gain** is the product of all the branch gains along that path (e.g., in Figure 3.2, the gain of path 1→2→4→5 is equal to $b_{21}b_{42}b_{54}$, and the gain of path 1→2→3→4→5 is equal to $b_{21}b_{32}b_{43}b_{54}$). The **loop gain** of a feedback loop is the product of the gains of the branches that the loop comprises (e.g., in Figure 3.2, the gain of loop 2→3→2 is equal to $b_{32}b_{23}$, and the gain of loop 2→4→3→2 is equal to $b_{42}b_{34}b_{23}$). The **signal-flow–graph gain** from source *i* to sink *j* is the amount by which the signal at node *j* changes in response to a unit change in the signal at node *i*. Note that we need to consider only one sink and one source at a time, because sources are superposable and sinks are independent of one another.

An important quantity associated with a linear signal-flow graph is the **signal-flow–graph determinant**; it is the denominator of each of a signal-flow–graph's gains. The determinant of a signal-flow graph is similar in many respects to the determinant of a

matrix. For instance, the determinant of a signal-flow graph is invariant if we renumber the nodes of that signal-flow graph [1, p. 1153] in the same way that the determinant of a matrix is invariant if we renumber the rows and columns of that matrix. The determinant of a signal-flow graph is given by

$$\Delta = 1 - \sum_j T_j + \sum_{i \not\leftrightarrow j} T_i T_j - \sum_{\substack{i \not\leftrightarrow j \\ i \not\leftrightarrow k \\ j \not\leftrightarrow k}} T_i T_j T_k + \ldots,$$

where T_i is the gain of the ith feedback loop, and $i \not\leftrightarrow j$ means that loop i and loop j do not touch each other. To illustrate the calculation of the signal-flow–graph determinant, I compute the determinant of the signal-flow graph shown in Figure 3.2. First, I identify all the feedback loops in the graph of Figure 3.2 and their respective gains; there are three such loops:

Loop	**Gain**
$2 \rightarrow 3 \rightarrow 2$	$T_1 = b_{32} b_{23}$
$3 \rightarrow 4 \rightarrow 3$	$T_2 = b_{43} b_{34}$
$2 \rightarrow 4 \rightarrow 3 \rightarrow 2$	$T_3 = b_{42} b_{34} b_{23}$

In this case, the calculation is straightforward, because all the feedback loops touch each other; the signal-flow–graph determinant, Δ, is given by

$$\begin{aligned} \Delta &= 1 - \sum_j T_j + \sum_{i \not\leftrightarrow j} T_i T_j - \ldots \\ &= 1 - \left(T_1 + T_2 + T_3 \right) \\ &= 1 - b_{32} b_{23} - b_{43} b_{34} - b_{42} b_{34} b_{23}. \end{aligned} \tag{3.3}$$

A fact about signal-flow–graph determinants that will prove useful involves signal-flow graphs whose feedback loops form nontouching subgraphs. We find the **loop subgraph** of any signal-flow graph by removing all branches that are not part of feedback loops. The useful fact is that *the determinant of the complete signal-flow graph is equal to the product of the determinants of each of the nontouching parts of the loop subgraph* [3, pp. 111–112]. This result is illustrated in Figure 3.3. An analogous fact about determinants of block triangular matrices is that the determinant of a block triangular matrix is the product of the determinants of the diagonal blocks.

Now, I am in a position to state Mason's gain formula.

Mason's gain formula: The gain of a signal-flow graph from source i to sink j is given by

$$G_{ji} = \frac{1}{\Delta}\sum_k G_k \Delta_k ,$$

where G_k is the gain of the kth forward path joining nodes i and j; Δ is the determinant of the signal-flow graph; and Δ_k is the **cofactor** of the kth forward path, and is defined to be the determinant of the part of the signal-flow graph that does not touch the kth forward path.

Proof: See Mason [2, pp. 925–926]. ■

Now, I illustrate the use of Mason's gain formula by calculating the gain of the graph of Figure 3.2 from node 1 to node 5; this result is equivalent to the result that I would obtain by solving Equation 3.2 for x_5 in terms of x_1 (eliminating x_2, x_3, and x_4), and calculating the ratio of x_5 to x_1. First, I identify all the forward paths from node 1 to node 5 in the signal-flow graph of Figure 3.2 and their respective gains; there are four such paths:

Path	**Gain**
$1\rightarrow 2\rightarrow 3\rightarrow 4\rightarrow 5$	$G_1 = b_{21}b_{32}b_{43}b_{54}$
$1\rightarrow 2\rightarrow 4\rightarrow 5$	$G_2 = b_{21}b_{42}b_{54}$
$1\rightarrow 2\rightarrow 3\rightarrow 5$	$G_3 = b_{21}b_{32}b_{53}$
$1\rightarrow 2\rightarrow 4\rightarrow 3\rightarrow 5$	$G_4 = b_{21}b_{42}b_{34}b_{53}$

The calculation of the cofactor of each forward path is simple, because all the feedback loops touch all the feedforward paths; consequently, $\Delta_k = 1$ for all k. I just calculated the determinant of this signal-flow graph; its value is given in Equation 3.3. Thus, by Mason's gain formula, I write that

$$\begin{aligned} G_{51} &= \frac{1}{\Delta}\sum_k G_k \Delta_k \\ &= \frac{1}{\Delta}\left(G_1\Delta_1 + G_2\Delta_2 + G_3\Delta_3 + G_4\Delta_4\right) \\ &= \frac{b_{21}b_{32}b_{43}b_{54} + b_{21}b_{42}b_{54} + b_{21}b_{32}b_{53} + b_{21}b_{42}b_{34}b_{53}}{1 - b_{32}b_{23} - b_{43}b_{34} - b_{42}b_{34}b_{23}} . \end{aligned}$$

In signal-flow graphs that have no feedback loops, the application of Mason's gain formula is particularly straightforward: We simply identify all the forward paths from the source to the sink for which we want to calculate the signal-flow–graph gain, and we add their respective gains.

3.2. Signal-Flow–Graph Analysis of Multiple-Input Translinear Element Networks

In this section, I show how to analyze the steady-state behavior of a MITE translinear circuit using linear signal-flow graphs. Here, I restrict my attention to those MITE networks in which all input MITEs are diode connected. In terms of the matrix analysis of Section 2.5, I require that all the entries along the main diagonal of the input connectivity matrix, $\mathbf{W}_{\text{in}}$, be nonzero. Aside from the fact my signal-flow–graph construction and analysis procedures require it, I have two reasons for restricting my attention to this class of MITE networks. First, MITE networks that do not have all input MITEs diode connected violate the sufficient conditions for asymptotic stability that I found in Section 2.6. Because these conditions are not necessary, a MITE network that violates them still can be stable, but we would have to address the question of stability on a case-by-case basis. Second, in Chapter 4, I show by construction that any expression that can be implemented by a MITE network can be embodied in a MITE network in which all input MITEs are diode connected. Consequently, we have no need to work with MITE networks that do not have diode-connected input MITEs.

Using the same terminology I did as in Section 2.5 and in Figure 2.9, I express the behavior of a MITE network comprising N input MITEs, Q_1 through Q_N, and M output MITEs, Q_{N+1} through Q_{N+M}, with $N+M$ coupled equations, one for each MITE, as follows:

$$I_n = \lambda_n \mathrm{I_s} \exp\left[\sum_{k=1}^{N} \frac{w_{nk} V_k}{\mathrm{U_T}}\right],\ 1 \le n \le N+M. \tag{3.4}$$

By introducing an auxiliary voltage variable, U_n, for each MITE, I rewrite Equation 3.4 as

$$U_n = \sum_{k=1}^{N} w_{nk} V_k\ ,\ 1 \le n \le N+M \tag{3.5}$$

and

$$U_n = \mathrm{U_T} \log \frac{I_n}{\lambda_n \mathrm{I_s}},\ 1 \le n \le N+M. \tag{3.6}$$

The system given by Equation 3.5 is linear, so it seems that I could simply draw the signal-flow graph that corresponds to it, use Mason's gain formula on it to solve the system, and express the resulting linear voltage equations in terms of the currents, I_1 through I_{N+M}, using the logarithmic relationships in Equation 3.6. However, the linear equations for U_n in Equation 3.5 do not have the appropriate **cause-and-effect form**. In other words, the system given by Equation 3.5 is formulated such that the voltages V_k are the causes, and the voltages U_n are the effects. However, a MITE network functions such

that input currents I_1 through I_N (or equivalently voltages U_1 through U_N) are causes giving rise to voltages V_1 through V_N, which are intermediate effects. In turn, these voltages give rise to output currents I_{N+1} through I_{N+M} (through the voltages U_{N+1} through U_{N+M}). In terms of the signal-flow graph for the system given by Equation 3.5, the U_n are all sink nodes; instead, I want U_1 through U_N to be source nodes and U_{N+1} through U_{N+M} to be sink nodes.

I can reformulate Equation 3.5 so that it has the proper cause-and-effect form as follows. First, I separate the equations for the inputs from the equations for the outputs, and write Equation 3.5 as

$$\begin{cases} U_n = \sum_{k=1}^{N} w_{nk} V_k, & 1 \le n \le N, \\ U_{N+m} = \sum_{k=1}^{N} w_{(N+m)k} V_k, & 1 \le m \le M. \end{cases}$$

Next, I break out the term $w_{nn}V_n$ from the summation for each U_n, $1 \le n \le N$, and rearrange to obtain

$$\begin{cases} w_{nn} V_n = U_n - \sum_{k \ne n} w_{nk} V_k, & 1 \le n,k \le N, \\ U_{N+m} = \sum_{k=1}^{N} w_{(N+m)k} V_k, & 1 \le m \le M. \end{cases} \tag{3.7}$$

Finally, I substitute[1] $V_n^* = w_{nn} V_n$, $1 \le n \le N$, into Equation 3.7, and get

$$\begin{cases} V_n^* = U_n - \sum_{k \ne n} \frac{w_{nk}}{w_{kk}} V_k^*, & 1 \le n,k \le N, \\ U_{N+m} = \sum_{k=1}^{N} \frac{w_{(N+m)k}}{w_{kk}} V_k^*, & 1 \le m \le M. \end{cases} \tag{3.8}$$

Equation 3.8 does have the right cause-and-effect form. To analyze the MITE network to which they correspond, I draw the signal-flow graph corresponding to Equation 3.8, and apply Mason's gain formula to calculate the signal-flow–graph gain between U_n and U_{N+m}, which I denote by Λ_{mn}, for each n between 1 and N and for each m between 1 and M. The result of this endeavor will be a set of M linear equations for each U_{N+m} in terms of the U_n as follows:

$$U_{N+m} = \sum_{n=1}^{N} \Lambda_{mn} U_n, \; 1 \le m \le M. \tag{3.9}$$

By substituting Equation 3.6 into Equation 3.9, I get

$$\mathrm{U_T} \log \frac{I_{N+m}}{\lambda_{N+m} \mathrm{I_s}} = \sum_{n=1}^{N} \Lambda_{mn} \mathrm{U_T} \log \frac{I_n}{\lambda_n \mathrm{I_s}}, \; 1 \le m \le M.$$

[1] I am grateful to Paul Hasler for suggesting this substitution.

If all MITEs are operating at the same temperature, the preceding equations become

$$\log \frac{I_{N+m}}{\lambda_{N+m} \mathrm{I_s}} = \sum_{n=1}^{N} \Lambda_{mn} \log \frac{I_n}{\lambda_n \mathrm{I_s}}, \; 1 \le m \le M. \tag{3.10}$$

Because $x \log y = \log y^x$ and $\log x + \log y = \log xy$, I write Equation 3.10 as

$$\log \frac{I_{N+m}}{\lambda_{N+m} \mathrm{I_s}} = \log \prod_{n=1}^{N} \left(\frac{I_n}{\lambda_n \mathrm{I_s}} \right)^{\Lambda_{mn}}, \; 1 \le m \le M. \tag{3.11}$$

By exponentiating both sides of Equation 3.11, I obtain

$$\frac{I_{N+m}}{\lambda_{N+m} \mathrm{I_s}} = \prod_{n=1}^{N} \left(\frac{I_n}{\lambda_n \mathrm{I_s}} \right)^{\Lambda_{mn}}, \; 1 \le m \le M,$$

which after rearranging becomes

$$I_{N+m} = \mathrm{I_s}^{1-\sum_{n=1}^{N} \Lambda_{mn}} \lambda_{N+m} \prod_{n=1}^{N} \lambda_n^{-\Lambda_{mn}} \prod_{n=1}^{N} I_n^{\Lambda_{mn}}, \; 1 \le m \le M. \tag{3.12}$$

Applying the definition of K_m from Equation 2.14 to Equation 3.12, I have that

$$I_{N+m} = \mathrm{I_s}^{1-\sum_{n=1}^{N} \Lambda_{mn}} K_m \prod_{n=1}^{N} I_n^{\Lambda_{mn}}, \; 1 \le m \le M,$$

which is identical to the result expressed in Equation 2.20 that I obtained in Section 2.5.

To make this process more concrete, I now apply it to the MITE network fragment shown in Figure 3.4. This circuit fragment comprises a single output MITE, Q_j, and four diode-connected input MITEs: Q_i, Q_k, Q_m, and Q_n. The full circuit could have additional input and output MITEs; however, I assume that all connections involving voltages V_i, V_k, V_m, and V_n are shown in Figure 3.4. This assumption suffices to guarantee that the signal-flow–graph determinant will factor into the product of the determinant of the signal-flow graph corresponding to the circuit fragment of Figure 3.4 and the determinant of the signal-flow graph corresponding to whatever circuitry is not shown. Likewise, the cofactor of the *k*th forward path in the graph corresponding to the circuit fragment shown will factor into the product of the cofactor of the *k*th forward path in the signal-flow graph of the part shown and the determinant of the rest of the graph.

Now, I can describe the steady-state behavior of the circuit fragment of Figure 3.4 with the following sets of equations:

$$\begin{cases} U_i = w_{ii} V_i + w_{ik} V_k + w_{in} V_n + \ldots, \\ U_k = w_{kk} V_k + w_{ki} V_i + \ldots, \\ U_m = w_{mm} V_m + w_{mn} V_n + \ldots, \\ U_n = w_{nn} V_n + \ldots, \\ U_j = w_{jm} V_m + w_{ji} V_i + \ldots, \end{cases} \tag{3.13}$$

and

$$\begin{cases} U_i = \mathrm{U_T} \log \dfrac{I_i}{\lambda_i \mathrm{I_s}}, \\ U_k = \mathrm{U_T} \log \dfrac{I_k}{\lambda_k \mathrm{I_s}}, \\ U_m = \mathrm{U_T} \log \dfrac{I_m}{\lambda_m \mathrm{I_s}}, \\ U_n = \mathrm{U_T} \log \dfrac{I_n}{\lambda_n \mathrm{I_s}}, \\ U_j = \mathrm{U_T} \log \dfrac{I_j}{\lambda_j \mathrm{I_s}}. \end{cases} \tag{3.14}$$

Next, I put Equation 3.13 in proper cause-and-effect form using the procedure just described, and I get

$$\begin{cases} V_i^* = U_i - \frac{w_{ik}}{w_{kk}} V_k^* - \frac{w_{in}}{w_{nn}} V_n^* - ..., \\ V_k^* = U_k - \frac{w_{ki}}{w_{ii}} V_i^* - ..., \\ V_m^* = U_m - \frac{w_{mn}}{w_{nn}} V_n^* - ..., \\ V_n^* = U_n - ..., \\ U_j = \frac{w_{jm}}{w_{mm}} V_m^* + \frac{w_{ji}}{w_{ii}} V_i^* + ... \end{cases} \tag{3.15}$$

The signal-flow graph corresponding to Equation 3.15 is shown in Figure 3.5a. There is a single feedback loop in this signal-flow graph, $V_i^* \to V_k^* \to V_i^*$, whose gain is given by

$$\left(-\frac{w_{ik}}{w_{kk}}\right)\left(-\frac{w_{ki}}{w_{ii}}\right) = \frac{w_{ik}}{w_{kk}} \frac{w_{ki}}{w_{ii}}.$$

Because there is only one feedback loop and, as I just argued, the determinant of the graph corresponding to the whole circuit is factorable, I express the overall determinant as

$$\Delta = \Delta_R \left(1 - \frac{w_{ik}}{w_{kk}} \frac{w_{ki}}{w_{ii}}\right),$$

where Δ_R is the determinant of whatever signal-flow graph corresponds to the rest of the circuit of Figure 3.4.

Now, I evaluate the signal-flow–graph gain from each source node shown to node U_j. There is a single forward path from node U_i to node U_j, $U_i \to V_i^* \to U_j$, whose gain is given by $\frac{w_{ji}}{w_{ii}}$. This path touches the feedback loop; consequently, the gain from node U_i to node U_j is given by

$$\Lambda_{ji} = \frac{\frac{w_{ji}}{w_{ii}} \Delta_R}{\Delta_R \left(1 - \frac{w_{ik}}{w_{kk}} \frac{w_{ki}}{w_{ii}}\right)} = \frac{\frac{w_{ji}}{w_{ii}}}{1 - \frac{w_{ik}}{w_{kk}} \frac{w_{ki}}{w_{ii}}}.$$

There is a single forward path from node U_k to node U_j, $U_k \to V_k^* \to V_i^* \to U_j$, whose gain is given by $-\frac{w_{ik}}{w_{kk}} \frac{w_{ji}}{w_{ii}}$. This path touches the feedback loop; consequently, the gain from node U_k to node U_j is given by

$$\Lambda_{jk} = \frac{-\frac{w_{ik}}{w_{kk}}\frac{w_{ji}}{w_{ii}}\Delta_R}{\Delta_R\left(1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}\right)} = \frac{-\frac{w_{ik}}{w_{kk}}\frac{w_{ji}}{w_{ii}}}{1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}}.$$

There is a single forward path from node U_m to node U_j, $U_m \rightarrow V_m^* \rightarrow U_j$, whose gain is given by $\frac{w_{jm}}{w_{mm}}$. This path does not touch the feedback loop; consequently, the gain from node U_m to node U_j is given by

$$\Lambda_{jm} = \frac{\frac{w_{jm}}{w_{mm}}\Delta_R\left(1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}\right)}{\Delta_R\left(1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}\right)} = \frac{w_{jm}}{w_{mm}}.$$

There are two forward paths from node U_n to node U_j: $U_n \rightarrow V_n^* \rightarrow V_m^* \rightarrow U_j$, whose gain is given by $-\frac{w_{mn}}{w_{nn}}\frac{w_{jm}}{w_{mm}}$ and $U_n \rightarrow V_n^* \rightarrow V_i^* \rightarrow U_j$, whose gain is given by $-\frac{w_{in}}{w_{nn}}\frac{w_{ji}}{w_{ii}}$. The former path does not touch the feedback loop whereas the latter path does; consequently, the gain from node U_n to node U_j is given by

$$\Lambda_{jn} = \frac{-\frac{w_{mn}}{w_{nn}}\frac{w_{jm}}{w_{mm}}\Delta_R\left(1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}\right)}{\Delta_R\left(1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}\right)} + \frac{-\frac{w_{in}}{w_{nn}}\frac{w_{ji}}{w_{ii}}\Delta_R}{\Delta_R\left(1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}\right)} = -\frac{w_{mn}}{w_{nn}}\frac{w_{jm}}{w_{mm}} - \frac{-\frac{w_{in}}{w_{nn}}\frac{w_{ji}}{w_{ii}}}{1-\frac{w_{ik}}{w_{kk}}\frac{w_{ki}}{w_{ii}}}.$$

Now, I simply superpose the sources to get the total response of U_j as follows:

$$U_j = \Lambda_{ji}U_i + \Lambda_{jk}U_k + \Lambda_{jm}U_m + \Lambda_{jn}U_n + ... \tag{3.16}$$

I then substitute Equation 3.14 into Equation 3.16 to obtain

$$\mathrm{U_T}\log\frac{I_j}{\lambda_j \mathrm{I_s}} = \Lambda_{ji}\mathrm{U_T}\log\frac{I_i}{\lambda_i \mathrm{I_s}} + \Lambda_{jk}\mathrm{U_T}\log\frac{I_k}{\lambda_k \mathrm{I_s}} + \Lambda_{jm}\mathrm{U_T}\log\frac{I_m}{\lambda_m \mathrm{I_s}} + \Lambda_{jn}\mathrm{U_T}\log\frac{I_n}{\lambda_n \mathrm{I_s}} + ...,$$

which I simplify to

$$\log\frac{I_j}{\lambda_j \mathrm{I_s}} = \log\left[\left(\frac{I_i}{\lambda_i \mathrm{I_s}}\right)^{\Lambda_{ji}}\left(\frac{I_k}{\lambda_k \mathrm{I_s}}\right)^{\Lambda_{jk}}\left(\frac{I_m}{\lambda_m \mathrm{I_s}}\right)^{\Lambda_{jm}}\left(\frac{I_n}{\lambda_n \mathrm{I_s}}\right)^{\Lambda_{jn}} \times ...\right]. \tag{3.17}$$

By exponentiating both sides of Equation 3.17 and rearranging, I obtain

$$I_j = \lambda_j \mathrm{I_s}\left(\frac{I_i}{\lambda_i \mathrm{I_s}}\right)^{\Lambda_{ji}}\left(\frac{I_k}{\lambda_k \mathrm{I_s}}\right)^{\Lambda_{jk}}\left(\frac{I_m}{\lambda_m \mathrm{I_s}}\right)^{\Lambda_{jm}}\left(\frac{I_n}{\lambda_n \mathrm{I_s}}\right)^{\Lambda_{jn}} \times ...,$$

which implies that

$$I_j \propto I_i^{\Lambda_{ji}} I_k^{\Lambda_{jk}} I_m^{\Lambda_{jm}} I_n^{\Lambda_{jn}}.$$

Now, I make four observations about the cause-and-effect form of Equation 3.8, which relate the voltages U_1 through U_N and V_1^* through V_N^* and about the structure of the resulting signal-flow graph. These observations allow me to simplify the structure and use of the signal-flow graph corresponding to Equation 3.8.

1. For each value of n between 1 and N, the node corresponding to U_n feeds exclusively into the node corresponding to V_n^* always through a branch gain of 1.

Because multiplying a path gain by a branch gain of 1 leaves the path gain unaltered, I can effectively collapse each node pair (U_n, V_n^*) into a single node indexed by n. Each of these new nodes serves both as a source from its own standpoint and as an intermediate node from the standpoint of other nodes in the graph that are farther away from a sink. In other words, this new node n assumes the roles formerly played both by node U_n and by node V_n^*.

2. Branches that start on a V^* node and end on another V^* node always have negative branch gains, whereas branches that start on a V^* node and end on a U node (i.e., a sink) always have positive branch gains. However, the magnitudes of the branch gains all have the same form: If i is the index of the starting node and j is the index of the ending node of a branch, then the magnitude of the branch gain is always given by $\frac{w_{ji}}{w_{ii}}$.
3. All forward paths comprise some number of branches between V^* nodes—say, n—and a single branch from a V^* node to a U sink node. Thus, a forward-path gain is given by the product of n negative numbers and a single positive number. Consequently, the sign of a forward-path gain is given by $1 \times (-1)^n$, which is +1 if the number of branches in the forward path is odd and –1 if the number of branches in the forward path is even.
4. All feedback loops occur among V^* nodes. Thus, the gain of a loop composed of n branches is given by the product of n negative numbers. Consequently, the sign of a loop gain is given by $(-1)^n$, which is +1 if the number of branches in the loop is even and –1 if the number of branches in the loop is odd.

The second observation suggests that I can discard the sign of each branch gain in the signal-flow graph; calculating a path gain or loop gain on this new signal-flow graph gives me the magnitude of the path gain or loop gain in the old signal-flow graph. Then, I recover the sign of a forward-path gain or loop gain using the third and fourth observations, respectively.

I refer to the signal-flow graph that results from these two simplifications (i.e., collapsing each (U_n, V_n^*) node pair, for each n between 1 and N, into a single node n, and discarding the sign of each branch gain) as a **reduced signal-flow graph**. Figure 3.5b shows the reduced signal-flow graph corresponding to the signal-flow graph shown in Figure 3.5a. When using Mason's gain formula on a reduced signal-flow graph, I treat each nonsink node both as a source node and as an intermediate node for other nonsink

nodes. I calculate the magnitude of forward-path gains and loop gains as the product of the branch gains that make up the forward path or loop. I assign the sign of a forward-path gain as follows: A path gain is positive if the forward path consists of an odd number of branches, and is negative if the forward path consists of an even number of branches. I assign the sign of a loop gain as follows: A loop gain is negative if the number of branches in the loop is odd, and is positive if the number of branches in the loop is even.

3.3. Reduced Signal-Flow Graph Construction

In this section, I describe how to construct the reduced signal-flow graph that corresponds to any given MITE network that has all input MITEs diode connected directly from the circuit schematic. The process is straightforward. First, I draw a node corresponding to each MITE in the circuit, and index it with the index of the corresponding MITE. I connect these nodes with branches according to the following rule: If MITE Q_i is diode connected and its output connects to the weighted sum of MITE Q_j, then I draw a branch from node i to node j, and label it with the branch gain $\frac{w_{ji}}{w_{ii}}$.

I use the MITE network fragment shown in Figure 3.4 to illustrate this process. First, I draw five nodes, one for each MITE, indexed, respectively, by i, j, k, m, and n. Next, I consider each diode-connected input MITE in turn to determine how to connect the nodes. I begin with MITE Q_i; its output node, V_i, couples into the weighted sum of both MITE Q_j and MITE Q_k. So, as shown in Figure 3.6, I draw a branch from node i to node j, labeling it with the branch gain $\frac{w_{ji}}{w_{ii}}$, and I draw a branch from node i to node k, labeling it with the branch gain $\frac{w_{ki}}{w_{ii}}$. Next, I turn to MITE Q_k; its output couples into the weighted sum of MITE Q_i. Consequently, I draw a branch from node k to node i, labeling it with the branch gain $\frac{w_{ik}}{w_{kk}}$. Next, I consider MITE Q_m; its output couples into the weighted sum of MITE Q_j. Thus, I draw a branch from node m to node j, labeling it with the branch gain $\frac{w_{jm}}{w_{mm}}$. Finally, I consider MITE Q_n; its output connects both to the weighted sum of MITE Q_i and to the weighted sum of MITE Q_m. So, I draw a branch from node n to node i, labeling it with the branch gain $\frac{w_{in}}{w_{nn}}$, and I draw a branch from node n to node m, labeling it with the branch gain $\frac{w_{mn}}{w_{nn}}$. The resulting reduced signal-flow graph is identical to the one shown in Figure 3.5b.

3.4. Examples of Construction and Analysis of Reduced Signal-Flow Graphs

In this section, I consider several simple examples of increasing complexity to illustrate the

process of constructing a reduced signal-flow graph from a MITE network schematic and using that graph to analyze the MITE network's steady-state behavior. In a great many cases, with a little practice, we can dispense with the actual construction of the reduced signal-flow graph and obtain the steady-state response of a MITE network directly by inspection of the circuit schematic.

3.4.1. Single-Branch Paths

Consider the MITE network shown in Figure 3.7a. There are three MITEs, so I begin construction of the reduced signal-flow graph by drawing three nodes, labeling them 1, 2, and 3. MITE Q_1 is diode connected through two unit weighting coefficients, and its output connects to the weighted sum of MITE Q_3 through one unit weighting coefficient, so I draw a branch from node 1 to node 3, labeling it with branch gain $\frac{1}{2}$. Likewise, MITE Q_2 is diode connected through two unit weighting coefficients, and its output connects to the weighted sum of MITE Q_3 through one unit weighting coefficient, so I draw a branch from node 2 to node 3, labeling it with a branch gain of $\frac{1}{2}$. The resulting reduced signal-flow graph is shown in Figure 3.7b.

To analyze the circuit, I first calculate the signal-flow–graph gain from node 1 to node 3. There is one forward path from node 1 to node 3 with path gain $+\frac{1}{2}$, so the graph gain from node 1 to node 3 is simply $+\frac{1}{2}$. Thus, I have that $I_3 \propto \sqrt{I_1}$. Next, I calculate the signal-flow–graph gain from node 2 to node 3. Again, there is a single forward path from node 2 to node 3 with path gain $+\frac{1}{2}$, so this graph gain is also $+\frac{1}{2}$. So, I have that $I_3 \propto \sqrt{I_2}$. Taken together, these results imply that

$$I_3 \propto \sqrt{I_1 I_2},$$

so the circuit of Figure 3.7a is a two-input geometric-mean circuit.

3.4.2. Two-Branch Paths

Consider the MITE network shown in Figure 3.8a. There are three MITEs, so I begin construction of the reduced signal-flow graph by drawing three nodes, labeling them 1, 2, and 3. MITE Q_1 is diode connected through one unit weighting coefficient, and its output connects to the weighted sum of MITE Q_3 through two unit weighting coefficients, so I draw a branch from node 1 to node 3, labeling it with branch gain $\frac{2}{1}$. MITE Q_2 is diode connected through two unit weighting coefficients, and its output connects to the weighted sum of MITE Q_1 through one unit weighting coefficient, so I draw a branch from node 2 to node 1, labeling it with a branch gain of $\frac{1}{2}$. The resulting reduced signal-flow graph is shown in Figure 3.8b.

To analyze the circuit, I first calculate the signal-flow–graph gain from node 1 to node 3. There is one forward path from node 1 to node 3 with path gain $+\frac{2}{1}$, so the graph gain from node 1 to node 3 is simply $+\frac{2}{1}$. Thus, I have that $I_3 \propto I_1^2$. Next, I calculate the signal-flow–graph gain from node 2 to node 3. Again, there is a single forward path from node 2 to node 3, 2→1→3, with path gain $-\left(\frac{1}{2}\right)\left(\frac{2}{1}\right) = -1$, so this signal-flow–graph gain is equal to –1. So, I have that $I_3 \propto I_2^{-1}$. Taken together, these results imply that

$$I_3 \propto \frac{I_1^2}{I_2},$$

so the circuit of Figure 3.8a is a squaring-reciprocal circuit.

3.4.3. Parallel Paths

Consider the MITE network shown in Figure 3.9a. There are four MITEs, so I begin construction of the reduced signal-flow graph by drawing four nodes, labeling them 1, 2, 3, and 4. MITE Q_1 is diode connected through one unit weighting coefficient, and its output connects to the weighted sum of MITE Q_4 through one unit weighting coefficient, so I draw a branch from node 1 to node 4, labeling it with branch gain $\frac{1}{1}$. Likewise, MITE Q_2 is diode connected through one unit weighting coefficient, and its output connects to the weighted sum of MITE Q_4 through one unit weighting coefficient, so I draw a branch from node 2 to node 4, labeling it with a branch gain of $\frac{1}{1}$. MITE Q_3 is diode connected through two unit weighting coefficients, and it connects to both the weighted sum of MITE Q_1 and the weighted sum of MITE Q_2 through one unit weighting coefficient each, so I draw a branch from node 3 to node 1, labeling it with a branch gain of $\frac{1}{2}$, and I draw a branch from node 3 to node 2, labeling it with a branch gain of $\frac{1}{2}$ as well. The resulting reduced signal-flow graph is shown in Figure 3.9b.

To analyze the circuit, I first calculate the signal-flow–graph gain from node 1 to node 4. There is one forward path from node 1 to node 4 with path gain $+\frac{1}{1}$, so the graph gain from node 1 to node 4 is simply +1. Thus, I have that $I_4 \propto I_1$. Next, I calculate the signal-flow–graph gain from node 2 to node 4. Again, there is a single forward path from node 2 to node 4, with path gain $+\frac{1}{1}$, so this graph gain is also equal to +1. So, I have that $I_4 \propto I_2$. Next, I calculate the graph gain from node 3 to node 4. There are two forward paths from node 3 to node 4: 3→1→4, with path gain $-\left(\frac{1}{2}\right)\left(\frac{1}{1}\right) = -\frac{1}{2}$, and 3→2→4, with path gain $-\left(\frac{1}{2}\right)\left(\frac{1}{1}\right) = -\frac{1}{2}$. Consequently, the signal-flow–graph gain from node 1 to node 4 is equal to $-\frac{1}{2} - \frac{1}{2} = -1$. So, I have that $I_4 \propto I_3^{-1}$. Taken together, these results imply that

$$I_4 \propto \frac{I_1 I_2}{I_3},$$

so the circuit of Figure 3.9a is a product-reciprocal circuit.

3.4.4. Reciprocal Connections

Consider the MITE network shown in Figure 3.10a. There are three MITEs, so I begin construction of the reduced signal-flow graph by drawing three nodes, labeling them 1, 2, and 3. MITE Q_1 is diode connected through three unit weighting coefficients, and its output connects to the weighted sum of MITE Q_3 through four unit weighting coefficients, so I draw a branch from node 1 to node 3, labeling it with branch gain $\frac{4}{3}$. MITE Q_1 also connects to the weighted sum of MITE Q_2 through two unit weighting coefficients, so I draw a branch from node 1 to node 2, labeling it with branch gain $\frac{2}{3}$. MITE Q_2 is diode connected through two unit weighting coefficients, and its output connects to the weighted sum of MITE Q_1 through one unit weighting coefficient, so I draw a branch from node 2 to node 1, labeling it with a branch gain of $\frac{1}{2}$. The resulting reduced signal-flow graph is shown in Figure 3.10b.

To analyze the circuit, I first note that there is a single feedback loop, 1→2→1, with loop gain $+\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{3}$. Next, I calculate the signal-flow–graph gain from node 1 to node 3. There is one forward path from node 1 to node 3 with path gain $+\frac{4}{3}$. This path touches the feedback loop, so the graph gain from node 1 to node 3 is given by

$$\Lambda_{31} = \frac{+\frac{4}{3}}{1-\frac{1}{3}} = \frac{4}{2} = 2.$$

Thus, I have that $I_3 \propto I_1^2$. Next, I calculate the signal-flow–graph gain from node 2 to node 3. Again, there is a single forward path from node 2 to node 3, 2→1→3, with path gain $-\left(\frac{1}{2}\right)\left(\frac{4}{3}\right) = -\frac{2}{3}$. This path also touches the feedback loop, so the signal-flow–graph gain from node 2 to node 3 is equal to

$$\Lambda_{32} = \frac{-\frac{2}{3}}{1-\frac{1}{3}} = -1.$$

Thus, I have that $I_3 \propto I_2^{-1}$. Taken together, these results imply that

$$I_3 \propto \frac{I_1^2}{I_2},$$

so the circuit of Figure 3.10a is also a squaring-reciprocal circuit.

3.5. References

1. S. J. Mason, "Feedback Theory—Some Properties of Signal Flow Graphs," *Proceedings of the IRE*, vol. 41, no. 9, pp. 1144–1156, 1953.

2. S. J. Mason, "Feedback Theory—Some Further Properties of Signal Flow Graphs," *Proceedings of the IRE*, vol. 44, no. 7, pp. 920–926, 1956.

3. S. J. Mason and H. J. Zimmerman, *Electronic Circuits, Signals, and Systems*, Cambridge, MA: MIT Press, 1970.

4. C. L. Coates, "Flow-Graph Solutions of Linear Algebraic Equations," *IRE Transactions on Circuit Theory*, vol. CT-6, no. 2, pp. 170–187, 1959.

5. Y. Chow, "Node Duplication—A Transformation of Signal-Flow Graphs," *IRE Transactions on Circuit Theory*, vol. CT-6, no. 2, pp. 233–234, 1959.

6. C. A. Desoer, "The Optimum Formula for the Gain of a Flow Graph or a Simple Derivation of Coates' Formula," *Proceedings of the IRE*, vol. 48, no. 5, pp. 883–889, 1960.

7. Y. Chow and E. Cassignol, *Linear Signal-Flow Graphs and Applications*, New York: Wiley, 1962.

8. L. P. A. Robichaud, M. Boisvert, and J. Robert, *Signal Flow Graphs and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1962.

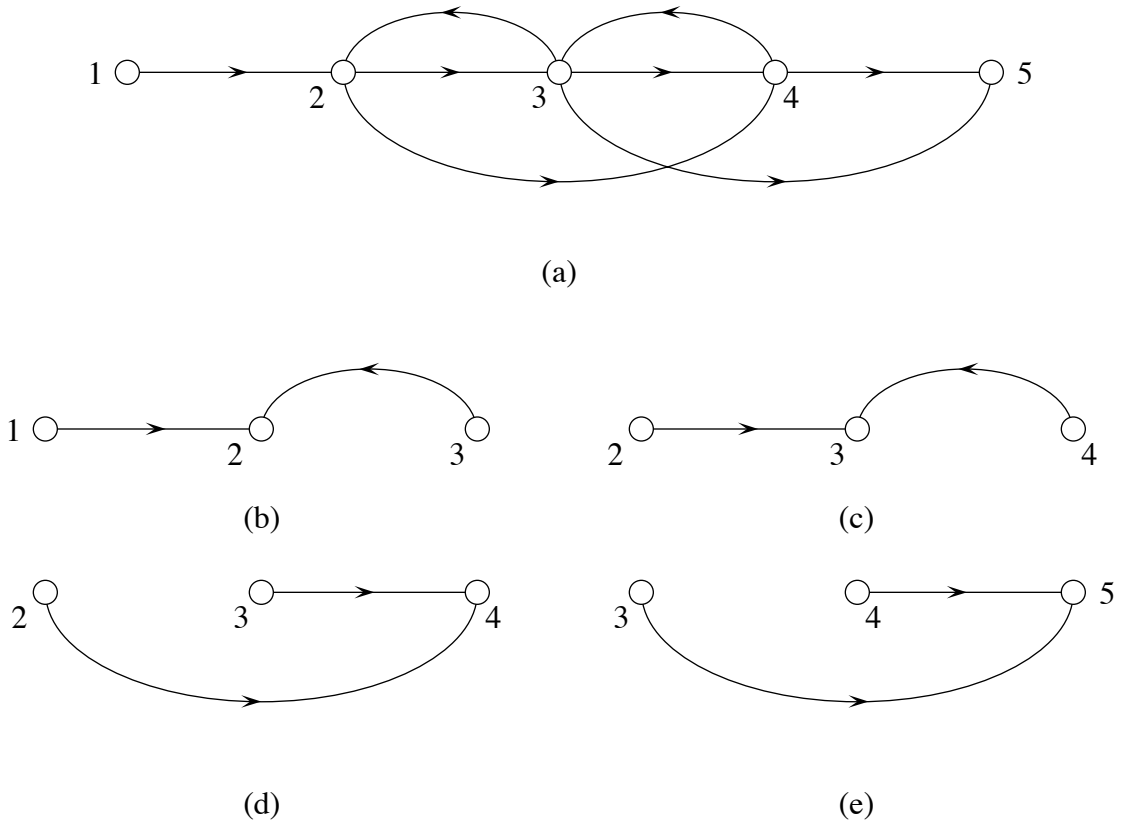


Figure 3.1. Signal-flow graphs that illustrate various terms and concepts. (a) A signal-flow graph with five nodes and eight branches. (*Source:* Signal-flow–graph structure adapted from S. J. Mason, "Feedback Theory—Some Further Properties of Signal Flow Graphs," *Proceedings of the IRE*, vol. 44, no. 7, Figure 3, p. 921, 1956.) If I denote the signal associated with node i by x_i, then this signal-flow graph implies the following equations relating node signals x_1 through x_5:

$$\begin{cases} x_1 = x_1, \\ x_2 = f_2(x_1, x_3), \\ x_3 = f_3(x_2, x_4), \\ x_4 = f_4(x_2, x_3), \\ x_5 = f_5(x_3, x_4). \end{cases}$$

I show various subgraphs corresponding to (b) the equation for x_2, (c) the equation for x_3, (d) the equation for x_4, and (e) the equation for x_5.

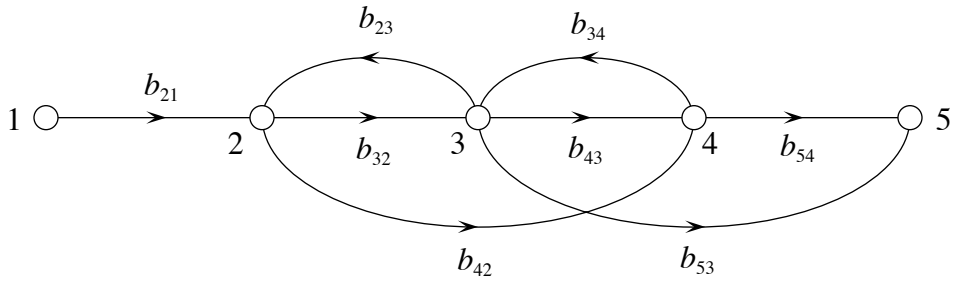
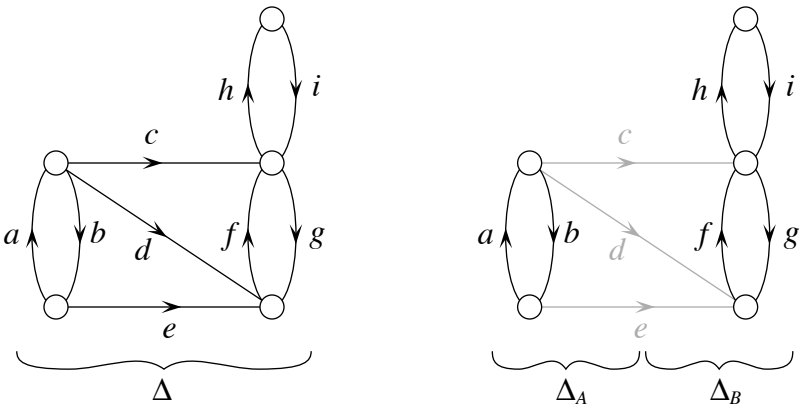


Figure 3.2. A linear signal-flow graph with five nodes and eight branches, with the same topology as that of Figure 3.1a. If I denote the signal associated with node i by x_i, and the branch gain of branch ji by b_{ji}, then this signal-flow graph corresponds to the following set of linear equations relating the node signals x_1 through x_5:

$$\begin{cases} x_1 = x_1, \\ x_2 = b_{21}x_1 + b_{23}x_3, \\ x_3 = b_{32}x_2 + b_{34}x_4, \\ x_4 = b_{42}x_2 + b_{43}x_3, \\ x_5 = b_{53}x_3 + b_{54}x_4. \end{cases}$$



(a) (b)

$$\Delta = 1 - (ab + fg + hi) + (abfg + abhi)$$

$$\Delta_A = 1 - ab \quad \Delta_B = 1 - fg - hi$$

$$\Delta = \Delta_A \Delta_B = (1 - ab)(1 - fg - hi)$$

Figure 3.3. The determinant of a signal-flow graph (a) whose loop subgraph (b) has two separate parts is factorable. (*Source:* Adapted from S. J. Mason and H. J. Zimmerman, *Electronic Circuits, Signals, and Systems*, Cambridge, MA: MIT Press, Figure 4–19, p. 111, 1970.)

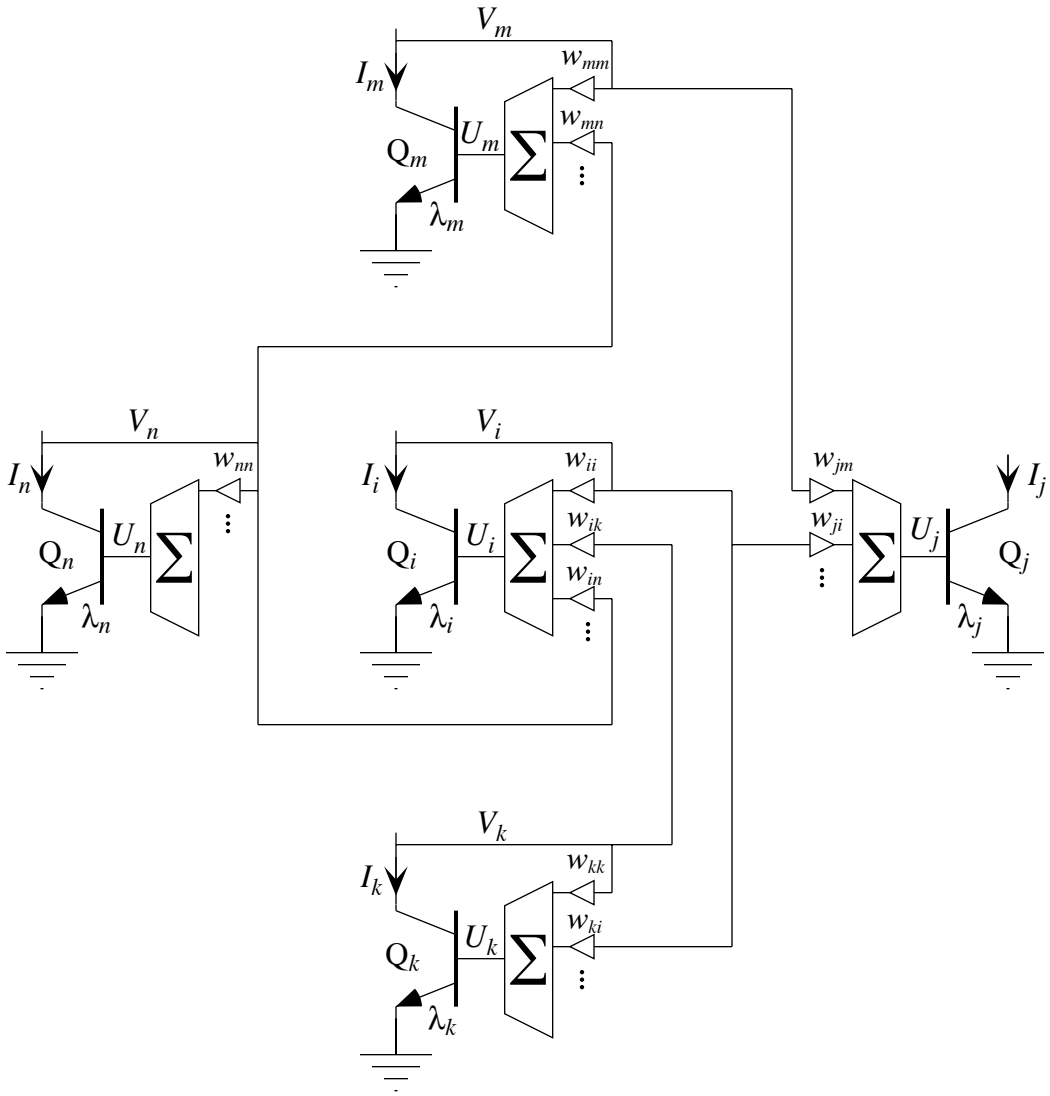


Figure 3.4. A MITE network fragment comprising a single output MITE, Q_j, and four diode-connected input MITEs: Q_i, Q_k, Q_m, and Q_n. The full circuit could have more input and output MITEs, but I assume that all connections involving V_i, V_k, V_m, and V_n are shown in the schematic. I use this circuit fragment to illustrate the signal-flow–graph analysis procedure.

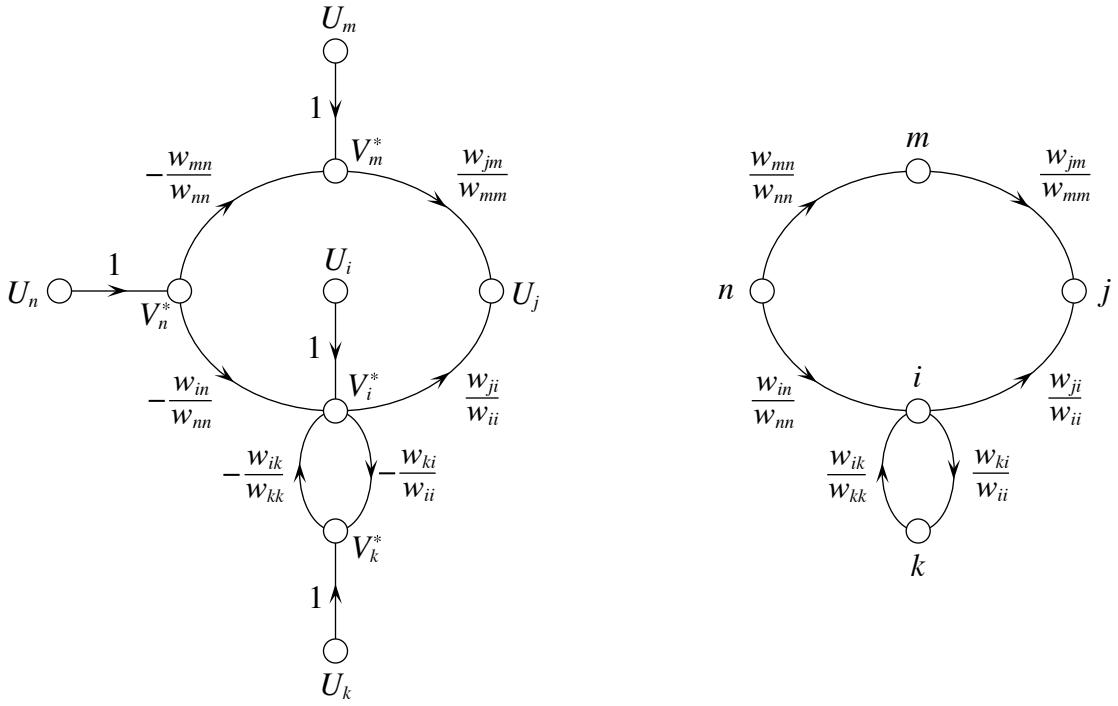
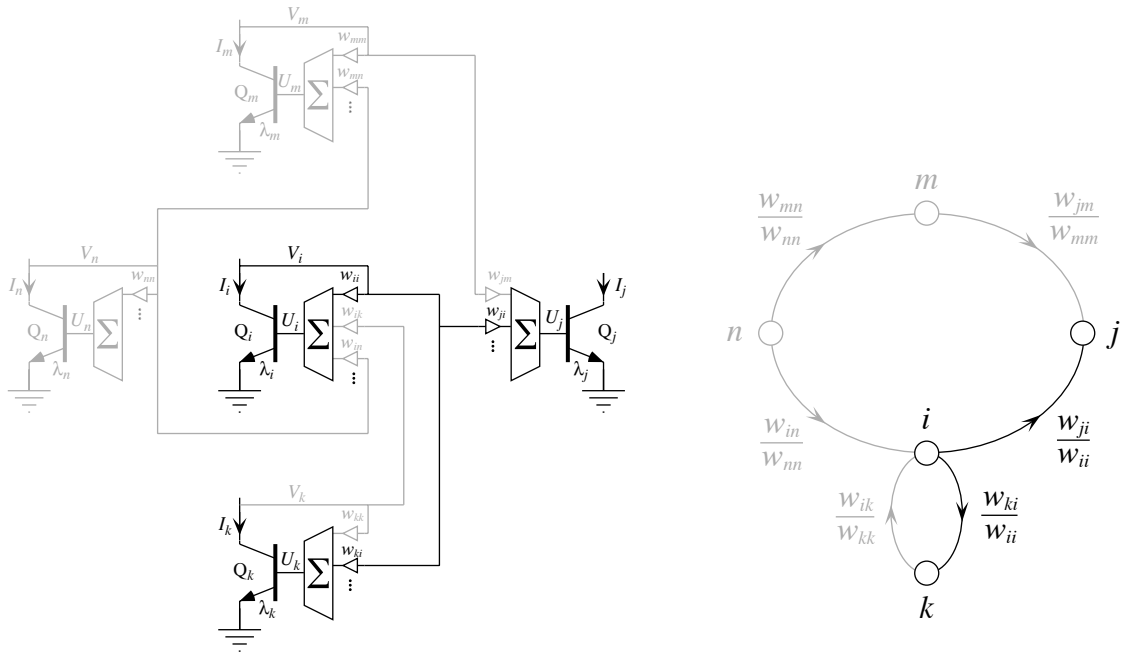
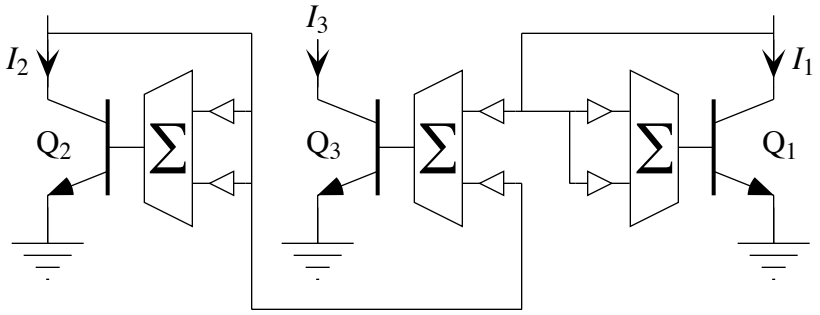


Figure 3.5. Signal-flow graphs corresponding to the MITE network fragment of Figure 3.4. (a) Signal-flow graph representation of Equation 3.15, which describe the circuit of Figure 3.4. (b) Reduced version of the signal-flow graph shown in part a. To obtain the reduced signal-flow graph from the full version, I collapse each pair of nodes (U_n, V_n^*) into a single node indexed by n and I take the absolute value of each branch gain. In the reduced signal-flow graph, each nonsink node is considered both as a source node from its own standpoint and as an intermediate node from the standpoint of nodes farther away from a sink. When operating on a reduced signal-flow graph, I calculate the magnitude of path gains and loop gains as the product of the branch gains comprising the path or loop. The sign of a path gain is positive if the path comprises an odd number of branches, and is negative if the path comprises an even number of branches. The sign of a loop gain is positive if the loop comprises an even number of branches, and is negative if the loop comprises an odd number of branches.

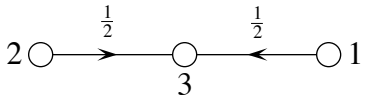


(a) (b)

Figure 3.6. The reduced signal-flow graph corresponding to a MITE network, constructed directly from the circuit schematic. (a) MITE network fragment from Figure 3.4 emphasizing MITEs Q_i, Q_j, and Q_k and the connections from MITE Q_i to both Q_j and Q_k. (b) Reduced signal-flow graph from Figure 3.5b emphasizing nodes i, j, and k and the connections from node i to both node j and node k. To construct a reduced signal-flow graph from a MITE circuit schematic, I first draw a node for each MITE, numbering it according to the number of the MITE. I connect these nodes with branches according to the following rule: If MITE Q_i is diode connected and its output couples into the weighted sum of MITE Q_j, then I draw a branch from node i to node j and label it with the branch gain $\frac{w_{ji}}{w_{ii}}$. For instance, in part a, MITE Q_i is diode connected and its output does couple into the weighted sum of MITE Q_j; therefore, in part b, I draw a branch from node i to node j, and label it with branch gain $\frac{w_{ji}}{w_{ii}}$. The output of MITE Q_i also couples into the weighted sum of Q_k, so I draw a branch from node i to node k and label it with branch gain $\frac{w_{ki}}{w_{ii}}$.

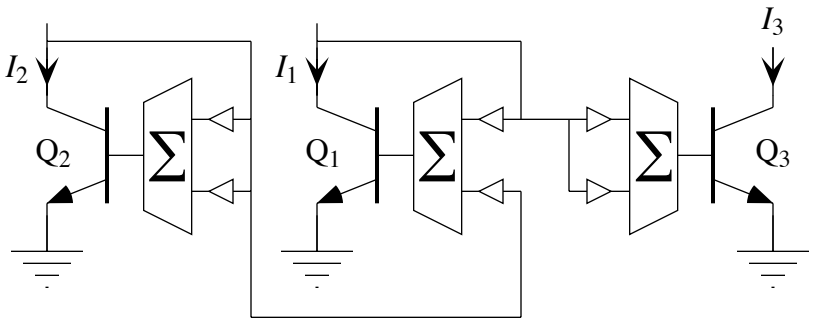


(a)

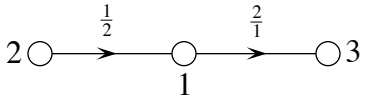


(b)

Figure 3.7. MITE network with two single-branch paths. (a) Circuit schematic. (b) Corresponding reduced signal-flow graph.

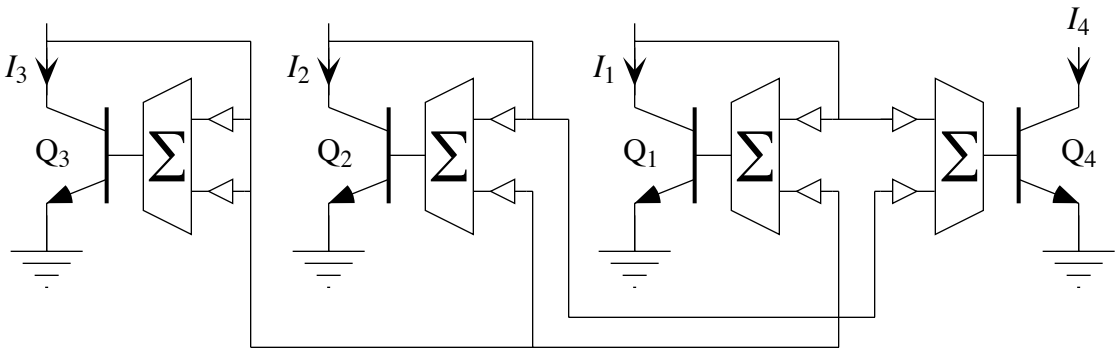


(a)

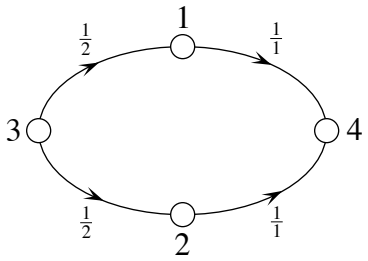


(b)

Figure 3.8. MITE network with a one single-branch path and one two-branch path. (a) Circuit schematic. (b) Corresponding reduced signal-flow graph.

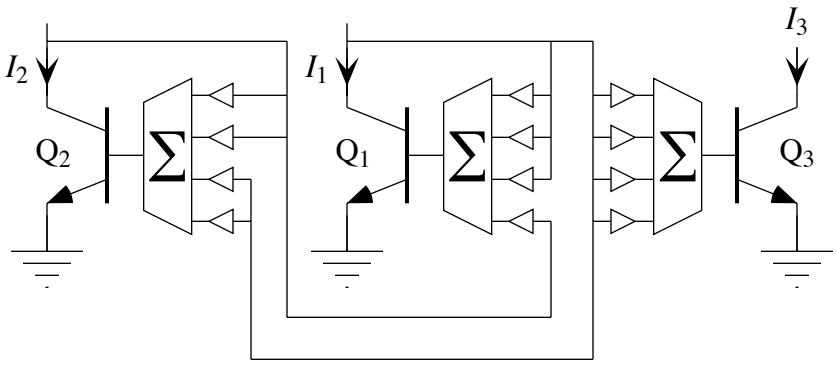


(a)

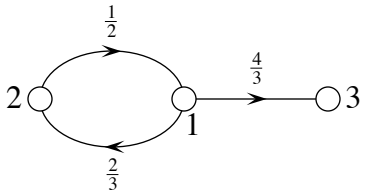


(b)

Figure 3.9. MITE network with parallel paths. (a) Circuit schematic. (b) Corresponding reduced signal-flow graph.



(a)



(b)

Figure 3.10. MITE network with reciprocal connections. (a) Circuit schematic. (b) Corresponding reduced signal-flow graph.

CHAPTER 4
SYNTHESIS OF MULTIPLE-INPUT TRANSLINEAR ELEMENT NETWORKS

In this chapter, I discuss several aspects of the synthesis of MITE networks from the specification of an expression relating an output current to a product of input currents, each of which is raised to an arbitrary rational power. In Section 4.1, I describe the synthesis problem. In Section 4.2, I show that the synthesis problem is underconstrained—in other words, for any expression of the appropriate form, there are many (in principle, a countable infinity of) MITE networks that embody the expression. Then, I discuss several criteria by which we can choose one from among the many possible circuit solutions for a given expression. In Section 4.3, I present two systematic procedures for synthesizing an asymptotically stable, single-output MITE network for an arbitrary product-of-powers relationship among any number of input currents. Given an expression to realize, and using the first procedure, we arrive at a MITE network that is relatively insensitive to mismatch in weighting-coefficient values. Using the second procedure, we obtain a MITE network that minimizes the number of inputs required for each MITE. In Section 4.4, I show how we can formulate the MITE-network–synthesis problem as a integer linear program, and hence, how we can use well-established algorithms for solving combinatorial optimization problems to synthesize MITE networks with the aid of a computer.

4.1. The Scope of the Synthesis Problem

In this section, I describe **the synthesis problem**. We are given an expression of the form

$$I_{N+1} = \prod_{n=1}^{N} I_n^{\Lambda_n}, \tag{4.1}$$

where I_{N+1} is the output current (i.e., the dependent variable), I_1 through I_N are input currents (i.e., the independent variables), and powers given by Λ_1 through Λ_N are dimensionless rational numbers (either positive or negative) satisfying

$$\sum_{n=1}^{N} \Lambda_n = 1.$$

Our objective in the synthesis problem is to construct a MITE network that embodies Equation 4.1. I could consider the general case of multiple output currents that depend on the same set of input currents. However, systematic synthesis procedures applicable to this general case are difficult to state simply and succinctly. Consequently, in Section 4.3, I consider only the single-output case. I address the multiple-output case in Section 4.4, when I frame the synthesis problem as an integer linear program with suitable constraints.

As I have defined it, the synthesis problem is the exact inverse of the analysis problem addressed both in Chapter 2 and in Chapter 3. In the analysis problem, we begin with a MITE network such as that shown in Figure 2.9, and we derive an expression (or a collection of expressions in the case of multiple output currents) of the form of Equation 4.1 that describe the steady-state behavior of the MITE network being analyzed. Contrariwise, in the synthesis problem, we begin with an expression (or perhaps with a collection of expressions) of the form of Equation 4.1, and we construct a MITE network of the form that is shown in Figure 2.9 whose steady-state behavior embodies the expressions to be realized.

Seevinck [1] has described various aspects of the systematic synthesis of translinear loop circuits. He adopts a higher-level view of the synthesis problem than I do in this chapter. He begins with a mathematical function of a dimensionless variable to be realized with translinear loop circuits. He discusses various techniques for decomposing and approximating this function as products of powers of linear combinations of dimensionless variables. Then, he identifies these dimensionless variables with current ratios or with modulation indices (i.e., dimensionless numbers, ranging either from 0 to 1 or from -1 to 1, scaling or *modulating* the value of a bias current). He converts the linear combinations of these dimensionless variables into strictly positive linear combinations of currents, which are sourced into or sunk from the input nodes of various cascades or parallel combinations of simple translinear loop circuits. The overall circuit output is taken to be a dimensionless quantity scaling a linear combination of the output currents of the various translinear loop circuits constituting the circuit. We can use many of the concepts, the approximation techniques, and the principles that Seevinck discusses directly in concert with the synthesis procedures that I develop in this chapter to realize complex analog signal-processing functions with MITE networks, in place of translinear loop circuits.

4.2. The Myriad Solutions to the Synthesis Problem

The synthesis problem, as I defined it in Section 4.1, is underconstrained. We can think of the synthesis of a MITE network as the problem of finding the elements of the $N \times N$ input

connectivity matrix, $\mathbf{W}_{\text{in}}$, and those of the $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, satisfying

$$\Lambda \mathbf{W}_{\text{in}} = \mathbf{W}_{\text{out}}, \tag{4.2}$$

where Λ is the given $M \times N$ matrix of powers to be implemented. Then, we have $N^2 + NM$ variables related to one another by NM equations. Consequently, there will be an infinity of possible solutions of Equation 4.2, and, hence, an infinite number of MITE networks embodying the power-law relations specified by Λ. The following theorem gives us another way to see that there is an infinity of MITE network implementations of any given set of power-law relations.

Theorem 4.1: If there exists a single MITE network embodying a given set of product-of-power-law relationships, then there exists a countable infinity of such MITE networks.

Proof: I assume that there exists at least one MITE network that embodies the product-of-power-law relations specified by the $M \times N$ matrix of powers, Λ. This MITE network is specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, such that $\Lambda = \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1}$. From this MITE network, I construct a new one as follows. First, I add a single weighting coefficient with value w to each MITE in the circuit. Then, I connect each of these new inputs to the jth input-node voltage, V_j. In terms of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$, I can express this transformation as

$$\mathbf{W}'_{\text{in}} = \mathbf{W}_{\text{in}} + w\mathbf{e}_j^{(N)}$$

and

$$\mathbf{W}'_{\text{out}} = \mathbf{W}_{\text{out}} + w\mathbf{e}_j^{(M)},$$

where $\mathbf{e}_j^{(n)}$ denotes an $n \times N$ matrix with 1s in the jth column and 0s everywhere else. In Appendix 4.A, I show that, if each row of Λ sums to unity (i.e., for each m between 1 and M, $\sum_{n=1}^{N} \Lambda_{mn} = 1$), then the matrix of powers, $\Lambda' = \mathbf{W}'_{\text{out}} \mathbf{W}'^{-1}_{\text{in}}$, for this new MITE network is the same as that of the original one. Thus, I have a second MITE network that embodies the original set of product-of-power-law relationships. I can then apply the same construction recursively to the new circuit, to get a third MITE network that implements the given relationships. I can apply this construction recursively an indefinite number of times; hence, I can construct a countable infinity of MITE networks embodying the same relationships as did the original MITE network. ■

Given that, if any MITE network exists that realizes a given set of product-of-

power-law relationships, then there is an infinity of MITE networks, each of which realizes the given set of product-of-power-law relationships, by what criteria are we to choose one from among them? Other considerations aside, it seems as though we would prefer a MITE network that has fewer inputs per MITE. In general, we must consider other criteria to make this choice. For some product-of-power-law relationships, there is even a degeneracy of MITE network realizations that comprise MITEs with the same number of inputs per MITE. For example, consider the four distinct MITE networks depicted in Figure 4.1. Each comprises four two-input MITEs, and each embodies the relationship

$$I_4 = \frac{I_1 I_2}{I_3}. \tag{4.3}$$

It is easy to verify that these MITE networks embody Equation 4.3 by inspecting their respective reduced signal-flow graphs, which are shown also in Figure 4.1.

The first MITE network realizing Equation 4.3, which is shown in Figure 4.1a, is a **two-layer MITE network** (I do not count the output MITEs as a layer). The first layer, which I call the **numerator layer**, comprises all the input MITEs whose currents appear in the numerator of the expression that is embodied by the network; in MITE network of Figure 4.1a, MITEs Q_1 and Q_2 make up the numerator layer. The second layer, which I call the **denominator layer**, comprises all the input MITEs whose currents appear in the denominator of the expression that the MITE network embodies; in the MITE network of Figure 4.1a, the denominator layer comprises MITE Q_3. In this case, two MITE inputs are unused (i.e., are connected to ground). Because Equation 4.3 is symmetric in I_1 and I_2, there is another MITE network embodying Equation 4.3 whose structure is identical to that shown in Figure 4.1a. I obtain this network by renumbering inputs 1 and 2.

The second MITE network realizing Equation 4.3, which is shown in Figure 4.1b, is another two-layer network. Again, the numerator layer is made up of MITEs Q_1 and Q_2, and the denominator layer comprises MITE Q_3. In this case, however, all MITE inputs are used (i.e., there is no MITE input connected to ground). Because this MITE network is symmetric in MITEs Q_1 and Q_2, there is not a second MITE network with this structure reflecting the symmetry of Equation 4.3 in I_1 and I_2.

The third MITE network realizing Equation 4.3, which is shown in Figure 4.1c, is a **cascade MITE network**. In this case, I arrange the input MITEs in a linear sequence, alternating between those MITEs whose currents appear in the numerator of the expression realized by the MITE network and those MITEs whose currents appear in the denominator of the expression. In the MITE network of Figure 4.1c, because there are two currents in

the numerator of Equation 4.3 and only one current in the denominator, this linear chain of input MITEs is three links long. In this case, two MITE inputs are unused, and I can again obtain a second MITE network with the same structure as that shown in Figure 4.1c by renumbering inputs 1 and 2.

The fourth MITE network realizing Equation 4.3, which is shown in Figure 4.1d, is another cascade MITE network. In this case, there is a direct connection from MITE Q_2 to MITE Q_4, but the alternating sequence of numerator input and denominator input that is characteristic of the cascade-network topology is present. In this MITE network, there are no unused MITE inputs, and I can again obtain a second MITE network with the same structure as that shown in Figure 4.1d by renumbering inputs 1 and 2.

How should I decide which of these four distinct alternative MITE-network structures is optimal? I can group these four MITE networks in at least two different ways. First, the MITE networks of Figures 4.1a and 4.1c are similar to each other in that they each have unused MITE inputs, whereas the MITE networks of Figures 4.1b and 4.1d have no unused MITE inputs. As I mentioned in Section 2.5, by leaving MITE inputs unused, I am, in a sense, wasting MITE transconductance. This wasted transconductance implies that the voltage swings on those input nodes with the unused MITE inputs will be larger for a given input-current change than they would be in an equivalent MITE network in which all MITE inputs are used. These larger voltage swings, in turn, require a higher power-supply voltage. For example, consider the voltage swing on MITE Q_3 for a given change in I_3 in the MITE networks of Figures 4.1a and 4.1b; this voltage swing in the MITE network of Figure 4.1a will be twice the corresponding voltage swing in the MITE network of Figure 4.1b. I can make a similar comparison for MITE Q_2 in the MITE networks of Figures 4.1c and 4.1d.

If I am wasting MITE transconductance by having unused inputs, then why should I have them at all? What purpose do they serve? The answer to these questions is not obvious from anything that I have discussed thus far; the need for these unused inputs arises from the ways in which I implement the weighted–voltage-summation operation (i.e., using either a resistive voltage divider or a capacitive voltage divider), as I discuss in Chapters 6 and 7. For example, in a capacitive voltage divider, the coupling strength between an input node and the summation node is proportional to the size of the capacitance with which the input node and the summation node are connected, and this coupling strength is normalized by the total capacitance connected to the summation node. Now, as designers, we would like to be able to treat these weighting coefficients as though they were not normalized by the total capacitance connected to the summation nodes, especially

if this total capacitance includes any parasitic capacitance, and it is nearly inevitable that it does. We can always accomplish this goal if, in constructing a MITE network, we use MITEs with identical complements of weighting coefficients. To show that we can, I make use of the following theorem.

Theorem 4.2: If all the weights of a MITE network are scaled by an arbitrary positive quantity, then the product-of-power-law relationships embodied in the network remain unchanged.

Proof: Suppose that I have a MITE network specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$. The product-of-power-law relationships embodied in this MITE network are given by $\Lambda = \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1}$. Now, suppose I scale each weighting coefficient in both $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ by some positive quantity, ω. I thus obtain new connectivity matrices given by

$$\mathbf{W}'_{\text{in}} = \omega \mathbf{W}_{\text{in}}$$

and

$$\mathbf{W}'_{\text{out}} = \omega \mathbf{W}_{\text{out}}.$$

Now, the product-of-power-law relationships realized by this new MITE network are given by

$$\begin{aligned} \Lambda' &= \mathbf{W}'_{\text{out}} \mathbf{W}'^{-1}_{\text{in}} \\ &= \omega \mathbf{W}_{\text{out}} \left(\omega \mathbf{W}_{\text{in}} \right)^{-1} \\ &= \omega \mathbf{W}_{\text{out}} \left(\frac{1}{\omega} \mathbf{W}_{\text{in}}^{-1} \right) \\ &= \frac{\omega}{\omega} \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1} \\ &= \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1} \\ &= \Lambda. \end{aligned}$$

Thus, I have that the product-of-power-law relationships embodied in a MITE network are invariant if I rescale all the weighting coefficients by the same positive quantity. ■

So, we can construct the connectivity matrices, $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$, to get a desired set of product-of-power-law relationships given by $\Lambda = \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1}$, treating the weighting coefficients as though they were completely independent of one another (i.e., were not normalized by the total weight connected to each MITE's summing node). We connect any unused MITE inputs to ground. If we have used MITEs with the same complement of

weighting coefficients, so the total weight connected to each MITE, w_T, is the same for each MITE, then we can apply Theorem 4.2 with $\omega = \frac{1}{w_T}$, and conclude that the actual MITE network, which has normalized weighting coefficients, will embody the same product-of-power-law relationships as we intended. As I mentioned in Section 2.1 and Section 2.7, in practice, we obtain the most accurate ratios of weighting coefficients if we restrict ourselves to rational numbers, and if we construct integer weighting coefficients by connecting that number of nominally identical unit cells in parallel with one another. In this case, using MITEs that have identical complements of weighting coefficients translates into using MITEs with the same number of unit inputs.

It so happens that we are never forced to leave any MITE inputs unused. In Appendix 4.B, I show that, if each of the rows of Λ sums to unity (i.e., for each m between 1 and M, $\sum_{n=1}^{N} \Lambda_{mn} = 1$), then we can connect all the unused MITE inputs to one of the input nodes without changing the product-of-power-law relationships embodied by the MITE network. In so doing, we should choose the input node of a MITE that has only self connections so that we are sure not to introduce any feedback loops into the MITE network by this transformation. The MITE networks of Figures 4.1b and 4.1d are related, respectively, to those of Figures 4.1a and 4.1c by this transformation. I use **completion** to denote this MITE network transformation.

Given this relationship of the MITE networks of Figures 4.1a and 4.1c to those of Figures 4.1b and 4.1d, it is not surprising that the second way that I can group these MITE networks is according to the maximum number of stages in each one. The MITE networks of Figures 4.1a and 4.1b are both two-layer networks, whereas the MITE networks of Figures 4.1c and 4.1d are both cascade networks, and each has paths comprising three branches. Intuitively, I expect that, if there were any mismatch in the values of the weighting coefficients in a MITE network, then the effect of this mismatch would be more likely to compound in long paths consisting of many branches than it would in short paths with fewer branches. I would thus expect that errors due to weighting-coefficient mismatch in the power-law relationships embodied in a MITE network would be larger in a network with longer paths than they would be in an equivalent network with shorter paths. Consequently, I would expect that the MITE networks of Figures 4.1a and 4.1b are, in some sense, more insensitive to component mismatch than are the MITE networks of Figures 4.1c and 4.1d.

In Appendix 4.C, I formalize this expectation by developing a simple model that allows us to calculate the effect of mismatch in the values of weighing coefficients in a

MITE network that has a given topology on the product-of-power-law relationships embodied by that MITE network. In this model, I assume that all the weighting coefficients in the MITE network are integral multiples of identical unit weighting coefficients, each with value *w*. I assume that each unit weighting coefficient is perturbed by a zero-mean Gaussian random variable that is statistically independent of the other perturbations. Further, I assume that the standard deviation of these random perturbations, σ, is small (on the order of a few percent or less) compared with the nominal value of each weighting coefficient, *w*. In Appendix 4.C, I show that, with these assumptions, the variances of the errors in the powers contained in Λ, for a given MITE network specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, are given by

$$E(\delta\Lambda \circ \delta\Lambda) = \frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1} \circ \mathbf{W}_{\text{in}}^{-1}\right), \tag{4.4}$$

where $\delta\Lambda$ is the matrix of errors in the powers contained in Λ, σ^2 is the variance of the random perturbations in the unit weighting coefficients, and $\mathbf{A} \circ \mathbf{B}$ denotes the **Hadamard product** (i.e., element-by-element product) of two matrices, $\mathbf{A}$ and $\mathbf{B}$, defined by $(\mathbf{A} \circ \mathbf{B})_{ij} \equiv a_{ij} b_{ij}$. If we have a number of alternative MITE-network implementations of a given set of product-of-power-law relationships, then we can use Equation 4.4 to evaluate how sensitive each topology is to mismatch in the values of the unit weighting coefficients, and thus can choose the one that is least sensitive.

To illustrate this application of Equation 4.4, I now compute $E(\delta\Lambda \circ \delta\Lambda)$ for each of the MITE networks shown in Figure 4.1. For simplicity, I take the value *w* to be unity. I specify the MITE network of Figure 4.1a by

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}.$$

From these matrices, I compute $E(\delta\Lambda \circ \delta\Lambda)$ for the MITE network of Figure 4.1a as

$$E(\delta\Lambda \circ \delta\Lambda) = \frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1} \circ \mathbf{W}_{\text{in}}^{-1}\right)$$

$$
\begin{aligned}
&= \sigma^2\left(\begin{bmatrix}1 & 1 & 0\end{bmatrix}+\begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\right)\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\\
&= \sigma^2\left(\begin{bmatrix}1 & 1 & 0\end{bmatrix}+\begin{bmatrix}1 & 1 & 2\end{bmatrix}\right)\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\\
&= \sigma^2\begin{bmatrix}2 & 2 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\\
&= \sigma^2\begin{bmatrix}2 & 2 & 4\end{bmatrix}. \qquad (4.5)
\end{aligned}
$$

Next, I specify the MITE network of Figure 4.1b by

$$\mathbf{W}_{\text{in}} = \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 2\end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix}1 & 1 & 0\end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix}1 & 0 & -\frac{1}{2}\\ 0 & 1 & -\frac{1}{2}\\ 0 & 0 & \frac{1}{2}\end{bmatrix} \text{ and } \Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix}1 & 1 & -1\end{bmatrix}.$$

From these matrices, I compute $E(\delta\Lambda \circ \delta\Lambda)$ for the MITE network of Figure 4.1b as

$$
\begin{aligned}
E(\delta\Lambda \circ \delta\Lambda) &= \frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1} \circ \mathbf{W}_{\text{in}}^{-1}\right)\\
&= \sigma^2\left(\begin{bmatrix}1 & 1 & 0\end{bmatrix}+\begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 2\end{bmatrix}\right)\begin{bmatrix}1 & 0 & \frac{1}{4}\\ 0 & 1 & \frac{1}{4}\\ 0 & 0 & \frac{1}{4}\end{bmatrix}\\
&= \sigma^2\left(\begin{bmatrix}1 & 1 & 0\end{bmatrix}+\begin{bmatrix}1 & 1 & 4\end{bmatrix}\right)\begin{bmatrix}1 & 0 & \frac{1}{4}\\ 0 & 1 & \frac{1}{4}\\ 0 & 0 & \frac{1}{4}\end{bmatrix}\\
&= \sigma^2\begin{bmatrix}2 & 2 & 4\end{bmatrix}\begin{bmatrix}1 & 0 & \frac{1}{4}\\ 0 & 1 & \frac{1}{4}\\ 0 & 0 & \frac{1}{4}\end{bmatrix}\\
&= \sigma^2\begin{bmatrix}2 & 2 & 2\end{bmatrix}. \qquad (4.6)
\end{aligned}
$$

Next, I specify the MITE network of Figure 4.1c by

$$\mathbf{W}_{\text{in}} = \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix}1 & 0 & 0\end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \text{ and } \Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}.$$

From these matrices, I compute $E(\delta\Lambda \circ \delta\Lambda)$ for the MITE network of Figure 4.1c as

$$\begin{aligned} E(\delta\Lambda \circ \delta\Lambda) &= \frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1} \circ \mathbf{W}_{\text{in}}^{-1}\right) \\ &= \sigma^2\left(\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\right)\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \\ &= \sigma^2\left(\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}\right)\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \\ &= \sigma^2\begin{bmatrix} 2 & 2 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \\ &= \sigma^2\begin{bmatrix} 2 & 6 & 4 \end{bmatrix}. \end{aligned} \tag{4.7}$$

Finally, I specify the MITE network of Figure 4.1d by

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} 1 & \frac{1}{2} & -1 \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 \end{bmatrix} \text{ and } \Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}.$$

From these matrices, I compute $E(\delta\Lambda \circ \delta\Lambda)$ for the MITE network of Figure 4.1d as

$$\begin{aligned} E(\delta\Lambda \circ \delta\Lambda) &= \frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1} \circ \mathbf{W}_{\text{in}}^{-1}\right) \\ &= \sigma^2\left(\begin{bmatrix} 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}\right)\begin{bmatrix} 1 & \frac{1}{4} & 1 \\ 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 1 \end{bmatrix} \\ &= \sigma^2\left(\begin{bmatrix} 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}\right)\begin{bmatrix} 1 & \frac{1}{4} & 1 \\ 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 1 \end{bmatrix} \end{aligned}$$

$$= \sigma^2 \begin{bmatrix} 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{4} & 1 \\ 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 1 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 2 & 2 & 4 \end{bmatrix}. \tag{4.8}$$

By comparing the norms of matrices 4.5, 4.6, 4.7, and 4.8, I can order the four MITE networks shown in Figure 4.1 according the overall sensitivity of each to mismatch in the unit weighting coefficients. The MITE network of Figure 4.1b is the least sensitive to mismatch. The MITE networks of Figures 4.1a and 4.1d are next, and they are both equally sensitive to mismatch. The MITE network of Figure 4.1c is the most sensitive to mismatch. Thus, with respect to weighting coefficient mismatch, I should prefer the MITE network of Figure 4.1b to the other three.

To reduce the sensitivity of a MITE network to component mismatch, we would like to make the average path length between each input MITE and the output MITE as small as possible. I show by construction in Section 4.3 that it is always possible to synthesize a two-layer MITE network to embody any single expression of the form of Equation 4.1. If there are any negative powers in the expression to be realized, we need at least two layers of input MITEs, because, as I showed in Section 2.3, we must employ a MITE as an inverting VIVO stage between the input MITE and the output MITE to get a negative power. In a two-layer network, if there is a large number of currents in the numerator of the expression to be realized, then there will be a large number of MITEs converging on the output MITE. This convergence, in turn, requires us to allocate a large number of inputs for the output MITE, and hence, for every MITE.

In Section 4.3, I show by construction that it is also always possible to synthesize a cascade MITE network to embody any single expression the form of Equation 4.1. In a cascade network, we build a linear chain of input MITEs alternating between currents in the numerator of the expression to be realized and those in the denominator. In such a structure, each input MITE will typically have some self connections and connections from one other input MITE. Consequently, we expect that a cascade network realizing a given expression could have fewer inputs per MITE than a two-layer structure embodying the same expression. However, because the path lengths in a cascade network are longer on average than in a two-layer network, we would expect that the cascade structure realizing a given product-of-power-law relationship will be more sensitive to weighting-coefficient mismatch than will a two-layer structure embodying the same expression.

4.3. Synthesis of Multiple-Input Translinear Element Networks

In this section, I describe two simple procedures for constructing a MITE network to embody an expression of the form

$$I_{N+1} = \prod_{n=1}^{N} I_n^{\Lambda_n}, \tag{4.9}$$

where I_{N+1} is the output current, I_1 through I_N are input currents (i.e., the independent variables), and powers given by Λ_1 through Λ_N are dimensionless rational numbers (either positive or negative) satisfying

$$\sum_{n=1}^{N} \Lambda_n = 1.$$

Because I have restricted the values of Λ_1 through Λ_N to the rational numbers, I have that

$$|\Lambda_n| = \frac{p_n}{q_n},$$

where p_n and q_n are positive integers for each n. Further, I assume that, for each n, p_n and q_n have no common divisors other than unity. Without loss of generality, I assume that Λ_1 through Λ_J are positive and that Λ_{J+1} through Λ_N are negative, so I_{N+1} is of the form

$$I_{N+1} = \frac{\prod_{n=1}^{J} I_n^{\Lambda_n}}{\prod_{n=J+1}^{N} I_n^{|\Lambda_n|}}.$$

I can always renumber the inputs so that the expression to be embodied in a MITE network is of this form. If I use the first procedure, then I obtain a two-layer MITE network. If I use the second, then I obtain a cascade MITE network. In each case, I show that the resulting MITE network meets the sufficient conditions for asymptotic stability that I obtained in Section 2.6. I illustrate each of these procedures for a simple example.

4.3.1. Construction of a Two-Layer Network

I begin the construction of a two-layer MITE network embodying Equation 4.9 by creating a MITE for the output current and labeling it Q_{N+1}. The remainder of the procedure is as follows:

1. *Constructing the numerator layer:* For each value of n between 1 and J, I perform the following steps. First, I create a MITE for the nth input current, and we label it Q_n. Then, I diode connect MITE Q_n through q_n unit weighting coefficients, and I connect the input node of MITE Q_n to MITE Q_{N+1} through p_n

unit weighting coefficients. In terms of the $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and the $N \times 1$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, making these connections corresponds to setting $(\mathbf{W}_{\text{in}})_{nn} = q_n$ and setting $(\mathbf{W}_{\text{out}})_{1n} = p_n$.

2. *Constructing the denominator layer:* For each value of n between $J+1$ and N, I perform the following steps. First, I create a MITE for the nth input current, and I label it Q_n. Then, I pick one of the MITEs in the numerator layer—say, MITE $\text{Q}_{n'}$. Let k denote the greatest common divisor of $p_{n'}q_n$ and $p_n q_{n'}$. Then, I diode connect MITE Q_n through $\frac{p_{n'}q_n}{k}$ unit weighting coefficients, and I connect the input node of MITE Q_n to MITE $\text{Q}_{n'}$ through $\frac{p_n q_{n'}}{k}$ unit weighting coefficients. In terms of $\mathbf{W}_{\text{in}}$, making these connections corresponds to setting $(\mathbf{W}_{\text{in}})_{nn} = \frac{p_{n'}q_n}{k}$ and setting $(\mathbf{W}_{\text{in}})_{n'n} = \frac{p_n q_{n'}}{k}$.
3. *Adding the required unused inputs:* I denote by K the largest number of inputs feeding into any MITE. If each MITE happens to have K inputs, I am done. Otherwise, I add sufficient grounded unit weighting coefficients to each MITE that they each have K inputs.
4. *Completing the MITE network:* I connect each of the grounded inputs that I added in step 3 to the input node of one of the MITEs in the denominator layer.

Because I number the numerator inputs from 1 to J and the denominator inputs from $J+1$ to N, at the end of step 3, the input connectivity matrix, $\mathbf{W}_{\text{in}}$, is upper triangular by construction. In particular, $\mathbf{W}_{\text{in}}$ has the form

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{21} \\ \mathbf{0} & \mathbf{W}_{22} \end{bmatrix},$$

where $\mathbf{W}_{11}$ is a $J \times J$ diagonal matrix containing the self-coupling strengths for each of the inputs in the numerator layer, $\mathbf{W}_{22}$ is an $(N-J)\times(N-J)$ diagonal matrix containing the self-coupling strengths for each of the inputs in the denominator layer, and $\mathbf{W}_{21}$ is a $J \times (N-J)$ matrix containing the feedforward connections from the denominator layer to the numerator layer. In step 4, if I choose to connect all the unused MITE inputs to the input node of MITE Q_N, then $\mathbf{W}_{\text{in}}$ remains upper triangular. If I choose any of the other MITEs in the denominator layer, $\mathbf{W}_{\text{in}}$ can loose its upper triangularity. However, I can always make it upper triangular again by exchanging with N the number of the input that I chose in step 4, and vice versa. Thus, the two-layer MITE network constructed by this procedure satisfies the sufficient condition for asymptotic stability that I found in Section 2.6.

4.3.2. Construction of a Cascade Network

I begin the construction of a cascade MITE network embodying Equation 4.9 by allocating a MITE for the output current, and labeling it Q_{N+1}. The remainder of the procedure is as follows:

1. I begin with an input current from the numerator of Equation 4.9—say the *n*th input current—and I allocate a MITE for it, labeling the MITE Q_n. I diode connect MITE Q_n through q_n unit weighting coefficients, and I connect the input node of MITE Q_n to MITE Q_{N+1} through p_n unit weighting coefficients. In terms of the $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and the $N \times 1$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, making these connections corresponds to setting $\left(\mathbf{W}_{\text{in}}\right)_{nn} = q_n$ and setting $\left(\mathbf{W}_{\text{out}}\right)_{1n} = p_n$.
2. I choose an input current from the denominator of Equation 4.9. If there are no more input currents in the denominator of Equation 4.9, then I proceed to step 5. Otherwise, if I choose the *n*th input current, then I allocate a MITE for it, labeling the MITE Q_n. Suppose that the MITE created in the preceding step was MITE $Q_{n'}$. Let *k* denote the greatest common divisor of $p_{n'}q_n$ and $p_n q_{n'}$. I then diode connect MITE Q_n through $\frac{p_{n'}q_n}{k}$ unit weighting coefficients, and I connect the input node of MITE Q_n to MITE $Q_{n'}$ through $\frac{p_n q_{n'}}{k}$ unit weighting coefficients. In terms of $\mathbf{W}_{\text{in}}$, making these connections corresponds to setting $\left(\mathbf{W}_{\text{in}}\right)_{nn} = \frac{p_{n'}q_n}{k}$ and setting $\left(\mathbf{W}_{\text{in}}\right)_{n'n} = \frac{p_n q_{n'}}{k}$.
3. I choose an input current from the numerator of Equation 4.9. If there are no more input currents in the numerator of Equation 4.9, then I proceed to step 6. Otherwise, if I choose the *n*th input current, then I allocate a MITE for it, labeling the MITE Q_n. Suppose that the MITE created in the preceding step was MITE $Q_{n'}$. Let *k* denote the greatest common divisor of $p_{n'}q_n$ and $p_n q_{n'}$. I then diode connect MITE Q_n through $\frac{p_{n'}q_n}{k}$ unit weighting coefficients, and I connect the input node of MITE Q_n to MITE $Q_{n'}$ through $\frac{p_n q_{n'}}{k}$ unit weighting coefficients. In terms of $\mathbf{W}_{\text{in}}$, making these connections corresponds to setting $\left(\mathbf{W}_{\text{in}}\right)_{nn} = \frac{p_{n'}q_n}{k}$ and setting $\left(\mathbf{W}_{\text{in}}\right)_{n'n} = \frac{p_n q_{n'}}{k}$.
4. I return to step 2.
5. For each of the remaining numerator inputs currents, I perform the following steps. For the *n*th input, I first allocate a new MITE, labeling it Q_n. I diode connect MITE Q_n through q_n unit weighting coefficients, and I connect the input node of MITE Q_n to MITE Q_{N+1} through p_n unit weighting coefficients. In terms of the

$N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and of the $N \times 1$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, making these connections corresponds to setting $\left(\mathbf{W}_{\text{in}}\right)_{nn} = q_n$ and setting $\left(\mathbf{W}_{\text{out}}\right)_{1n} = p_n$. Once I exhaust the numerator input currents, I proceed to step 7.

6. For each of the remaining denominator input currents, I perform the following steps. For the *n*th input, I first allocate a new MITE, labeling it Q_n. Suppose that the MITE created in step 1 was MITE $\mathrm{Q}_{n'}$. Let k denote the greatest common divisor of $p_{n'}q_n$ and $p_n q_{n'}$. I then diode connect MITE Q_n through $\frac{p_{n'}q_n}{k}$ unit weighting coefficients, and I connect the input node of MITE Q_n to MITE $\mathrm{Q}_{n'}$ through $\frac{p_n q_{n'}}{k}$ unit weighting coefficients. In terms of $\mathbf{W}_{\text{in}}$, making these connections corresponds to setting $\left(\mathbf{W}_{\text{in}}\right)_{nn} = \frac{p_{n'}q_n}{k}$ and setting $\left(\mathbf{W}_{\text{in}}\right)_{n'n} = \frac{p_n q_{n'}}{k}$.
7. I denote by K the largest number of inputs feeding into any MITE. If each MITE happens to have K inputs, I am done. Otherwise, I add sufficient grounded unit weighting coefficients to each MITE that they each have K inputs.
8. I complete the network by connecting each of the grounded inputs that I added in step 7 to the input node of the final MITE created in the construction process.

In general, the input connectivity matrix, $\mathbf{W}_{\text{in}}$, resulting from the procedure just described will not be upper triangular. However, at step 7, if I renumber the inputs according to the order in which I chose them in the construction of the MITE network, the input connectivity matrix will be upper triangular, because all the connections represented by off-diagonal elements are made to MITEs with lower numbers, and, hence, appear above the main diagonal in $\mathbf{W}_{\text{in}}$. Thus, the cascade MITE network constructed by this procedure satisfies the sufficient condition for asymptotic stability that we found in Section 2.6.

4.3.3. An Illustrative Example

In this section, I illustrate each of the two MITE-network construction procedures described in Sections 4.3.1 and 4.3.2 by applying them in turn to generate MITE networks embodying the following relationship:

$$I_6 = \frac{I_1^{\frac{1}{2}} I_2^{\frac{3}{2}} I_3^2}{I_4^2 I_5}. \tag{4.10}$$

For this expression, I have that

$$\Lambda = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} & 2 & -2 & -1 \end{bmatrix}. \tag{4.11}$$

First, using the procedure described in Section 4.3.1, I construct a two-layer MITE

network realizing Equation 4.10. This process is illustrated in Figures 4.2, 4.3, and 4.4. I begin by allocating a MITE corresponding to the output current, I_6, and labeling it Q_6. Then, I construct the numerator layer as shown in Figure 4.2. I start with input current I_1, and I allocate another MITE, Q_1, for it. Because I_1 is raised to the $\frac{1}{2}$ power in Equation 4.10, I diode connect MITE Q_1 through two unit inputs, and I connect the input node of MITE Q_1 to MITE Q_6 through one unit input, as shown in Figure 4.2a. Next, I consider input current I_2, to which I allocate another MITE, Q_2. Because I_2 is raised to the $\frac{3}{2}$ power in Equation 4.10, I diode connect MITE Q_2 through two unit inputs, and I connect the input node of MITE Q_2 to MITE Q_6 through three unit inputs, as shown in Figure 4.2b. Finally, I consider input current I_3; I allocate another MITE, Q_3, for it. Because I_3 is squared in Equation 4.10, I diode connect MITE Q_3 through one unit input, and I connect the input node of MITE Q_3 to MITE Q_6 through two unit inputs.

Next, I construct the denominator layer of the two-layer MITE network realizing Equation 4.10, as shown in Figure 4.3. I begin with input current I_4; I allocate a new MITE, Q_4, for it. I choose to connect MITE Q_4 to MITE Q_6 through MITE Q_2. Because I_2 is raised to the $\frac{3}{2}$ power and I_4 is raised to the $-\frac{2}{1}$ power, I need to find the greatest common divisor of $2\times 2 = 4$ and $3\times 1 = 3$, which is equal to 1. Then, I diode connect MITE Q_4 through $\frac{3\times 1}{1} = 3$ unit inputs, and I connect the input node of MITE Q_4 to MITE Q_2 through $\frac{2\times 2}{1} = 4$ unit inputs, as shown in Figure 4.3a. Next, I consider input current I_5; I allocate another MITE, Q_5, for it. I choose to connect MITE Q_5 to MITE Q_6 through MITE Q_3. Because I_3 is raised to the $\frac{2}{1}$ power and I_5 is raised to the $-\frac{1}{1}$ power, I need to find the greatest common divisor of $2\times 1 = 2$ and $1\times 1 = 1$, which is equal to 1. Then, I diode connect MITE Q_5 through $\frac{2\times 1}{1} = 2$ unit inputs, and I connect the input node of MITE Q_4 to MITE Q_2 through $\frac{1\times 1}{1} = 1$ unit input, as shown in Figure 4.3b.

Finally, I count the largest number of inputs possessed by any MITE in the network of Figure 4.3b; this number is 6. Consequently, I add four grounded inputs to MITEs Q_1, Q_3, and Q_5, and I add three grounded inputs to MITE Q_4, as shown in Figure 4.4a. The resulting MITE network embodies Equation 4.10, but some of the transconductance of MITEs Q_1, Q_3, Q_4, and Q_5 is wasted. I complete this MITE network by connecting all the unused inputs added in Figure 4.4a to the input node of MITE Q_5. The resulting MITE network also embodies Equation 4.10, but there is no wasted MITE transconductance.

Next, using the procedure described in Section 4.3.2, I construct a cascade MITE network that embodies Equation 4.10. This process is illustrated in Figures 4.5, 4.6, and 4.7. Again, I begin by allocating a MITE corresponding to the output current, I_6, and labeling it Q_6. I choose input current I_1 from the numerator of Equation 4.10; I allocate a

MITE, Q_1, for it. Because I_1 is raised to the $\frac{1}{2}$ power, I diode connect MITE Q_1 through two unit inputs, and I connect the input node of MITE Q_1 to MITE Q_6 through one unit input, as shown in Figure 4.5a. Next, I choose input current I_5 from the denominator of Equation 4.10; I allocate a MITE, Q_5, for it. Because I_1 is raised to the $\frac{1}{2}$ power and I_5 is raised to the $-\frac{1}{1}$ power, I need to find the greatest common divisor of $1 \times 1 = 1$ and $2 \times 1 = 2$, which is equal to 1. So, I diode connect MITE Q_5 through $\frac{1 \times 1}{1} = 1$ unit input, and I connect the input node of MITE Q_5 to MITE Q_1 through $\frac{2 \times 1}{1} = 2$ unit inputs, as shown in Figure 4.5b. Next, I choose input current I_3 from the numerator of Equation 4.10; I allocate a MITE, Q_3, for it. Because I_5 is raised to the $-\frac{1}{1}$ power and I_3 is raised to the $\frac{2}{1}$ power, I need to find the greatest common divisor of $2 \times 1 = 2$ and $1 \times 1 = 1$, which is equal to 1. So, I diode connect MITE Q_3 through $\frac{1 \times 1}{1} = 1$ unit input, and I connect the input node of MITE Q_3 to MITE Q_5 through $\frac{2 \times 1}{1} = 2$ unit inputs, as shown in Figure 4.5c. Next, I choose input current I_4 from the denominator of Equation 4.10; I allocate a MITE, Q_4, for it. Because I_3 is raised to the $\frac{2}{1}$ power and I_4 is raised to the $-\frac{2}{1}$ power, I need to find the greatest common divisor of $2 \times 1 = 2$ and $2 \times 1 = 2$, which is equal to 2. So, I diode connect MITE Q_4 through $\frac{2 \times 1}{2} = 1$ unit input, and I connect the input node of MITE Q_4 to MITE Q_3 through $\frac{2 \times 1}{2} = 1$ unit input, as shown in Figure 4.6a. Finally, I choose input current I_2 from the numerator of Equation 4.10, and I allocate a MITE, Q_2, for it. Because I_4 is raised to the $-\frac{2}{1}$ power and I_2 is raised to the $\frac{3}{2}$ power, I need to find the greatest common divisor of $3 \times 1 = 3$ and $2 \times 2 = 4$, which is equal to 1. So, I diode connect MITE Q_2 through $\frac{2 \times 2}{1} = 4$ unit inputs, and I connect the input node of MITE Q_2 to MITE Q_3 through $\frac{3 \times 1}{1} = 3$ unit inputs, as shown in Figure 4.6b.

Finally, I count the largest number of inputs possessed by any MITE in the network of Figure 4.6b; this number is 4. Consequently, I add one grounded input to MITE Q_5, I add two grounded inputs to MITE Q_3, and I add three grounded inputs to MITE Q_6. The resulting MITE network, which is shown in Figure 4.6c, embodies Equation 4.10, but some of the transconductance of MITEs Q_3, Q_5, and Q_6 is wasted. I complete this MITE network by connecting all the unused inputs added in Figure 4.6c to the input node of MITE Q_2. The resulting MITE network, which is shown in Figure 4.7, also embodies Equation 4.10, but there is no wasted MITE transconductance.

4.4. The Synthesis Problem as an Integer Linear Program

As I mentioned in Section 4.2, I can treat the synthesis problem mathematically as the problem of assigning nonnegative values to the unknown elements of the $N \times N$ input

connectivity matrix, $\mathbf{W}_{\text{in}}$, and to those of the $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, given the $M \times N$ values of the power matrix, Λ, such that

$$\Lambda = \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1}, \tag{4.11}$$

and such that

$$\sum_{n=1}^{N} \Lambda_{mn} = 1, \quad m = 1,\ldots,M. \tag{4.12}$$

I can transform Equation 4.11 into a linear system of equations in standard form by postmultiplying both sides of Equation 4.11 by $\mathbf{W}_{\text{in}}$ and then subtracting $\mathbf{W}_{\text{out}}$ from both sides of Equation 4.11 to obtain

$$\Lambda \mathbf{W}_{\text{in}} - \mathbf{W}_{\text{out}} = \mathbf{0}. \tag{4.13}$$

Provided that $\mathbf{W}_{\text{in}}$ is nonsingular, if I obtain a solution to Equation 4.13, it will also satisfy Equation 4.11. However, in Equation 4.13, because there are $N^2 + NM$ variables related by NM linear equations, I cannot simply solve for the unknown elements of $\mathbf{W}_{\text{in}}$ and those of $\mathbf{W}_{\text{out}}$.

Here, I restrict the elements of Λ to be rational numbers, which corresponds to restricting my attention to MITE networks that use integral numbers of unit weighting coefficients. For convenience, I take the nominal value of each unit weighting coefficient to be unity. Under these conditions, the synthesis problem becomes the problem of assigning nonnegative integer values to the unknown elements of $\mathbf{W}_{\text{in}}$ and to those of $\mathbf{W}_{\text{out}}$ such that Equation 4.13 is satisfied. If I can identify any additional linear constraints on the components of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ and if I can find a linear objective function, which, when minimized subject to all of the constraints, yields a "good" MITE network, then I have framed the synthesis problem as an **integer linear program** [2, 3] and I can use well-established algorithms to synthesize MITE networks. Note that, with this approach, I am not restricted to having a single output current and that an integer–linear-programming algorithm can produce a single optimized MITE network that simultaneously implements multiple expressions of overlapping sets of input currents. In this case, I can share the input MITEs corresponding to the common input currents and I am not required to supply multiple copies of any input currents. This situation is good both from the standpoint of reducing circuit complexity and from the standpoint of reducing offsets associated with making multiple copies of an input current. Instead, if I were to design a number of independent single-output MITE networks, each implementing a different expression, I could have unnecessary hardware and I would need to supply multiple copies of the input currents that are common to the multiple expressions.

As I noted in Section 4.2, in the end, I must have the same number of inputs per MITE in a MITE network, and, if I do not use all of them, then I am wasting MITE transconductance. I can use these observations to further constrain the integer linear program as follows. I introduce another nonnegative integer variable, K, which represents the number of inputs per MITE, into the integer linear program. The conditions that each MITE possess the same number of inputs and that all MITE inputs are connected to some input node (i.e., are not unused) are equivalent to constraining each of the rows of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to sum to K. Thus, I introduce an additional linear $N+M$ constraint equations given by

$$\begin{cases} \sum_{k=1}^{N} \left(\mathbf{W}_{\text{in}}\right)_{nk} - K = 0, \quad n = 1,\ldots,N, \\ \sum_{k=1}^{N} \left(\mathbf{W}_{\text{out}}\right)_{mk} - K = 0, \quad m = 1,\ldots,M. \end{cases} \tag{4.14}$$

Further, as I noted in Section 4.2, I would prefer a MITE network with fewer inputs per MITE; consequently, I can choose K to be a simple linear objective function that I seek to minimize.

Instead, if I choose not to constrain the rows of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to sum to the same number, then I can use the sum of all the components of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ given by

$$\left\|\mathbf{W}_{\text{in}}\right\|_1 + \left\|\mathbf{W}_{\text{out}}\right\|_1 = \sum_{n=1}^{N}\sum_{k=1}^{N}\left(\mathbf{W}_{\text{in}}\right)_{nk} + \sum_{m=1}^{M}\sum_{k=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{mk},$$

where $\|\bullet\|_1$ denotes the L_1 matrix norm. When I have a suitable MITE network, I can add grounded MITE inputs where they are required. To utilize all of the available MITE transconductance, I can apply the completion transformation that I described in Section 4.2 to the final MITE network.

When I have some *a priori* knowledge about the structure of the MITE network that I would like to synthesize, I can use this knowledge to further constrain the integer linear program by judiciously numbering the input currents and by constraining some of the components of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to be zero. For example, suppose that I would like to embody a single expression in a two-layer MITE network. Suppose that there are J currents in the numerator and $N-J$ currents in the denominator of the expression. If I number the inputs such that all of the numerator inputs have lower numbers than all of the denominator inputs, if I apply Equation 4.14, and if I use K as the objective function, then I can impose the following structure on $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$:

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{0} & \mathbf{W}_{21} \end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix} \mathbf{W}_{31} & \mathbf{0} \end{bmatrix},$$

where $\mathbf{W}_{11}$ is a $J \times J$ diagonal matrix containing the self connections in the numerator layer of MITEs; $\mathbf{W}_{12}$ is a $J \times (N-J)$ full matrix containing the connections from the denominator layer to the numerator layer; $\mathbf{W}_{21}$ is a $(N-J) \times (N-J)$ diagonal matrix containing the self connections in the denominator layer of MITEs; and $\mathbf{W}_{31}$ is a $1 \times J$ matrix containing the connections from the numerator layer input MITEs to the output MITE. In this case, I have $N + J(N-J) + J + 1$ variables related by $2N+1$ linear constraint equations. It is interesting to note that, when $J = 1$ and when $J = N$, I have the same number of variables as equations, and hence I can solve the system explicitly. The former case corresponds to having one current in the numerator of the expression to be embodied in a MITE network. The latter case corresponds to having all input currents in the numerator of the expression to be embodied in a MITE network.

It is also possible to construct a nonlinear objective function that I can use to measure the merit of a MITE network and that I can minimize subject to the constraints expressed in Equations 4.13 and 4.14. In this case, I would have to use some more general constrained-optimization algorithm. For example, I might consider using some matrix norm of the error matrix given by Equation 4.4, so that the resulting MITE network is minimally sensitive to weighting-coefficient mismatch. However, if I choose this objective function, then the number of inputs in a MITE network will grow without bound in order to reduce the sensitivity of the MITE network to weighting-coefficient mismatch. To see why the number of inputs per MITE will grow without bound in order to minimize the variance of the errors in the power laws for a given level of weighting-coefficient mismatch, I make use of the following theorem.

Theorem 4.3: If all the weights of a MITE network are scaled by some integer, *n*, then the variances in the errors of the power laws of the resulting MITE network are $\frac{1}{n}$ times the variances in the errors of the power laws of the original MITE network.

Proof: Suppose that I have a MITE network specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$. The product-of-power-law relationships implemented by this MITE network are given by the $M \times N$ matrix of powers, $\Lambda = \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1}$. Now, suppose that I multiply each element of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ by some positive integer, *n*, to get new input and output connectivity matrices given by

$$\mathbf{W}'_{\text{in}} = n\mathbf{W}_{\text{in}} \quad \text{and} \quad \mathbf{W}'_{\text{out}} = n\mathbf{W}_{\text{out}}.$$

The inverse of the new input connectivity matrix is then given by

$$\mathbf{W}_{\text{in}}'^{-1} = \frac{1}{n}\mathbf{W}_{\text{in}}^{-1}.$$

By Theorem 4.2 in Section 4.2, I have that the power matrix of the new MITE network, Λ', is the same as that of the original MITE network. Now, I consider the quantity

$$\begin{aligned} E(\delta\Lambda' \circ \delta\Lambda') &= \frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}}' + (\Lambda' \circ \Lambda')\mathbf{W}_{\text{in}}'\right)\left(\mathbf{W}_{\text{in}}'^{-1} \circ \mathbf{W}_{\text{in}}'^{-1}\right) \\ &= \frac{\sigma^2}{w}\left(n\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)(n\mathbf{W}_{\text{in}})\right)\left(\left(\frac{1}{n}\mathbf{W}_{\text{in}}^{-1}\right) \circ \left(\frac{1}{n}\mathbf{W}_{\text{in}}^{-1}\right)\right) \\ &= \frac{n}{n^2}\frac{\sigma^2}{w}\left(\mathbf{W}_{\text{out}} + (\Lambda \circ \Lambda)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1} \circ \mathbf{W}_{\text{in}}^{-1}\right) \\ &= \frac{1}{n}E(\delta\Lambda \circ \delta\Lambda), \end{aligned}$$

which is just what I set out to show. ■

Thus, if a constrained optimization algorithm is trying to adjust the values of the components of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$, so as to minimize some norm of the matrix of error variances, it can make the value of this matrix norm arbitrarily close to zero by scaling all the weights by some large number while simultaneously satisfying all of the constraints. Therefore, the weights, and hence, the number of inputs per MITE will increase without bound. Consequently, if I use Equation 4.4 to construct an objective function, so as to penalize MITE networks with higher sensitivity to weighting-coefficient mismatch, I should also penalize networks with a large number of inputs per MITE. I might consider using an objective function like

$$\lambda_1\left(\|\mathbf{W}_{\text{in}}\| + \|\mathbf{W}_{\text{out}}\|\right) + \lambda_2\left\|E(\delta\Lambda \circ \delta\Lambda)\right\|,$$

where λ_1 reflects the cost of adding more MITE inputs to the network, λ_2 reflects the cost of being sensitive to weighting-coefficient mismatch, and $\|\bullet\|$ denotes some matrix norm.

4.5. Appendix 4.A

Suppose that I have a MITE network specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, such that the powers in each row of the $M \times N$ matrix given by $\Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}$ sum to unity (i.e., for each m between 1 and M, $\sum_{n=1}^{N}\Lambda_{mn} = 1$). In this appendix, I construct a new MITE network from the original MITE network that has an input connectivity matrix, $\mathbf{W}_{\text{in}}'$, given by

$$\mathbf{W}_{\text{in}}' = \mathbf{W}_{\text{in}} + w\mathbf{e}_j^{(N)},$$

and an output connectivity matrix, $\mathbf{W}_{\text{out}}'$, given by

$$\mathbf{W}'_{\text{out}} = \mathbf{W}_{\text{out}} + w\mathbf{e}_j^{(M)},$$

where $\mathbf{e}_j^{(n)}$ denotes an $n \times N$ matrix with 1s in the jth column and 0s everywhere else. I show that the product-of-power-law relationships implemented by the new MITE network, given by $\Lambda' = \mathbf{W}'_{\text{out}}\mathbf{W}'^{-1}_{\text{in}}$, are the same as those of the original MITE network, given by $\Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}$. The new connectivity matrices, $\mathbf{W}'_{\text{in}}$ and $\mathbf{W}'_{\text{out}}$, are those that I obtain by adding a weighting coefficient with value w to each MITE in the original network and connecting each of these weighting coefficients to the jth input-node voltage, V_j.

Now, I consider the quantity

$$\begin{aligned}
\mathbf{W}'^{-1}_{\text{in}} &= \left(\mathbf{W}_{\text{in}} + w\mathbf{e}_j^{(N)}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\mathbf{W}_{\text{in}}\left(\mathbf{W}_{\text{in}} + w\mathbf{e}_j^{(N)}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{W}_{\text{in}}\left(\mathbf{W}_{\text{in}} + w\mathbf{e}_j^{(N)}\right)^{-1}\right) \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\left(\mathbf{W}_{\text{in}} + w\mathbf{e}_j^{(N)}\right)\mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} + w\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - w\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} + \left(w\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^2 - \left(w\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^3 + \ldots\right) \\
&= \mathbf{W}_{\text{in}}^{-1} - w\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} + w^2\mathbf{W}_{\text{in}}^{-1}\left(\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^2 \\
&\qquad - w^3\mathbf{W}_{\text{in}}^{-1}\left(\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^3 + \ldots . \qquad (4.15)
\end{aligned}$$

If I denote by ω_{ij} the ijth element of the inverse of the original input connectivity matrix, $\left(\mathbf{W}_{\text{in}}^{-1}\right)_{ij}$, then, I have that

$$\begin{aligned}
\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} &= \begin{bmatrix} & 1 & \\ 0 & \vdots & 0 \\ & 1 & \end{bmatrix}\begin{bmatrix} \omega_{11} & \cdots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \cdots & \omega_{NN} \end{bmatrix} \\
&= \begin{bmatrix} \omega_{j1} & \cdots & \omega_{jN} \\ \vdots & \ddots & \vdots \\ \omega_{j1} & \cdots & \omega_{jN} \end{bmatrix}.
\end{aligned}$$

Further, I have that

$$\begin{aligned}
\left(\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^2 &= \mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\,\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} \\
&= \begin{bmatrix} \omega_{j1} & \cdots & \omega_{jN} \\ \vdots & \ddots & \vdots \\ \omega_{j1} & \cdots & \omega_{jN} \end{bmatrix}\begin{bmatrix} \omega_{j1} & \cdots & \omega_{jN} \\ \vdots & \ddots & \vdots \\ \omega_{j1} & \cdots & \omega_{jN} \end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} \omega_{j1}\sum_{n=1}^{N}\omega_{jn} & \cdots & \omega_{jN}\sum_{n=1}^{N}\omega_{jn} \\ \vdots & \ddots & \vdots \\ \omega_{j1}\sum_{n=1}^{N}\omega_{jn} & \cdots & \omega_{jN}\sum_{n=1}^{N}\omega_{jn} \end{bmatrix} \\
&= \left(\sum_{n=1}^{N}\omega_{jn}\right)\begin{bmatrix} \omega_{j1} & \cdots & \omega_{jN} \\ \vdots & \ddots & \vdots \\ \omega_{j1} & \cdots & \omega_{jN} \end{bmatrix} \\
&= \left(\sum_{n=1}^{N}\omega_{jn}\right)\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1},
\end{aligned}
$$

from which it follows that

$$
\left(\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right)^k = \left(\sum_{n=1}^{N}\omega_{jn}\right)^{k-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}.
$$

Using this result, I can write Equation 4.15 as

$$
\begin{aligned}
\mathbf{W}_{\text{in}}'^{-1} &= \mathbf{W}_{\text{in}}^{-1} - w\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} + w^2\left(\sum_{n=1}^{N}\omega_{jn}\right)\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} \\
&\qquad - w^3\left(\sum_{n=1}^{N}\omega_{jn}\right)^2\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} + \ldots \\
&= \mathbf{W}_{\text{in}}^{-1} - w\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\left(1 + w\left(\sum_{n=1}^{N}\omega_{jn}\right) - w^2\left(\sum_{n=1}^{N}\omega_{jn}\right)^2 + \ldots\right) \\
&= \mathbf{W}_{\text{in}}^{-1} - \frac{w}{1+w\sum_{n=1}^{N}\omega_{jn}}\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}.
\end{aligned}
$$

Next, I consider the quantity

$$
\begin{aligned}
\Lambda' &= \mathbf{W}_{\text{out}}'\mathbf{W}_{\text{in}}'^{-1} \\
&= \left(\mathbf{W}_{\text{out}} + w\mathbf{e}_j^{(M)}\right)\left(\mathbf{W}_{\text{in}}^{-1} - \frac{w}{1+w\sum_{n=1}^{N}\omega_{jn}}\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1}\right) \\
&= \Lambda + w\mathbf{e}_j^{(M)}\mathbf{W}_{\text{in}}^{-1} - \frac{w}{1+w\sum_{n=1}^{N}\omega_{jn}}\Lambda\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} - \frac{w^2}{1+w\sum_{n=1}^{N}\omega_{jn}}\mathbf{e}_j^{(M)}\mathbf{W}_{\text{in}}^{-1}\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + w\mathbf{e}_j^{(M)}\mathbf{W}_{\text{in}}^{-1} - \frac{w}{1+w\sum_{n=1}^{N}\omega_{jn}}\Lambda\mathbf{e}_j^{(N)}\mathbf{W}_{\text{in}}^{-1} - \frac{w^2\sum_{n=1}^{N}\omega_{jn}}{1+w\sum_{n=1}^{N}\omega_{jn}}\mathbf{e}_j^{(M)}\mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + w\left(\mathbf{e}_j^{(M)} - \frac{1}{1+w\sum_{n=1}^{N}\omega_{jn}}\Lambda\mathbf{e}_j^{(N)} - \frac{w\sum_{n=1}^{N}\omega_{jn}}{1+w\sum_{n=1}^{N}\omega_{jn}}\mathbf{e}_j^{(M)}\right)\mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + w\left(\frac{1}{1+w\sum_{n=1}^{N}\omega_{jn}}\mathbf{e}_j^{(M)} - \frac{1}{1+w\sum_{n=1}^{N}\omega_{jn}}\Lambda\mathbf{e}_j^{(N)}\right)\mathbf{W}_{\text{in}}^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&= \Lambda + \frac{w}{1 + w\sum_{n=1}^{N}\omega_{jn}}\left(\mathbf{e}_j^{(M)} - \Lambda\mathbf{e}_j^{(N)}\right)\mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + \frac{w}{1 + w\sum_{n=1}^{N}\omega_{jn}}\begin{bmatrix} 0 & 1-\sum_{n=1}^{N}\Lambda_{1n} & 0 \\ & \vdots & \\ & 1-\sum_{n=1}^{N}\Lambda_{1M} & \end{bmatrix}\mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + \mathbf{0} \\
&= \Lambda .
\end{aligned}
$$

Thus, I have that the product-of-power-law relationships implemented by the transformed MITE network are identical to those embodied in the original MITE network.

4.6. Appendix 4.B

Suppose that I have a MITE network specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, such that the powers in each row of the $M \times N$ matrix given by $\Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}$ sum to unity (i.e., for each m between 1 and M, $\sum_{n=1}^{N}\Lambda_{mn} = 1$). In this appendix, I construct a new MITE network from the original MITE network that has an input connectivity matrix, $\mathbf{W}'_{\text{in}}$, given by

$$\mathbf{W}'_{\text{in}} = \mathbf{W}_{\text{in}} + \mathbf{A},$$

and an output connectivity matrix, $\mathbf{W}'_{\text{out}}$, given by

$$\mathbf{W}'_{\text{out}} = \mathbf{W}_{\text{out}} + \mathbf{B},$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & w^* - \sum_{n=1}^{N} w_{1n} & 0 \\ & \vdots & \\ & w^* - \sum_{n=1}^{N} w_{Nn} & \end{bmatrix},$$

and

$$\mathbf{B} = \begin{bmatrix} 0 & w^* - \sum_{n=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{1n} & 0 \\ & \vdots & \\ & w^* - \sum_{n=1}^{N}\left(\mathbf{W}_{\text{out}}\right)_{Mn} & \end{bmatrix},$$

and

$$w^* = \max_{1 \le k \le M+N}\left\{\sum_{n=1}^{N} w_{kn}\right\}.$$

I show that the product-of-power-law relationships implemented by the new MITE network, given by $\Lambda' = \mathbf{W}'_{\text{out}}\mathbf{W}'^{-1}_{\text{in}}$, are the same as those of the original MITE network, given by $\Lambda = \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}$. The new connectivity matrices, $\mathbf{W}'_{\text{in}}$ and $\mathbf{W}'_{\text{out}}$, are those that I

obtain by taking any weighting coefficients that are connected to ground and connecting them all to the jth input-node voltage, V_j.

Now, I consider the quantity

$$\begin{aligned}
\mathbf{W}_{\text{in}}'^{-1} &= \left(\mathbf{W}_{\text{in}} + \mathbf{A}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\mathbf{W}_{\text{in}}\left(\mathbf{W}_{\text{in}} + \mathbf{A}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{W}_{\text{in}}\left(\mathbf{W}_{\text{in}} + \mathbf{A}\right)^{-1}\right) \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\left(\mathbf{W}_{\text{in}} + \mathbf{A}\right)\mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} + \mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \mathbf{A}\mathbf{W}_{\text{in}}^{-1} + \left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)^2 - \left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)^3 + \ldots\right). \qquad (4.16)
\end{aligned}$$

If I denote by ω_{ij} the ijth element of the inverse of the original input connectivity matrix, $\left(\mathbf{W}_{\text{in}}^{-1}\right)_{ij}$, then I have that

$$\begin{aligned}
\mathbf{A}\mathbf{W}_{\text{in}}^{-1} &= \begin{bmatrix} 0 & w^* - \sum_{n=1}^{N} w_{1n} & 0 \\ & \vdots & \\ & w^* - \sum_{n=1}^{N} w_{Nn} & \end{bmatrix}\begin{bmatrix} \omega_{11} & \cdots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \cdots & \omega_{NN} \end{bmatrix} \\
&= \begin{bmatrix} \omega_{11}\left(w^* - \sum_{n=1}^{N} w_{1n}\right) & \cdots & \omega_{1N}\left(w^* - \sum_{n=1}^{N} w_{1n}\right) \\ \vdots & \ddots & \vdots \\ \omega_{N1}\left(w^* - \sum_{n=1}^{N} w_{Nn}\right) & \cdots & \omega_{NN}\left(w^* - \sum_{n=1}^{N} w_{Nn}\right) \end{bmatrix}.
\end{aligned}$$

Consequently, I have that the pqth element of $\left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)^2$

$$\begin{aligned}
\left(\left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)^2\right)_{pq} &= \left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)_{pq} \\
&= \omega_{jq}\left(w^* - \sum_{n=1}^{N} w_{pn}\right)\sum_{k=1}^{N}\omega_{jk}\left(w^* - \sum_{n=1}^{N} w_{kn}\right) \\
&= \omega_{jq}\left(w^* - \sum_{n=1}^{N} w_{pn}\right)\left(w^*\sum_{k=1}^{N}\omega_{jk} - \sum_{k=1}^{N}\omega_{jk}\sum_{n=1}^{N} w_{kn}\right) \\
&= \omega_{jq}\left(w^* - \sum_{n=1}^{N} w_{pn}\right)\left(w^*\sum_{k=1}^{N}\omega_{jk} - \sum_{n=1}^{N}\sum_{k=1}^{N} w_{kn}\omega_{jk}\right) \\
&= \omega_{jq}\left(w^* - \sum_{n=1}^{N} w_{pn}\right)\left(w^*\sum_{k=1}^{N}\omega_{jk} - \sum_{n=1}^{N}\delta_{jn}\right) \\
&= \omega_{jq}\left(w^* - \sum_{n=1}^{N} w_{pn}\right)\left(w^*\sum_{k=1}^{N}\omega_{jk} - 1\right) \\
&= \left(w^*\sum_{k=1}^{N}\omega_{jk} - 1\right)\left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)_{pq}.
\end{aligned}$$

Applying the preceding equation recursively, I have that

$$\left(\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right)^k = \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right)^{k-1} \mathbf{A}\mathbf{W}_{\text{in}}^{-1}.$$

Using this result, I can write Equation 4.16 as

$$\begin{aligned}
\mathbf{W}_{\text{in}}'^{-1} &= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \mathbf{A}\mathbf{W}_{\text{in}}^{-1} + \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right)\mathbf{A}\mathbf{W}_{\text{in}}^{-1} \right. \\
&\qquad \left. - \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right)^2 \mathbf{A}\mathbf{W}_{\text{in}}^{-1} + \ldots\right) \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \mathbf{A}\mathbf{W}_{\text{in}}^{-1}\left(1 - \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right) + \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right)^2 - \ldots\right)\right) \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \mathbf{A}\mathbf{W}_{\text{in}}^{-1}\left(\frac{1}{1 + w^* \sum_{k=1}^{N} \omega_{jk} - 1}\right)\right) \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right) \\
&= \mathbf{W}_{\text{in}}^{-1} - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \mathbf{W}_{\text{in}}^{-1}\mathbf{A}\mathbf{W}_{\text{in}}^{-1}.
\end{aligned}$$

Next, I consider the quantity

$$\begin{aligned}
\Lambda' &= \mathbf{W}_{\text{out}}' \mathbf{W}_{\text{in}}'^{-1} \\
&= \left(\mathbf{W}_{\text{out}} + \mathbf{B}\right)\left(\mathbf{W}_{\text{in}}^{-1} - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \mathbf{W}_{\text{in}}^{-1}\mathbf{A}\mathbf{W}_{\text{in}}^{-1}\right) \\
&= \Lambda - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \Lambda\mathbf{A}\mathbf{W}_{\text{in}}^{-1} + \mathbf{B}\mathbf{W}_{\text{in}}^{-1} - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \mathbf{B}\mathbf{W}_{\text{in}}^{-1}\mathbf{A}\mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}}\left(\Lambda\mathbf{A} - \left(w^* \sum_{k=1}^{N} \omega_{jk}\right)\mathbf{B} + \mathbf{B}\mathbf{W}_{\text{in}}^{-1}\mathbf{A}\right)\mathbf{W}_{\text{in}}^{-1}. \qquad (4.17)
\end{aligned}$$

To further simplify Equation 4.17, I consider the matrix quantities in the parentheses separately. First, I consider the quantity

$$\begin{aligned}
\Lambda\mathbf{A} &= \begin{bmatrix} \Lambda_{11} & \cdots & \Lambda_{1N} \\ \vdots & \ddots & \vdots \\ \Lambda_{N1} & \cdots & \Lambda_{NN} \end{bmatrix}\begin{bmatrix} 0 & w^* - \sum_{n=1}^{N} w_{1n} & 0 \\ & \vdots & \\ & w^* - \sum_{n=1}^{N} w_{Nn} & \end{bmatrix} \\
&= \begin{bmatrix} 0 & w^* \sum_{k=1}^{N} \Lambda_{1k} - \sum_{k=1}^{N} \Lambda_{1k} \sum_{n=1}^{N} w_{kn} & 0 \\ & \vdots & \\ & w^* \sum_{k=1}^{N} \Lambda_{Mk} - \sum_{k=1}^{N} \Lambda_{Mk} \sum_{n=1}^{N} w_{kn} & \end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} & w^* \sum_{k=1}^N \Lambda_{1k} - \sum_{n=1}^N \sum_{k=1}^N \Lambda_{1k} w_{kn} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \Lambda_{Mk} - \sum_{n=1}^N \sum_{k=1}^N \Lambda_{Mk} w_{kn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* \sum_{k=1}^N \Lambda_{1k} - \sum_{n=1}^N \sum_{k=1}^N \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1j} \omega_{jk} w_{kn} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \Lambda_{Mk} - \sum_{n=1}^N \sum_{k=1}^N \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mj} \omega_{jk} w_{kn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* \sum_{k=1}^N \Lambda_{1k} - \sum_{n=1}^N \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1j} \sum_{k=1}^N \omega_{jk} w_{kn} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \Lambda_{Mk} - \sum_{n=1}^N \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mj} \sum_{k=1}^N \omega_{jk} w_{kn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* \sum_{k=1}^N \Lambda_{1k} - \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1j} \sum_{n=1}^N \delta_{jn} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \Lambda_{Mk} - \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mj} \sum_{n=1}^N \delta_{jn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* \sum_{k=1}^N \Lambda_{1k} - \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1j} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \Lambda_{Mk} - \sum_{j=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mj} & \end{bmatrix}. \qquad (4.18)
\end{aligned}
$$

Next, I consider the quantity

$$
\begin{aligned}
\mathbf{B}\mathbf{W}_{\text{in}}^{-1}\mathbf{A} &= \begin{bmatrix} & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mn} & \end{bmatrix} \begin{bmatrix} \omega_{11} & \cdots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \cdots & \omega_{NN} \end{bmatrix} \begin{bmatrix} & w^* - \sum_{n=1}^N w_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^N w_{Nn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mn} & \end{bmatrix} \begin{bmatrix} & w^* \sum_{k=1}^N \omega_{1k} - \sum_{k=1}^N \omega_{1k} \sum_{n=1}^N w_{kn} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \omega_{Nk} - \sum_{k=1}^N \omega_{Nk} \sum_{n=1}^N w_{kn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mn} & \end{bmatrix} \begin{bmatrix} & w^* \sum_{k=1}^N \omega_{1k} - \sum_{n=1}^N \sum_{k=1}^N \omega_{1k} w_{kn} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \omega_{Nk} - \sum_{n=1}^N \sum_{k=1}^N \omega_{Nk} w_{kn} & \end{bmatrix} \\
&= \begin{bmatrix} & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^N \left(\mathbf{W}_{\text{out}}\right)_{Mn} & \end{bmatrix} \begin{bmatrix} & w^* \sum_{k=1}^N \omega_{1k} - \sum_{n=1}^N \delta_{1n} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^N \omega_{Nk} - \sum_{n=1}^N \delta_{Nn} & \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{bmatrix} & w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{Mn} & \end{bmatrix} \begin{bmatrix} & w^* \sum_{k=1}^{N} \omega_{1k} - 1 & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^{N} \omega_{Nk} - 1 & \end{bmatrix} \\
&= \begin{bmatrix} & \left(w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{1n}\right)\left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right) & \\ 0 & \vdots & 0 \\ & \left(w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{Mn}\right)\left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right) & \end{bmatrix} \\
&= \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right) \begin{bmatrix} & w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{Mn} & \end{bmatrix} \\
&= \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1\right) \mathbf{B}. \qquad (4.19)
\end{aligned}
$$

Substituting Equations 4.18 and 4.19 into Equation 4.17, I obtain

$$
\begin{aligned}
\Lambda' &= \Lambda - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \left(\begin{bmatrix} & w^* \sum_{k=1}^{N} \Lambda_{1k} - \sum_{j=1}^{N} (\mathbf{W}_{\text{out}})_{1j} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^{N} \Lambda_{Mk} - \sum_{j=1}^{N} (\mathbf{W}_{\text{out}})_{Mj} & \end{bmatrix} \right. \\
&\qquad \left. - \left(w^* \sum_{k=1}^{N} \omega_{jk} \right) \mathbf{B} + \left(w^* \sum_{k=1}^{N} \omega_{jk} - 1 \right) \mathbf{B} \right) \mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda - \frac{1}{w^* \sum_{k=1}^{N} \omega_{jk}} \left(\begin{bmatrix} & w^* \sum_{k=1}^{N} \Lambda_{1k} - \sum_{j=1}^{N} (\mathbf{W}_{\text{out}})_{1j} & \\ 0 & \vdots & 0 \\ & w^* \sum_{k=1}^{N} \Lambda_{Mk} - \sum_{j=1}^{N} (\mathbf{W}_{\text{out}})_{Mj} & \end{bmatrix} \right. \\
&\qquad \left. - \begin{bmatrix} & w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{1n} & \\ 0 & \vdots & 0 \\ & w^* - \sum_{n=1}^{N} (\mathbf{W}_{\text{out}})_{Mn} & \end{bmatrix} \right) \mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + \frac{w^*}{w^* \sum_{k=1}^{N} \omega_{jk}} \begin{bmatrix} & 1 - \sum_{k=1}^{N} \Lambda_{1k} & \\ 0 & \vdots & 0 \\ & 1 - \sum_{k=1}^{N} \Lambda_{Mk} & \end{bmatrix} \mathbf{W}_{\text{in}}^{-1} \\
&= \Lambda + \mathbf{0} \\
&= \Lambda .
\end{aligned}
$$

Thus, I have that the product-of-power-law relationships implemented by the transformed MITE network are identical to those embodied in the original MITE network.

4.7. Appendix 4.C

Suppose that I have a MITE network specified by an $N \times N$ input connectivity matrix, $\mathbf{W}_{\text{in}}$, and by an $M \times N$ output connectivity matrix, $\mathbf{W}_{\text{out}}$, such that each element of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ is an integral multiple of some unit weighting coefficient with value w. I assume that the nominal value of each unit weighting coefficient is perturbed by a small zero-mean Gaussian random variable. In this appendix, I derive a simple approximate formula relating the variance of the error in each of the powers contained in the $M \times N$ matrix given by $\Lambda = \mathbf{W}_{\text{out}} \mathbf{W}_{\text{in}}^{-1}$ to the variance of the random perturbations in the values of the unit weighting coefficients. If we have a number of alternate MITE-network topologies that implement a given set of product-of-power-law relationships, we can use this formula to evaluate how sensitive each topology is to component mismatch.

I begin by supposing that each element of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ is some integral number of unit weighting coefficients, each with nominal value w, so I can write the ijth component of either $\mathbf{W}_{\text{in}}$ or $\mathbf{W}_{\text{out}}$ as

$$w_{ij} = \sum_{n=1}^{n_{ij}} w = n_{ij} w .$$

Further, I suppose that each of these unit weighting coefficients is perturbed by a zero-mean Gaussian random whose standard deviation, σ, is small compared to w (i.e., on the order of a few percent or less), and that each of these random perturbation is statistically independent of all the others. Thus, I can write the perturbed ijth component of either $\mathbf{W}_{\text{in}}$ or $\mathbf{W}_{\text{out}}$ as

$$\begin{aligned} w'_{ij} &= \sum_{n=1}^{n_{ij}} \left(w + \delta w_{ij}^{(n)} \right) \\ &= n_{ij} w + \sum_{n=1}^{n_{ij}} \delta w_{ij}^{(n)} \\ &= w_{ij} + \delta w_{ij}, \end{aligned}$$

where $\delta w_{ij}^{(n)}$ denotes the random perturbation in the nth unit weighting coefficient making up w_{ij}, and δw_{ij} denotes the total random perturbation in w_{ij}. Consequently, I can write the perturbed input connectivity matrix, $\mathbf{W}'_{\text{in}}$, as

$$\mathbf{W}'_{\text{in}} = \mathbf{W}_{\text{in}} + \delta \mathbf{W}_{\text{in}},$$

and the perturbed output connectivity matrix, $\mathbf{W}'_{\text{out}}$, as

$$\mathbf{W}'_{\text{out}} = \mathbf{W}_{\text{out}} + \delta \mathbf{W}_{\text{out}},$$

where $\delta \mathbf{W}_{\text{in}}$ and $\delta \mathbf{W}_{\text{out}}$ are matrices containing the random perturbations.

Now, I denote by δw_{ij} the ijth component of either $\delta \mathbf{W}_{\text{in}}$ and $\delta \mathbf{W}_{\text{out}}$, and I compute the mean of δw_{ij} as follows:

$$\begin{aligned} E\left(\delta w_{ij}\right) &= E\left(\sum_{n=1}^{n_{ij}} \delta w_{ij}^{(n)}\right) \\ &= \sum_{n=1}^{n_{ij}} E\left(\delta w_{ij}^{(n)}\right) \\ &= 0. \end{aligned}$$

Because the mean of δw_{ij} is zero, I compute the variance of δw_{ij} as follows:

$$\begin{aligned} E\left(\delta w_{ij}^2\right) &= E\left(\left(\sum_{n=1}^{n_{ij}} \delta w_{ij}^{(n)}\right)\left(\sum_{n'=1}^{n_{ij}} \delta w_{ij}^{(n')}\right)\right) \\ &= E\left(\sum_{n=1}^{n_{ij}} \sum_{n'=1}^{n_{ij}} \delta w_{ij}^{(n)} \delta w_{ij}^{(n')}\right) \\ &= E\left(\sum_{n=1}^{n_{ij}} \left(\delta w_{ij}^{(n)}\right)^2 + \sum_{n=1}^{n_{ij}} \sum_{n' \neq n} \delta w_{ij}^{(n)} \delta w_{ij}^{(n')}\right) \\ &= \sum_{n=1}^{n_{ij}} E\left(\left(\delta w_{ij}^{(n)}\right)^2\right) + \sum_{n=1}^{n_{ij}} \sum_{n' \neq n} E\left(\delta w_{ij}^{(n)} \delta w_{ij}^{(n')}\right) \\ &= \sum_{n=1}^{n_{ij}} \sigma^2 + \sum_{n=1}^{n_{ij}} \sum_{n' \neq n} E\left(\delta w_{ij}^{(n)}\right) E\left(\delta w_{ij}^{(n')}\right) \\ &= n_{ij} \sigma^2 \\ &= \frac{\sigma^2}{w} w_{ij}. \end{aligned}$$

Thus, I can express the variance of the random perturbations contained in $\delta \mathbf{W}_{\text{in}}$ and $\delta \mathbf{W}_{\text{out}}$ in matrix notation as

$$E\left(\delta \mathbf{W}_{\text{in}} \circ \delta \mathbf{W}_{\text{in}}\right) = \frac{\sigma^2}{w} \mathbf{W}_{\text{in}},$$

and

$$E\left(\delta \mathbf{W}_{\text{out}} \circ \delta \mathbf{W}_{\text{out}}\right) = \frac{\sigma^2}{w} \mathbf{W}_{\text{out}},$$

where $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product (i.e., element-by-element product) of two matrices, $\mathbf{A}$ and $\mathbf{B}$, defined by $(\mathbf{A} \circ \mathbf{B})_{ij} \equiv a_{ij} b_{ij}$.

Now, I consider the quantity

$$\begin{aligned} \mathbf{W}_{\text{in}}'^{-1} &= \left(\mathbf{W}_{\text{in}} + \delta \mathbf{W}_{\text{in}}\right)^{-1} \\ &= \mathbf{W}_{\text{in}}^{-1} \mathbf{W}_{\text{in}} \left(\mathbf{W}_{\text{in}} + \delta \mathbf{W}_{\text{in}}\right)^{-1} \\ &= \mathbf{W}_{\text{in}}^{-1} \left(\mathbf{W}_{\text{in}} \left(\mathbf{W}_{\text{in}} + \delta \mathbf{W}_{\text{in}}\right)^{-1}\right) \\ &= \mathbf{W}_{\text{in}}^{-1} \left(\left(\mathbf{W}_{\text{in}} + \delta \mathbf{W}_{\text{in}}\right) \mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\ &= \mathbf{W}_{\text{in}}^{-1} \left(\mathbf{I} + \delta \mathbf{W}_{\text{in}} \mathbf{W}_{\text{in}}^{-1}\right)^{-1} \end{aligned}$$

$$= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1} + \left(\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)^2 - \left(\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)^3 + \ldots\right). \tag{4.20}$$

Next, using Equation 4.20, I compute the matrix of perturbed powers, Λ', as

$$\begin{aligned}
\Lambda' &= \mathbf{W}'_{\text{out}}\mathbf{W}'^{-1}_{\text{in}} \\
&= \left(\mathbf{W}_{\text{out}} + \delta\mathbf{W}_{\text{out}}\right)\mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1} + \left(\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)^2 - \left(\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)^3 + \ldots\right) \\
&= \left(\Lambda + \delta\mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}\right)\left(\mathbf{I} - \delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1} + \left(\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)^2 - \left(\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)^3 + \ldots\right) \\
&\approx \Lambda + \delta\mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} - \Lambda\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1},
\end{aligned}$$

neglecting terms of second order and above. I would like to compute the variance of the errors in the powers contained in Λ, which are given, to first order, by

$$\begin{aligned}
\delta\Lambda &= \Lambda' - \Lambda \\
&\approx \left(\delta\mathbf{W}_{\text{out}} - \Lambda\delta\mathbf{W}_{\text{in}}\right)\mathbf{W}_{\text{in}}^{-1}.
\end{aligned} \tag{4.21}$$

Now, I want to use Equation 4.21 to compute the variance of the errors in the powers contained in Λ. To do so, I need to compute $E(\delta\Lambda) \circ E(\delta\Lambda)$ and $E(\delta\Lambda \circ \delta\Lambda)$. First, I use Equation 4.21 to compute the quantity

$$\begin{aligned}
E(\delta\Lambda) &= E\left(\left(\delta\mathbf{W}_{\text{out}} - \Lambda\delta\mathbf{W}_{\text{in}}\right)\mathbf{W}_{\text{in}}^{-1}\right) \\
&= E\left(\left(\delta\mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} - \Lambda\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right)\right) \\
&= E\left(\delta\mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}\right) - E\left(\Lambda\delta\mathbf{W}_{\text{in}}\mathbf{W}_{\text{in}}^{-1}\right) \\
&= E\left(\delta\mathbf{W}_{\text{out}}\right)\mathbf{W}_{\text{in}}^{-1} - \Lambda E\left(\delta\mathbf{W}_{\text{in}}\right)\mathbf{W}_{\text{in}}^{-1} \\
&= \mathbf{0}\mathbf{W}_{\text{in}}^{-1} - \Lambda\mathbf{0}\mathbf{W}_{\text{in}}^{-1} \\
&= \mathbf{0}.
\end{aligned}$$

Consequently, $E(\delta\Lambda) \circ E(\delta\Lambda) = \mathbf{0}$ and, thus, the variance of the errors is simply given by $E(\delta\Lambda \circ \delta\Lambda)$. Next, denoting by ω_{ij} the ijth component of $\mathbf{W}_{\text{in}}^{-1}$, I compute the mnth component of $\delta\Lambda \circ \delta\Lambda$ as

$$\begin{aligned}
\left(\delta\Lambda\right)^2_{mn} &= \left(\sum_{k=1}^{N}\omega_{kn}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk} - \sum_{i=1}^{N}\omega_{in}\sum_{j=1}^{N}\Lambda_{mj}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\right) \\
&\qquad \times\left(\sum_{k'=1}^{N}\omega_{k'n}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk'} - \sum_{i'=1}^{N}\omega_{i'n}\sum_{j'=1}^{N}\Lambda_{mj'}\left(\delta\mathbf{W}_{\text{in}}\right)_{j'i'}\right) \\
&= \sum_{k=1}^{N}\sum_{k'=1}^{N}\omega_{kn}\omega_{k'n}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk'} + \sum_{i=1}^{N}\sum_{i'=1}^{N}\omega_{in}\omega_{i'n}\sum_{j=1}^{N}\sum_{j'=1}^{N}\Lambda_{mj}\Lambda_{mj'}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{in}}\right)_{j'i'} \\
&\qquad -2\sum_{k=1}^{N}\sum_{i=1}^{N}\omega_{kn}\omega_{in}\sum_{j=1}^{N}\Lambda_{mj}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk} \\
&= \sum_{k=1}^{N}\omega_{kn}^2\left(\delta\mathbf{W}_{\text{out}}\right)^2_{mk} + \sum_{k=1}^{N}\sum_{k'\neq k}\omega_{kn}\omega_{k'n}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk'}
\end{aligned}$$

$$+\sum_{i=1}^{N}\omega_{in}^{2}\sum_{j=1}^{N}\Lambda_{mj}^{2}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}^{2}+\sum_{i=1}^{N}\sum_{i'\neq i}\omega_{in}\omega_{i'n}\sum_{j=1}^{N}\sum_{j'\neq j}\Lambda_{mj}\Lambda_{mj'}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{in}}\right)_{j'i'}$$
$$-2\sum_{k=1}^{N}\sum_{i=1}^{N}\omega_{kn}\omega_{in}\sum_{j=1}^{N}\Lambda_{mj}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}. \tag{4.22}$$

Next, I take the expected value of Equation 4.22 and obtain

$$\begin{aligned}
E\left(\delta\Lambda\right)_{mn}^{2} = {} & E\Bigg(\sum_{k=1}^{N}\omega_{kn}^{2}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}^{2}+\sum_{k=1}^{N}\sum_{k'\neq k}\omega_{kn}\omega_{k'n}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk'} \\
& +\sum_{i=1}^{N}\omega_{in}^{2}\sum_{j=1}^{N}\Lambda_{mj}^{2}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}^{2}+\sum_{i=1}^{N}\sum_{i'\neq i}\omega_{in}\omega_{i'n}\sum_{j=1}^{N}\sum_{j'\neq j}\Lambda_{mj}\Lambda_{mj'}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{in}}\right)_{j'i'} \\
& -2\sum_{k=1}^{N}\sum_{i=1}^{N}\omega_{kn}\omega_{in}\sum_{j=1}^{N}\Lambda_{mj}\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}\Bigg) \\
= {} & \sum_{k=1}^{N}\omega_{kn}^{2}E\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}^{2}+\sum_{k=1}^{N}\sum_{k'\neq k}\omega_{kn}\omega_{k'n}E\left(\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk'}\right) \\
& +\sum_{i=1}^{N}\omega_{in}^{2}\sum_{j=1}^{N}\Lambda_{mj}^{2}E\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}^{2}+\sum_{i=1}^{N}\sum_{i'\neq i}\omega_{in}\omega_{i'n}\sum_{j=1}^{N}\sum_{j'\neq j}\Lambda_{mj}\Lambda_{mj'}E\left(\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{in}}\right)_{j'i'}\right) \\
& -2\sum_{k=1}^{N}\sum_{i=1}^{N}\omega_{kn}\omega_{in}\sum_{j=1}^{N}\Lambda_{mj}E\left(\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}\right) \\
= {} & \sum_{k=1}^{N}\omega_{kn}^{2}E\left(\delta\mathbf{W}_{\text{out}}\right)_{mk}^{2}+\sum_{i=1}^{N}\omega_{in}^{2}\sum_{j=1}^{N}\Lambda_{mj}^{2}E\left(\delta\mathbf{W}_{\text{in}}\right)_{ji}^{2} \\
= {} & \sum_{k=1}^{N}\omega_{kn}^{2}\frac{\sigma^{2}}{w}\left(\mathbf{W}_{\text{out}}\right)_{mk}+\sum_{i=1}^{N}\omega_{in}^{2}\sum_{j=1}^{N}\Lambda_{mj}^{2}\frac{\sigma^{2}}{w}\left(\mathbf{W}_{\text{in}}\right)_{ji} \\
= {} & \frac{\sigma^{2}}{w}\sum_{k=1}^{N}\left(\left(\mathbf{W}_{\text{out}}\right)_{mk}+\sum_{j=1}^{N}\Lambda_{mj}^{2}\left(\mathbf{W}_{\text{in}}\right)_{jk}\right)\omega_{kn}^{2}.
\end{aligned} \tag{4.23}$$

Finally, I express Equation 4.23 in matrix notation as

$$E\left(\delta\Lambda\circ\delta\Lambda\right)=\frac{\sigma^{2}}{w}\left(\mathbf{W}_{\text{out}}+\left(\Lambda\circ\Lambda\right)\mathbf{W}_{\text{in}}\right)\left(\mathbf{W}_{\text{in}}^{-1}\circ\mathbf{W}_{\text{in}}^{-1}\right). \tag{4.24}$$

If we have a number of alternative MITE-network implementations of a given set of product-of-power-law relationships, we can use Equation 4.24 to evaluate how sensitive each topology is to component mismatch. Then, we can choose the one that is least sensitive to mismatch in the values of the weighting coefficients.

4.8. References

1. E. Seevinck, *Analysis and Synthesis of Translinear Integrated Circuits*, Amsterdam, The Netherlands: Elsevier, 1988.

2. H. M. Salkin, *Integer Programming*, Reading, MA: Addison-Wesley, 1975.

3. T. C. Hu, *Integer Programming and Network Flows*, Reading, MA: Addison-Wesley, 1969.

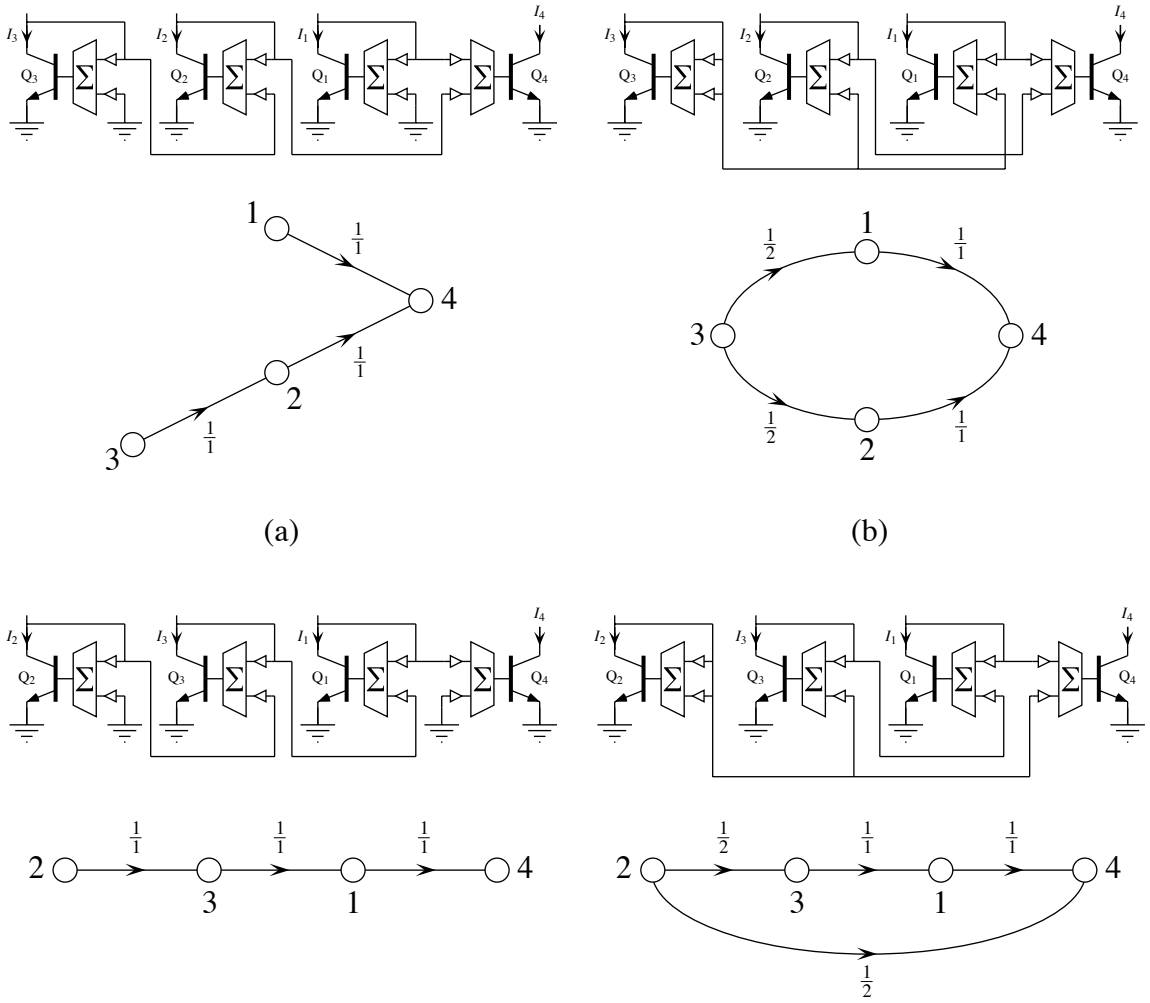


Figure 4.1. Four distinct MITE networks, each of which implements a product-reciprocal relationship, and their associated reduced signal-flow graphs. Each MITE network comprises four two-input MITEs. (a) Two-layer MITE network in which MITEs Q_1 and Q_2 form the numerator layer of inputs, and MITE Q_3 forms the denominator layer of inputs and its associated reduced signal-flow graph. (b) Completed two-layer MITE network and its associated reduced signal-flow graph. (c) Cascade MITE network in which numerator inputs and denominator inputs are alternated with each other, and its associated reduced signal-flow graph. (d) Completed cascade MITE network and its associated reduced signal-flow graph.

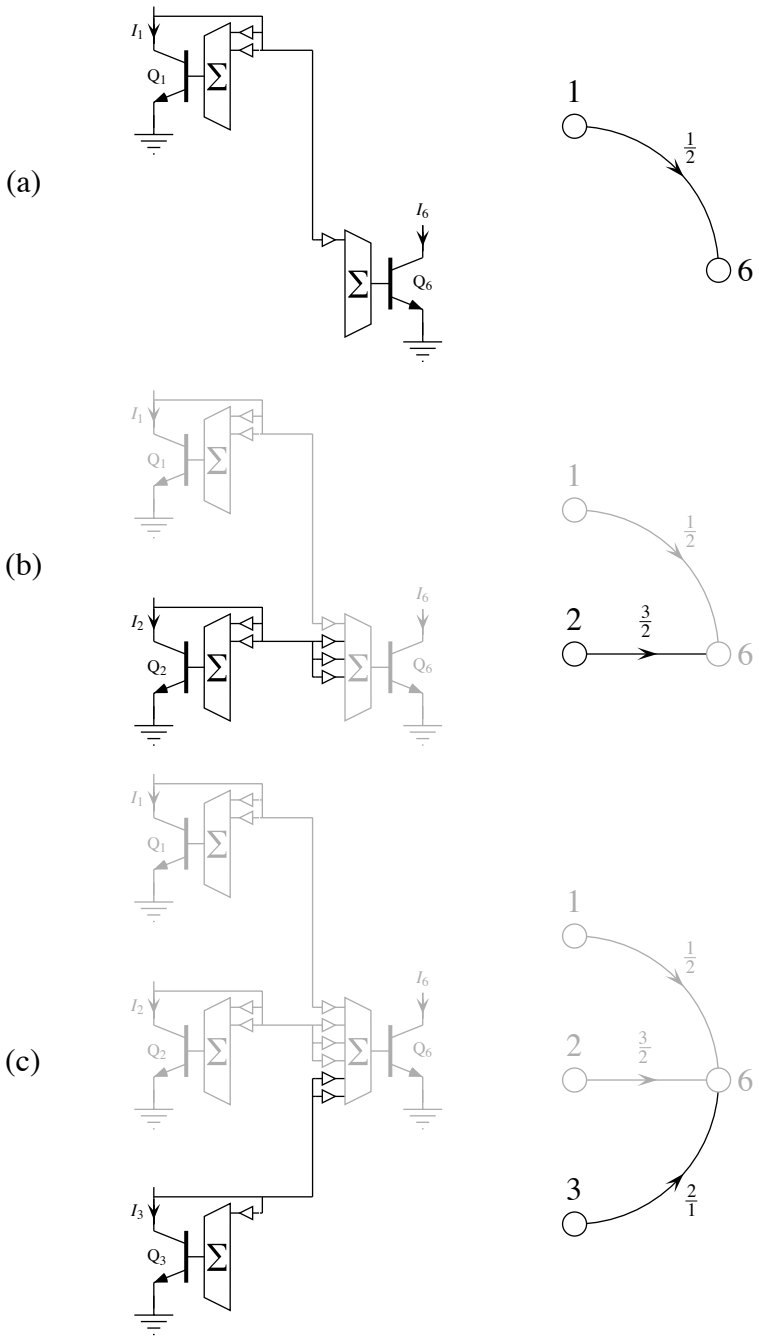


Figure 4.2. Construction of the numerator layer of a two-layer MITE network realizing Equation 4.10. (a) I allocate MITE Q_1 for input current I_1. I_1 is raised to the $\frac{1}{2}$ power, so I diode connect MITE Q_1 through two inputs, and I connect its input node to MITE Q_6 through one input. (b) I allocate MITE Q_2 for I_2. I_2 is raised to the $\frac{3}{2}$ power, so I diode connect MITE Q_2 through two inputs, and I connect its input node to MITE Q_6 through three inputs. (c) I create MITE Q_3 for I_3. I_3 is squared, so I diode connect MITE Q_1 through one input, and I connect its input node to MITE Q_6 through two inputs.

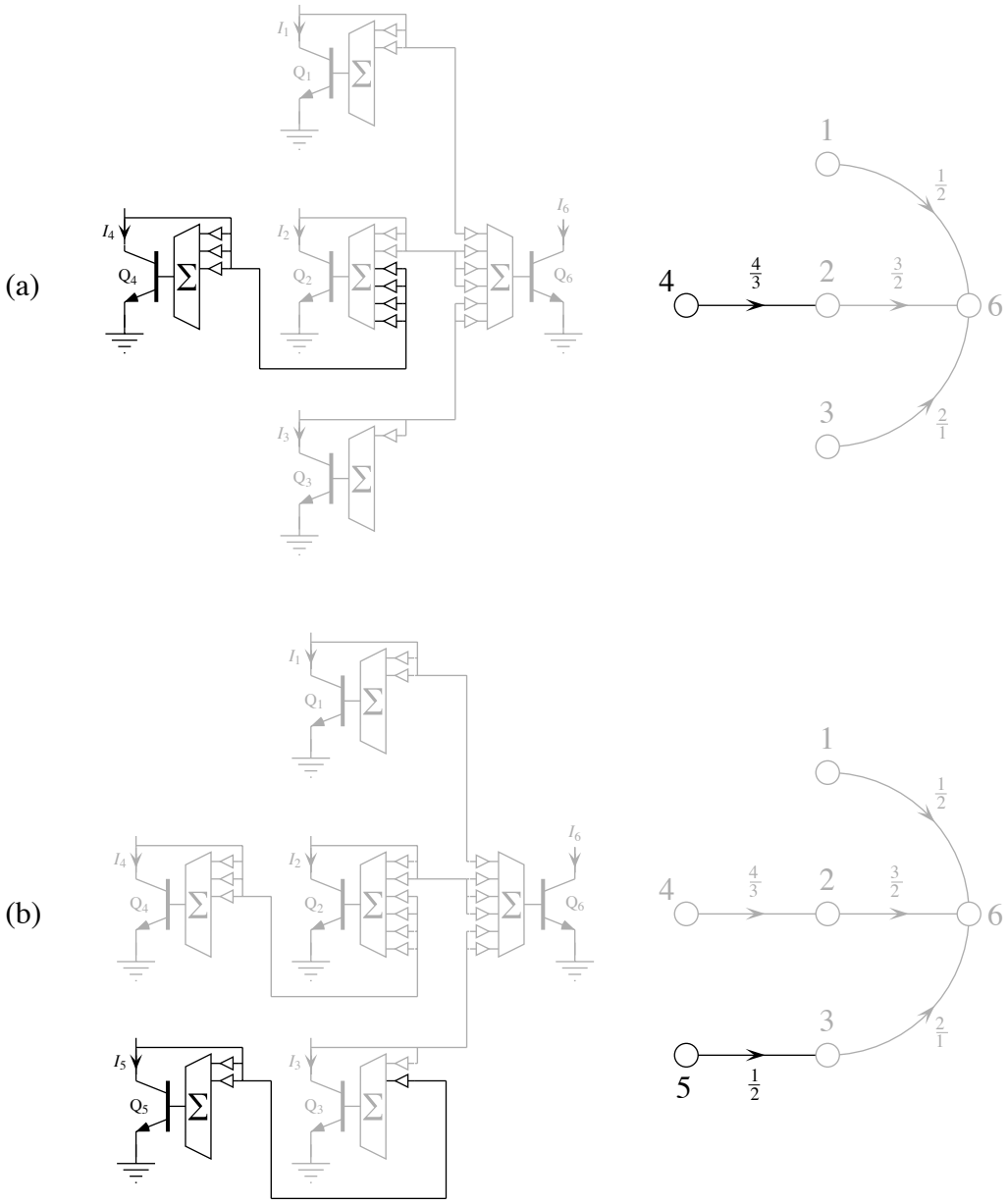


Figure 4.3. Continued construction of the denominator layer of the two-layer MITE network realizing Equation 4.10 that I began in Figure 4.2. (a) I create MITE Q_4 for input current I_4. I choose to connect MITE Q_4 to MITE Q_6 through MITE Q_2. The greatest common divisor between $2 \times 2 = 4$ and $3 \times 1 = 3$ is 1, so I diode connect MITE Q_4 through $\frac{3 \times 1}{1} = 3$ inputs, and I connect its input node to MITE Q_2 through $\frac{2 \times 2}{1} = 4$ inputs. (b) I create MITE Q_5 for I_5. I choose to connect MITE Q_5 to MITE Q_6 through MITE Q_3. The greatest common divisor between $2 \times 1 = 2$ and $1 \times 1 = 1$ is 1, so I diode connect MITE Q_5 through $\frac{2 \times 1}{1} = 2$ inputs, and I connect its input node to MITE Q_3 through $\frac{1 \times 1}{1} = 1$ input.

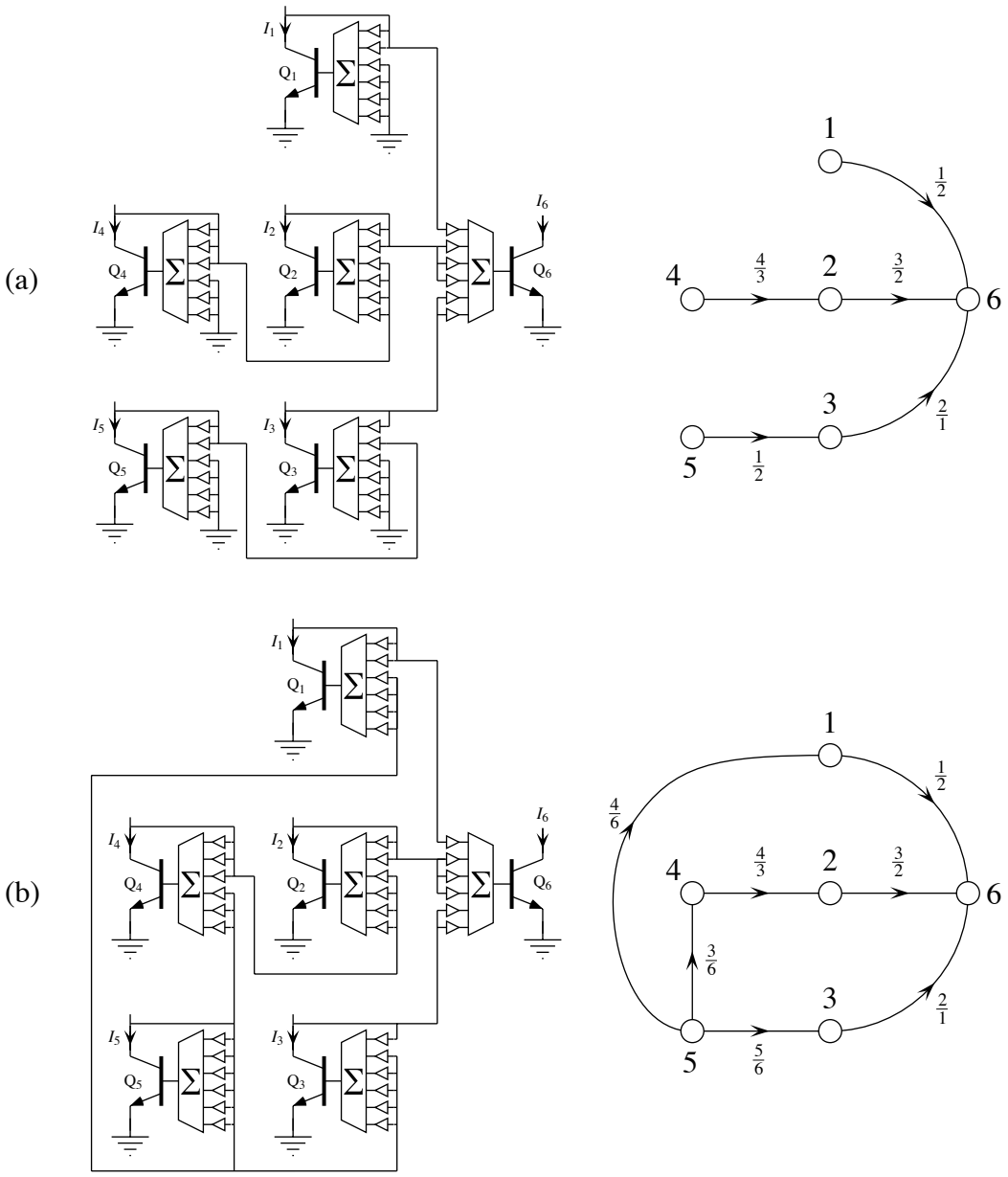


Figure 4.4. Completion of the two-layer MITE network that embodies Equation 4.10, began in Figure 4.2. (a) The largest number of MITE inputs in the MITE network of Figure 4.3b is 6. Consequently, I add four grounded inputs to MITEs Q_1, Q_3, and Q_5 and I add three grounded inputs to MITE Q_4. This MITE network embodies Equation 4.10, but some of the transconductance of MITEs Q_1, Q_3, Q_4, and Q_5 is wasted. (b) I complete this MITE network by connecting all the unused inputs added in part a to the input node of MITE Q_5. The resulting network also embodies Equation 4.10, but there is no wasted MITE transconductance.

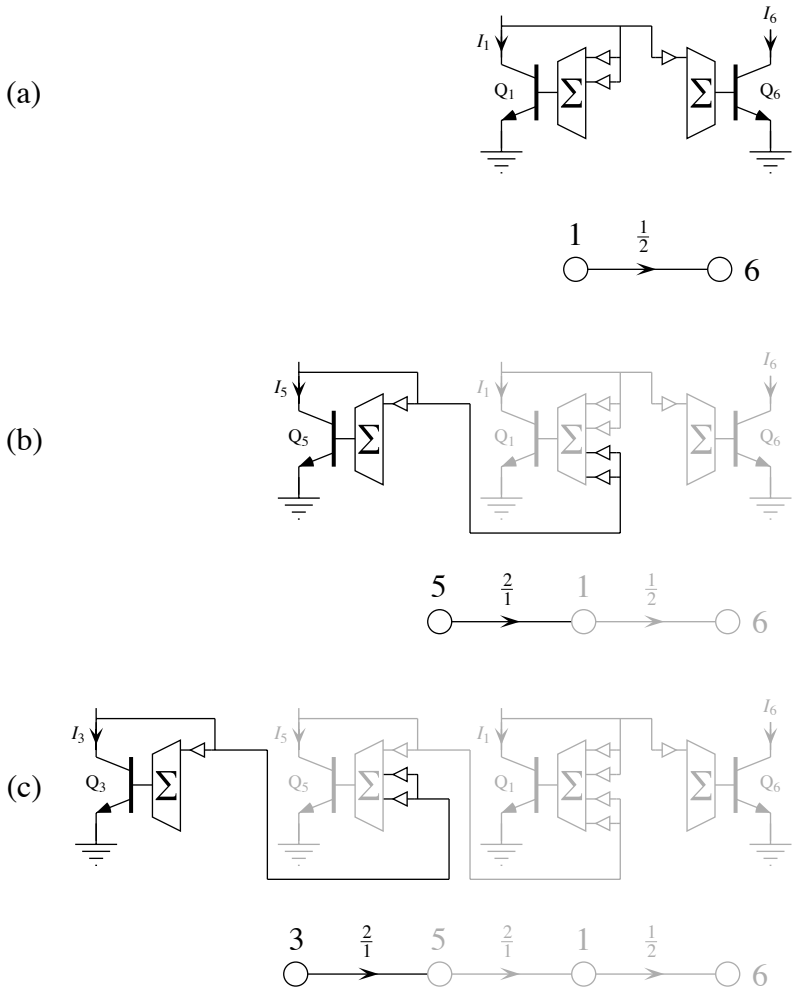


Figure 4.5. Construction of a cascade MITE network embodying Equation 4.10. (a) I begin by creating MITE Q_1 for input current I_1. I_1 is raised to the $\frac{1}{2}$ power, so I diode connect MITE Q_1 through two inputs, and I connect its input node to MITE Q_6 through one input. (b) Next, I choose I_5 from the denominator of Equation 4.10, and I create MITE Q_5 for it. The greatest common divisor between $1\times1=1$ and $2\times1=2$ is 1, so I diode connect MITE Q_5 through $\frac{1\times1}{1}=1$ input, and I connect its input node to MITE Q_1 through $\frac{2\times1}{1}=2$ inputs. (c) Next, I choose I_3 from the numerator of Equation 4.10, and I create MITE Q_3 corresponding to it. The greatest common divisor between $2\times1=2$ and $1\times1=1$ is 1, so I diode connect MITE Q_3 through $\frac{1\times1}{1}=1$ input, and I connect its input node to MITE Q_5 through $\frac{2\times1}{1}=2$ inputs.

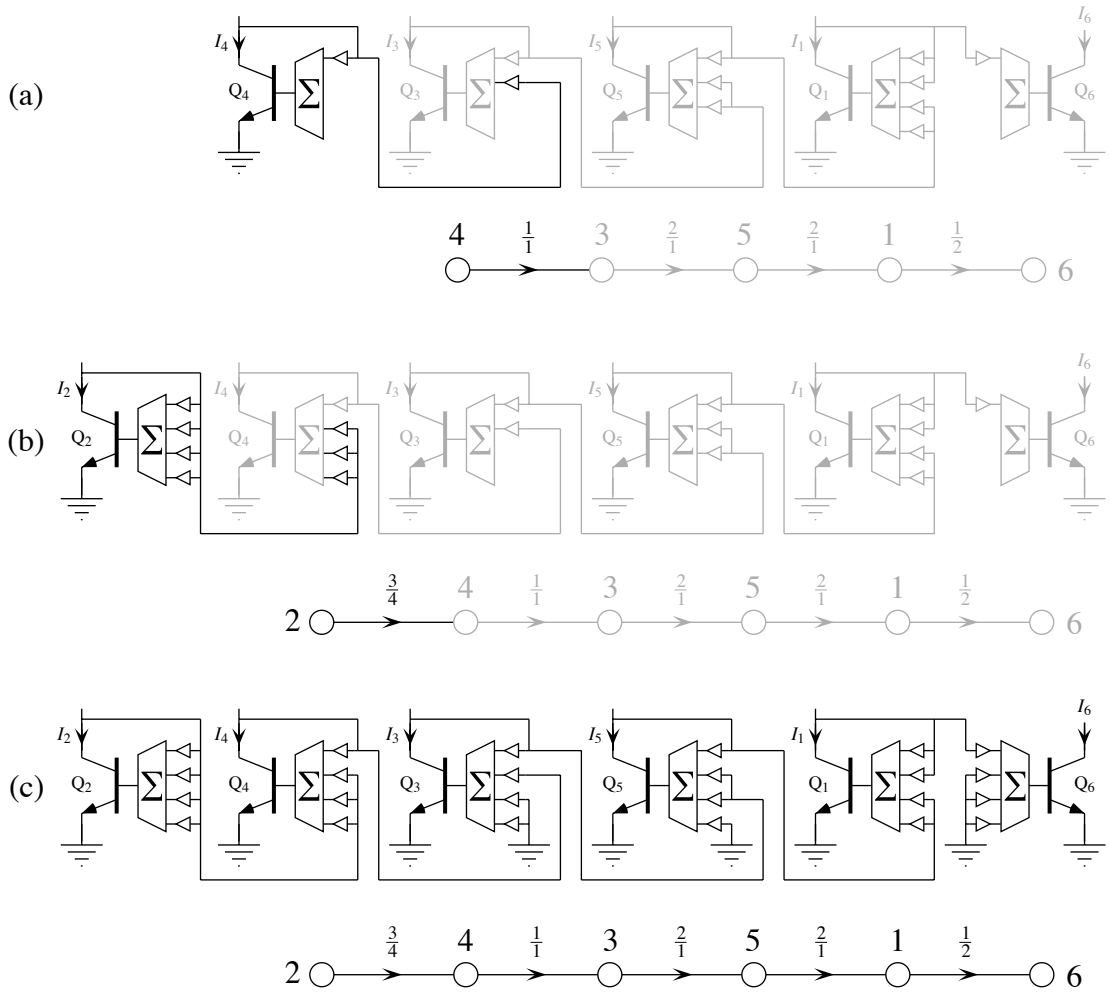


Figure 4.6. Continued construction of the cascade MITE network embodying Equation 4.10 that I started in Figure 4.5. (a) Next, I choose I_4 from the denominator of Equation 4.10, and I create MITE Q_4 for it. The greatest common divisor between $2 \times 1 = 2$ and $2 \times 1 = 2$ is 2, so I diode connect MITE Q_4 through $\frac{2 \times 1}{2} = 1$ input, and I connect its input node to MITE Q_3 through $\frac{2 \times 1}{2} = 1$ input. (b) Finally, I choose I_2 from the numerator of Equation 4.10; I create MITE Q_2 for it. The greatest common divisor between $3 \times 1 = 3$ and $2 \times 2 = 4$ is 1, so I diode connect MITE Q_2 through $\frac{2 \times 2}{1} = 4$ inputs, and I connect its input node to MITE Q_1 through $\frac{3 \times 1}{1} = 3$ inputs. (c) The largest number of MITE inputs in the MITE network of part b is 4. Consequently, I add one grounded input to MITE Q_5, I add two grounded inputs to MITE Q_3, and I add three grounded inputs to MITE Q_6. The resulting MITE network embodies Equation 4.10, but some of the transconductance of MITEs Q_3, Q_5, and Q_6 is wasted.

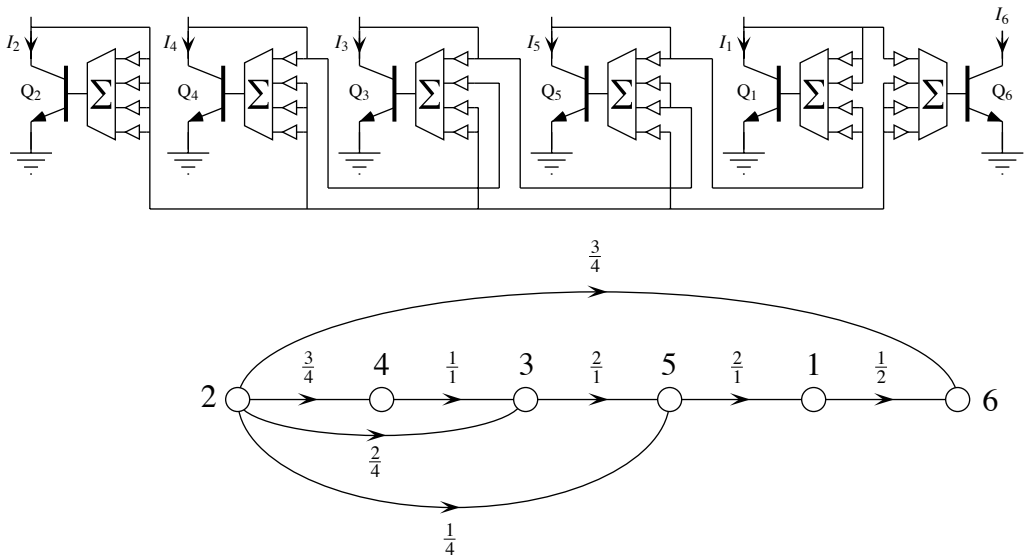


Figure 4.7. Completion of the cascade MITE network embodying Equation 4.10, started in Figure 4.5. I complete the MITE network of Figure 4.6c by connecting all the unused inputs added to the network in Figure 4.6c to the input node of MITE Q_2. This network also embodies Equation 4.10, but there is no wasted MITE transconductance.

CHAPTER 5
THE SUBTHRESHOLD FLOATING-GATE MOS TRANSISTOR

In this chapter, I discuss the subthreshold operation of the *K*-input **floating-gate MOS** (**FGMOS**) transistor. In Section 5.1, I give an overview of the the FGMOS transistor. In Section 5.2, I develop a simple first-order model for the subthreshold *K*-input FGMOS transistor based on a capacitive voltage divider. In Section 5.3, to verify the simple subthreshold FGMOS transistor model, I present experimental data from a four-input FGMOS transistor that was fabricated in a standard 2-μm double-poly CMOS process through MOSIS. Here, I also present experimental data showing the most significant second-order effect (from the standpoint of the using a subthreshold FGMOS transistor to implement a MITE) in these devices; that is, I present data showing that transconductance as a function of current level in these devices deviates slightly from linearity. In Section 5.4, I present data on the matching of small poly1-poly2 capacitors in the 2-μm double-poly CMOS process that is commonly available through MOSIS.

5.1. Floating-Gate MOS Transistors: They're Not Just for Storing Information Anymore

In 1967, Kahng and Sze [1] reported the first floating-gate structure as a mechanism for nonvolatile information storage. Since then, FGMOS transistors have been used widely to store digital information for long periods in structures such as EPROMs, EEPROMs, and flash memories. FGMOS transistors have also been used for long-term nonvolatile information storage devices for analog applications. The quantity of charge stored on the insulated gate of this device has been used both as a free parameter that is adapted to increase the precision of analog circuits [2–4], and as a weight value that is updated according to various learning algorithms in neural-network hardware implementations [5–8].

In 1983, Wada and his associates [9, 10] described a new EEPROM addressing structure called a dual–control-gate EEPROM in which two **control gates** couple into the floating gate of an EEPROM transistor with equal strengths. In this EEPROM cell, one

control gate receives a column-select signal and the other receives a row-select signal. Only the cell receiving two positive inputs develops a floating-gate voltage that is high enough for the writing process (i.e., for Fowler–Nordheim tunneling) to occur. In 1992, based on a generalization of the dual-control-gate EEPROM addressing scheme, Shibata and Ohmi [11] proposed a marvelous new way of thinking about FGMOS transistors, in which multiple control gates couple into a floating gate capacitvely, and thereby establish the floating-gate voltage as a weighted summation of the input voltages via a capacitive voltage divider. The floating-gate voltage, in turn, modulates the current flowing in a channel in the silicon below the floating gate. Shibata and Ohmi [11] dubbed these compound devices **neuron MOS** (**neuMOS** or **νMOS**) **transistors**, based on a loose analogy between the function performed by these devices and by cells in the nervous system. Yang, Andreou, and Boahen [12–14] refer to such devices as **multiple-input floating-gate MOS** (**FGMOS**) **transistors**. Ramírez-Angulo [15] calls them **multiple-input floating-gate** (**MIFG**) **transistors**. I follow the naming convention of Yang, Andreou, and Boahen, by referring to such devices as FGMOS transistors.

Figure 5.1 depicts a K-input, n-channel FGMOS (nFGMOS) transistor. Figure 5.1a shows a typical layout view of a such a device with K nominally identical control gates in a double-poly, n-well CMOS process. In such a process, we typically make the floating gate from the first level of polysilicon (**poly1**), and we make the control gates from the second level of polysilicon (**poly2**). As designers, we can proportionally change the coupling strength of a control gate by changing the area of overlap between poly1 and poly2 for that control gate. Note that we do not need to work in a double-poly process to make these devices; we require only a means of making a linear floating capacitor. For example, certain submicron CMOS processes presently offer a layer—usually called capwell—that results in a buried n^+ region under a relatively thin oxide; we should be able to use such a layer to form control gates. Figure 5.1b shows the circuit symbol that I use to represent such a device schematically. I indicate the capacitance of the kth control gate by a C_k nearby the control-gate symbol. To prevent unnecessary clutter in circuit schematics, if I have K control gates, each with nominal capacitance C, then I do not show the C with each control gate in the schematic. I indicate the value of the net charge stored on the floating gate by a Q near the floating-gate symbol.

The way of thinking about FGMOS transistors introduced by Shibata and Ohmi has led to a number of interesting analog and digital information-processing circuits, including simple D/A converters [11], a multiple-input floating-gate differential amplifier [13, 14], a

four-quadrant voltage multiplier [15], reconfigurable digital logic based on threshold logic units [16, 17], a neural network [18], a voltage winner-take-all circuit [19], a current soft-maximum/winner-take-all circuit [20], and FGMOS translinear circuits (i.e., MITE networks comprising subthreshold FGMOS transistors) [21]. Because of Shibata and Ohmi's conceptual advance, FGMOS transistors are not just being used for storing information anymore—they are also being used as information processing devices.

5.2. Subthreshold Floating-Gate MOS Transistor Model

In their original paper describing the νMOS transistor concept [11], Shibata and Ohmi present a simple above-threshold model for the multiple-input FGMOS transistor. In various contexts [12–14], Yang, Andreou, and Boahen present both a subthreshold model and an above-threshold model for the multiple-input FGMOS transistor. In this section, I derive a simple first-order model of subthreshold operation of a K-input nFGMOS transistor; this model is similar to that presented by Andreou and Boahen [12]. I include it primarily to show that we can use the subthreshold FGMOS transistor to implement a MITE, and to identify the limitations of this single-device MITE implementation. I obtain a model for a K-input, p-channel FGMOS (pFGMOS) transistor by following an identical procedure. The resulting model equation is similar to that of the nFGMOS transistor, except that all the terminal voltages have the opposite sign and are measured relative to the power supply voltage, V_{DD}.

I begin by constructing a lumped circuit model of the K-input nFGMOS transistor that is shown in Figure 5.1a. I connect together K capacitors, C_1 through C_K, in parallel to model the K control-gate capacitances. I then connect the gate of an nMOS transistor to the node to which all the control-gate capacitors connect; this node is the floating gate. To complete the equivalent circuit, I add three parasitic capacitances. I include a capacitance from the floating gate to the bulk (i.e., to the substrate), C_b. This parasitic capacitance occurs because the floating gate is suspended over the silicon substrate (i.e., a conductor) and is separated from it by a relatively thin layer of dielectric material (i.e., SiO_2); hence, the floating gate and the silicon substrate form a parallel-plate capacitor whose capacitance is proportional to the total area of the floating gate. I also include a capacitance from the floating gate both to the source, $C_{fg\text{-}s}$, and to the drain, $C_{fg\text{-}d}$, of the nMOS transistor. In subthreshold, there is little mobile charge in the channel (i.e., there is no inversion layer), so this parasitic capacitance arises because the source and drain active regions diffuse slightly underneath the floating gate during fabrication, thereby creating a small region of

overlap between the floating gate and both the source and the drain. These capacitances are linearly proportional to the width of the *n*MOS transistor. The final circuit model is shown in Figure 5.2a.

The channel-surface potential determines the amount of current that will flow in the channel of the *n*MOS transistor. To determine the surface potential, I replace the *n*MOS transistor in the circuit of Figure 5.2a with an equivalent capacitor circuit, as shown in Figure 5.2b. The floating gate couples into the channel surface through the oxide capacitance, C_{ox}. The channel surface, in turn, couples into the substrate through a (nonlinear) depletion-layer capacitance, C_{dep}. Without loss of generality, I measure all voltages relative to the substrate, and take substrate to be ground.

I assume that a net charge of Q is stored on the floating gate, and I apply conservation of charge to the floating gate to obtain

$$Q = C_{\text{ox}}\left(V_{\text{fg}} - \psi_{\text{s}}\right) + C_{\text{b}}V_{\text{fg}} - \sum_{k=1}^{K} C_k\left(V_k - V_{\text{fg}}\right) - C_{\text{fg-s}}\left(V_{\text{s}} - V_{\text{fg}}\right) - C_{\text{fg-d}}\left(V_{\text{d}} - V_{\text{fg}}\right), \quad (5.1)$$

where V_{fg} is the floating-gate voltage, ψ_{s} is the channel-surface potential, V_{s} is the source voltage, V_{d} is the drain voltage, and V_k is the kth control-gate voltage. By rearranging Equation 5.1 and solving for the floating-gate voltage, I obtain

$$V_{\text{fg}} = \frac{C_{\text{ox}}}{C_{\text{T}}^{*}}\psi_{\text{s}} + \frac{Q}{C_{\text{T}}^{*}} + \sum_{k=1}^{K}\frac{C_k}{C_{\text{T}}^{*}}V_k + \frac{C_{\text{fg-s}}}{C_{\text{T}}^{*}}V_{\text{s}} + \frac{C_{\text{fg-d}}}{C_{\text{T}}^{*}}V_{\text{d}}, \quad (5.2)$$

where

$$C_{\text{T}}^{*} = C_{\text{ox}} + C_{\text{b}} + C_{\text{fg-s}} + C_{\text{fg-d}} + \sum_{k=1}^{K} C_k .$$

I assume that the surface potential, ψ_{s}, of the *n*MOS transistor is a slowly varying function of its gate voltage [22], so I can expand ψ_{s} in a Taylor series and keep only the constant and first-order terms. So, I write

$$\psi_{\text{s}} \approx \psi_0 + \kappa V_{\text{fg}}, \quad (5.3)$$

where

$$\kappa = \frac{C_{\text{ox}}}{C_{\text{ox}} + C_{\text{dep}}}.$$

By substituting Equation 5.2 into Equation 5.3 and rearranging, I obtain

$$\psi_{\text{s}} = \frac{C_{\text{T}}^{*}}{C_{\text{T}}}\psi_0 + \kappa\frac{Q}{C_{\text{T}}} + \kappa\sum_{k=1}^{K}\frac{C_k}{C_{\text{T}}}V_k + \kappa\frac{C_{\text{fg-s}}}{C_{\text{T}}}V_{\text{s}} + \kappa\frac{C_{\text{fg-d}}}{C_{\text{T}}}V_{\text{d}}, \quad (5.4)$$

where

$$C_{\text{T}} = \left(C_{\text{ox}} \parallel C_{\text{dep}}\right) + C_{\text{b}} + C_{\text{fg-s}} + C_{\text{fg-d}} + \sum_{k=1}^{K} C_k$$

and

$$x \parallel y \equiv \frac{xy}{x+y}.$$

In the subthreshold regime, there is an exceedingly small amount of mobile charge in the channel, compared to the amount of fixed charge in the depletion layer beneath the channel; hence, the channel-surface potential is spatially uniform. The electric field on a mobile charge in the channel is proportional to the gradient of the surface potential. A uniform channel-surface potential implies an exceedingly small electric field; consequently, current flow in subthreshold is primarily by diffusion, rather than by drift. So, I write

$$I_{\mathrm{d}} = -\mathrm{q}DW\frac{\partial N}{\partial x},$$

where W is the width of the channel, q is the charge of an electron, D is the diffusion coefficient of channel electrons, and $N(x)$ is the electron concentration along the channel. Because no carriers are lost in the channel, the carrier concentration is a linear function of position along the channel. So, I have that

$$I_{\mathrm{d}} = -\mathrm{q}DW\frac{N_{\mathrm{d}} - N_{\mathrm{s}}}{L}, \tag{5.5}$$

where L is the length of the channel, and N_{s} and N_{d} are the electron concentrations at the source end and at the drain end of the channel, respectively. The electron concentrations at the source and drain ends of the channel, in turn, are given by the Boltzmann distribution as

$$N_{\mathrm{s}} = N_0\exp\left[\frac{\psi_{\mathrm{s}} - V_{\mathrm{s}}}{\mathrm{U_T}}\right] \text{ and } N_{\mathrm{d}} = N_0\exp\left[\frac{\psi_{\mathrm{s}} - V_{\mathrm{d}}}{\mathrm{U_T}}\right], \tag{5.6}$$

where N_0 is the effective density of states at the channel surface and $\mathrm{U_T}$ is the thermal voltage, $\frac{\mathrm{kT}}{\mathrm{q}}$. By substituting Equation 5.6 into Equation 5.5, I obtain

$$\begin{aligned} I_{\mathrm{d}} &= \mathrm{qD}\frac{W}{L}\left(N_0\exp\left[\frac{\psi_{\mathrm{s}} - V_{\mathrm{s}}}{\mathrm{U_T}}\right] - N_0\exp\left[\frac{\psi_{\mathrm{s}} - V_{\mathrm{d}}}{\mathrm{U_T}}\right]\right) \\ &= \frac{W}{L}\mathrm{qD}N_0\exp\left[\frac{\psi_{\mathrm{s}}}{\mathrm{U_T}}\right]\left(\exp\left[-\frac{V_{\mathrm{s}}}{\mathrm{U_T}}\right] - \exp\left[-\frac{V_{\mathrm{d}}}{\mathrm{U_T}}\right]\right). \end{aligned} \tag{5.7}$$

By substituting Equation 5.4 into Equation 5.7 and rearranging, I obtain

$$\begin{aligned} I_{\mathrm{d}} = \frac{W}{L}\mathrm{I}_0\exp\left[\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_{\mathrm{T}}}\frac{V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-s}}}{C_{\mathrm{T}}}\frac{V_{\mathrm{s}}}{\mathrm{U_T}} + \frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{V_{\mathrm{d}}}{\mathrm{U_T}}\right] \\ \times\left(\exp\left[-\frac{V_{\mathrm{s}}}{\mathrm{U_T}}\right] - \exp\left[-\frac{V_{\mathrm{d}}}{\mathrm{U_T}}\right]\right), \end{aligned} \tag{5.8}$$

where

$$\mathrm{I}_0 = \mathrm{qD}N_0\exp\left[\frac{C_{\mathrm{T}}^*}{C_{\mathrm{T}}}\frac{\psi_0}{\mathrm{U_T}}\right],$$

and

$$\mathrm{Q_T} = \frac{C_\mathrm{T}\mathrm{U_T}}{\kappa}.$$

If I ground the source of the *n*FGMOS transistor, then Equation 5.8 becomes

$$I_\mathrm{d} = \frac{W}{L}\mathrm{I}_0 \exp\left[\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}}\frac{V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_\mathrm{T}}\frac{V_\mathrm{d}}{\mathrm{U_T}}\right]\left(1-\exp\left[-\frac{V_\mathrm{d}}{\mathrm{U_T}}\right]\right). \tag{5.9}$$

If I both ground the source and operate the *n*FGMOS transistor with its drain voltage, V_d, more than a few $\mathrm{U_T}$ above ground, then the *n*FGMOS transistor is in saturation and Equation 5.9 reduces to

$$I_\mathrm{d} = \frac{W}{L}\mathrm{I}_0 \exp\left[\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}}\frac{V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_\mathrm{T}}\frac{V_\mathrm{d}}{\mathrm{U_T}}\right]. \tag{5.10}$$

Finally, if each control gate capacitance has a nominal value of *C*, Equation 5.10 becomes

$$I_\mathrm{d} = \frac{W}{L}\mathrm{I}_0 \exp\left[\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\frac{\kappa C}{C_\mathrm{T}}\sum_{k=1}^{K}\frac{V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_\mathrm{T}}\frac{V_\mathrm{d}}{\mathrm{U_T}}\right]. \tag{5.11}$$

Figure 5.3 shows two views of a *K*-input *p*FGMOS transistor in a double-poly, *n*-well CMOS process. If I were to repeat the derivation just outlined for the *K*-input, *n*FGMOS transistor for the *K*-input *p*FGMOS transistor, I would obtain the following model equation corresponding to Equation 5.8:

$$I_\mathrm{d} = \frac{W}{L}\mathrm{I}_0 \exp\left[-\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}}\frac{\mathrm{V_{DD}}-V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-s}}}{C_\mathrm{T}}\frac{\mathrm{V_{DD}}-V_\mathrm{s}}{\mathrm{U_T}}+\frac{\kappa C_{\mathrm{fg-d}}}{C_\mathrm{T}}\frac{\mathrm{V_{DD}}-V_\mathrm{d}}{\mathrm{U_T}}\right]$$
$$\times\left(\exp\left[-\frac{\mathrm{V_{DD}}-V_\mathrm{s}}{\mathrm{U_T}}\right]-\exp\left[-\frac{\mathrm{V_{DD}}-V_\mathrm{d}}{\mathrm{U_T}}\right]\right). \tag{5.12}$$

If I both connect the source of the *p*FGMOS transistor to $\mathrm{V_{DD}}$ and operate the transistor with its drain voltage, V_d, more than a few $\mathrm{U_T}$ below $\mathrm{V_{DD}}$, then the *p*FGMOS transistor is in saturation and Equation 5.12 reduces to

$$I_\mathrm{d} = \frac{W}{L}\mathrm{I}_0 \exp\left[-\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\frac{\kappa C}{C_\mathrm{T}}\sum_{k=1}^{K}\frac{\mathrm{V_{DD}}-V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_\mathrm{T}}\frac{\mathrm{V_{DD}}-V_\mathrm{d}}{\mathrm{U_T}}\right]. \tag{5.13}$$

In Section 5.3, I compare Equation 5.13 with experimental measurements from a four-input *p*FGMOS transistor with nominally identical control gates that was fabricated in a 2-μm double-poly CMOS process through MOSIS.

5.3. Measurements from a Subthreshold Floating-Gate MOS Transistor

In this section, I present experimental measurements taken from a four-input *p*FGMOS transistor with four nominally identical control gates that was fabricated in a standard 2-μm

double-poly CMOS process available through MOSIS. I test the validity of the subthreshold FGMOS transistor model that I developed in Section 5.2 by fitting Equation 5.13 to the data.

Figure 5.4 shows measurements of drain current from a four-input *p*FGMOS transistor as a function of various combinations of the four control-gate voltages. To obtain the data shown in Figure 5.4a, for each n between 1 and 4, I measured the drain current while I swept n of the four control gates from 0 to 3 volts below $\mathrm{V_{DD}}$. For each sweep, I connected the remaining $4-n$ control gates to $\mathrm{V_{DD}}$ and I fixed the drain voltage at 5 volts below $\mathrm{V_{DD}}$. Under these conditions, Equation 5.13 becomes

$$\begin{aligned} I_{\mathrm{d}} &= \frac{W}{L}\mathrm{I_0}\exp\left[-\frac{Q}{\mathrm{Q_T}}\right]\exp\left[n\frac{\kappa C}{C_{\mathrm{T}}}\frac{V_1}{\mathrm{U_T}}+(4-n)\frac{\kappa C}{C_{\mathrm{T}}}\frac{\mathrm{V_{DD}}-\mathrm{V_{DD}}}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{5\mathrm{V}}{\mathrm{U_T}}\right] \\ &= \frac{W}{L}\mathrm{I_0}\exp\left[-\frac{Q}{\mathrm{Q_T}}\right]\exp\left[n\frac{\kappa C}{C_{\mathrm{T}}}\frac{V_1}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{5\mathrm{V}}{\mathrm{U_T}}\right] \\ &= \mathrm{I_0'}\exp\left[n\frac{\kappa C}{C_{\mathrm{T}}}\frac{V_1}{\mathrm{U_T}}\right] \\ &= \mathrm{I_0'}\exp\left[n\frac{V_1}{\mathrm{V_C}}\right], \end{aligned} \tag{5.14}$$

where

$$\mathrm{I_0'}=\frac{W}{L}\mathrm{I_0}\exp\left[-\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{5\mathrm{V}}{\mathrm{U_T}}\right] \text{ and } \mathrm{V_C}=\frac{C_{\mathrm{T}}\mathrm{U_T}}{\kappa C}. \tag{5.15}$$

The solid lines in Figure 5.4a show theoretical fits of Equation 5.14 to the data with $\mathrm{V_C}$=173 mV and $\mathrm{I_0'}$=171 fA. In all four cases, the data and theoretical fits agree well over a current range from about 20 pA to 20 nA.

To obtain the data shown in Figure 5.4b, I measured the *p*FGMOS transistor's drain current while I swept one of the four control gates from 0 to 3 volts below $\mathrm{V_{DD}}$ for seven different values of the voltage of the remaining three control gates. Again, for each of these sweeps, I fixed the drain voltage at 5 volts below $\mathrm{V_{DD}}$. Under these conditions, Equation 5.13 becomes

$$\begin{aligned} I_{\mathrm{d}} &= \frac{W}{L}\mathrm{I_0}\exp\left[-\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\frac{\kappa C}{C_{\mathrm{T}}}\frac{V_1+3V_2}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{5\mathrm{V}}{\mathrm{U_T}}\right] \\ &= \mathrm{I_0'}\exp\left[\frac{\kappa C}{C_{\mathrm{T}}}\frac{V_1+3V_2}{\mathrm{U_T}}\right] \\ &= \mathrm{I_0'}\exp\left[\frac{V_1+3V_2}{\mathrm{V_C}}\right], \end{aligned} \tag{5.16}$$

where $\mathrm{I_0'}$ and $\mathrm{V_C}$ are defined in Equation 5.15. The solid lines in Figure 5.4b show

theoretical fits of Equation 5.16 to the data with V_C=173 mV and I'_0=171 fA. Note that these fit parameter values are the same as those that I used to fit the data shown in Figure 5.4a. For each of the seven curves, the data and theoretical fit agree well over a current range from about 20 pA to 20 nA.

Figure 5.5 shows measurements of the *p*FGMOS transistor's drain current as a function of drain voltage. To obtain the data shown in Figure 5.5a, I measured the drain current while I swept the drain voltage from 0 to 5 volts below V_{DD} for five different values of the voltage on the four control gates. Under these conditions, if we denote by V_g the voltage on each of the four control gates, then Equation 5.13 becomes

$$\begin{aligned} I_d &= \frac{W}{L} I_0 \exp\left[-\frac{Q}{Q_T}\right] \exp\left[\frac{\kappa C}{C_T}\frac{4V_g}{U_T}\right] \exp\left[\frac{\kappa C_{fg-d}}{C_T}\frac{V_d}{U_T}\right] \\ &= I'_0 \exp\left[-\frac{\kappa C_{fg-d}}{C_T}\frac{5V}{U_T}\right] \exp\left[\frac{4V_g}{V_C}\right] \exp\left[\frac{\kappa C_{fg-d}}{C_T}\frac{V_d}{U_T}\right] \\ &= I'_0 \exp\left[\frac{4V_g}{V_C}\right] \exp\left[\frac{V_d - 5V}{V_A}\right], \end{aligned} \tag{5.17}$$

where I'_0 and V_C are defined in Equation 5.15 and

$$V_A = \frac{C_T U_T}{\kappa C_{fg-d}}. \tag{5.18}$$

The solid lines in Figure 5.5a show theoretical fits of Equation 5.17 to the data with V_C=173 mV, I'_0=171 fA, and V_A=2.88 V. Note again that the values of V_C and I'_0 are the same as those that I used to fit all the data shown in Figure 5.4. Again, the data and theoretical curves agree well over a current range spanning 20 pA to 20 nA.

Figure 5.5b shows a plot of measured drain current as a function of drain voltage for the four-input *p*FGMOS transistor on a linear scale. I adjusted the voltage on the four control gates such that the transistor sourced about 3 nA when the drain was 5 volts below V_{DD}. At about 100 mV below V_{DD}, the *p*FGMOS transistor "saturates;" however, the drain current then increases by a factor of six as the drain voltage changes from 100 mV to 5 volts below V_{DD}. The best-fit exponential curve to these data has a slope of 2.93 V/*e*-fold. The exponential dependence of the drain current on the drain voltage implies that FGMOS transistors make poor current sources. From Equation 5.18, we have that the slope of this exponential dependence is proportional to C_T and is inversely proportional to C_{fg-d}. In principle, we can increase V_A as much as we like by making the coupling capacitors all proportionally larger (thereby increasing C_T), or by making the transistor narrower (thereby decreasing C_{fg-d}), or by using both techniques. However, fighting an

exponential in this manner is almost always a losing battle: If we make the transistor narrow, we lose the high end of our exponential current range; if we make the coupling capacitors too large, we reduce the bandwidth of our circuits.

For comparison, Figure 5.5b shows a plot of measured drain current as a function of drain voltage for a normal *p*MOS transistor on the same chip as the *p*FGMOS transistor and with geometry identical to that of the four-input *p*FGMOS transistor. To obtain this curve, I adjusted the gate voltage of the *p*MOS transistor such that it passed about 3 nA when its drain was 5 volts below V_{DD}, and then I swept the drain voltage from 0 to 5 volts below V_{DD} while measuring the drain current. The measured current–voltage curve is slightly concave down, but, if I fit a straight line to it over drain voltages ranging from 100 mV up to 2 V, then I get an Early voltage of about 20 V for this transistor. Thus, while the *p*MOS transistor is not a perfect current source, it is much better current source than is the *p*FGMOS transistor.

I can increase the output resistance of a FGMOS transistor dramatically by connecting a second transistor with a fixed gate voltage in series with the FGMOS transistor, as shown in Figures 5.5b and 5.13; this well-known circuit configuration is called a **cascode**. The cascode transistor acts as a source follower with a constant input voltage, V_{cas}; thus, it fixes the drain voltage of the FGMOS transistor (i.e., fixes the source-follower's output voltage), thereby reducing the change in current through both transistors resulting from a change in the drain voltage of the cascode transistor. Figure 5.5b shows a plot of measured drain current as a function of drain voltage for the four-input *p*FGMOS transistor cascoded with the *p*MOS transistor. In this case, I set V_{cas} to 1.1 V below V_{DD} and I set the voltage of all four control gates such that the cascoded *p*FGMOS transistor passed 3 nA when its drain was 5 V below V_{DD}. Then, I swept the cascode transistor's drain voltage from 0 to 5 volts below V_{DD} while measuring the current through the transistors. From these data, we can see that it takes slightly more drain voltage for the cascoded *p*FGMOS transistor to saturate than it did for either the *p*FGMOS transistor or the *p*MOS transistor (i.e., about 300 mV instead of 100 mV), but once the cascoded *p*FGMOS transistor has saturated, the current–voltage curve is flat to within 0.1%. In Appendix 5.A, I show that the effective Early voltage of a FGMOS transistor that is cascoded by a MOS transistor with an Early voltage of V_0 is given by

$$V_0' = \left(\frac{C_T}{\kappa C_{fg-d}} + 1 \right) V_0,$$

which I can express in terms of V_A, which I defined in Equation 5.18, as

$$V_0' = \left(\frac{V_A}{U_T} + 1 \right) V_0 .$$

Using this expression, I calculate that the cascoded *p*FGMOS transistor should have an effective Early voltage of about

$$V_0' = \left(\frac{2.93\text{V}}{0.0258\text{V}} + 1 \right) 20\text{V} \approx 2300\text{V},$$

which is consistent with its current–voltage curve being flat to within 0.1%.

Figure 5.6 shows measured normalized transconductance plotted as a function of drain current for a *p*MOS transistor and the four-input *p*FGMOS transistor. I obtained these data from measurements of drain current as I swept the gate voltage (in the case of the *p*MOS transistor) and the voltage on all four control gates (in the case of the *p*FGMOS transistor) in 1-mV increments throughout the subthreshold regime. For each of these sweeps, I computed the slope of a least-squares, straight-line fit to the log of the drain current over a ±5 mV range of gate voltages surrounding each point on the curve. Doing this procedure is equivalent to calculating the transconductance of the device at each point and dividing by the DC current level. To see this equivalence, I write

$$\frac{\partial}{\partial V_g} \log I_d = \frac{1}{I_d} \frac{\partial I_d}{\partial V_g} = \frac{g_m}{I_d}.$$

I multiplied the resulting curves by the thermal voltage, U_T, to obtain dimensionless quantities corresponding to κ for the *p*MOS transistor and to $\kappa \frac{C}{C_T}$ for the *p*FGMOS transistor. For devices with ideal exponential current–voltage characteristics, these curves would be flat (i.e., would be constant). At currents below 100 pA, each curve falls off due to leakage currents. At currents above 100 nA, each curve falls off because the transistor is beginning to go above threshold. Over the three decades of current between 100 pA and 100 nA, the normalized transconductance curves are flat to within about 5% of their peak values. Over the two-decade range from about 400 pA to 40 nA, the normalized transconductance curves are flat to within about 1% of their peak values.

5.4. The Matching of Small Capacitors for Analog VLSI

In this section, I describe a DC technique for assessing capacitor mismatch based on the fact that, in subthreshold, the slope of a FGMOS transistor's current–voltage characteristic plotted on semilog axes is directly proportional to the capacitance of the control gate. Thus, for a given FGMOS transistor operating at a constant temperature, mismatch in these slopes directly reflect mismatch in the coupling capacitors. I also present data on the matching of

small (ranging from 6 μm per side to 20 μm per side) square poly1-poly2 capacitors fabricated in a standard 2-μm double-poly CMOS process that is commonly available through MOSIS. My colleagues and I originally published this technique and some of these data in 1996 [23]; a similar method of assessing capacitor matching was proposed independently and published nearly simultaneously by Tuinhout and his colleagues [24]. Their measurement technique, called the **floating-gate capacitance measurement method**, is based on measuring mismatch in the voltage gains of an above-threshold, multiple-input floating-gate source follower.

As analog circuits have been integrated with digital circuits on CMOS chips [25], the capacitor has come to dominate analog circuit design. In many cases, resistors of suitable values are unavailable in CMOS processes, and inductors are generally unattractive for use in the design of all but RF or microwave circuits. As a result, many analog circuits rely on capacitor matching to achieve high accuracy; these include analog and digital signal-processing circuits made from FGMOS transistors [11, 13–21], switched-capacitor circuits [26, 27], many D/A and A/D converters [28, 29], and precision-gain amplifiers [30]. In 1994, McNutt and his colleagues [31] reported layout techniques for achieving capacitor matching to within 0.1% or less. These techniques involve interleaving square identical unit cells between 20 μm and 40 μm per side in clever alternating patterns, and using dummy capacitors on the boundaries of capacitor arrays. Such techniques are well suited to situations in which a relatively small number of high-precision analog circuits is required on a single mixed-signal VLSI chip. However, as analog VLSI information processing systems develop [22], small cell sizes will prevent us from using such large unit cells. Furthermore, because there is a large wiring overhead associated with connecting together the interleaved unit cells, we will not be able to use such common-centroid layout techniques and simultaneously achieve small cell sizes. Consequently, we want to find out how much precision is achievable using small capacitors (i.e., using capacitors less than 20 μm per side).

5.4.1. Experimental Method to Assess Capacitor Mismatch

Figure 5.7 illustrates schematically the experimental method that I use to asses mismatch among K nominally identical capacitors. For each value of k between 1 and K, using a decoder and CMOS transmission gates, I connect the kth control gate of a K-input FGMOS transistor (I show an nFGMOS transistor in Figure 5.7) to one voltage source, V_s, and the remaining $K-1$ control gates to another voltage source, V_c. I set the value of V_c such that the FGMOS transistor is operating in subthreshold. Then, for each k between

1 and K, I measure the FGMOS transistor's drain current, I_{d}, as I sweep V_{s} over some voltage range so that I_{d} changes by about one decade. From Equation 5.10, under these conditions I_{d} has the form

$$I_{\mathrm{d}} = I_{\mathrm{c}} \exp\left[\frac{\kappa C_k}{C_{\mathrm{T}}} \frac{V_{\mathrm{s}}}{\mathrm{U}_{\mathrm{T}}}\right],$$

where

$$I_{\mathrm{c}} = \frac{W}{L} \mathrm{I}_0 \exp\left[\frac{Q}{\mathrm{Q}_{\mathrm{T}}}\right] \exp\left[\frac{\kappa V_{\mathrm{c}}}{C_{\mathrm{T}} \mathrm{U}_{\mathrm{T}}} \sum_{i \neq k} C_i\right] \exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}} \frac{V_{\mathrm{d}}}{\mathrm{U}_{\mathrm{T}}}\right].$$

Thus, the slope of the I_{d} versus V_{s} curve plotted on a semilog scale is directly proportional to C_k. Because the K current–voltage characteristics come from the same FGMOS transistor over the same range of drain currents, the values of κ and C_{T} are identical across all K measurements. Consequently, assuming the temperature remains fixed during the course of the experiment, mismatch in the slopes of these K current–voltage characteristics will directly reflect mismatch in the capacitances of the K control gates.

5.4.2. Experimental Results and Discussion

I designed a chip with eight 25-input *n*FGMOS transistors that was fabricated in Oribit's 2-μm double-poly CMOS process through MOSIS. The 25 control gates of each transistor were addressable through a 5-bit decoder and CMOS transmission gates, as shown in Figure 5.7. The control gates of each *n*FGMOS transistor were identically drawn squares arranged in a linear array with the minimum spacing allowed by design rules, as shown in Figure 5.8. The first *n*FGMOS transistor's control gates were 6 μm per side. The second *n*FGMOS transistor's control gates were 8 μm per side, and so on in 2-μm increments up to 20 μm per side for the eighth *n*FGMOS transistor.

For each of the four chips that I received from MOSIS, I swept the control gate of each *n*FGMOS transistor from ground to 5 V in 100 mV increments while measuring the transistor's drain current. I extracted the slopes of the resulting exponential current–voltage curves by linear regression. For each capacitor size and for each chip, I normalized the distribution of slopes to have a mean of unity by dividing each by its sample mean. I call the resulting normalized quantities **relative capacitances**. For each capacitor size, I then pooled the relative-capacitance distributions from the four chips.

Figures 5.9, 5.10, and 5.11 show the relative-capacitance distributions for each capacitor size in two different ways. For each capacitor size, the plot on the left shows relative capacitance as a function of position in the array, and the plot on the right shows a

histogram of the relative-capacitance distribution. Note that, for each capacitor size, the capacitors on the ends of each array (i.e., the capacitors at positions 1 and 25) are systematically lower than the rest (ranging from almost 2% in the 6-μm case to 0.5% in the 20-μm case). The capacitors on the ends of each array have only one neighbor, whereas the rest of the capacitors have two. These data corroborate the well-known fact that we can improve matching by adding **dummy capacitors** to make the surroundings of each functional capacitor nominally identical [24, 25, 27]. For example, if we designate the capacitors in positions 1 and 25 in each of the four arrays of 6-μm capacitors as dummys and throw them out of the distribution, the worst-case mismatch goes from nearly 4% to about 2.5%. So it appears that we can improve the matching of even the smallest of capacitors markedly by using dummy capacitors to make the surround of each functional capacitor nominally identical.

Figure 5.12 summarizes all the capacitor matching data shown in Figures 5.9, 5.10, and 5.11; it shows a plot of **coefficient of variation**, which is defined as the standard deviation of a sample divided by its mean, of each pooled relative-capacitance distribution as a function of drawn capacitor area both including (o) and excluding (×) the capacitors on the ends (i.e., the capacitors at positions 1 and 25) of each array. The coefficient of variation is a relative measure of the width of a distribution. By excluding the capacitors on the ends of each array from each distribution, we get an idea of the improvement in matching that we achieve by using dummy capacitors as just mentioned.

In 1982, Shyu and his colleagues published a study modeling random errors in MOS capacitors [32]. They show that edge variations (e.g., variations in the lithographic process) introduce a relative capacitance error that scales as

$$\frac{\sigma_c}{\overline{C}} \propto C^{-\frac{3}{4}},$$

whereas area variations (e.g., variations in oxide thickness or dielectric constant) introduce a relative capacitance error that scales as

$$\frac{\sigma_c}{\overline{C}} \propto C^{-\frac{1}{2}}.$$

For small capacitors, which have a higher edge-to-area ratio, we expect edge variations to be the dominant error process. For large capacitors, which have a lower edge-to-area ratio, we expect area variations to be the dominant error process. The solid line in Figure 5.12 shows a –0.72 power-law fit to the first four points of the curve marked with circles, indicating that edge variations are the dominant error process for the smallest (i.e., from 6-μm per side to 12-μm per side) capacitors tested. It is difficult to say anything conclusive

about the larger capacitors tested. There is more variation in the data and they do not follow a –0.5 power law relationship. It could be that the measurement system that I used has a noise floor, and hence, is incapable of accurately resolving relative-capacitance mismatch much below 0.2%. Tuinhout and his colleagues [24], claim to be able to resolve relative-capacitance mismatch down to a level of 50 ppm (i.e., 0.005%) using a similar technique, so, even if there is a measurement system noise floor at the 0.2% level, I should be able to improve the method to resolve finer levels of mismatch by controlling for temperature variation, by collecting more data, and by averaging the data more.

5.5. Appendix 5.A

In this appendix, I derive a model of the current–voltage characteristic of a cascoded *n*FGMOS transistor operating in the subthreshold region, as shown in Figure 5.13. I assume that the drain voltage, V_{d}, of the cascode transistor, $\mathrm{M_{cas}}$, is sufficiently far above ground that both transistors are in saturation. For this calculation, I assume that the drain current of $\mathrm{M_{cas}}$ is given by

$$I_{\mathrm{d}} = \mathrm{I_{cas}} \exp\left[\frac{\kappa_{\mathrm{cas}} V_{\mathrm{cas}}}{\mathrm{U_T}}\right] \exp\left[-\frac{V}{\mathrm{U_T}}\right] \exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right], \tag{5.19}$$

where $\mathrm{I_{cas}}$ is the subthreshold pre-exponential scaling current of $\mathrm{M_{cas}}$, κ_{cas} is the back-gate coefficient of $\mathrm{M_{cas}}$, and $\mathrm{V_0}$ is the Early voltage of $\mathrm{M_{cas}}$. In Equation 5.19, I model the Early effect (i.e., channel-length modulation) with $\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right]$, rather than with the usual factor of $1+\frac{V_{\mathrm{d}}}{\mathrm{V_0}}$, for three reasons.[1] First, for small values of x, $1+x \approx \exp[x]$; we normally use this approximation in the other direction, but nothing stops us from using it in this direction. So, for moderately large values of $\mathrm{V_0}$, $1+\frac{V_{\mathrm{d}}}{\mathrm{V_0}}$ will be very nearly equal to $\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right]$. Second, for MOS transistors with short channels, the factor $\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right]$ better accounts for the drain-induced barrier-lowering (DIBL) effect than does $1+\frac{V_{\mathrm{d}}}{\mathrm{V_0}}$. Third, when we are working with exponential current–voltage relationships, if we account for the Early effect with $\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right]$, we can usually get closed-form solutions to the equations that we need to solve. If we instead use $1+\frac{V_{\mathrm{d}}}{\mathrm{V_0}}$, we nearly always get transcendental equations that are difficult to handle analytically.

I model the current through the *n*FGMOS transistor, $\mathrm{M_{FG}}$, with Equation 5.10. For convenience, I define

$$I_\nu = \frac{W}{L}\mathrm{I_0} \exp\left[\frac{Q}{\mathrm{Q_T}}\right] \exp\left[\sum_{k=1}^{K} \frac{\kappa C_k}{C_{\mathrm{T}}} \frac{V_k}{\mathrm{U_T}}\right],$$

[1] This insightful way of modeling the Early effect was suggested by Paul Hasler.

so that I can write Equation 5.10 as

$$I_{\mathrm{d}} = I_{\nu} \exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{V}{\mathrm{U_T}}\right]. \tag{5.20}$$

Because transistors $\mathrm{M_{cas}}$ and $\mathrm{M_{FG}}$ are connected in series, the current, I_{d}, flows through each transistor. Thus, I can equate Equations 5.19 and 5.20 to get

$$I_{\nu} \exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{V}{\mathrm{U_T}}\right] = \mathrm{I_{cas}} \exp\left[\frac{\kappa_{\mathrm{cas}} V_{\mathrm{cas}}}{\mathrm{U_T}}\right]\exp\left[-\frac{V}{\mathrm{U_T}}\right]\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right],$$

which I rearrange to obtain

$$\exp\left[\left(1+\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\right)\frac{V}{\mathrm{U_T}}\right] = \frac{\mathrm{I_{cas}}}{I_{\nu}}\exp\left[\frac{\kappa_{\mathrm{cas}} V_{\mathrm{cas}}}{\mathrm{U_T}}\right]\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right]. \tag{5.21}$$

By raising both sides of Equation 5.21 to the

$$\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}} + \kappa C_{\mathrm{fg-d}}}$$

power, I obtain

$$\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{V}{\mathrm{U_T}}\right] = \left(\frac{\mathrm{I_{cas}}}{I_{\nu}}\right)^{\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}} \exp\left[\frac{\kappa_{\mathrm{cas}}\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{cas}}}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right]. \tag{5.22}$$

Substituting Equation 5.22 into Equation 5.20, I obtain an expression for the drain current of a cascoded *n*FGMOS transistor:

$$\begin{aligned} I_{\mathrm{d}} &= I_{\nu}^{1-\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}} \mathrm{I_{cas}}^{\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}} \exp\left[\frac{\kappa_{\mathrm{cas}}\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{cas}}}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right] \\ &= I_{\nu}^{\frac{C_{\mathrm{T}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}} \mathrm{I_{cas}}^{\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}} \exp\left[\frac{\kappa_{\mathrm{cas}}\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{cas}}}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{d}}}{\mathrm{V_0}}\right] \\ &= \left(\frac{W}{L}\right)^{\frac{C_{\mathrm{T}}}{C'_{\mathrm{T}}}} \mathrm{I'_0} \exp\left[\frac{Q}{\mathrm{Q'_T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C'_{\mathrm{T}}}\frac{V_k}{\mathrm{U_T}}\right]\exp\left[\frac{V_{\mathrm{d}}}{\mathrm{V'_0}}\right], \end{aligned} \tag{5.23}$$

where

$$C'_{\mathrm{T}} = C_{\mathrm{T}} + \kappa C_{\mathrm{fg\text{-}d}}, \quad \mathrm{Q'_T} = \frac{C'_{\mathrm{T}}\mathrm{U_T}}{\kappa}, \quad \mathrm{V'_0} = \left(\frac{C_{\mathrm{T}}}{\kappa C_{\mathrm{fg\text{-}d}}}+1\right)\mathrm{V_0},$$

and

$$\mathrm{I'_0} = \mathrm{I_0}\left(\frac{\mathrm{I_{cas}}}{\mathrm{I_0}}\right)^{\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}} \exp\left[\frac{\kappa_{\mathrm{cas}}\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}+\kappa C_{\mathrm{fg-d}}}\frac{V_{\mathrm{cas}}}{\mathrm{U_T}}\right].$$

Equation 5.23 has the same form as does Equation 5.10. For the cascoded *n*FGMOS transistor, the slope of the exponential dependence of the drain current on the drain voltage has been reduced by a factor of approximately $\frac{\mathrm{V_0}}{\mathrm{U_T}}$, which can be several hundreds, over

that of the uncascoded *n*FGMOS transistor. Hence, the cascode connection can greatly increase the output resistance of the *n*FGMOS transistor. A similar calculation can be done for a cascoded *p*FGMOS transistor, and will yield identical results.

5.6. References

1. D. Kahng and S. M. Sze, "A Floating-Gate and Its Application to Memory Devices," *The Bell System Technical Journal*, vol. 46, no. 4, pp. 1288–1295, 1967.

2. E. Säckinger and W. Guggenbühl, "An Analog Trimming Circuit Based on a Floating-Gate Device," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 6, pp. 1437–1440, 1988.

3. L. R. Carley, "Trimming Analog Circuits Using Floating-Gate Analog MOS Memory," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 6, pp. 1569–1575, 1989.

4. D. B. Kirk, *Accurate and Precise Computation Using Analog VLSI with Applications to Computer Graphics and Neural Networks*, Ph.D. Thesis, Department of Computer Science, California Institute of Technology, Pasadena, CA, 1993.

5. M. Holler, S. Tam, H. Castro, and R. Benson, "An Electronically Trainable Artificial Neural Network (ETANN) With 10240 'Floating-Gate' Synapses," in *Proceedings of the 1989 International Conference on Neural Networks*, Washington, D.C., vol. 2, pp. 191–196, June 1989.

6. J. Anderson, J. C. Platt, and D. B. Kirk, "An Analog VLSI Chip for Radial Basis Functions," in S. J. Hanson, J. D. Cowan, and C. L. Giles, eds., *Advances in Neural Information Processing Systems 5*, San Mateo, CA: Morgan Kaufmann, pp. 765–772, 1993.

7. E. Vittoz, H. Oguey, M. A. Maher, O. Nys, E. Dijkstra, and M. Chevroulet, "Analog Storage of Adjustable Synaptic Weights," in U. Ramacher and U. Rückert, eds., *VLSI Design of Neural Networks*, Boston: Kluwer, pp. 47–63, 1991.

8. P. Hasler, C. Diorio, B. A. Minch, and C. Mead, "Single Transistor Learning Synapses," in G. Tesauro, D. S. Touretzky, and T. K. Leen, eds., *Advances in Neural Information Processing Systems 7*, Cambridge, MA: MIT Press, pp. 817–824, 1995.

9. K. Hieda, M. Wada, T. Shibata, and H. Iizuka, "A New EEPROM Cell with Dual Control Gate Structure," in *Digest of Technical Papers of the 1983 Symposium on VLSI Technology*, pp. 114–115, September 1983.

10. K. Hieda, M. Wada, T. Shibata, and H. Iizuka, "Optimum Design of Dual-Control Gate Cell for High-Density EEPROM's," *IEEE Transactions on Electron Devices*, vol. ED-32, no. 9, pp. 1776–1780, 1985.

11. T. Shibata and T. Ohmi, "A Functional MOS Transistor Featuring Gate-Level Weighted Sum and Threshold Operations," *IEEE Transactions on Electron Devices*, vol. 39, no. 6, pp. 1444–1455, 1992.

12. A. G. Andreou and K. A. Boahen, "Neural Information Processing II," in M. Ismail and T. Fiez, eds., *Analog VLSI Signal and Information Processing*, New York: McGraw-Hill, pp. 358–413, 1994.

13. K. Yang, *An Investigation of MOSFET and Floating-Gate MOSFET Devices and Circuits for Analog VLSI*, Ph.D. Thesis, Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD, 1994.

14. K. Yang and A. G. Andreou, "A Multiple Input Differential Amplifier Based on Charge Sharing on a Floating-Gate MOSFET," *Journal of Analog Integrated Circuits and Signal Processing*, vol. 6, no. 3, pp. 197–208, 1994.

15. J. Ramírez-Angulo, "±0.75V BiCMOS Four Quadrant Analog Multiplier with Rail-Rail Input Signal-Swing," in *Proceedings of the 1996 IEEE International Symposium on Circuits and Systems*, Atlanta, GA, vol. 1, pp. 242–245, May 1996.

16. T. Shibata and T. Ohmi, "Neuron MOS Binary-Logic Integrated Circuits—Part I: Design Fundamentals and Soft-Hardware-Logic Circuit Implementation," *IEEE Transactions on Electron Devices*, vol. 40, no. 3, pp. 570–576, 1993.

17. T. Shibata and T. Ohmi, "Neuron MOS Binary-Logic Integrated Circuits—Part II: Simplifying Techniques of Circuit Configuration and Their Practical Applications," *IEEE Transactions on Electron Devices*, vol. 40, no. 5, pp. 570–576, 1993.

18. T. Yamashita, T. Shibata, and T. Ohmi, "Neuron MOS Winner-Take-All Circuit and Its Application to Associative Memory," in *Digest of Technical Papers of the 1993 IEEE International Solid-State Circuits Conference*, San Francisco, CA, pp. 236–237, February 1993.

19. T. Shibata, H. Kosaka, H. Ishii, and T. Ohmi, "A Neuron-MOS Neural Network Using Self-Learning-Compatible Synapses," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 8, pp. 913–922, 1995.

20. B. A. Minch, C. Diorio, P. Hasler, and C. Mead, "A νMOS Soft-Maximum Current Mirror," in *Proceedings of the 1995 IEEE International Symposium on Circuits and Systems*, Seattle, WA, vol. 3, pp. 2249–2252, May 1995.

21. B. A. Minch, C. Diorio, P. Hasler, and C. Mead, "Translinear Circuits Using Subthreshold Floating-Gate MOS Transistors," *Journal of Analog Integrated Circuits and Signal Processing*, vol. 9, no. 2, pp. 167–179, 1996.

22. C. Mead, *Analog VLSI and Neural Systems*, Reading, MA: Addison-Wesley, 1989.

23. B. A. Minch, C. Diorio, P. Hasler, and C. Mead, "The Matching of Small Capacitors for Analog VLSI," in *Proceedings of the 1996 IEEE International Symposium on Circuits and Systems*, Atlanta, GA, vol. 1, pp. 239–241, May 1996.

24. H. P. Tuinhout, H. Elzinga, J. T. Brugman, and F. Postma, "The Floating Gate Measurement Technique for Characterization of Capacitor Matching," *IEEE Transactions on Semiconductor Manufacturing*, vol. 9, no. 1, pp. 2–8, 1996.

25. E. A. Vittoz, "The Design of High-Performance Analog Circuits on Digital CMOS Chips," *IEEE Journal of Solid-State Circuits*, vol. SC-20, no. 3, pp. 657–665, 1985.

26. P. E. Allen and E. Sánches-Sinencio, *Switched Capacitor Circuits*, New York: Van Nostrand Reinhold, 1984.

27. R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, New York: Wiley, 1986.

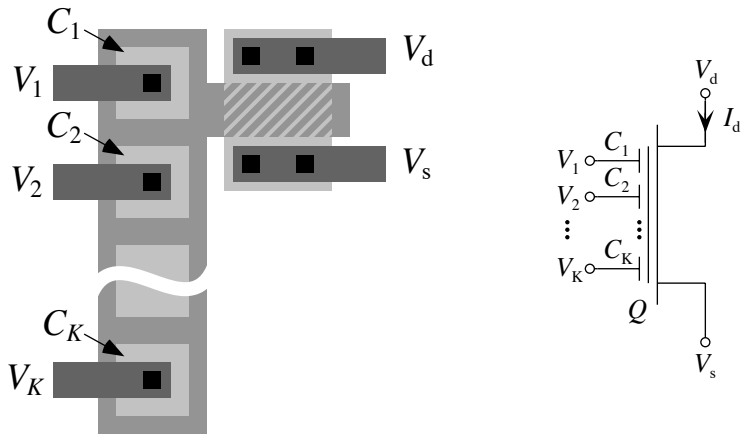
28. J. L. McCreary and P. R. Gray, "All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques—Part I," *IEEE Journal of Solid-State Circuits*, vol. SC-10, no. 6, pp. 371–379, 1975.

29. R. E. Suarez, P. R. Gray, and D. A. Hodges, "All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques—Part II," *IEEE Journal of Solid-State Circuits*, vol. SC-10, no. 6, pp. 379–385, 1975.

30. R. McCharles and D. A. Hodges, "Charge Circuits for Analog LSI," *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 7, pp. 490–497, 1978.

31. M. J. McNutt, S. LeMarquis, and J. L. Dunkley, "Systematic Capacitance Matching Errors and Corrective Layout Procedures," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 5, pp. 611–616, 1994.

32. J.-B. Shyu, G. C. Temes, and K. Yao, "Random Errors in MOS Capacitors," *IEEE Journal of Solid-State Circuits*, vol. SC-17, no. 6, pp. 1070–1076, 1982.



(a) (b)

Figure 5.1. Two views of a *K*-input, *n*-channel floating-gate MOS (*n*FGMOS) transistor. (a) A typical *n*FGMOS transistor layout in a double-poly, *n*-well CMOS process. We use the first layer of polysilicon (poly1) to form the floating gate. We use the second layer of polysilicon (poly2) to form the control gates; these control gates capacitively couple into the floating gate to modulate its voltage. We can proportionally change the coupling strength between a control gate and the floating gate by changing the area of overlap between poly1 and poly2 for that control gate. (b) A circuit symbol for a *K*-input *n*FGMOS transistor. I show the capacitance of the *k*th control gate on the circuit symbol by a nearby C_k. If all control-gate capacitances are nominally identical with value *C*, I omit the *C* from the schematic to prevent unnecessary clutter. I indicate the value of the net charge stored on the floating gate by a nearby *Q* as shown.

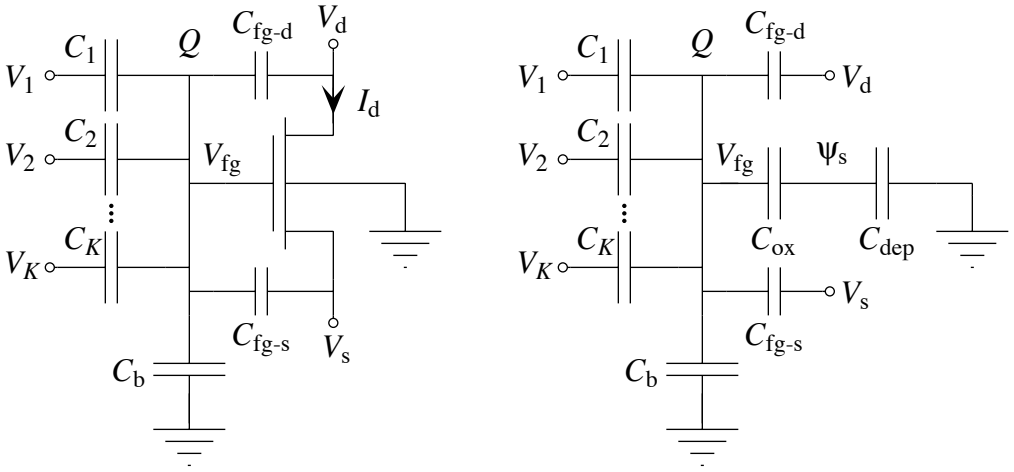
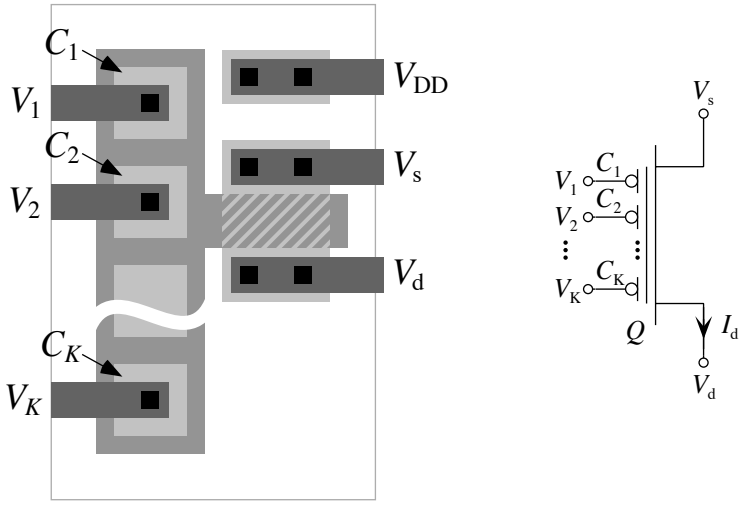


Figure 5.2. Capacitive-divider model for a K-input nFGMOS transistor. (a) A lumped circuit model for the nFGMOS transistor shown in Figure 5.1a. The K control-gate voltages, V_1 through V_K, capacitively couple into the floating gate through capacitors, C_1 through C_K, respectively. I denote by C_b the parasitic capacitance to the substrate beneath the floating gate. The source and drain voltages, V_s and V_d, couple into the floating gate through parasitic overlap capacitances, $C_{fg\text{-}s}$ and $C_{fg\text{-}d}$, respectively. I denote by Q the net charge stored on the floating gate. (b) The lumped circuit model of part a with the nMOS transistor replaced itself by an equivalent capacitive divider model. I denote by C_{ox} the gate-oxide capacitance of the nMOS transistor. I denote by C_{dep} the equivalent (nonlinear) capacitance of the depletion layer beneath the channel of the nMOS transistor. I denote by ψ_s the surface potential of the nMOS transistor.



(a) (b)

Figure 5.3. Two views of a K-input, p-channel floating-gate MOS (pFGMOS) transistor. (a) A typical pFGMOS transistor layout in a double-poly, n-well CMOS process. To minimize sensitivity to power-supply fluctuations, I keep as much of the floating gate area as possible over the n-well. (b) A circuit symbol for a K-input pFGMOS transistor. If no substrate connection is shown, I assume that the well is connected to the power-supply voltage. Again, I show the capacitance of the kth control gate on the circuit symbol by a nearby C_k. If all control-gate capacitances are nominally identical with value C, I omit the C from the schematic to prevent unnecessary clutter. I indicate the value of the net charge stored on the floating gate by a nearby Q as shown.

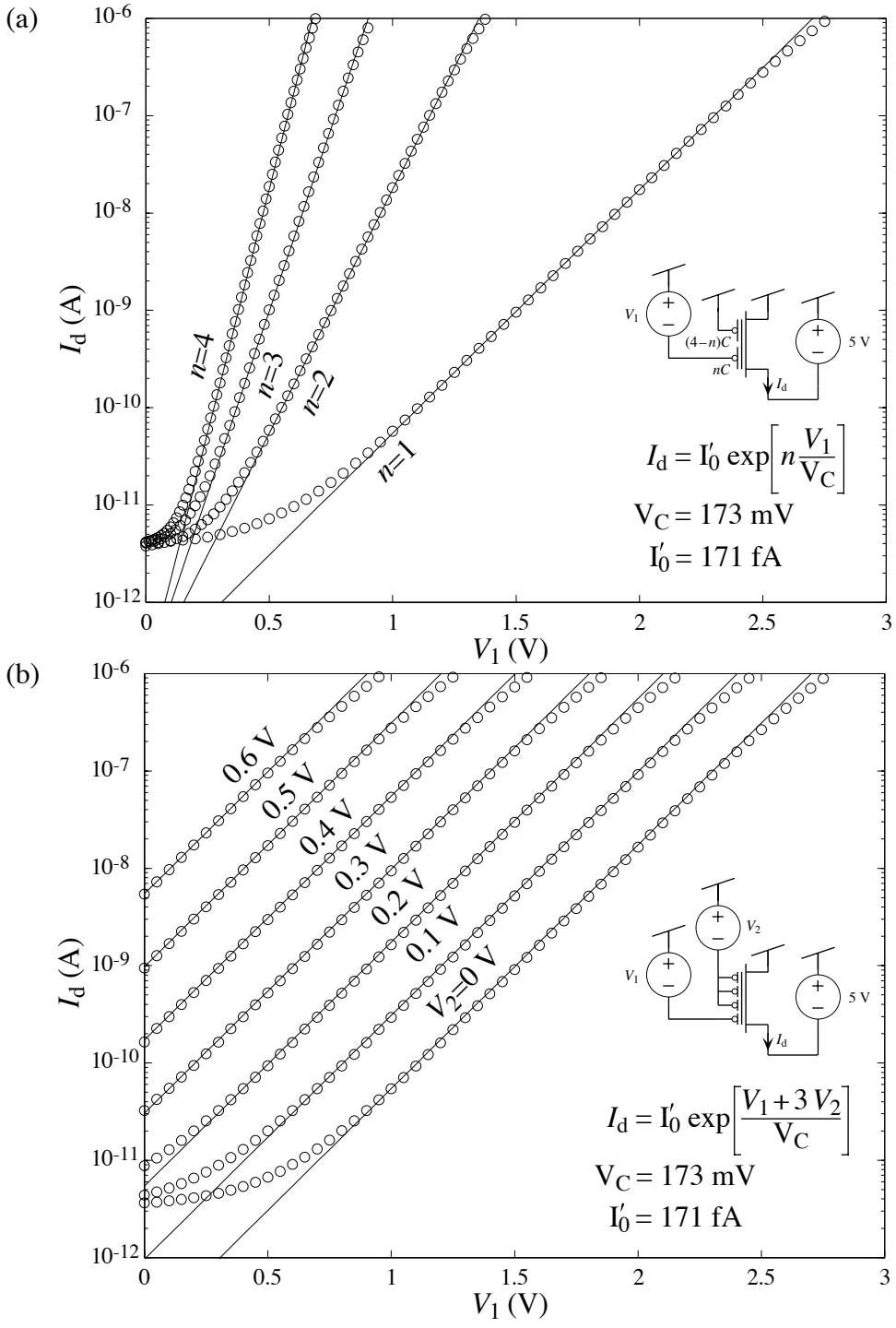


Figure 5.4. Measurements of a four-input *p*FGMOS transistor's drain current (a) sweeping *n* control gates with the others connected to V_{DD}, and (b) sweeping one control gate for various values of the remaining three. Solid lines show fits of Equation 5.13 to the data; note that the fit parameters I'_0 and V_C are the same for all curves in parts a and b.

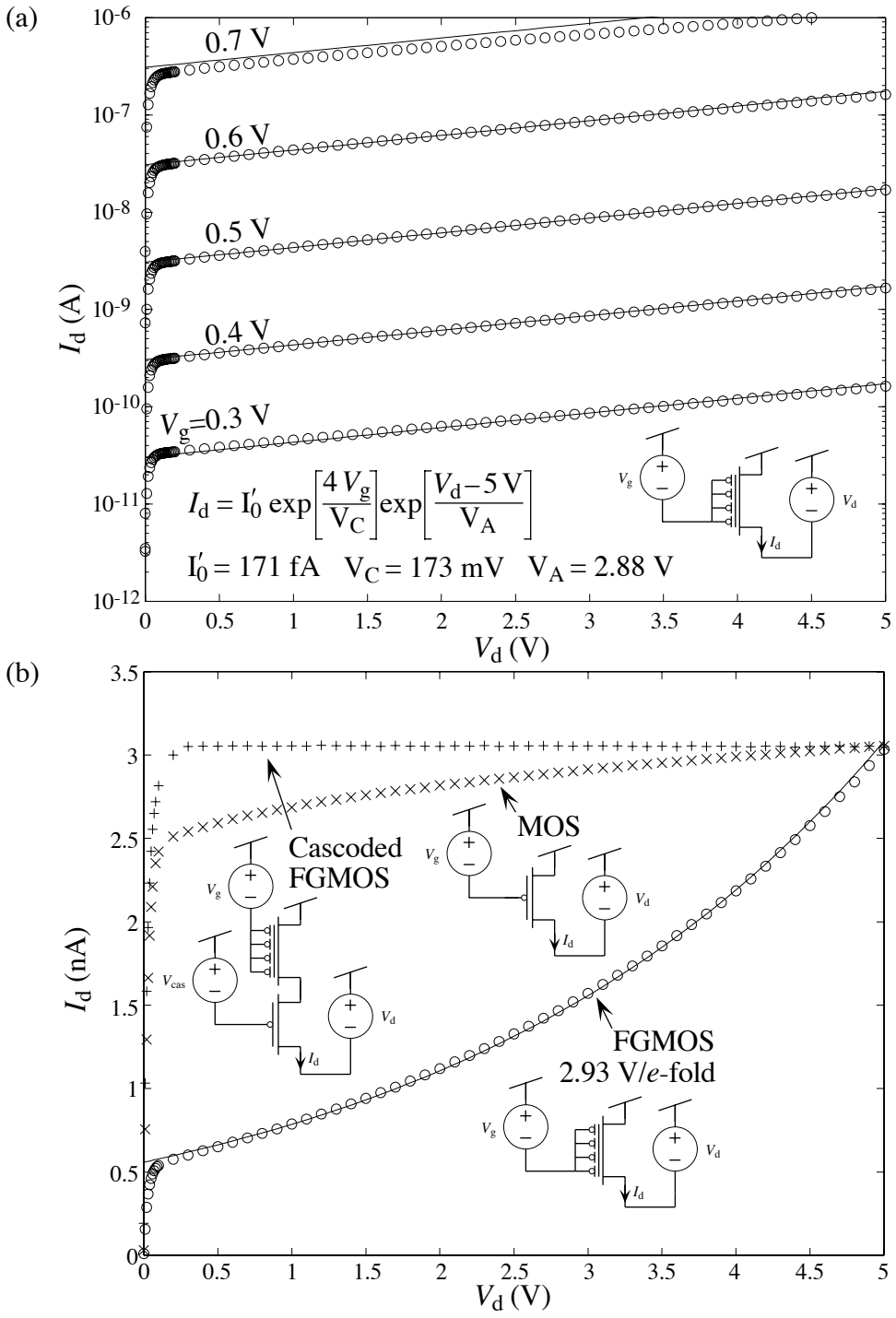


Figure 5.5. Measurements of a four-input *p*FGMOS transistor's drain current for (a) several values of V_g plotted on semilog axes, and (b) for a single value of V_g plotted on a linear scale with corresponding curves for a *p*MOS transistor and a cascoded *p*FGMOS transistor. Solid lines show fits of Equation 5.13 to the data.

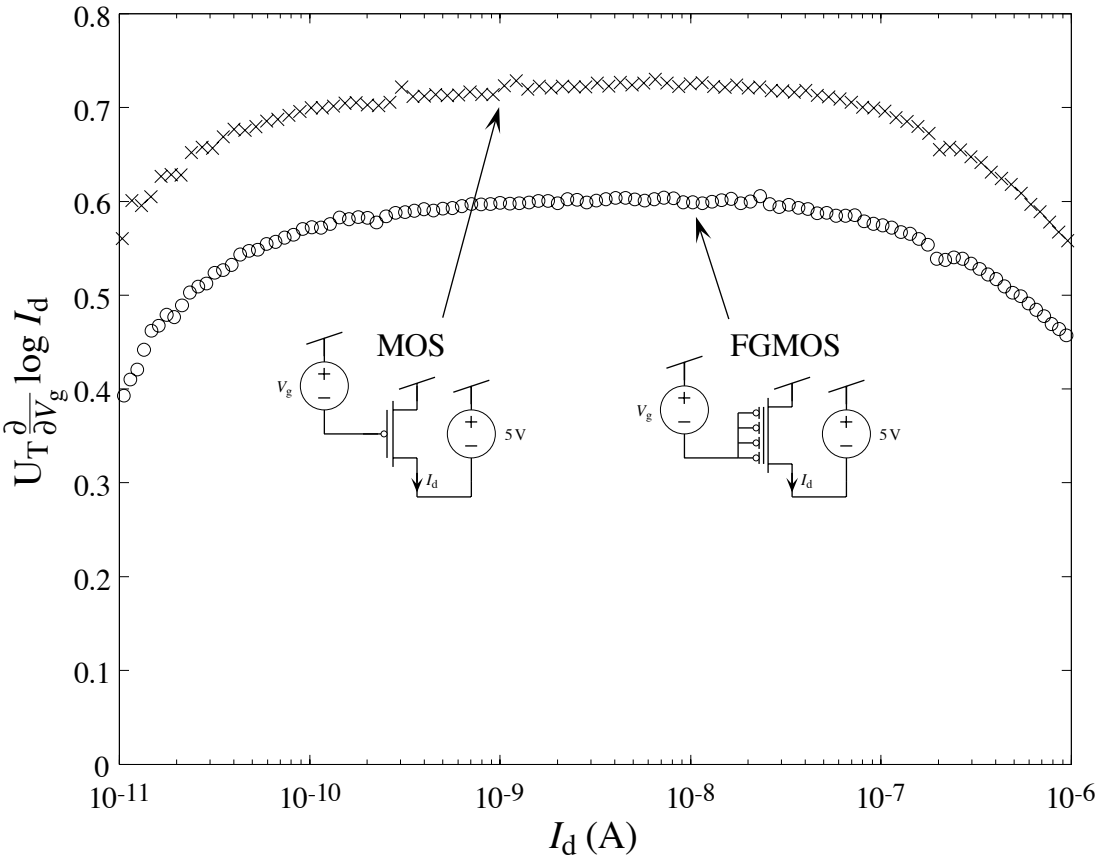


Figure 5.6. Measured normalized transconductance plotted as a function of drain current for a *p*MOS transistor and a four-input *p*FGMOS transistor. For devices with ideal exponential current–voltage characteristics, these curves would be flat (i.e., would be constant). At currents below 100 pA, the curves each fall off because of leakage currents. At currents above 100 nA, the curves each fall off because the transistors are going above threshold. Over these three decades of current, the normalized transconductance curves are flat to within about 5% of their peak values. Over the two-decade range from about 400 pA to 40 nA, the normalized transconductance curves are flat to within about 1% of their peak values.

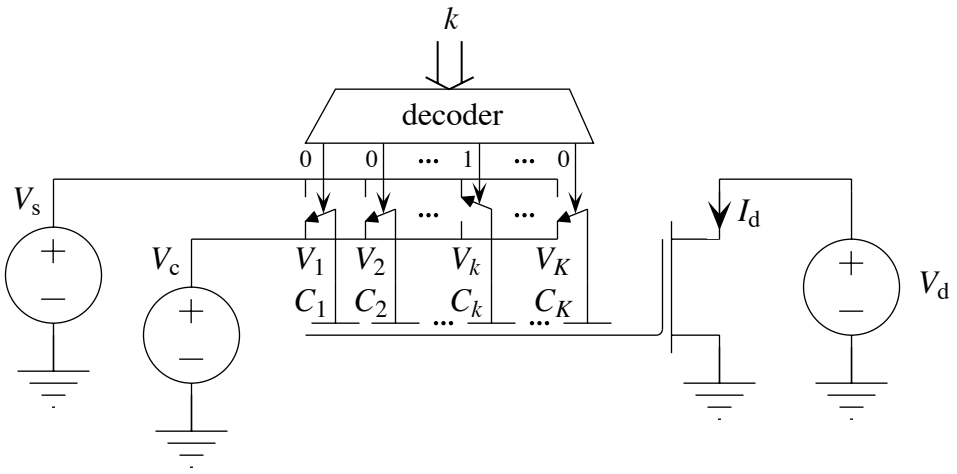
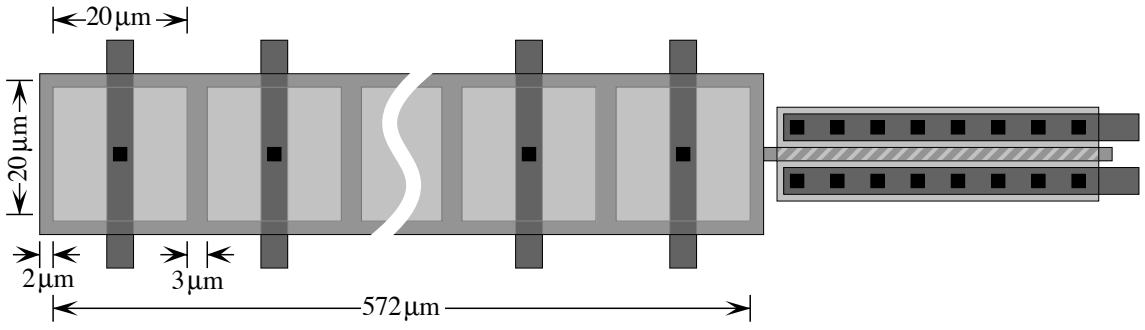
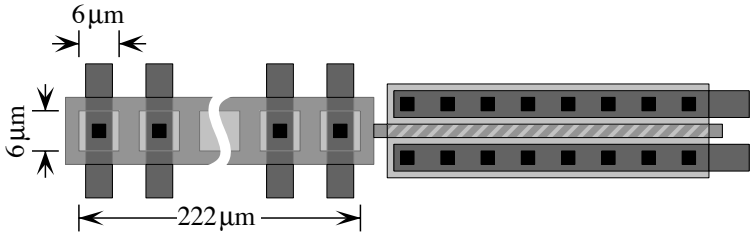


Figure 5.7. Experimental setup for measuring relative mismatch in nominally identical capacitors C_1 through C_K. For each value of k between 1 and K, I measure I_d as I sweep V_s over some range of voltage. The value of V_c is set such that the nFGMOS transistor is operating in subthreshold. The slope of I_d plotted as a function of V_k on semilog axes is directly proportional to C_k; hence, mismatch in the values of these slopes directly reflects mismatch in capacitors C_1 through C_K.



(a)



(b)

Figure 5.8. Layout view of the linear arrays of 25 square (a) 20-µm capacitors and (b) 6-µm capacitors with their respective *n*FGMOS transistors.

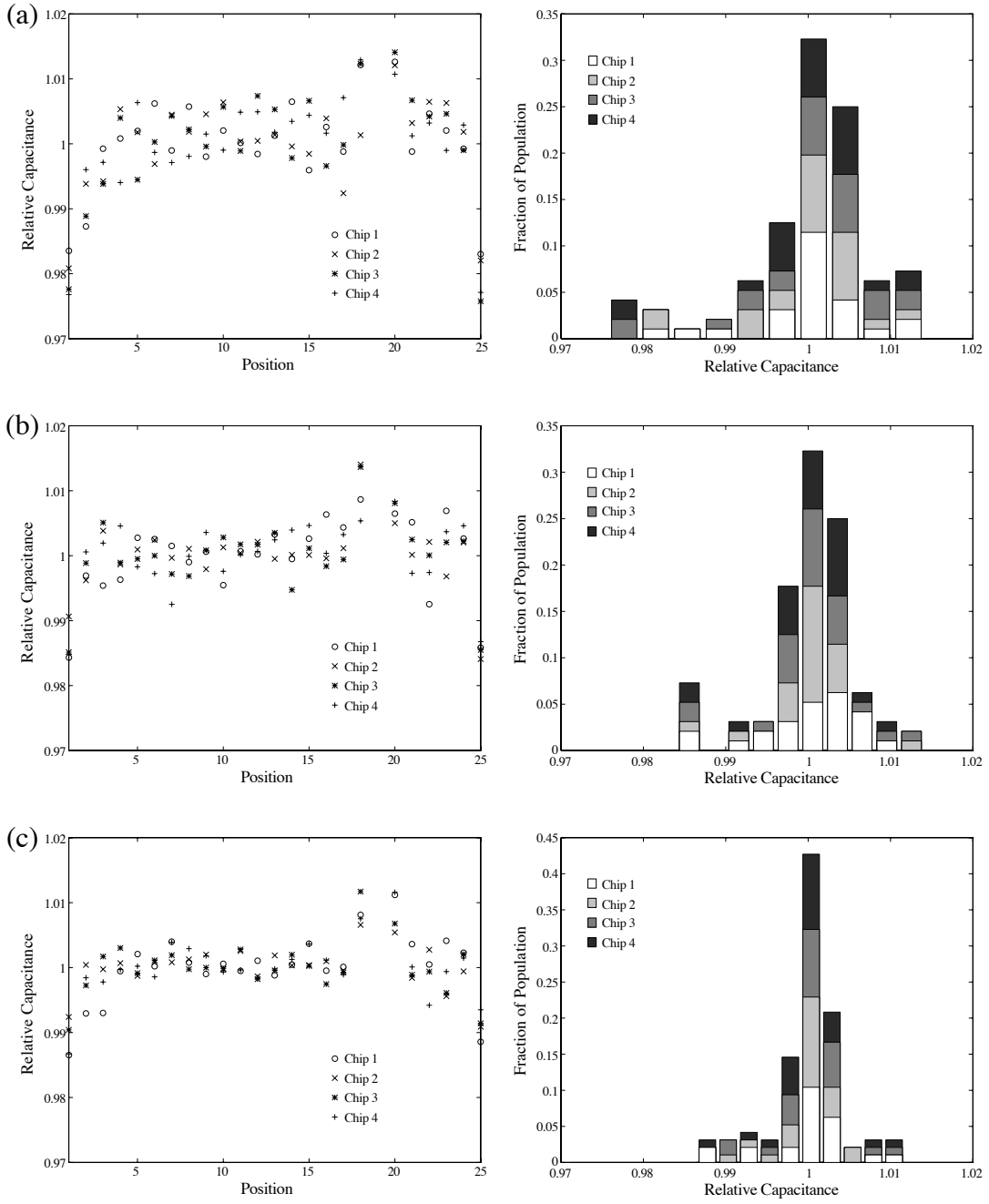


Figure 5.9. Relative-capacitance distributions for (a) 6-µm capacitors, (b) 8-µm capacitors, and (c) 10-µm capacitors. The plots on the left shows relative capacitance as a function of position in the array. Note that the capacitors on the right and left edges of the array (i.e., at positions 1 and 25) are systematically lower than the rest. These capacitors have only one neighbor, whereas the rest have two. The plots on the right show histograms of the relative-capacitance distributions.

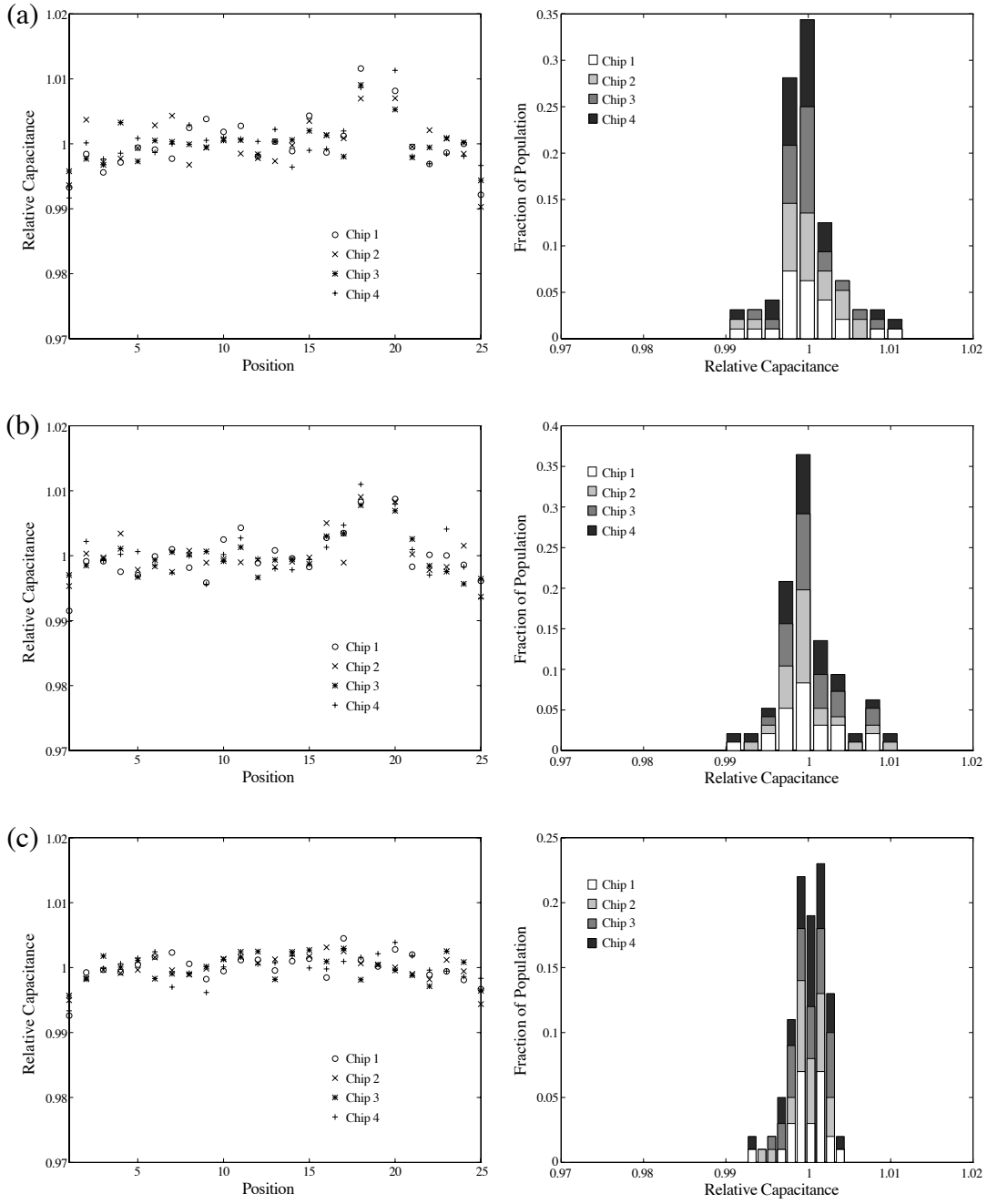


Figure 5.10. Relative-capacitance distributions for (a) 12-µm capacitors, (b) 14-µm capacitors, and (c) 16-µm capacitors. The plots on the left shows relative capacitance as a function of position in the array. Note that the capacitors on the right and left edges of the array (i.e., at positions 1 and 25) are systematically lower than the rest. These capacitors have only one neighbor, whereas the rest have two. The plots on the right show histograms of the relative-capacitance distributions.

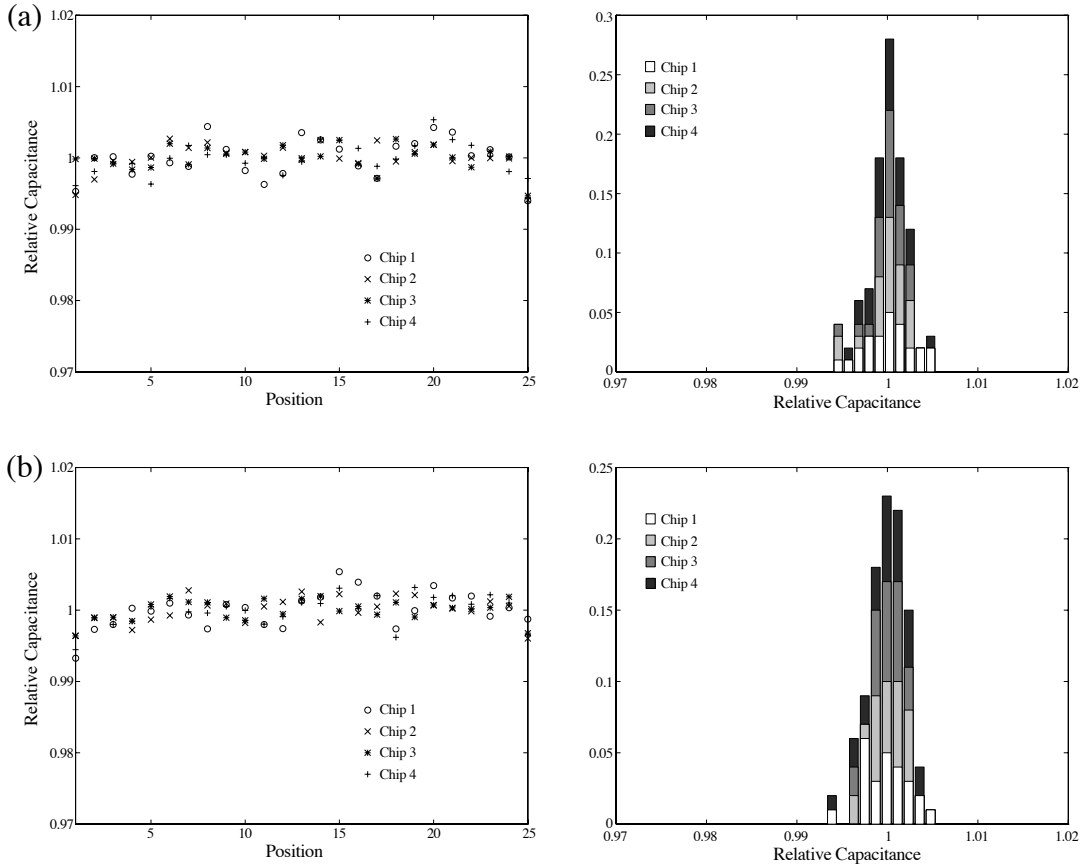


Figure 5.11. Relative-capacitance distributions for (a) 18-μm capacitors and (b) 20-μm capacitors. The plots on the left shows relative capacitance as a function of position in the array. Note that the capacitors on the right and left edges of the array (i.e., at positions 1 and 25) are systematically lower than the rest. These capacitors have only one neighbor, whereas the rest have two. The plots on the right show histograms of the relative-capacitance distributions.

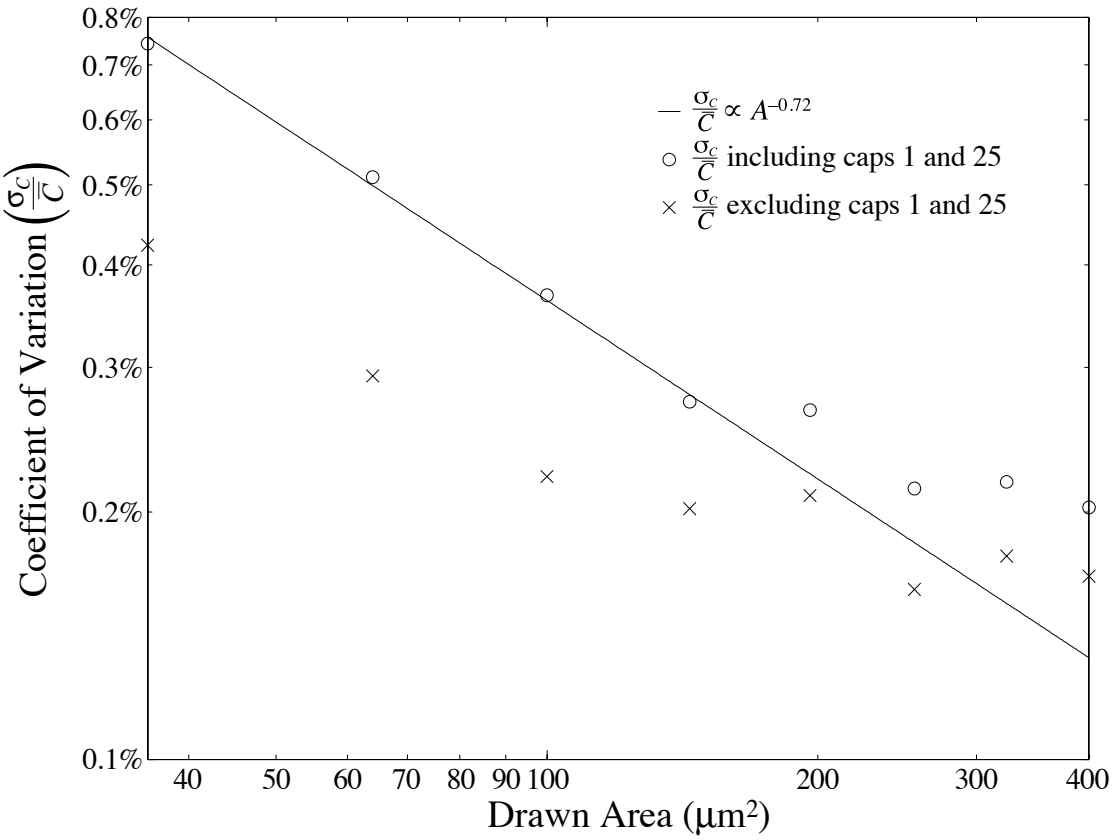


Figure 5.12. Plot showing coefficient of variation of the relative-capacitance distributions as a function of drawn area both including (O) and excluding (×) the capacitors on the ends (i.e., the capacitors at positions 1 and 25) of each array. The solid line shows a –0.72 power-law fit to the first four circles; a power law of –0.75 indicates that edge variations are the dominant mismatch mechanism [32].

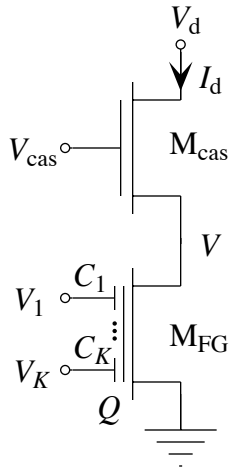


Figure 5.13. Schematic of a cascoded *n*FGMOS transistor. I assume that the drain of the cascode transistor, $\mathrm{M_{cas}}$, is sufficiently above ground that both transistors are saturated. For this configuration, I typically fix V_{cas} at about 1.1 volts above ground.

CHAPTER 6
THE SUBTHRESHOLD FLOATING-GATE MOS TRANSISTOR AS A MULTIPLE-INPUT TRANSLINEAR ELEMENT

In this chapter, I discuss the use as MITEs of the *K*-input subthreshold FGMOS transistor and of the cascoded *K*-input subthreshold FGMOS transistor. In Section 6.1, using the subthreshold FGMOS transistor model developed in Section 5.2, I identify and discuss two important aspects of using these devices as MITEs, and two of the largest nonidealities of these devices from the standpoint of MITE implementation. In Section 6.2, I present experimental measurements from 12 different MITE networks breadboarded both from four-input subthreshold *p*FGMOS transistors and from four-input cascoded subthreshold *p*FGMOS transistors that were fabricated in a standard 2-μm double-poly CMOS process through MOSIS.

6.1. The Subthreshold Floating-Gate MOS Transistor as a Multiple-Input Translinear Element

In Section 5.2, I showed that the drain current of a saturated *K*-input FGMOS transistor with a grounded source and operating in its subthreshold region is given by

$$I_{\mathrm{d}} = \frac{W}{L}\mathrm{I}_0 \exp\left[\frac{Q}{\mathrm{Q_T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_{\mathrm{T}}}\frac{V_k}{\mathrm{U_T}}\right]\exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_{\mathrm{T}}}\frac{V_{\mathrm{d}}}{\mathrm{U_T}}\right], \tag{6.1}$$

where W is the width of the channel, L is the length of the channel, I_0 is the pre-exponential scaling current, Q is the net charge stored on the floating gate, $\mathrm{Q_T}$ is the thermal charge given by $\frac{C_{\mathrm{T}}\mathrm{U_T}}{\kappa}$, κ measures the efficiency with which the floating gate modulates the FGMOS transistor's channel-surface potential, C_{T} is the total capacitance of the floating gate, $\mathrm{U_T}$ is the thermal voltage, C_k is the capacitance of the *k*th control gate, V_k is the *k*th control-gate voltage, $C_{\mathrm{fg-d}}$ is the parasitic drain-overlap capacitance, and V_{d} is the drain voltage. In Section 2.1, I defined the current–voltage relationship of an ideal MITE as

$$I = \lambda \mathrm{I_s} \exp\left[\sum_{k=1}^{K} \frac{w_k V_k}{\mathrm{U_T}}\right], \tag{6.2}$$

where λ is a dimensionless positive quantity that scales the output current proportionally, $\mathrm{I_s}$ is a pre-exponential scaling current, w_k is a dimensionless positive quantity that scales the kth input voltage, and V_k is the kth input voltage. If we neglect the last factor in Equation 6.1, which arises from the parasitic drain-overlap capacitance, we can easily see that a single K-input subthreshold FGMOS transistor implements a MITE with the following identifications between elements of Equation 6.1 and those of Equation 6.2:

$$\mathrm{I_s} \equiv \mathrm{I_0}, \quad \lambda \equiv \frac{W}{L} \exp\left[\frac{Q}{\mathrm{Q_T}}\right], \text{ and } \quad w_k \equiv \frac{\kappa C_k}{C_\mathrm{T}}.$$

Thus, the weighting coefficients are given by $\frac{\kappa C_k}{C_\mathrm{T}}$. Now, κ is a device parameter; it is a function of floating-gate-to-bulk potential and its value varies from run to run. Nonetheless, for two transistors on a single chip operating at comparable gate-to-bulk potentials, the values of κ will match well (a coefficient of variation of 0.3% is typical [1]). The total capacitance of the floating gate, C_T, includes various parasitic capacitances including the stray capacitance from the floating gate to the substrate and the (nonlinear) gate capacitance of the FGMOS transistor. Because capacitors match well in CMOS processes, as designers, we should be able to design our FGMOS transistors with well-matched (i.e., matched to within 1%) total capacitances by careful layout practices. Note that having matched total floating-gate capacitances implies that the geometry of each floating gate should be nominally identical, that each FGMOS transistor should have an identical set of control-gate capacitors, that the each FGMOS transistor channel should have identical geometry (so that all the gate capacitances are well matched), and that all parasitic capacitances that couple into each of the floating gates should be comparable. If we make C_T well matched for all FGMOS transistors, and if all FGMOS transistors are operating at roughly the same gate-to-bulk potential, the value of $\frac{\kappa}{C_\mathrm{T}}$ will be nearly identical for all FGMOS transistors. Then, by Theorem 4.2 in Section 4.2, we can scale each weighting coefficient by $\omega = \frac{C_\mathrm{T}}{\kappa}$ without changing the MITE-network power-law relationships, and, hence, we can conclude that the power-law relationships embodied in a MITE network made with nominally identical FGMOS transistors depend primarily on the values of the control-gate capacitances—these capacitances are precisely what we, as designers, have direct control over—and not on device parameters or parasitic capacitances. In other words, if we use FGMOS transistors with well-matched total floating-gate capacitances, then we can treat the elements of $\mathbf{W}_\mathrm{in}$ and $\mathbf{W}_\mathrm{out}$ as though they depended on only the

control-gate capacitances.

If we use only floating-gate devices that have nominally identical total floating-gate capacitances, then, as I just mentioned, we must make our FGMOS transistor channels with identical geometry. Thus, all FGMOS transistor will have identical $\frac{W}{L}$ ratios. So, for a given MITE network implemented with nominally identical subthreshold FGMOS transistors, if each row of the matrix of powers, Λ, sums to unity, then the constant of proportionality for the mth output current will be given by

$$\begin{aligned} K_m &= \lambda_{N+m} \prod_{n=1}^{N} \lambda_n^{-\Lambda_{mn}} \\ &= \frac{W}{L} \exp\left[\frac{Q_{N+m}}{\mathrm{Q_T}}\right] \prod_{n=1}^{N} \left(\frac{W}{L} \exp\left[\frac{Q_n}{\mathrm{Q_T}}\right]\right)^{-\Lambda_{mn}} \\ &= \frac{W}{L} \exp\left[\frac{Q_{N+m}}{\mathrm{Q_T}}\right] \prod_{n=1}^{N} \left(\frac{W}{L}\right)^{-\Lambda_{mn}} \exp\left[-\Lambda_{mn} \frac{Q_n}{\mathrm{Q_T}}\right] \\ &= \left(\frac{W}{L}\right)^{1-\sum_{n=1}^{N} \Lambda_{mn}} \exp\left[\frac{1}{\mathrm{Q_T}} \left(Q_{N+m} - \sum_{n=1}^{N} \Lambda_{mn} Q_n\right)\right] \\ &= \exp\left[\frac{1}{\mathrm{Q_T}} \left(Q_{N+m} - \sum_{n=1}^{N} \Lambda_{mn} Q_n\right)\right]. \end{aligned} \tag{6.3}$$

Thus, the mth output current is scaled by a dimensionless positive quantity that depends on the amount of charge stored on the floating gate of the mth output FGMOS transistor relative to a weighted summation of the amounts of charge stored on the floating gates of the input FGMOS transistors that factor into the mth output current. The charge stored on the floating gates can be modulated by Fowler–Nordheim tunneling [2], by subthreshold channel hot-electron injection [3, 4], by short-wave UV photoinjection [5, 6], or by a combination of these processes. Consequently, we can use the stored charge to compensate for scale-factor errors resulting from device mismatch, or to store weights that might be required in certain applications. If the charge stored on the floating gates is modulated in a sensible manner as a function of the history of the inputs to the circuit, then we can use these circuits as components in the construction of a variety of learning systems [3, 4]. Note that, if we arrange all the floating-gate charges such that they are balanced (i.e., so that $Q_1 = \ldots = Q_{N+M} = Q$), then Equation 6.3 becomes

$$\begin{aligned} K_m &= \exp\left[\frac{1}{\mathrm{Q_T}} \left(Q_{N+m} - \sum_{n=1}^{N} \Lambda_{mn} Q_n\right)\right] \\ &= \exp\left[\frac{Q}{\mathrm{Q_T}} \left(1 - \sum_{n=1}^{N} \Lambda_{mn}\right)\right] \end{aligned}$$

$$= \exp[0]$$
$$= 1,$$

independent of temperature. We can easily balance the floating-gate charge by exposing the FGMOS transistors to short-wave UV light for about 20 minutes with no power applied to the circuits; in this situation, any floating-gate charge imbalances set up potential differences that drive currents carried by photoexcited electrons through the oxide such that the charge imbalances are reduced. At equilibrium, no charge imbalances should persist.

We now return to account for the effects of the parasitic drain-overlap capacitance. The drain-overlap capacitance plays a role in input FGMOS transistors different from the role it plays in output FGMOS transistors. For an input FGMOS transistor, the drain-overlap capacitance acts as a small amount of extra self-coupling in its diode connection. In Appendix 6.A, I show that, to first order, these extra self-coupling terms perturb the power laws contained in Λ according to

$$\Lambda' \approx \Lambda\left(\mathbf{I} - \varepsilon \mathbf{W}_{\text{in}}^{-1}\right), \tag{6.4}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix, and ε is a measure of the size of $C_{\text{fg-d}}$ in the same units in which the weighting coefficients are expressed. Roughly speaking, Equation 6.4 implies that each of the powers in Λ is reduced by a factor on the order of $1 - \frac{C_{\text{fg-d}}}{C}$, where C is the size of the the unit control-gate capacitance. For the mth output FGMOS transistor, I account for the extra factor in Equation 4.1 resulting from the drain-overlap capacitance by including it in λ_{N+m}. Thus, the scale factor on the mth output current given by Equation 6.3 is modified as follows:

$$K_m = \exp\left[\frac{1}{\mathrm{U_T}}\left(Q_{N+m} - \sum_{n=1}^{N} \Lambda_{mn} Q_n\right)\right] \exp\left[\frac{\kappa C_{\text{fg-d}}}{C_{\mathrm{T}}} \frac{V_{N+m}}{\mathrm{U_T}}\right],$$

which means that the mth output current depends exponentially on the mth output voltage. As designers, we can weaken this exponential dependence by making the coupling capacitors all proportionally larger (thereby increasing C_{T}), by making the FGMOS transistors narrower (thereby decreasing $C_{\text{fg-d}}$), or by using both techniques. However, as I mentioned in Section 5.3, fighting an exponential in this manner is almost always like fighting a losing battle. If we make the transistor excessively narrow, we lose the high end of our exponential current range. If we make the coupling capacitors too large, we reduce the bandwidth of our circuits for a given set of input currents. By cascoding all the FGMOS transistors, we can reduce the effects of the drain-overlap capacitance in both the input and the output FGMOS transistors to negligible proportions. In Appendix 5.A, I showed that the current–voltage characteristic of a cascoded subthreshold FGMOS

transistor is of the same form as that of a subthreshold FGMOS transistor, except that the exponential dependence of the drain current on the drain voltage of a cascoded subthreshold FGMOS transistor is reduced by a factor of approximately $\frac{V_0}{U_T}$ (which can be equal to several hundreds), where V_0 is the Early voltage of the cascode transistor, over that of an uncascoded FGMOS transistor.

In Section 5.3, I showed a plot of normalized transconductance as a function of current level in a subthreshold FGMOS transistor and in a subthreshold MOS transistor. Ideally, these curves would be flat; however, they are slightly concave down. After about one decade of current on either side of their peak values, the normalized transconductance curves fall off by about 1%. After about one and one-half decades of current on either side of their peak values, these curves fall off by about 5%. This slight curvature introduces small deviations from ideal power-law behavior (i.e., introduces a slight curvature of the steady-state transfer curves on a log-log scale) for MITE networks implemented with subthreshold FGMOS transistors. After the drain-overlap capacitance, this variation of normalized transconductance with current level is the largest nonideality of the subthreshold FGMOS transistor from the standpoint of MITE implementation. Note that translinear loop circuits comprising subthreshold MOS transistors [1] (even those built from MOS transistors with their sources connected to their local substrates) suffer from this same nonideality.

6.2. Experimental Results

In this section, I present experimental data from 12 MITE networks breadboarded from four-input subthreshold *p*FGMOS transistors with nominally identical control gates that were fabricated in Orbit's 2-μm double-poly CMOS process available through MOSIS. For each circuit, I show data both with and without cascode transistors. I nulled the initial charge imbalances on the floating gates by shorting together all of the chip's pins and exposing the chip to short-wave UV light for about 20 minutes; in this situation, any floating-gate charge imbalances set up potential differences that drive currents carried by photoexcited electrons through the oxide, so that the charge imbalances are reduced. The *p*FGMOS transistors that I used to breadboard the circuits for this section are identical to those from which I obtained the experimental data shown in Section 5.3. Consequently, for the noncascoded versions of each circuit, I use the ratio of the values of V_A and V_C from Figures 5.4 and 5.5 (which is equal to about 6%) to estimate the perturbations in the power laws resulting from the drain-overlap capacitance, $C_{fg\text{-}d}$.

6.2.1. Two-Input Geometric-Mean Circuit

Consider the circuit shown in Figure 6.1. It consists of three two-input *p*FGMOS transistors operating in their subthreshold regions. Because the *p*FGMOS transistors that I used to breadboard this MITE network each have four nominally identical control gates, each control gate shown in Figure 6.1 actually comprises two of four control gates. Taking the elements of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to be integer numbers of nominally identical control gates and neglecting the drain-overlap capacitance, for the circuit shown in Figure 6.1, I write that

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix} 2 & 2 \end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

and

$$\begin{aligned} \Lambda &= \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} \\ &= \begin{bmatrix} 2 & 2 \end{bmatrix}\begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

Consequently, the MITE network of Figure 6.1 ideally embodies the expression

$$I_3 = K\sqrt{I_1 I_2},$$

where

$$K = \exp\left[\frac{1}{\mathrm{Q_T}}\left(Q_3 - \tfrac{1}{2}Q_1 - \tfrac{1}{2}Q_2\right)\right].$$

Because I balanced the floating-gate charge by short-wave UV light exposure, I expect that $Q_1 = Q_2 = Q_3 = Q$, which implies that

$$\begin{aligned} K &= \exp\left[\frac{Q}{\mathrm{Q_T}}\left(1 - \tfrac{1}{2} - \tfrac{1}{2}\right)\right] \\ &= \exp[0] \\ &= 1, \end{aligned}$$

and that

$$I_3 = \sqrt{I_1 I_2}. \tag{6.5}$$

Thus, the circuit of Figure 6.1 is a two-input geometric-mean circuit.

To account for the effect of the parasitic drain-overlap capacitance on the power laws embodied in the circuit of Figure 6.1, using Equation 6.4, I write that

$$\Lambda' \approx \Lambda\left(\mathbf{I} - \varepsilon\mathbf{W}_{\text{in}}^{-1}\right)$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \varepsilon \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \right)$$
$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1-\frac{\varepsilon}{4} & 0 \\ 0 & 1-\frac{\varepsilon}{4} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{2}\left(1-\frac{\varepsilon}{4}\right) & \frac{1}{2}\left(1-\frac{\varepsilon}{4}\right) \end{bmatrix},$$

which, for ε=0.06, is equal to $\begin{bmatrix} 0.4925 & 0.4925 \end{bmatrix}$.

Figure 6.2 shows measured data from the circuit of Figure 6.1. Because the drain-overlap capacitance causes the output current to depend exponentially on the output voltage, I adjusted the value of the output voltage, V_3, such that M_3 sourced about 3.14 nA of current when I_1 and I_2 were both set to 3.14 nA. The circles shown in Figure 6.2a represent measured values of I_3 plotted as a function of I_1 over the four-decade current range from 30 pA to 300 nA for five different values of I_2 ranging from 63.2 pA to 157 nA. The circles shown in Figure 6.2b represent measured values of I_3 plotted as a function of I_2 over the four-decade current range from 30 pA to 300 nA for five different values of I_1 ranging from 63.2 pA to 157 nA. In both plots, solid lines show values of Equation 6.5 calculated for the values of I_1 and I_2 at each point, and dashed lines show fits with the power laws adjusted from 0.5 to 0.4925 to account for the drain-overlap capacitance. The data and fits agree well over much of the current range shown. Deviations at high current levels result from one or more of the *p*FGMOS transistors going above threshold. Deviations at low current levels result from leakage currents.

Figure 6.3 shows the two-input geometric-mean circuit made from cascoded subthreshold *p*FGMOS transistors. The cascode transistors mitigate the effects of the drain-overlap capacitance nearly completely. Figure 6.4 shows measured data from the circuit of Figure 6.3. To obtain these data, I set the cascode bias voltage, V_{cas}, to 1.1 volts below V_{DD}, and I set the output voltage, V_3, to 5 volts below V_{DD}. The circles shown in Figure 6.4a represent measured values of I_3 plotted as a function of I_1 over the four-decade current range from 100 pA to 1 μA for five different values of I_2 ranging from 207 pA to 519 nA. The circles shown in Figure 6.4b represent measured values of I_3 plotted as a function of I_2 over the four-decade current range from 100 pA to 1 μA for five different values of I_1 ranging from 207 pA to 519 nA. In both plots, solid lines show values of Equation 6.5 calculated from the values of I_1 and I_2 at each point. The data and fits agree well over much of the current range shown. Deviations at high current levels result from one or more of the *p*FGMOS transistors going above threshold. Deviations at low current levels result either from leakage currents or from one of the cascode transistors going out of

saturation.

6.2.2. Squaring-Reciprocal Circuit

Consider the circuit shown in Figure 6.5. It consists of three two-input *p*FGMOS transistors operating in their subthreshold regions. Here too, each control gate shown in Figure 6.5 actually comprises two of four control gates. Taking the elements of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to be integer numbers of nominally identical control gates and neglecting the drain-overlap capacitance, for the circuit shown in Figure 6.5, I write that

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} 2 & 2 \\ 0 & 4 \end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix} 4 & 0 \end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ 0 & \frac{1}{4} \end{bmatrix}$$

and

$$\begin{aligned} \Lambda &= \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} \\ &= \begin{bmatrix} 4 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ 0 & \frac{1}{4} \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 \end{bmatrix}. \end{aligned}$$

Consequently, the MITE network of Figure 6.5 ideally embodies the expression

$$I_3 = K\frac{I_1^2}{I_2},$$

where

$$K = \exp\left[\frac{1}{\text{Q}_\text{T}}\left(Q_3 - 2Q_1 + Q_2\right)\right].$$

Because I balanced the floating-gate charge by short-wave UV light exposure, I expect that $Q_1 = Q_2 = Q_3 = Q$, which implies that

$$\begin{aligned} K &= \exp\left[\frac{Q}{\text{Q}_\text{T}}(1-2+1)\right] \\ &= \exp[0] \\ &= 1, \end{aligned}$$

and that

$$I_3 = \frac{I_1^2}{I_2}. \tag{6.6}$$

Thus, the circuit of Figure 6.5 is a squaring-reciprocal circuit.

To account for the effect of the parasitic drain-overlap capacitance on the power

laws embodied in the circuit of Figure 6.5, using Equation 6.4, I write that

$$\begin{aligned}\Lambda' &\approx \Lambda\left(\mathbf{I}-\varepsilon\mathbf{W}_{\text{in}}^{-1}\right)\\ &= \begin{bmatrix}2 & -1\end{bmatrix}\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}-\varepsilon\begin{bmatrix}\frac{1}{2} & -\frac{1}{4}\\ 0 & \frac{1}{4}\end{bmatrix}\right)\\ &= \begin{bmatrix}2 & -1\end{bmatrix}\begin{bmatrix}1-\frac{\varepsilon}{2} & \frac{\varepsilon}{4}\\ 0 & 1-\frac{\varepsilon}{4}\end{bmatrix}\\ &= \begin{bmatrix}2\left(1-\frac{\varepsilon}{2}\right) & -1\left(1-\frac{3}{4}\varepsilon\right)\end{bmatrix},\end{aligned}$$

which, for ε=0.06, is equal to $\begin{bmatrix}1.94 & -0.955\end{bmatrix}$.

Figure 6.6 shows measured data from the circuit of Figure 6.5. Because the drain-overlap capacitance causes the output current to depend exponentially on the output voltage, I again adjusted the value of the output voltage, V_3, such that M_3 sourced about 3.14 nA of current when I_1 and I_2 were both set to 3.14 nA. The circles shown in Figure 6.6a represent measured values of I_3 plotted as a function of I_1 over the four-decade current range from 30 pA to 300 nA for five different values of I_2 ranging from 63.2 pA to 157 nA. The circles shown in Figure 6.6b represent measured values of I_3 plotted as a function of I_2 over the four-decade current range from 30 pA to 300 nA for five different values of I_1 ranging from 63.2 pA to 157 nA. In both plots, solid lines show values of Equation 6.6 calculated for the values of I_1 and I_2 at each point, and dashed lines show fits with the power laws adjusted from $\begin{bmatrix}2 & -1\end{bmatrix}$ to $\begin{bmatrix}1.94 & -0.955\end{bmatrix}$ to account for the drain-overlap capacitance. The data and fits agree reasonably well over much of the current range shown. Deviations at high current levels result from one or more of the *p*FGMOS transistors going above threshold. Deviations at low current levels result from leakage currents.

Figure 6.7 shows the squaring-reciprocal circuit made from cascoded subthreshold *p*FGMOS transistors. The cascode transistors mitigate the effects of the drain-overlap capacitance nearly completely. Figure 6.8 shows measured data from the circuit of Figure 6.7. To obtain these data, I set the cascode bias voltage, V_{cas}, to 1.1 volts below V_{DD}, and I set the output voltage, V_3, to 5 volts below V_{DD}. The circles shown in Figure 6.8a represent measured values of I_3 plotted as a function of I_1 over the four-decade current range from 100 pA to 1 μA for five different values of I_2 ranging from 207 pA to 519 nA. The circles shown in Figure 6.8b represent measured values of I_3 plotted as a function of I_2 over the four-decade current range from 100 pA to 1 μA for five different values of I_1 ranging from 207 pA to 519 nA. In both plots, solid lines show values of Equation 6.6 calculated for the values of I_1 and I_2 at each point. The data and fits agree well over much of the current range shown. Deviations at high current levels result from one or more of the

*p*FGMOS transistors going above threshold. Deviations at low current levels result either from leakage currents or from one of the cascode transistors going out of saturation.

6.2.3. Power-Law Circuits

Consider the circuit shown in Figure 6.9. It consists of three four-input *p*FGMOS transistors operating in their subthreshold regions. Taking the elements of $\mathbf{W}_{in}$ and $\mathbf{W}_{out}$ to be integer numbers of nominally identical control gates and neglecting the drain-overlap capacitance, for the circuit shown in Figure 6.9, I write that

$$\mathbf{W}_{in} = \begin{bmatrix} 2 & 2 \\ 0 & 4 \end{bmatrix} \text{ and } \mathbf{W}_{out} = \begin{bmatrix} 3 & 1 \end{bmatrix},$$

which imply that

$$\mathbf{W}_{in}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ 0 & \frac{1}{4} \end{bmatrix}$$

and

$$\begin{aligned} \Lambda &= \mathbf{W}_{out}\mathbf{W}_{in}^{-1} \\ &= \begin{bmatrix} 3 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ 0 & \frac{1}{4} \end{bmatrix} \\ &= \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \end{bmatrix}. \end{aligned}$$

Consequently, the MITE network of Figure 6.9 ideally embodies the expression

$$I_3 = K I_1^{\frac{3}{2}} I_2^{\frac{1}{2}},$$

where

$$K = \exp\left[\frac{1}{\mathrm{Q_T}}\left(Q_3 - \tfrac{3}{2}Q_1 + \tfrac{1}{2}Q_2\right)\right].$$

Because I balanced the floating-gate charge by short-wave UV light exposure, I expect that $Q_1 = Q_2 = Q_3 = Q$, which implies that

$$\begin{aligned} K &= \exp\left[\frac{Q}{\mathrm{Q_T}}\left(1 - \tfrac{3}{2} + \tfrac{1}{2}\right)\right] \\ &= \exp[0] \\ &= 1, \end{aligned}$$

and that

$$I_3 = I_1^{\frac{3}{2}} I_2^{\frac{1}{2}}.$$

Thus, the circuit of Figure 6.9 is a $\frac{3}{2}$-power-law circuit.

To account for the effect of the parasitic drain-overlap capacitance on the power laws embodied in the circuit of Figure 6.9, using Equation 6.4, I write that

$$\begin{aligned}\Lambda' &\approx \Lambda\left(\mathbf{I}-\varepsilon\mathbf{W}_{\text{in}}^{-1}\right)\\ &= \left[\tfrac{3}{2} \quad -\tfrac{1}{2}\right]\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}-\varepsilon\begin{bmatrix}\frac{1}{2} & -\frac{1}{4}\\ 0 & \frac{1}{4}\end{bmatrix}\right)\\ &= \left[\tfrac{3}{2} \quad -\tfrac{1}{2}\right]\begin{bmatrix}1-\frac{\varepsilon}{2} & \frac{\varepsilon}{4}\\ 0 & 1-\frac{\varepsilon}{4}\end{bmatrix}\\ &= \left[\tfrac{3}{2}\left(1-\tfrac{\varepsilon}{2}\right) \quad -\tfrac{1}{2}(1-\varepsilon)\right],\end{aligned}$$

which, for ε=0.06, is equal to $\begin{bmatrix}1.455 & -0.47\end{bmatrix}$.

In general, I can make a $\frac{p}{q}$-power-law circuit similar to the $\frac{3}{2}$-power-law circuit shown in Figure 6.9 by diode connecting M_1 through q control gates and by connecting p control gates of M_3 to V_1 and by connecting any remaining control gates to V_2. I denote this circuit configuration by the ratio p:q. Taking the elements of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to be integer numbers of nominally identical control gates and neglecting the drain-overlap capacitance, for the p:q circuit configuration, I write that

$$\mathbf{W}_{\text{in}} = \begin{bmatrix}q & 4-q\\ 0 & 4\end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix}p & 4-p\end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix}\frac{1}{q} & \frac{1}{4}-\frac{1}{q}\\ 0 & \frac{1}{4}\end{bmatrix}$$

and

$$\begin{aligned}\Lambda &= \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}\\ &= \begin{bmatrix}p & 4-p\end{bmatrix}\begin{bmatrix}\frac{1}{q} & \frac{1}{4}-\frac{1}{q}\\ 0 & \frac{1}{4}\end{bmatrix}\\ &= \left[\tfrac{p}{q} \quad 1-\tfrac{p}{q}\right].\end{aligned}$$

Consequently, the p:q MITE-network configuration ideally embodies the expression

$$I_3 = KI_1^{\frac{p}{q}}I_2^{1-\frac{p}{q}},$$

where

$$K = \exp\left[\frac{1}{\mathrm{Q_T}}\left(Q_3-\tfrac{p}{q}Q_1-\left(1-\tfrac{p}{q}\right)Q_2\right)\right].$$

Because I balanced the floating-gate charge by short-wave UV light exposure, I expect that $Q_1 = Q_2 = Q_3 = Q$, which implies that

$$\begin{aligned}K &= \exp\left[\frac{Q}{\mathrm{Q_T}}\left(1-\tfrac{p}{q}-\left(1-\tfrac{p}{q}\right)\right)\right]\\ &= \exp[0]\\ &= 1,\end{aligned}$$

and that

$$I_3 = I_1^{\frac{p}{q}} I_2^{1-\frac{p}{q}}. \tag{6.7}$$

Thus, the p:q MITE-network configuration is a $\frac{p}{q}$-power-law circuit.

To account for the effect of the parasitic drain-overlap capacitance on the power laws embodied in the p:q MITE-network configuration, using Equation 6.4, I write that

$$\begin{aligned}
\Lambda' &\approx \Lambda\left(\mathbf{I} - \varepsilon \mathbf{W}_{\text{in}}^{-1}\right) \\
&= \begin{bmatrix} \frac{p}{q} & 1-\frac{p}{q} \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \varepsilon \begin{bmatrix} \frac{1}{q} & \frac{1}{4}-\frac{1}{q} \\ 0 & \frac{1}{4} \end{bmatrix} \right) \\
&= \begin{bmatrix} \frac{p}{q} & 1-\frac{p}{q} \end{bmatrix} \begin{bmatrix} 1-\frac{\varepsilon}{q} & \frac{\varepsilon}{q}-\frac{\varepsilon}{4} \\ 0 & 1-\frac{\varepsilon}{4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{p}{q}\left(1-\frac{\varepsilon}{q}\right) & 1-\frac{p}{q}+\frac{p}{q}\frac{\varepsilon}{q}-\frac{\varepsilon}{4} \end{bmatrix}.
\end{aligned}$$

Figure 6.10 shows measured data from nine different p:q configurations of the circuit of Figure 6.9. Because the drain-overlap capacitance causes the output current to depend exponentially on the output voltage, for each circuit configuration, I adjusted the value of the output voltage, V_3, such that M_3 sourced about 3.14 nA of current when I_1 and I_2 were both set to 3.14 nA. The circles shown in Figure 6.10a represent measured values of I_3 plotted as a function of I_1 over the four-decade current range from 30 pA to 300 nA for I_2 set equal to 3.14 nA for each of the nine different circuit configurations. The circles shown in Figure 6.10b represent measured values of I_3 plotted as a function of I_2 over the four-decade current range from 30 pA to 300 nA for I_1 set equal to 3.14 nA for each of the nine different circuit configurations. In both plots, solid lines show values of Equation 6.7 calculated using the values of p and q indicated by the ratio labeling each curve for the values of I_1 and I_2 at each point, and dashed lines show fits with the power laws adjusted from $\left[\frac{p}{q} \quad 1-\frac{p}{q}\right]$ to $\left[\frac{p}{q}\left(1-\frac{\varepsilon}{q}\right) \quad 1-\frac{p}{q}+\frac{p}{q}\frac{\varepsilon}{q}-\frac{\varepsilon}{4}\right]$ with ε=0.06 to account for the drain-overlap capacitance. The data and fits agree reasonably well over much of the current range shown. Deviations at high current levels result from one or more of the pFGMOS transistors going above threshold. Deviations at low current levels result from leakage currents.

Figure 6.11 shows the $\frac{3}{2}$-power-law circuit made from cascoded subthreshold pFGMOS transistors. The cascode transistors mitigate the effects of the drain-overlap capacitance nearly completely. Figure 6.12 shows measured data from nine different p:q configurations of the circuit of Figure 6.11. To obtain these data, I set the cascode bias voltage, V_{cas}, to 1.1 volts below V_{DD}, and I set the output voltage, V_3, to 5 volts below V_{DD}. The circles shown in Figure 6.12a represent measured values of I_3 plotted as a

function of I_1 over the four-decade current range from 100 pA to 1 μA for I_2 set equal to 10.3 nA for each of the nine different circuit configurations. The circles shown in Figure 6.12b represent measured values of I_3 plotted as a function of I_2 over the four-decade current range from 100 pA to 1 μA for I_1 set equal to 10.3 nA for each of the nine different circuit configurations. In both plots, solid lines show values of Equation 6.7 calculated using the values of *p* and *q* indicated by the ratio labeling each curve for the values of I_1 and I_2 at each point. The data and fits agree well over much of the current range shown. Deviations at high current levels result from one or more of the *p*FGMOS transistors going above threshold. Deviations at low current levels result either from leakage currents or from one of the cascode transistors going out of saturation.

6.2.4. Product-Reciprocal Circuit

Consider the circuit shown in Figure 6.13. It consists of four two-input *p*FGMOS transistors operating in their subthreshold regions. Here again, each control gate shown in Figure 6.13 actually comprises two of four control gates. Taking the elements of $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{out}}$ to be integer numbers of nominally identical control gates and neglecting the drain-overlap capacitance, for the circuit shown in Figure 6.13, I write that

$$\mathbf{W}_{\text{in}} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } \mathbf{W}_{\text{out}} = \begin{bmatrix} 2 & 2 & 0 \end{bmatrix},$$

which imply that

$$\mathbf{W}_{\text{in}}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$$

and

$$\begin{aligned} \Lambda &= \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1} \\ &= \begin{bmatrix} 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & \frac{1}{4} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}. \end{aligned}$$

Consequently, the MITE network of Figure 6.13 ideally embodies the expression

$$I_4 = K\frac{I_1 I_2}{I_3},$$

where

$$K = \exp\left[\frac{1}{\mathrm{Q_T}}\left(Q_4 - Q_1 - Q_2 + Q_3\right)\right].$$

Because I balanced the floating-gate charge by short-wave UV light exposure, I expect that $Q_1 = Q_2 = Q_3 = Q_4 = Q$, which implies that

$$\begin{aligned} K &= \exp\left[\frac{Q}{\mathrm{Q_T}}(1-1-1+1)\right] \\ &= \exp[0] \\ &= 1, \end{aligned}$$

and that

$$I_4 = \frac{I_1 I_2}{I_3}. \tag{6.8}$$

Thus, the circuit of Figure 6.13 is a product-reciprocal circuit.

To account for the effect of the parasitic drain-overlap capacitance on the power laws embodied in the circuit of Figure 6.13, using Equation 6.4, I write that

$$\begin{aligned} \Lambda' &\approx \Lambda\left(\mathbf{I} - \varepsilon\mathbf{W}_{\mathrm{in}}^{-1}\right) \\ &= \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \varepsilon\begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & \frac{1}{4} \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1-\frac{\varepsilon}{2} & 0 & \frac{\varepsilon}{4} \\ 0 & 1-\frac{\varepsilon}{2} & \frac{\varepsilon}{4} \\ 0 & 0 & 1-\frac{\varepsilon}{4} \end{bmatrix} \\ &= \begin{bmatrix} 1\left(1-\frac{\varepsilon}{2}\right) & 1\left(1-\frac{\varepsilon}{2}\right) & -1\left(1-\frac{3}{4}\varepsilon\right) \end{bmatrix}, \end{aligned}$$

which, for ε=0.06, is equal to $\begin{bmatrix} 0.97 & 0.97 & -0.955 \end{bmatrix}$.

Figures 6.14 and 6.15 show measured data from the circuit of Figure 6.13. Because the drain-overlap capacitance causes the output current to depend exponentially on the output voltage, I adjusted the value of the output voltage, V_4, such that M_4 sourced about 3.20 nA of current when I_1 and I_2 were both set to 3.20 nA. The circles shown in Figure 6.14a represent measured values of I_4 plotted as a function of I_1 over the four-decade current range from 30 pA to 300 nA for five different values of I_2 ranging from 65.3 pA to 161 nA and for I_3 set equal to 3.20 nA. The circles shown in Figure 6.14b represent measured values of I_4 plotted as a function of I_2 over the four-decade current range from 30 pA to 300 nA for five different values of I_1 ranging from 65.3 pA to 161 nA and for I_3 set equal to 3.20 nA. The circles shown in Figure 6.15 represent measured values of I_4 plotted as a function of I_3 over the four-decade current range from 30 pA to

300 nA for five different values of I_1 ranging from 65.3 pA to 161 nA and for I_2 set equal to 3.20 nA. In all three plots, solid lines show values of Equation 6.8 calculated for the values of I_1, I_2, and I_3 at each point, and dashed lines show fits with the power laws adjusted from $[1 \quad 1 \quad -1]$ to $[0.97 \quad 0.97 \quad -0.955]$ to account for the drain-overlap capacitance. The data and fits agree reasonably well over much of the current range shown. Deviations at high current levels result from one or more of the *p*FGMOS transistors going above threshold. Deviations at low current levels result from leakage currents.

Figure 6.16 shows the product-reciprocal circuit made from cascoded subthreshold *p*FGMOS transistors. The cascode transistors mitigate the effects of the drain-overlap capacitance nearly completely. Figures 6.17 and 6.18 show measured data from the circuit of Figure 6.16. To obtain these data, I set the cascode bias voltage, V_{cas}, to 1.1 volts below V_{DD}, and I set the output voltage, V_3, to 5 volts below V_{DD}. The circles shown in Figure 6.17a represent measured values of I_4 plotted as a function of I_1 over the four-decade current range from 100 pA to 1 μA for five different values of I_2 ranging from 204 pA to 504 nA and for I_3 set equal to 10.0 nA. The circles shown in Figure 6.17b represent measured values of I_4 plotted as a function of I_2 over the four-decade current range from 100 pA to 1 μA for five different values of I_1 ranging from 204 pA to 504 nA and for I_3 set equal to 10.0 nA. The circles shown in Figure 6.18 represent measured values of I_4 plotted as a function of I_3 over the four-decade current range from 100 pA to 1 μA for five different values of I_1 ranging from 204 pA to 504 nA and for I_2 set equal to 10.0 nA. In all three plots, solid lines show values of Equation 6.8 calculated for the values of I_1, I_2, and I_3 at each point. The data and fits agree well over much of the current range shown. Deviations at high current levels result from one or more of the *p*FGMOS transistors going above threshold. Deviations at low current levels result either from leakage currents or from one of the cascode transistors going out of saturation.

6.3. Appendix 6.A

In this appendix, I calculate to first order the perturbations in the power laws embodied in a MITE network comprising subthreshold FGMOS transistors resulting from the presence of the parasitic drain-overlap capacitance, C_{fg-d}, in each of the input FGMOS transistors. For each of the input FGMOS transistors, the drain-overlap capacitance is effectively a small amount of extra self-coupling in parallel with all the other self-coupling weighing coefficients that we, as designers, specified in the input connectivity matrix. So, I can account for the presence of the overlap capacitance by adding a small quantity, ε, to each

of the diagonal elements of the input connectivity matrix, $\mathbf{W}_{\text{in}}$. This small quantity is a measure of the size of $C_{\text{fg-d}}$ in the same units in which the other weighting coefficients are expressed; for example, if the elements of $\mathbf{W}_{\text{in}}$ are integral numbers of unit capacitors, each with nominal value C, then ε is given by

$$\varepsilon = \frac{C_{\text{fg-d}}}{C}.$$

Consequently, I begin by writing a perturbed input connectivity matrix, $\mathbf{W}'_{\text{in}}$, as

$$\mathbf{W}'_{\text{in}} = \mathbf{W}_{\text{in}} + \varepsilon\mathbf{I},$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. Now, I compute the inverse of the perturbed input connectivity matrix as

$$\begin{aligned}
\mathbf{W}'^{-1}_{\text{in}} &= \left(\mathbf{W}_{\text{in}} + \varepsilon\mathbf{I}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\mathbf{W}_{\text{in}}\left(\mathbf{W}_{\text{in}} + \varepsilon\mathbf{I}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{W}_{\text{in}}\left(\mathbf{W}_{\text{in}} + \varepsilon\mathbf{I}\right)^{-1}\right) \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\left(\mathbf{W}_{\text{in}} + \varepsilon\mathbf{I}\right)\mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\
&= \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} + \varepsilon\mathbf{W}_{\text{in}}^{-1}\right)^{-1} \\
&\approx \mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \varepsilon\mathbf{W}_{\text{in}}^{-1}\right),
\end{aligned}$$

neglecting terms of second order and higher. So, to obtain the perturbed power laws, I write

$$\begin{aligned}
\Lambda' &= \mathbf{W}_{\text{out}}\mathbf{W}'^{-1}_{\text{in}} \\
&\approx \mathbf{W}_{\text{out}}\mathbf{W}_{\text{in}}^{-1}\left(\mathbf{I} - \varepsilon\mathbf{W}_{\text{in}}^{-1}\right) \\
&= \Lambda\left(\mathbf{I} - \varepsilon\mathbf{W}_{\text{in}}^{-1}\right),
\end{aligned}$$

which is what I set out to compute.

6.4. References

1. A. G. Andreou and K. A. Boahen, "Translinear Circuits in Subthreshold MOS," *Journal of Analog Integrated Circuits and SIgnal Processing*, vol. 9, no. 2, pp. 141–166, 1996.

2. M. Lenzlinger and E. H. Snow, "Fowler–Nordheim Tunneling into Thermally Grown SiO_2," *Journal of Applied Physics*, vol. 40, no. 1, pp. 278–283, 1969.

3. P. Hasler, *Foundations of Learning in Analog VLSI*, Ph.D. Thesis, Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA, 1997.

4. C. Diorio, *Neurally Inspired Silicon Learning: From Synapse Transistors to Learning Arrays*, Ph.D. Thesis, Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 1997.

5. D. A. Kerns, J. E. Tanner, M. A. Sivilotti, and J. Luo, "CMOS UV-Writable Non-Volatile Analog Storage," in C. Squin, ed., *Advanced Research in VLSI: Proceedings of the UC Santa Cruz Conference 1991*, Cambridge, MA: MIT Press, pp. 245–261, 1991.

6. R. G. Benson and D. A. Kerns, "UV-Activated Conductances Allow for Multiple Time-Scale Learning," *IEEE Transactions on Neural Networks*, vol. 4, no. 3, pp. 434–440, 1993.

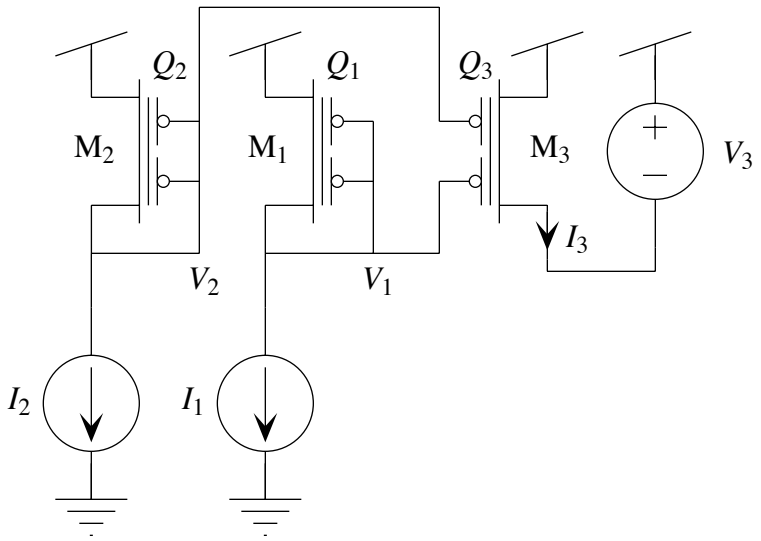


Figure 6.1. A two-input geometric-mean circuit comprising three two-input subthreshold *p*FGMOS transistors. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. The output voltage, V_3, was set such that M_3 sourced about 3.14 nA of current when I_1 and I_2 were set to 3.14 nA.

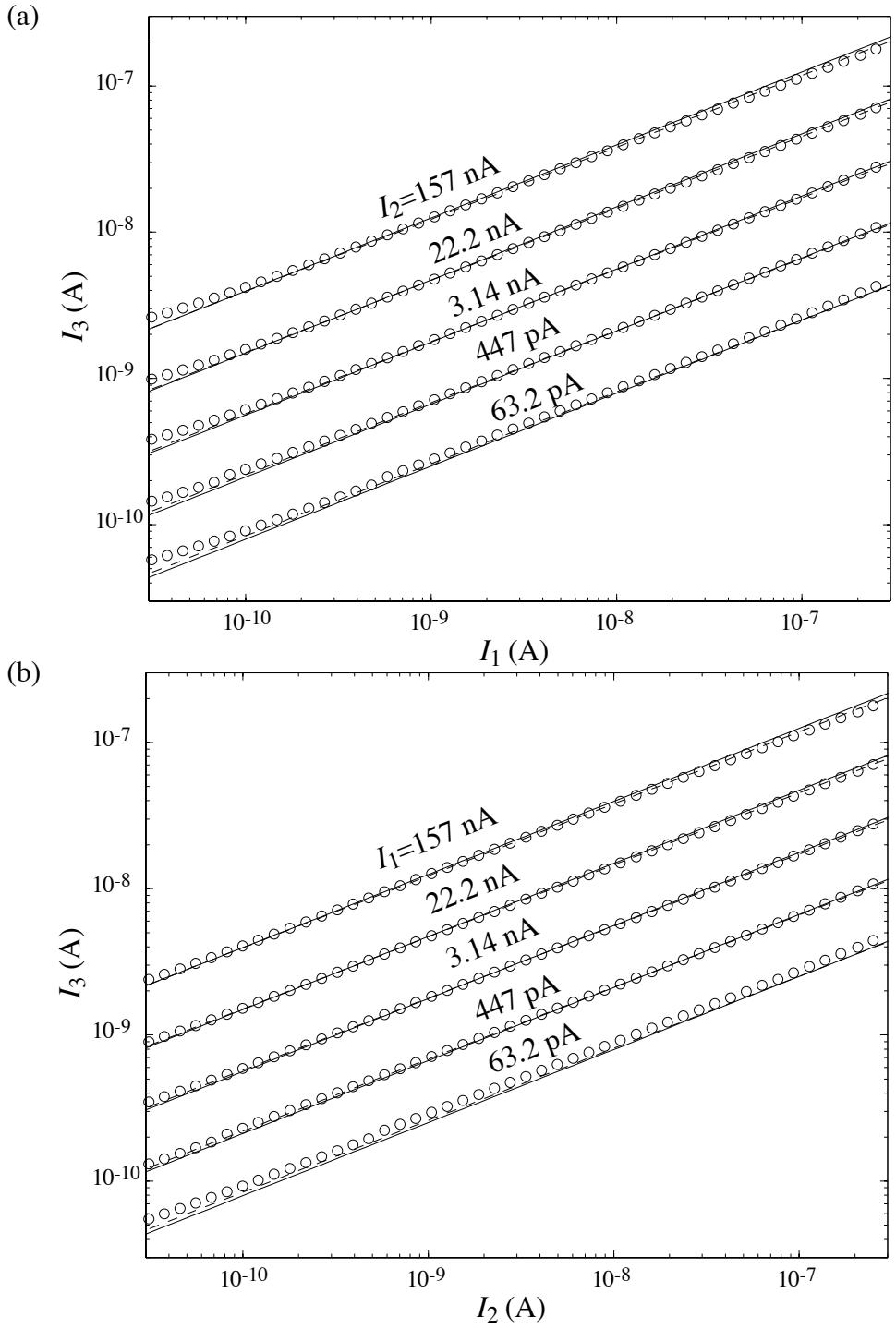


Figure 6.2. Measured data from the circuit of Figure 6.1. Circles are measured values of I_3 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1. Solid lines show the ideal expression, $I_3 = \sqrt{I_1 I_2}$, calculated for the values of I_1 and I_2 at each point. Dashed lines show fits adjusted to account for $C_{\text{fg-d}}$.

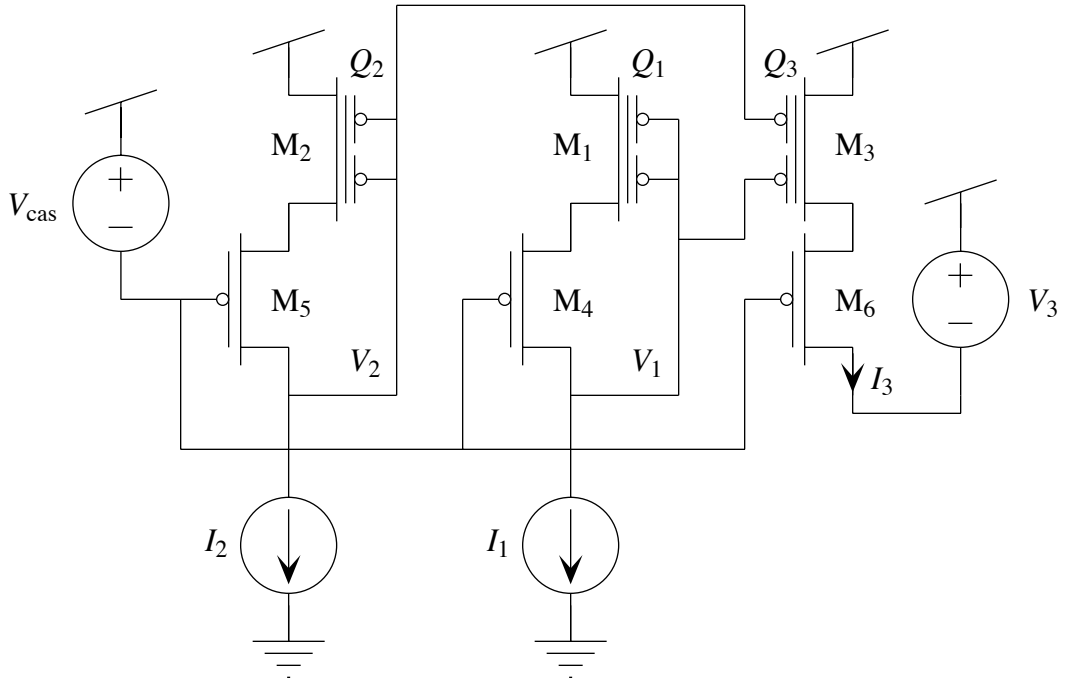


Figure 6.3. A two-input geometric-mean circuit comprising three cascoded two-input subthreshold *p*FGMOS transistors. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. The output voltage, V_3, was set to 5 volts below V_{DD}. The cascode bias voltage, V_{cas}, was set to 1.1 volts below V_{DD}.

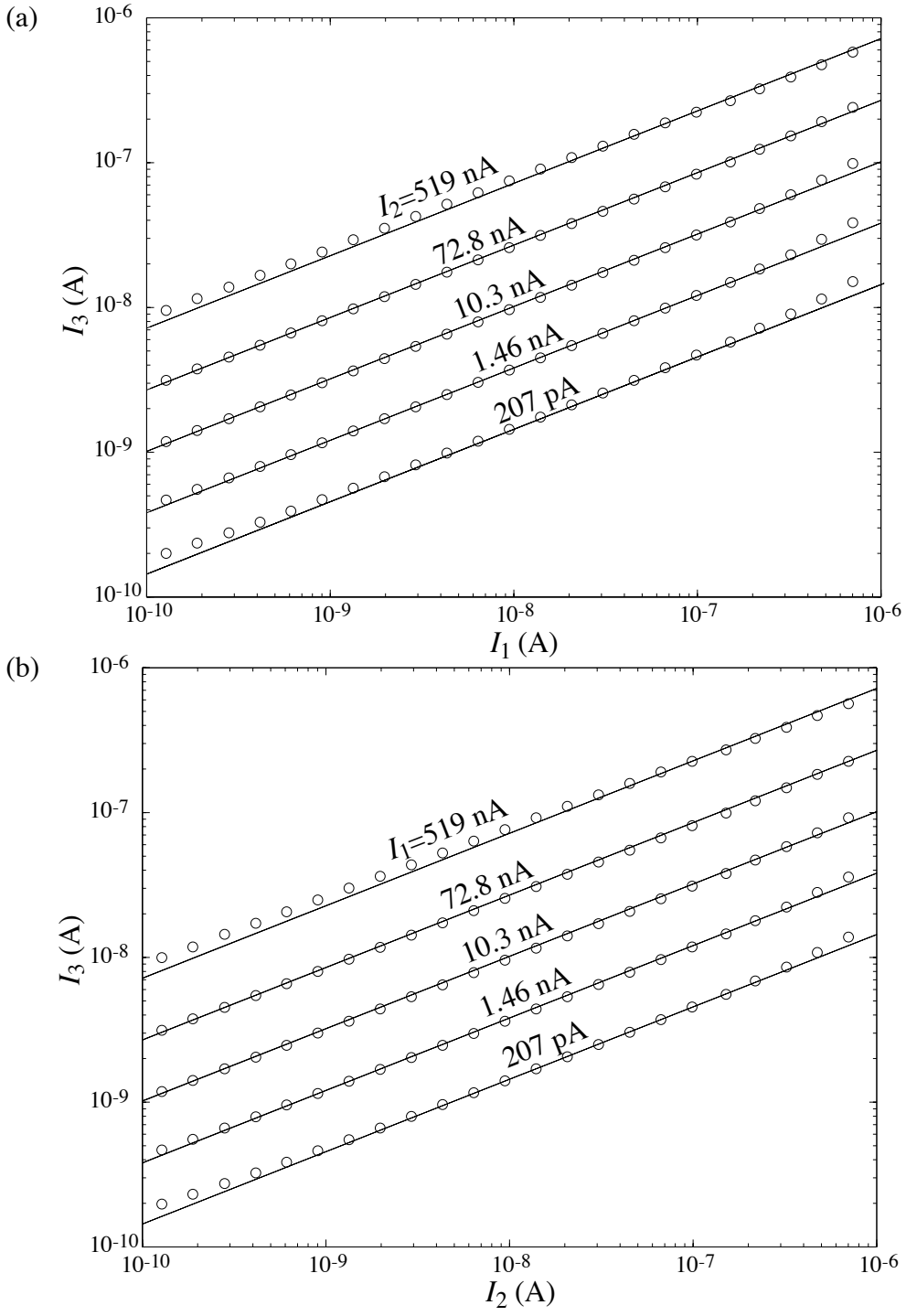


Figure 6.4. Measured data from the circuit of Figure 6.3. Circles are measured values of I_3 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1. Solid lines show the ideal expression, $I_3 = \sqrt{I_1 I_2}$, calculated for the values of I_1 and I_2 at each point.

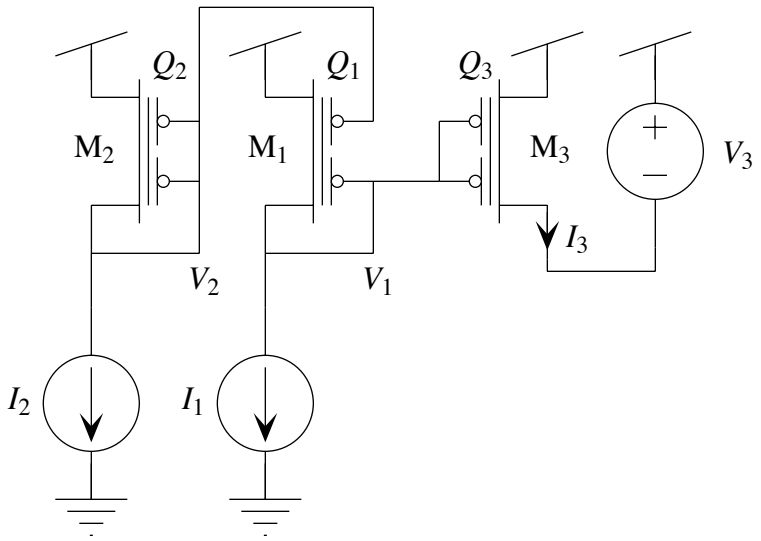


Figure 6.5. A squaring-reciprocal circuit comprising three two-input subthreshold *p*FGMOS transistors. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. The output voltage, V_3, was set such that M_3 sourced about 3.14 nA of current when I_1 and I_2 were set to 3.14 nA.

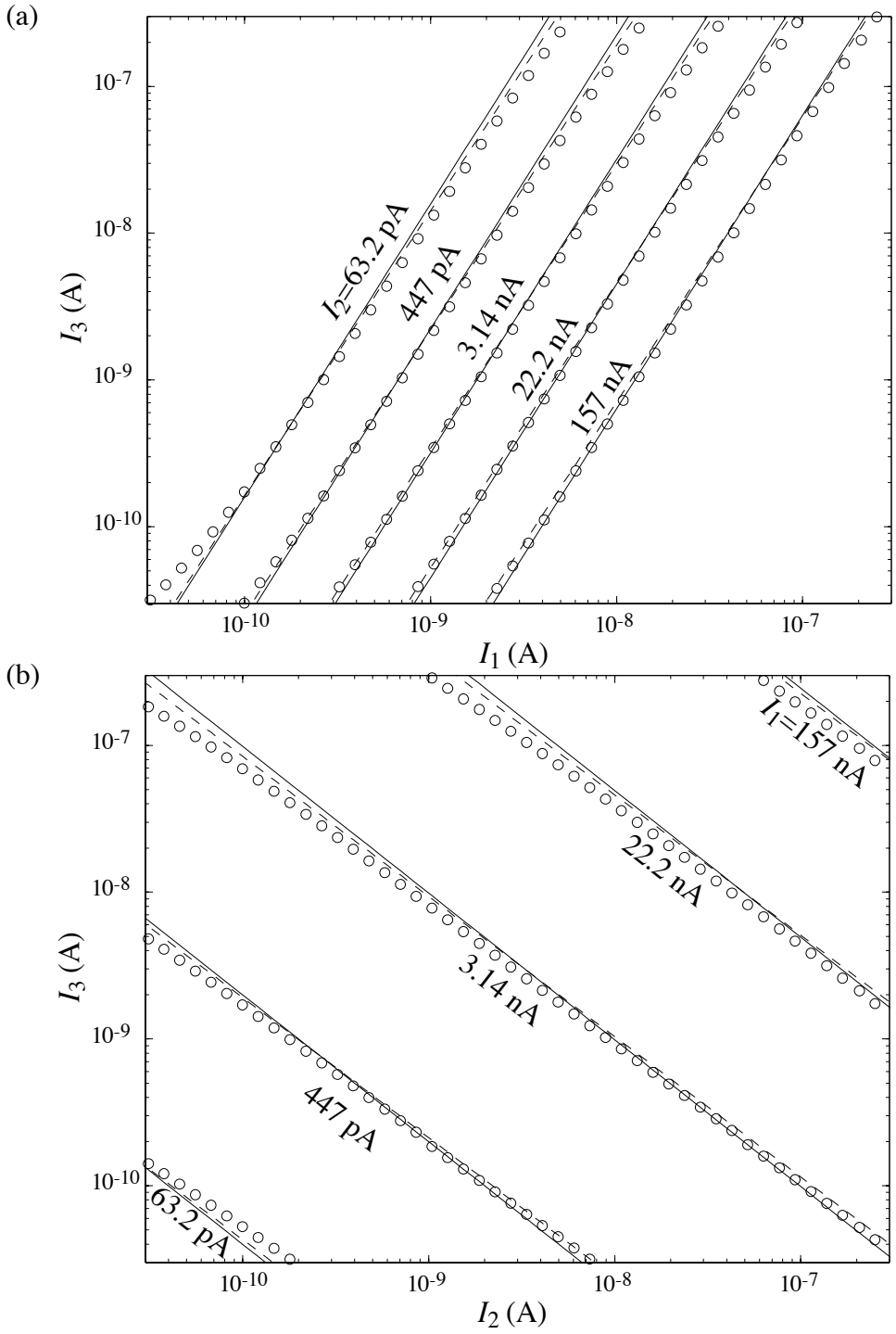


Figure 6.6. Measured data from the circuit of Figure 6.5. Circles are measured values of I_3 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1. Solid lines show the ideal expression, $I_3 = I_1^2 \div I_2$, calculated for the values of I_1 and I_2 at each point. Dashed lines show fits adjusted to account for $C_{\text{fg-d}}$.

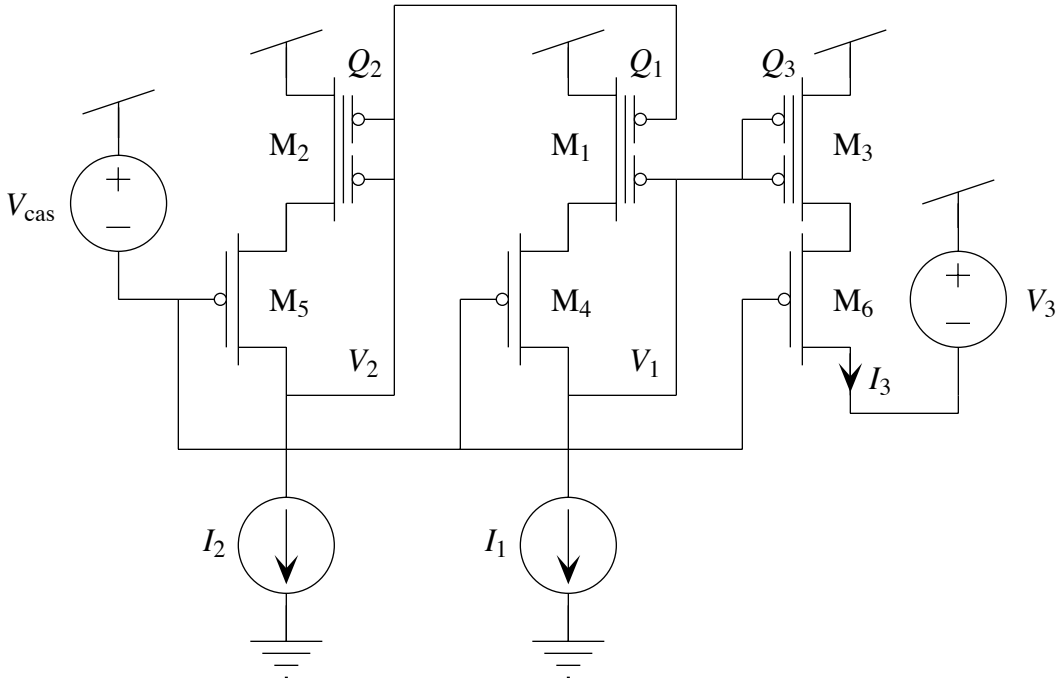


Figure 6.7. A squaring-reciprocal circuit comprising three cascoded two-input subthreshold *p*FGMOS transistors. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. The output voltage, V_3, was set to 5 volts below V_{DD}. The cascode bias voltage, V_{cas}, was set to 1.1 volts below V_{DD}.

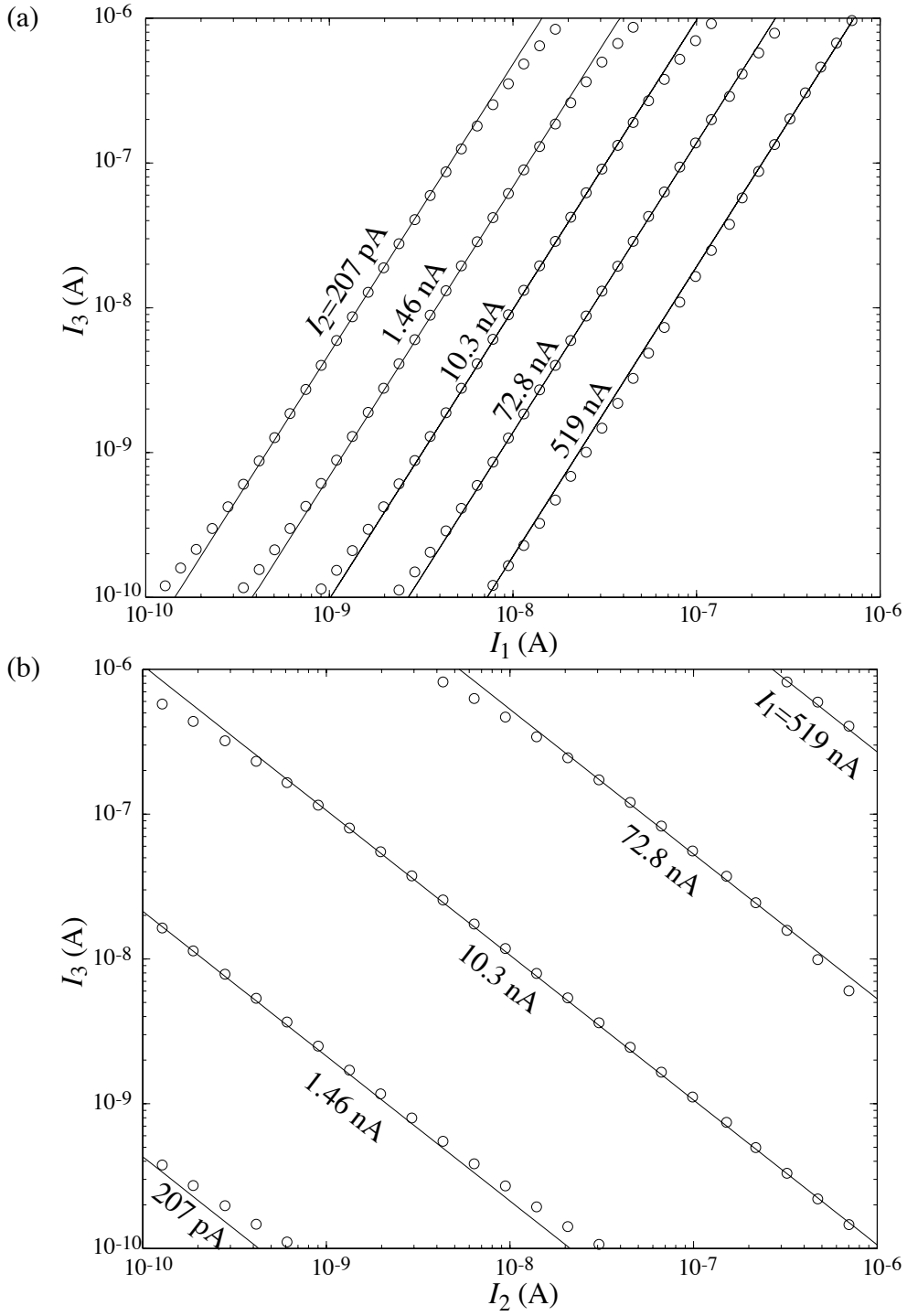


Figure 6.8. Measured data from the circuit of Figure 6.7. Circles are measured values of I_3 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1. Solid lines show the ideal expression, $I_3 = I_1^2 \div I_2$, calculated for the values of I_1 and I_2 at each point.

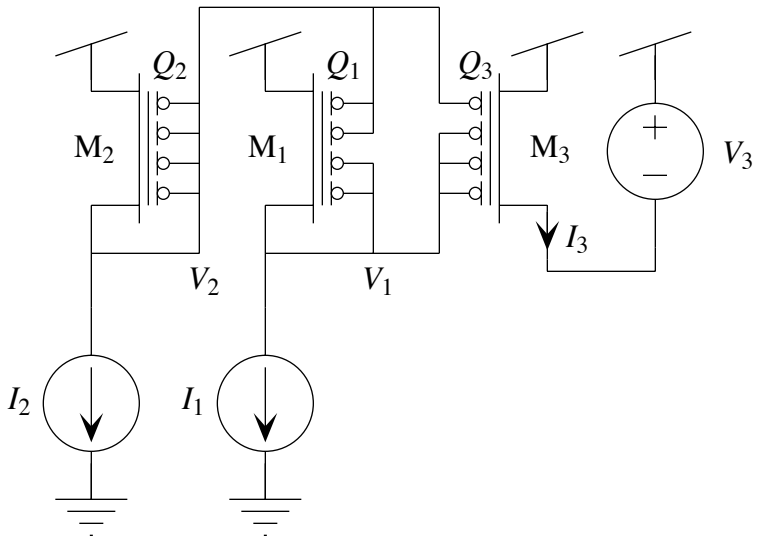


Figure 6.9. A $\frac{3}{2}$-power-law circuit comprising three four-input subthreshold *p*FGMOS transistors. In general, we make a $\frac{p}{q}$-power-law circuit by diode connecting M_1 through q control gates and by connecting p control gates of M_3 to V_1. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. For each configuration, p:q, of the circuit, the output voltage, V_3, was set such that M_3 sourced about 3.14 nA of current when I_1 and I_2 were set to 3.14 nA.

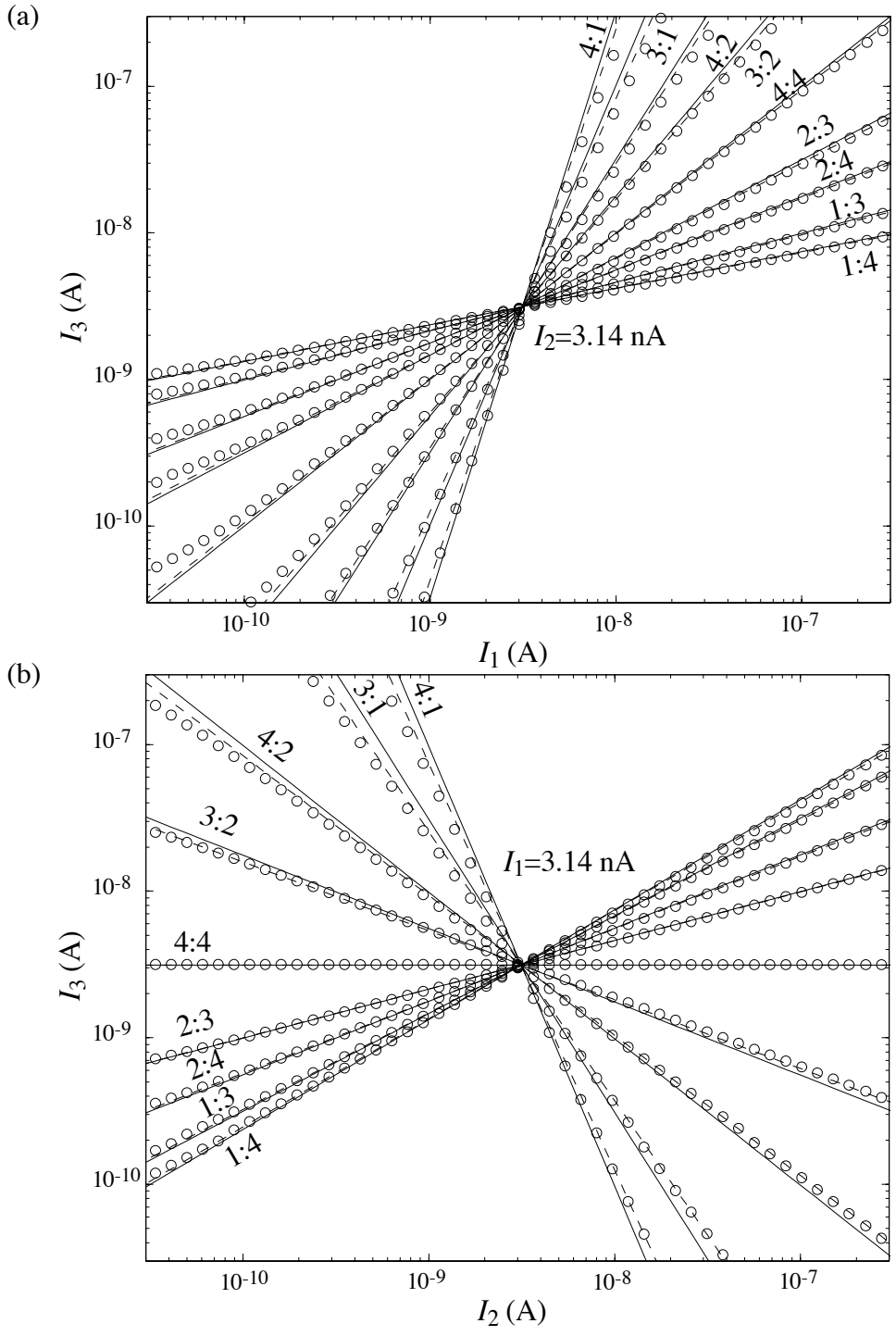


Figure 6.10. Measured data from nine configurations of the circuit of Figure 6.9. Circles are measured values of I_3 plotted as a function of (a) I_1 with I_2=3.14 nA, and (b) I_2 with I_1=3.14 nA. For curve p:q, the solid line shows the ideal expression, $I_3 = I_1^{p/q} I_2^{1-p/q}$, calculated for the values of I_1 and I_2 at each point. Dashed lines show $C_{\text{fg-d}}$-adjusted fits.

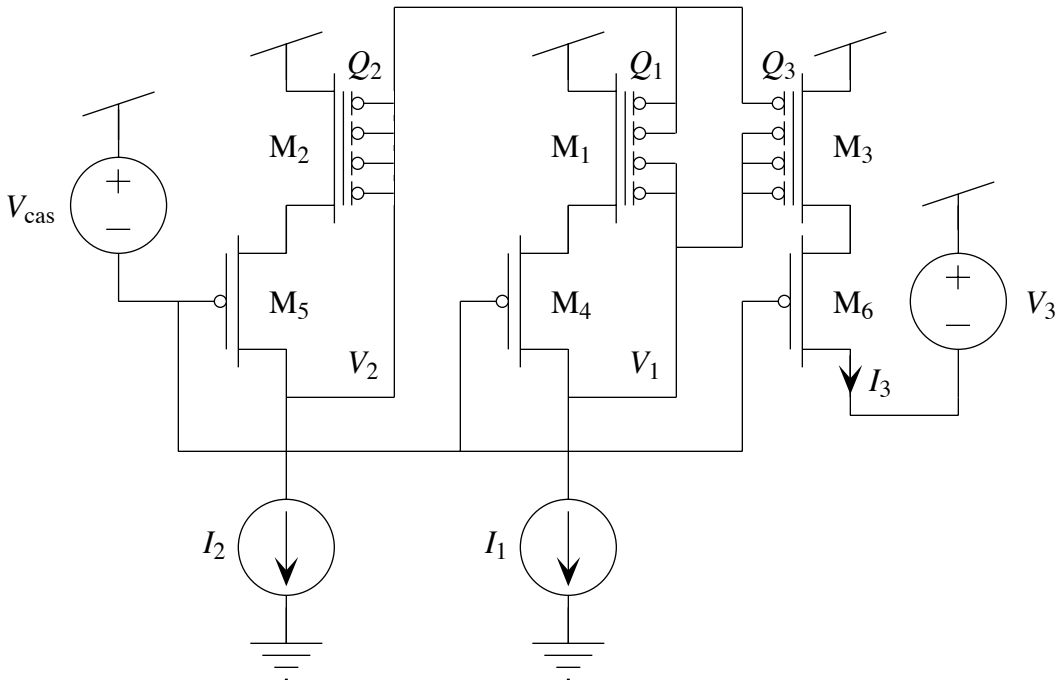


Figure 6.11. A $\frac{3}{2}$-power-law circuit comprising three cascoded four-input subthreshold *p*FGMOS transistors. In general, we make a $\frac{p}{q}$-power-law circuit by diode connecting M_1 through q control gates and by connecting p control gates of M_3 to V_1. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. For each configuration, p:q, of the circuit, the output voltage, V_3, was set to 5 volts below V_{DD}. The cascode bias voltage, V_{cas}, was set to 1.1 volts below V_{DD}.

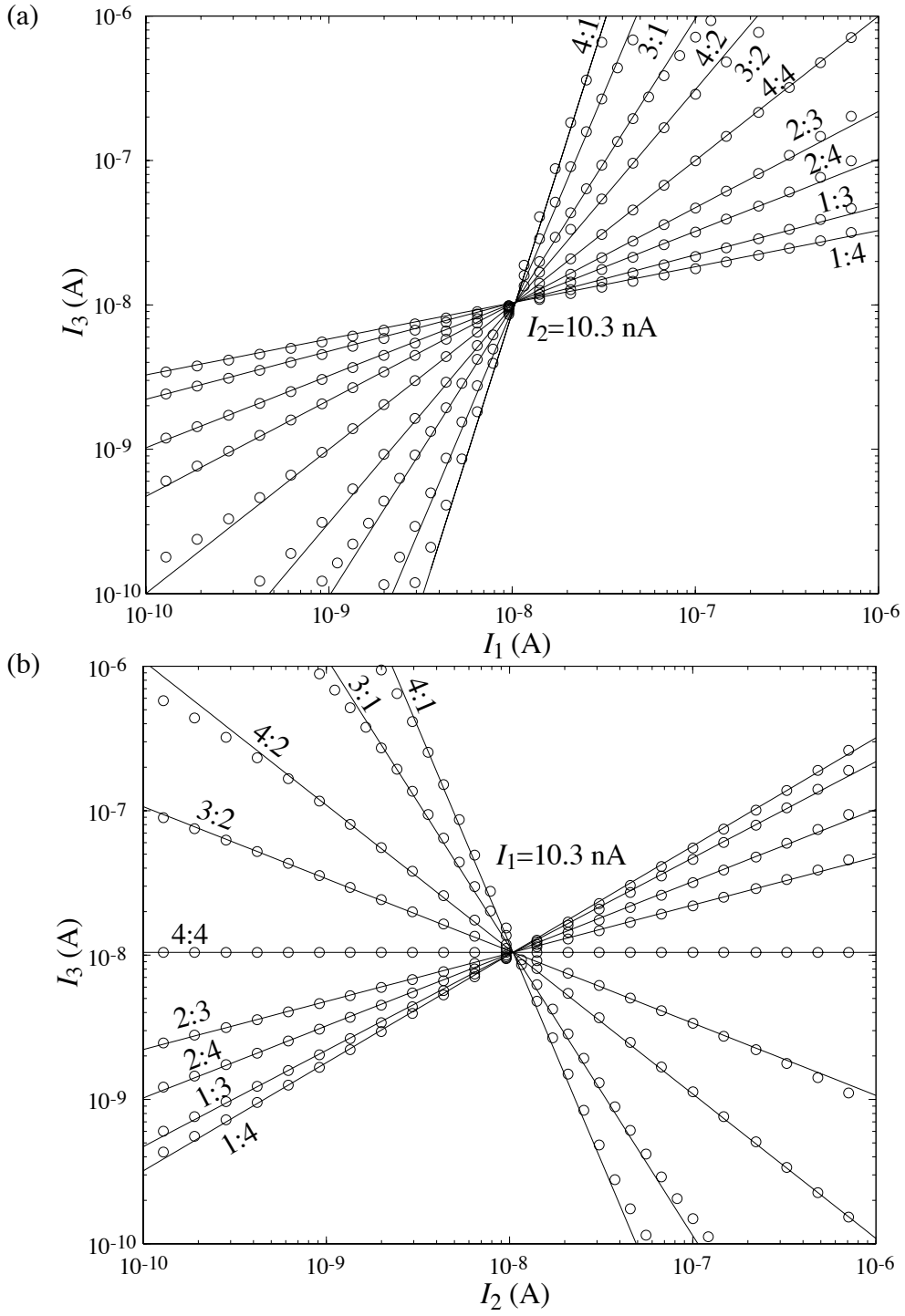


Figure 6.12. Measured data from nine configurations of the circuit of Figure 6.11. Circles are measured values of I_3 plotted as a function of (a) I_1 with I_2=10.3 nA, and (b) I_2 with I_1=10.3 nA. For curve p:q, the solid line shows the ideal expression, $I_3 = I_1^{p/q} I_2^{1-p/q}$, calculated for the values of I_1 and I_2 at each point.

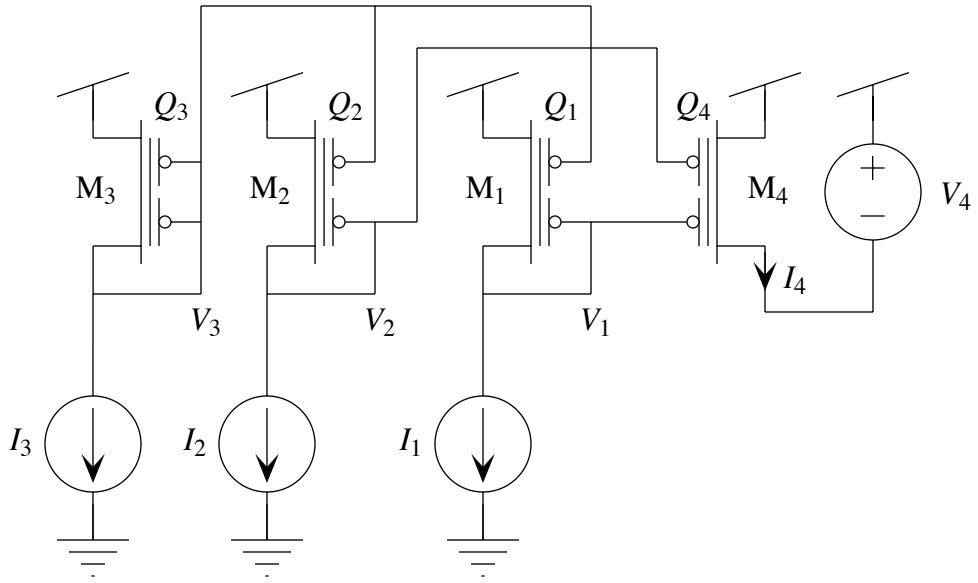


Figure 6.13. A product-reciprocal circuit comprising four two-input subthreshold *p*FGMOS transistors. The values of floating-gate charges Q_1, Q_2, Q_3, and Q_4 were equalized by short-wave UV photoinjection. The output voltage, V_4, was set such that M_4 sourced about 3.20 nA of current when I_1, I_2, and I_3 were set to about 3.20 nA.

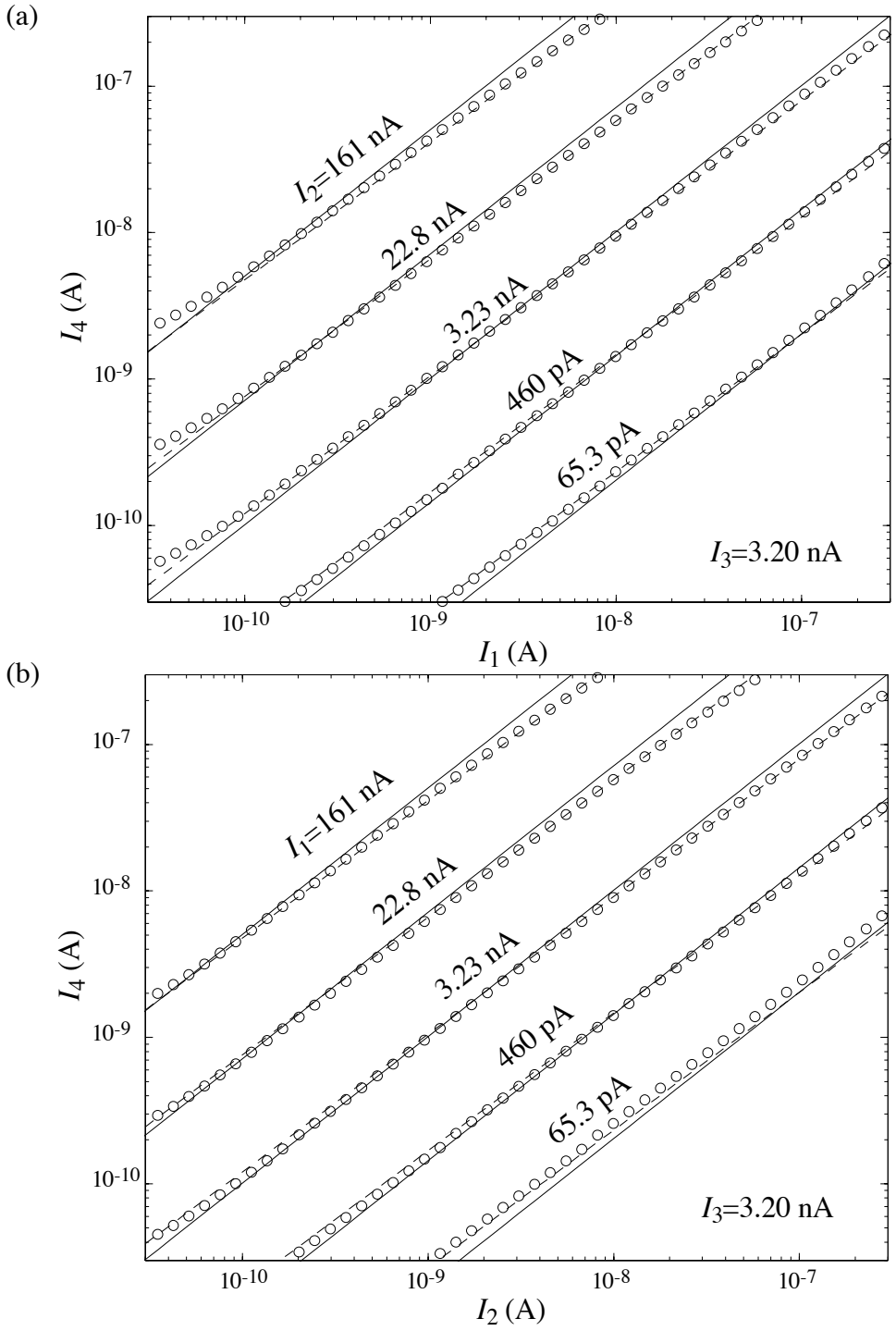


Figure 6.14. Measured data from the circuit of Figure 6.13. Circles are measured values of I_4 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1 with I_3=3.20 nA. Solid lines show the ideal expression, $I_4 = I_1 I_2 \div I_3$, calculated for the values of I_1, I_2, and I_3 at each point. Dashed lines show $C_{\text{fg-d}}$-adjusted fits.

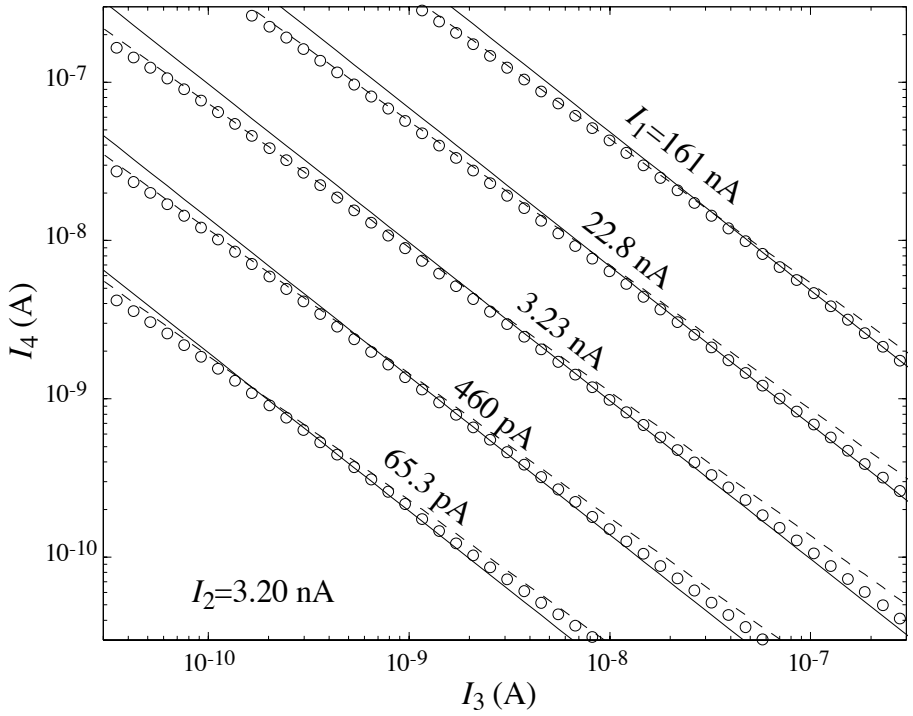


Figure 6.15. Measured data from the circuit of Figure 6.13. Circles are measured values of I_4 plotted as a function of (a) I_3 for various values of I_1 with I_2=3.20 nA. Solid lines show the ideal expression, $I_4 = I_1 I_2 \div I_3$, calculated for the values of I_1, I_2, and I_3 at each point. Dashed lines show fits adjusted to account for $C_{\text{fg-d}}$.

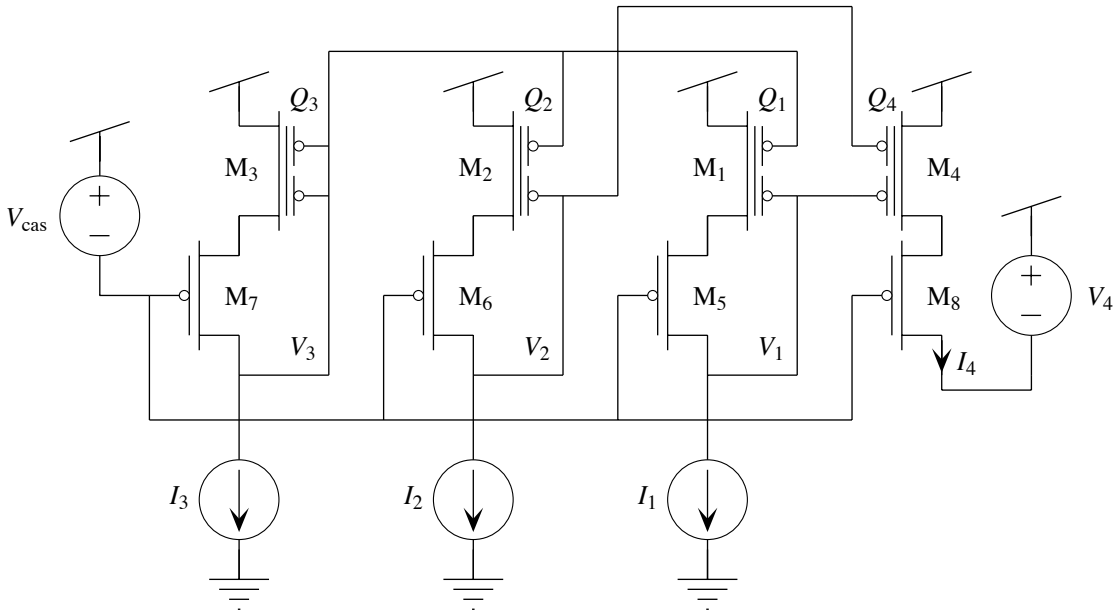


Figure 6.16. A product-reciprocal circuit comprising four cascoded two-input subthreshold *p*FGMOS transistors. The values of floating-gate charges Q_1, Q_2, Q_3, and Q_4 were equalized by short-wave UV photoinjection. The output voltage, V_4, was set to 5 volts below V_{DD}. The cascode bias voltage, V_{cas}, was set to 1.1 volts below V_{DD}.

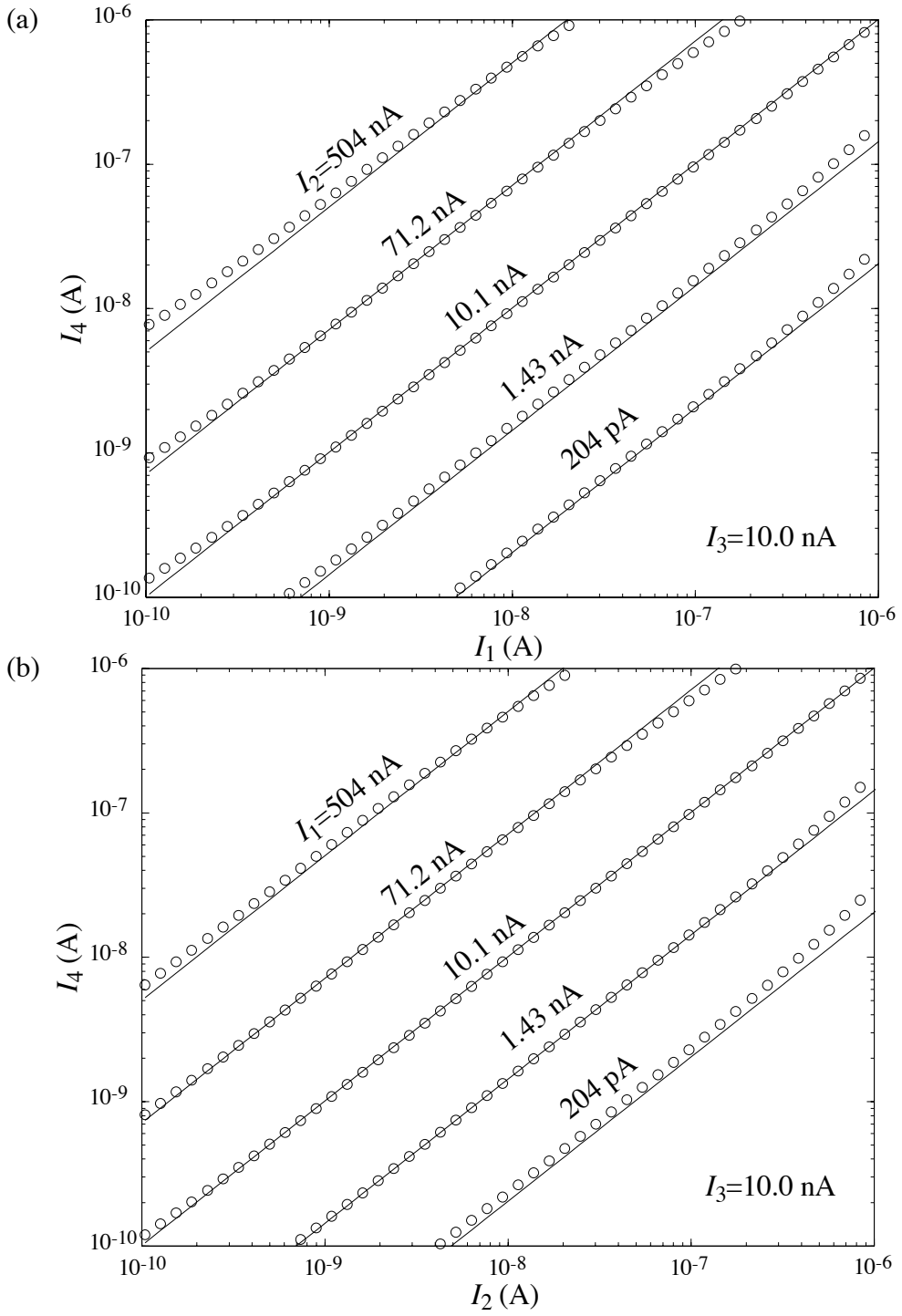


Figure 6.17. Measured data from the circuit of Figure 6.16. Circles are measured values of I_4 plotted as a function of (a) I_1 for various values of I_2 with I_3=10.0 nA, and (b) I_2 for various values of I_1 with I_3=10.0 nA. Solid lines show the ideal expression, $I_4 = I_1 I_2 \div I_3$, calculated for the values of I_1, I_2, and I_3 at each point.

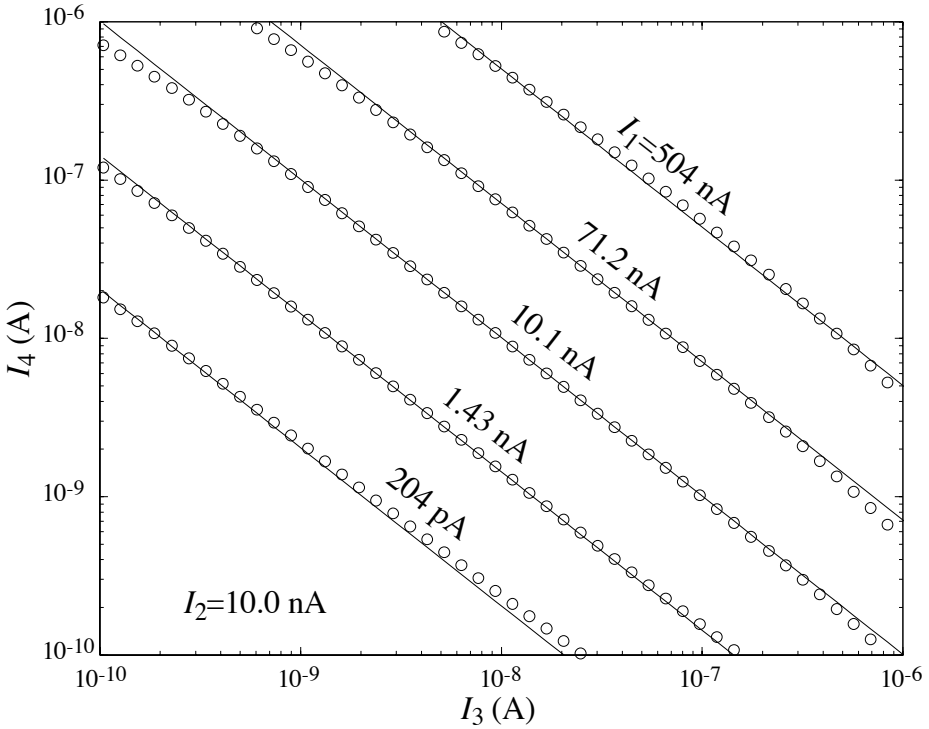


Figure 6.18. Measured data from the circuit of Figure 6.16. Circles are measured values of I_4 plotted as a function of I_3 for various values of I_1 with I_2=10.0 nA. Solid lines show the ideal expression, $I_4 = I_1 I_2 \div I_3$, calculated for the values of I_1, I_2, and I_3 at each point.

(This page has been left blank intentionally.)

CHAPTER 7
CONCLUSIONS AND FUTURE RESEARCH

In this chapter, I summarize the primary contributions that I have made in the research described in the preceding chapters, and I indicate possible directions for future research that proceed naturally from this body of work.

7.1. Summary of Primary Contributions

In this thesis, I have made the following 13 primary contributions.

1. In Chapter 2, I defined the ideal multiple-input translinear element (MITE). The MITE is a primitive circuit element that generates a current that is exponential in a weighted sum of its K input voltages.
2. In Chapter 2, I intuitively described the operation of a class of circuits comprising MITEs, called MITE networks. These circuits embody product-of-power-law relationships accurately in the current signal domain.
3. In Chapter 2, I developed a simple matrix-based analysis procedure for determining the steady-state behavior of (i.e., for determining the product-of-power-law relationships realized by) any given MITE network.
4. In Chapter 2, I delineated sufficient conditions on the input connectivity matrix of a MITE network that guarantee asymptotic stability of the MITE network's steady state, even in the presence of feedback loops. These sufficient conditions depend on only the structure of the input connectivity matrix; they do not depend on the units in which its elements are expressed, on the parasitic node capacitances, or on the input-current levels.
5. In Chapters 2 and 4, I showed constructively that, in a sense, the class of MITE networks and the class of translinear loop circuits are equivalent.
6. In Chapter 3, using the theory of linear signal-flow graphs, I rigorously developed a by-inspection analysis procedure for determining the product-of-power-law relationships embodied in a MITE network.
7. In Chapter 4, I gave two simple procedures for synthesizing an asymptotically stable MITE network that implements a given product-of-power-law relationship.

8. In Chapter 4, I delineated three useful MITE-network transformations that leave the product-of-power-law relationships embodied in a MITE network invariant.
9. In Chapter 4, I discussed how the MITE-network–synthesis problem can be framed as a constrained optimization problem.
10. In Chapter 4, I derived a formula relating the degree to which the product-of-power-law relationships embodied in a MITE network are perturbed given a certain level of mismatch in the unit MITE weighting coefficients. We can use the results of this mismatch analysis as a rational basis for choosing between alternate MITE-network topologies that each implement a given set of product-of-power-law relationships.
11. In Chapter 5, I derived a simple model both for a K-input subthreshold floating-gate MOS (FGMOS) transistor and for a saturated cascoded K-input subthreshold FGMOS transistor. I presented experimental data, which corroborate the model, from a four-input subthreshold FGMOS transistor that was fabricated in Orbit's 2-μm double-poly CMOS process that is available through MOSIS.
12. In Chapter 5, I described an experimental technique, based on the subthreshold FGMOS transistor model, for measuring capacitor mismatch. I presented data on the matching of small square capacitors in Orbit's 2-μm double-poly CMOS process that is available through MOSIS.
13. In Chapter 6, I discussed using both the K-input subthreshold FGMOS transistor and the cascoded K-input subthreshold FGMOS transistor as MITEs. I presented experimental results from 12 MITE networks breadboarded both from subthreshold FGMOS transistors and from cascoded subthreshold FGMOS transistors.

7.2. Directions for Future Research

I believe that the class of MITE networks has a great deal of potential; however, much work remains to be done before it grows from being merely an interesting class of circuits with a rich mathematical structure to being a class of industrial-strength circuits that can be purchased off the proverbial shelf. This remaining work includes, for each of the various MITE implementations, characterizing things such as the sensitivity of MITE networks to variations in temperature, the noise spectrum and dynamic range of MITE networks, and the MITE-network temporal response and bandwidth as a function of input-current level. For MITE networks implemented with FGMOS transistors, we will also be need to develop a reliable, inexpensive process, in a commercial production environment, for adjusting the

stored floating-gate charge; this process would ideally modulate the floating-gate charge in a closed-loop fashion so as to increase the precision of such circuits. Such a process would be an attractive alternative to the laser-trimmed–resistor techniques that are in commercial use today at the wafer-test stage of production. In this context, it would also be helpful to have a computer program for MITE-network synthesis that would take the specification of a function to be realized in a MITE network and would produce a layout for the optimal circuit given a set of design constraints. Such an automated synthesis tool should be straightforward to design based on the material that I developed in Chapter 4. It should be possible to develop a measure, for a fixed set of input currents, of the noise sensitivity of a MITE network with a given topology. This noise-sensitivity analysis should be similar to the mismatch analysis that I derived in Appendix 4.C. We could use such a noise-sensitivity measure in the construction of an objective function for use in a MITE-network–synthesis computer program.

In Subsections 7.2.1 and 7.2.2, I indicate other possible directions for future research.

7.2.1. Other Multiple-Input Translinear Element Implementations

Figure 7.1 shows six different MITE implementations. At the top of Figure 7.1 is the resistive-divider–bipolar MITE implementation that I mentioned in Section 2.4 and for which I showed experimental data from a breadboarded squaring-reciprocal circuit. In this MITE implementation, the weighting coefficients are set by resistive divider ratios. The input voltages must be buffered into the resistive networks so, in a MITE network, they neither supply current to, nor sink current from, the input nodes. This resistive-divider-bipolar circuit is a good MITE implementation over those collector currents for which the base resistance of the bipolar transistor is much greater than the resistances in the resistive voltage divider. When the base resistance becomes comparable to the resistances in the resistive voltage divider, the base voltage is clamped by the base-emitter junction and the collector current then increases approximately linearly, instead of exponentially, with the input voltages.

Proceeding around Figure 7.1 in the clockwise direction, we encounter the K-input subthreshold FGMOS transistor and the cascoded K-input subthreshold FGMOS transistor. In Section 6.1, I showed that both of these primitives are valid MITE implementations over the subthreshold range of drain currents (i.e., from about 1 pA to 1 μA). In Section 6.2, I showed experimental data from a wide range of MITE networks

made from each of these primitive elements. For each of these MITE implementations, the weighting coefficients are set by capacitive divider ratios. In Section 6.1, I discussed the effects of the parasitic drain-overlap capacitance on MITE networks comprising uncascoded subthreshold FGMOS transistors; I showed that we can reduce these effects to negligible proportions by adding a cascode transistor to each FGMOS transistor.

At the bottom of Figure 7.1 is a two-transistor MITE implementation comprising a *K*-input subthreshold FGMOS transistor and a bipolar transistor. Intuitively, this bipolar-FGMOS MITE implementation works as follows. The subthreshold FGMOS transistor makes a current that is exponential in the weighted sum of the input voltages; the weighting coefficients are again set by capacitive divider ratios. The bipolar transistor then acts as a current-gain stage by multiplying the subthreshold FGMOS transistor current by the bipolar's forward current gain. Because the drain of the FGMOS transistor is held at a fixed potential, this MITE implementation is insensitive to the parasitic drain-overlap capacitance problem. However, there is a parasitic source-overlap capacitance that could cause a problem for this MITE implementation similar to that caused by the drain-overlap capacitance for the single subthreshold FGMOS transistor MITE implementation. I show in Appendix 7.A that the source-overlap capacitance does not cause such a problem.

Figure 7.2 shows measured data from a four-input bipolar-FGMOS MITE that was fabricated in Orbit's 2-μm double-poly CMOS process that is available through MOSIS. To obtain these data, for each *n* between 1 and 4, I measured the collector current of the bipolar transistor while I swept *n* of the *n*FGMOS transistor's control gates from ground to 5 volts above ground. For each sweep, I connected the remaining $4-n$ control gates to ground. The solid lines show least-squares best-fit lines to the data plotted on a semilog scale. Note that the slopes, which are indicated along with each curve in Figure 7.2, are nearly in a ratio of 1:2:3:4. Note also that this two-transistor circuit is a good MITE implementation over approximately seven and one-half decades of collector current.

Figure 7.3 shows a squaring-reciprocal circuit made from three two-input bipolar-FGMOS MITEs. I breadboarded this circuit using three four-input bipolar-FGMOS MITEs that were fabricated in Orbit's 2-μm double-poly CMOS process. Then, I balanced the charge on the floating-gates by exposing them to short-wave UV light for about 20 minutes. After balancing the floating-gate charge, I measured the DC characteristics of this MITE network. The resulting data are shown in Figure 7.4. The circles shown in Figure 7.4a represent measured values of I_3 plotted as a function of I_1 over the nearly 7.5-decade current range from 10 pA to 200 μA for nine different values of I_2 ranging from 16.6 pA to

92.0 μA. The circles shown in Figure 7.4b represent measured values of I_3 plotted as a function of I_2 over the current range of nearly seven and one-half decades from 10 pA to 200 μA for nine different values of I_1 ranging from 16.6 pA to 92.0 μA. In both plots, solid lines show values of the ideal theoretical expression,

$$I_3 = \frac{I_1^2}{I_2},$$

calculated for the values of I_1 and I_2 at each point. The data and fits agree well over much of the current range shown.

The final two MITE implementations shown in Figure 7.1 are similar to each other. Each comprises three transistors: a two-transistor FGMOS source-follower and a third transistor that has an exponential current–voltage characteristic. Intuitively, the floating-gate voltage develops as a weighted sum of the *K* input voltages via a capacitive voltage divider. In the source-follower configuration, the FGMOS transistor's source voltage is approximately a linear function of the floating-gate voltage. Consequently, because the floating-gate voltage is a weighted sum of the input voltages, this source voltage is also a weighted sum of the input voltages. The third transistor then makes a current that is exponential in this source voltage. In the first case, the exponential element is a sub-threshold MOS transistor, whereas, in the second case, the exponential element is a bipolar transistor. Note that these MITE implementations do not suffer from the drain-overlap capacitance problem and that the source-overlap capacitance will manifest itself in much the same way as it does in the bipolar-FGMOS MITE.

Because the source-follower circuit configuration does not depend on the form of the current–voltage relationship of the MOS transistor, these three-transistor circuits are good MITE implementations even when we bias the FGMOS source follower with an above-threshold current. For the version of the circuit with the subthreshold MOS transistor, biasing the FGMOS source follower with an above-threshold current allows us to make the output MOS transistor as wide as necessary to get a larger range of exponential currents without having to make the FGMOS transistor, and, hence, the floating-gate capacitance large. The above-threshold bias gives the FGMOS source follower enough bandwidth to drive the large gate capacitance of a wide output MOS transistor. The version of the circuit with the bipolar output transistor is a valid MITE implementation only when the base current is negligible compared with the source-follower bias current. Thus, for the version of the circuit with the bipolar transistor, biasing the FGMOS source follower with above-threshold currents allows us to operate this MITE at high current levels and, thus, potentially with very high bandwidths.

For each of the five FGMOS-transistor–based MITE implementations just described, we can use the floating-gate charge to store adaptable weights for building learning systems or to compensate for scale-factor errors resulting from device mismatch. Note that none of the FGMOS-based MITE implementations, except for the single subthreshold FGMOS transistor, is affected adversely by the parasitic source/drain-overlap capacitances.

7.2.2. Continuous Floating-Gate Charge Adaptation

During the past several years, my colleagues (Chris Diorio and Paul Hasler) and I have been characterizing Fowler–Nordheim tunneling and subthreshold channel hot-electron injection in standard CMOS processes to develop a floating-gate technology with which we can build adaptive or learning information-processing systems [1–9]. Our goal has been to utilize the nonlinearities and dynamics of the native physical processes and devices in the CMOS process so that we can build information-processing systems that adapt or learn on a slow time scale as a natural part of their behavior with little extra circuit overhead.

Of particular note are the continuous-time adaptive FGMOS circuit techniques described by Paul Hasler [8, esp. Chapters 4 and 5]. In these circuits, we bias some or all of the FGMOS transistors such that they are capable of subthreshold channel hot-electron injection. We continuously remove electrons from these floating gates, usually with Fowler–Nordheim tunneling. These circuits have a stable equilibrium for which this continuous tunneling current is balanced by a hot-electron–injection current of equal magnitude at each of the adapting floating gates. When such a circuit is driven away from this equilibrium—say, by a changing input—an imbalance between the hot-electron injection and tunneling currents on each of the disturbed floating gates charges (if the tunneling current exceeds the hot-electron–injection current) or discharges (if the hot-electron–injection current exceeds the tunneling current) the floating gates until the equilibrium is reestablished. As designers, we choose this equilibrium to achieve some desired baseline of circuit operation. Some interesting competitive and cooperative learning behaviors emerge as a natural result of the native nonlinearities and feedback processes in circuits with multiple coupled adapting FGMOS transistors.

An interesting and potentially fruitful line of future research would be to apply these continuous floating-gate charge adaptation techniques to MITE networks made from floating-gate MOS transistors. Because the floating-gate charge directly sets multiplicative scale factors on the product-of-power-law relationships embodied in MITE networks, these continuous floating-gate charge adaptation techniques would naturally confer on such

circuits—with almost no circuit overhead—an automatic-gain-control function, which operates on a slow time scale in parallel with their nonlinear signal-processing function.

Using such charge-adaptation techniques, we should also be able to implement a variety of neural-network structures. One example of such a system is a sigma-pi neural network; this network is capable of learning, with only a single layer of units, to perform nontrivial pattern-classification tasks. A sigma-pi unit takes a weighted sum (whence *sigma*) of products (whence *pi*) of various combinations of its inputs. One straightforward implementation of a sigma-pi unit is as follows. If input signals are represented as currents, we can implement the required products of inputs using subthreshold FGMOS MITE networks. The charge stored on the floating gates in these circuits provide nonvolatile weights on the output currents. With the weights represented this way, learning is accomplished by appropriate modulation of the charge stored on each of the floating gates in the sigma-pi unit. Fowler–Nordheim tunneling and subthreshold channel hot-electron injection could serve as learning mechanisms for such a sigma-pi unit. Finally, because the weighted outputs are represented as currents, we can sum them using a single wire. To make a single-layer sigma-pi neural network, we simply replicate these units. Many other neurally inspired adaptive analog information-processing systems are possible; FGMOS transistors are ideal elements from which to construct such systems.

7.3. Appendix 7.A

In this appendix, I calculate the current–voltage relationship of the bipolar-FGMOS MITE that is shown at the bottom of Figure 7.1. As shown in Figure 7.5, I denote by V the base voltage of the bipolar transistor. I model the large-signal current–voltage relationship of the bipolar transistor with

$$I_{\mathrm{b}} = \mathrm{I_s} \exp\left[\frac{V}{\mathrm{U_T}}\right], \tag{7.1}$$

and

$$I_{\mathrm{c}} = \beta I_{\mathrm{b}}, \tag{7.2}$$

where I_{b} is the base current of the bipolar transistor, $\mathrm{I_s}$ is the saturation current of the bipolar transistor, $\mathrm{U_T}$ is the thermal voltage, I_{c} is the collector current of the bipolar transistor, and β is the forward current gain of the bipolar transistor. From Equation 5.8, the drain current through the subthreshold FGMOS transistor is given by

$$I_{\mathrm{d}} = \mathrm{I_0'} \exp\left[\frac{Q}{\mathrm{Q_T}}\right] \exp\left[\sum_{k=1}^{K} \frac{\kappa C_k}{C_{\mathrm{T}}} \frac{V_k}{\mathrm{U_T}}\right] \exp\left[\left(\frac{\kappa C_{\mathrm{fg-s}}}{C_{\mathrm{T}}} - 1\right) \frac{V}{\mathrm{U_T}}\right], \tag{7.3}$$

where

$$\mathrm{I}_0' = \frac{W}{L}\mathrm{I}_0 \exp\left[\frac{\kappa C_{\mathrm{fg-d}}}{C_\mathrm{T}}\frac{\mathrm{V}_\mathrm{DD}}{\mathrm{U}_\mathrm{T}}\right].$$

By KCL, the drain current of the FGMOS transistor will be equal to the base current of the bipolar. Consequently, using Equations 7.1 and 7.3, I have that

$$\mathrm{I}_\mathrm{s}\exp\left[\frac{V}{\mathrm{U}_\mathrm{T}}\right] = \mathrm{I}_0'\exp\left[\frac{Q}{\mathrm{Q}_\mathrm{T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}}\frac{V_k}{\mathrm{U}_\mathrm{T}}\right]\exp\left[\left(\frac{\kappa C_{\mathrm{fg-s}}}{C_\mathrm{T}} - 1\right)\frac{V}{\mathrm{U}_\mathrm{T}}\right]. \tag{7.4}$$

By rearranging Equation 7.4, I obtain

$$\exp\left[\left(2 - \frac{\kappa C_{\mathrm{fg-s}}}{C_\mathrm{T}}\right)\frac{V}{\mathrm{U}_\mathrm{T}}\right] = \frac{\mathrm{I}_0'}{\mathrm{I}_\mathrm{s}}\exp\left[\frac{Q}{\mathrm{Q}_\mathrm{T}}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}}\frac{V_k}{\mathrm{U}_\mathrm{T}}\right]. \tag{7.5}$$

If I define $C_\mathrm{T}' = 2C_\mathrm{T} - \kappa C_{\mathrm{fg-s}}$, then by raising both sides of Equation 7.5 to the $\frac{C_\mathrm{T}}{C_\mathrm{T}'}$ power, I get

$$\exp\left[\frac{V}{\mathrm{U}_\mathrm{T}}\right] = \left(\frac{\mathrm{I}_0'}{\mathrm{I}_\mathrm{s}}\right)^{\frac{C_\mathrm{T}}{C_\mathrm{T}'}}\exp\left[\frac{Q}{\mathrm{Q}_\mathrm{T}'}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}'}\frac{V_k}{\mathrm{U}_\mathrm{T}}\right], \tag{7.6}$$

where

$$\mathrm{Q}_\mathrm{T}' = \frac{C_\mathrm{T}'\mathrm{U}_\mathrm{T}}{\kappa}.$$

Substituting Equation 7.6 into Equation 7.1, I obtain the following expression for the base current of the bipolar transistor:

$$I_\mathrm{b} = \mathrm{I}_\mathrm{s}\left(\frac{\mathrm{I}_0'}{\mathrm{I}_\mathrm{s}}\right)^{\frac{C_\mathrm{T}}{C_\mathrm{T}'}}\exp\left[\frac{Q}{\mathrm{Q}_\mathrm{T}'}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}'}\frac{V_k}{\mathrm{U}_\mathrm{T}}\right],$$

which, in conjunction with Equation 7.2, implies that the collector current of the bipolar transistor is given by

$$I_\mathrm{c} = \beta\mathrm{I}_\mathrm{s}\left(\frac{\mathrm{I}_0'}{\mathrm{I}_\mathrm{s}}\right)^{\frac{C_\mathrm{T}}{C_\mathrm{T}'}}\exp\left[\frac{Q}{\mathrm{Q}_\mathrm{T}'}\right]\exp\left[\sum_{k=1}^{K}\frac{\kappa C_k}{C_\mathrm{T}'}\frac{V_k}{\mathrm{U}_\mathrm{T}}\right]. \tag{7.7}$$

Thus, the source-overlap capacitance causes the effective total floating-gate capacitance to appear slightly smaller than it would without this parasitic capacitance.

7.4. References

1. P. Hasler, C. Diorio, B. A. Minch, and C. Mead, "Single Transistor Learning Synapses," in Gerald Tesauro, David S. Touretzky, and Todd K. Leen, eds., *Advances in Neural Information Processing Systems 7,* Cambridge, MA: MIT Press, pp. 817–824, 1995.

2. P. Hasler, C. Diorio, B. A. Minch, and C. Mead, "Single Transistor Learning Synapses with Long Term Storage," in *Proceedings of the 1995 International Symposium on Circuits and Systems,* Seattle, WA, vol. 3, pp. 1660–1663, May 1995.

3. C. Diorio, S. Mahajan, P. Hasler, B. A. Minch, and C. Mead, "A High-Resolution Non-Volatile Analog Memory Cell," in *Proceedings of the 1995 International Symposium on Circuits and Systems,* Seattle, WA, vol. 3, pp. 2233–2236, May 1995.

4. P. Hasler, B. A. Minch, C. Diorio, and C. Mead, "An Autozeroing Amplifier Using pFET Hot-Electron Injection," in *Proceedings of the 1996 International Symposium on Circuits and Systems,* Atlanta, GA, vol. 3, pp. 325–328, May 1996.

5. C. Diorio, P. Hasler, B. A. Minch, and C. Mead, "A Single-Transistor Silicon Synapse," *IEEE Transactions on Electron Devices,* vol. 43, no. 11, pp. 1972–1980, 1996.

6. C. Diorio, P. Hasler, B. A. Minch, and C. Mead, "A Complementary Pair of Four-Terminal Silicon Synapses," *Analog Integrated Circuits and Signal Processing,* (in press).

7. P. Hasler, B. A. Minch, C. Diorio, and C. Mead "An Autozeroing Floating-Gate Amplifier," *IEEE Transactions on Circuits and Systems I: Analog and Digital Signal Processing*, (in press).

8. P. Hasler, *Foundations of Learning in Analog VLSI*, Ph.D. Thesis, Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA, 1997.

9. C. Diorio, *Neurally Inspired Silicon Learning: From Synapse Transistors to Learning Arrays*, Ph.D. Thesis, Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 1997.

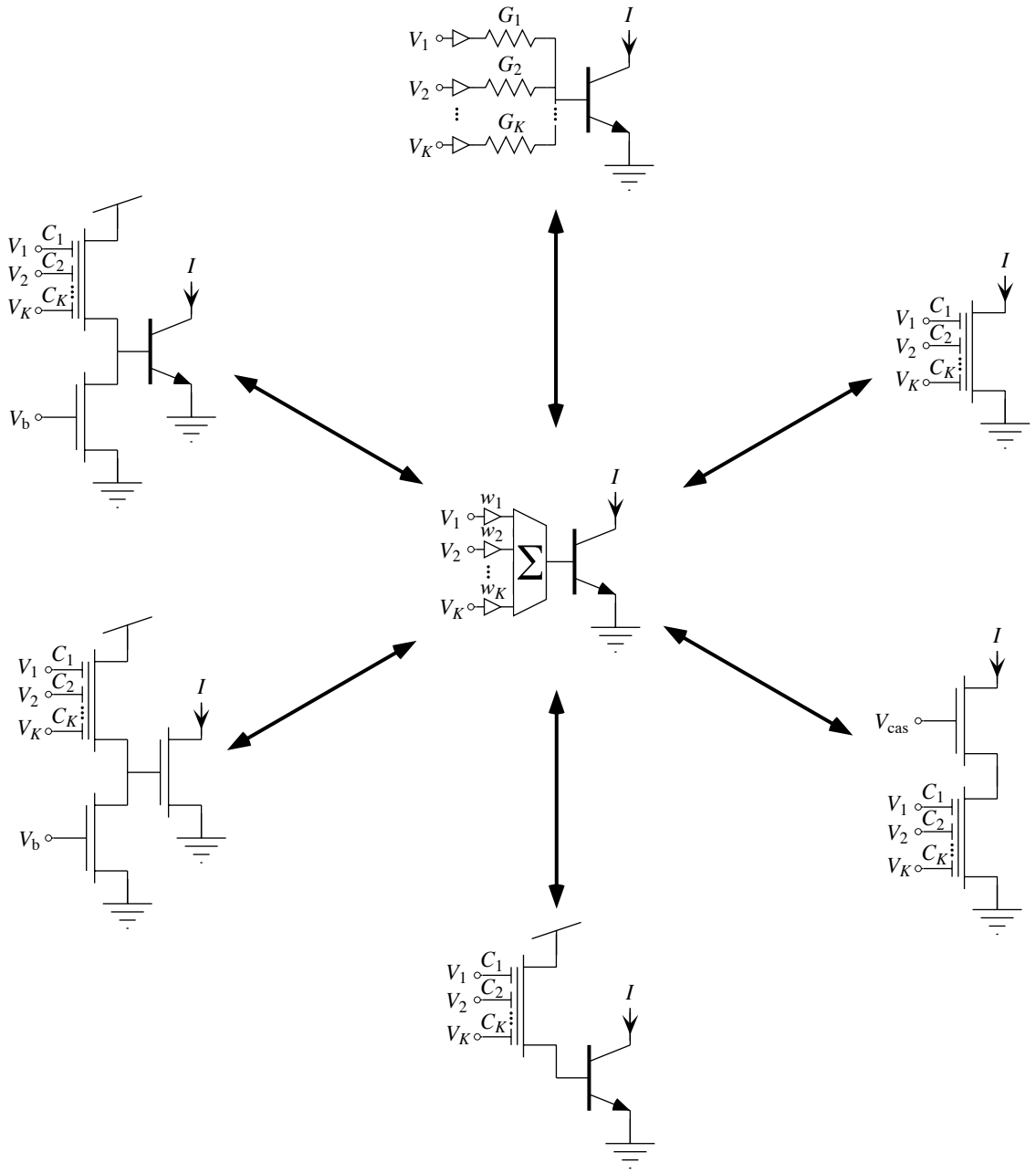


Figure 7.1. Six possible implementations of the ideal MITE. For the five FGMOS-transistor–based implementations, we can use the floating-gate charge to store adaptable weights to build learning systems or to compensate for scale-factor errors resulting from device mismatch. All the FGMOS-based MITE implementations—except for the single subthreshold FGMOS transistor—are unaffected by the parasitic source/drain-overlap capacitances.

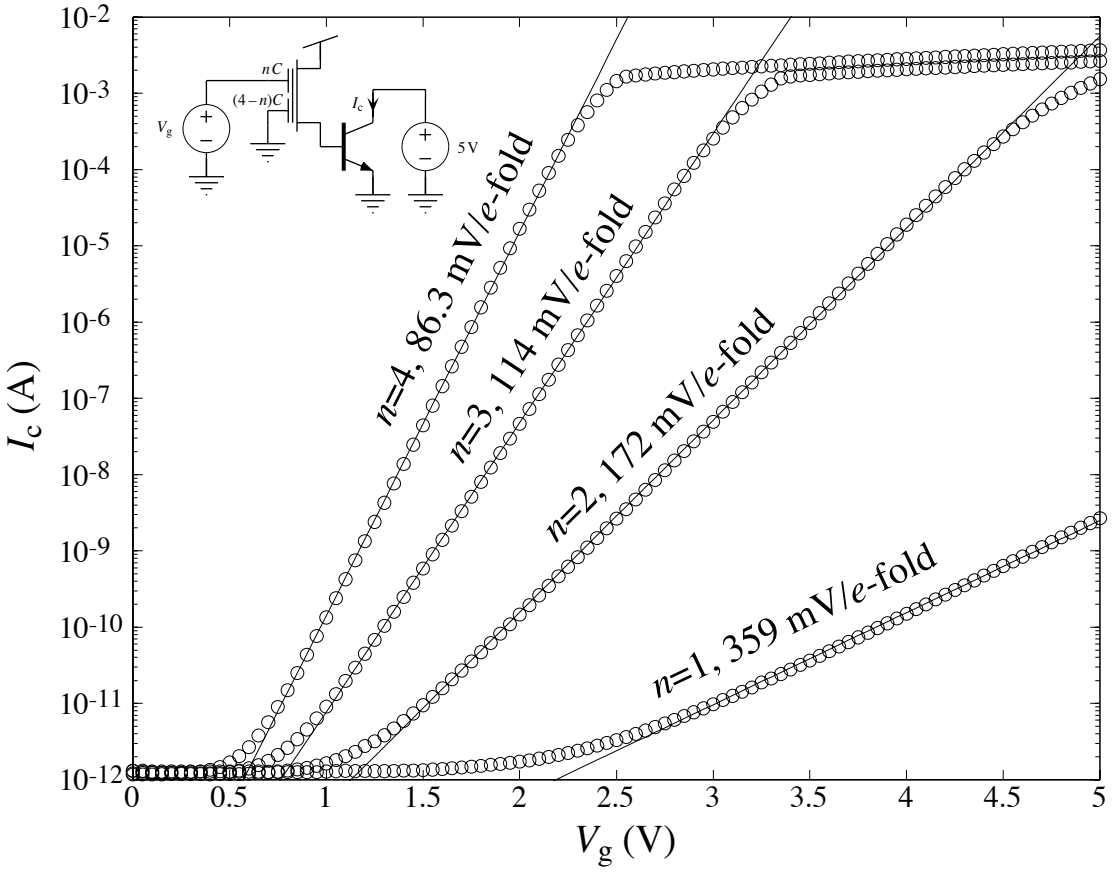


Figure 7.2. Measurements of a four-input bipolar-FGMOS MITE's output current sweeping n control gates with the remaining $4-n$ control gates connected to ground. The solid lines show least-squares best-fit lines to the data plotted on a semilog scale. The slopes, which are indicated along with each curve, are in a ratio of approximately 1:2:3:4.

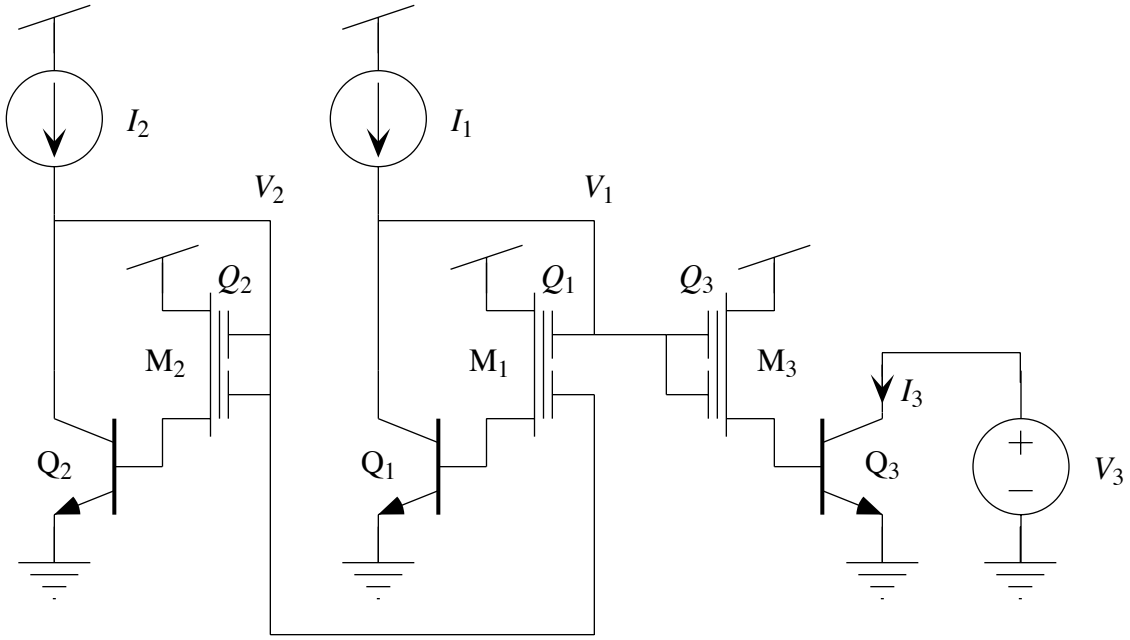


Figure 7.3. A squaring-reciprocal circuit comprising three two-input bipolar-FGMOS MITEs. The values of floating-gate charges Q_1, Q_2, and Q_3 were equalized by short-wave UV photoinjection. The output voltage, V_3, was set to 5 volts above ground.

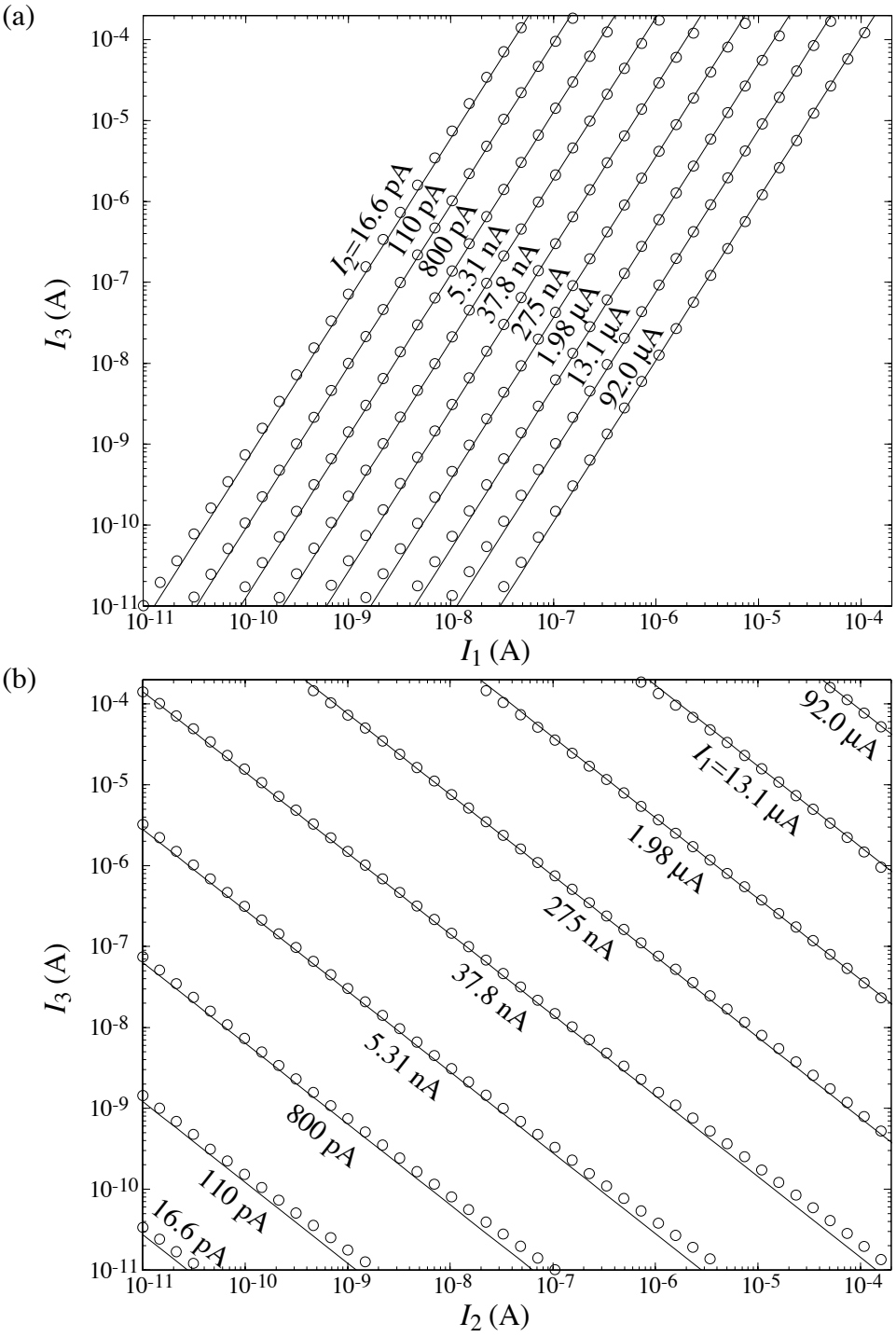


Figure 7.4. Measured data from the circuit of Figure 7.3. Circles are measured values of I_3 plotted as a function of (a) I_1 for various values of I_2, and (b) I_2 for various values of I_1. Solid lines show the ideal expression, $I_3 = I_1^2 \div I_2$, calculated for the values of I_1 and I_2 at each point.

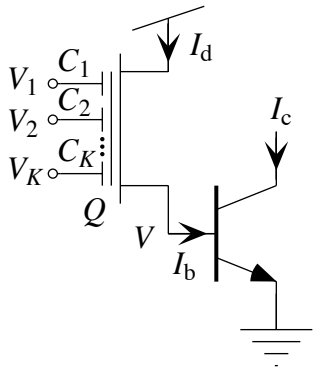


Figure 7.5. Schematic of a bipolar-FGMOS MITE. I assume that the collector voltage of the bipolar is sufficiently far above ground that the bipolar transistor is not saturated.