Analog VLSI Circuits for Sensorimotor Feedback

Thesis by

Stephen P. DeWeerth

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

California Institute of Technology
Pasadena, California
1991
(Defended 27 May 1991)

Acknowledgments

I owe a great debt to my advisor, Carver Mead, for his support and encouragement during my years of graduate school. He guided me through both his words and his examples to an deeper understanding of the process of doing research. I would also like to thank my committee, Yaser Abu-Mostafa, Jim Bower, Al Barr, Joel Burdick, and Fred Culick, for their suggestions and guidance throughout my graduate career and especially in the preparation of this thesis. I feel honored to have had the chance to interact with and learn from these people.

I thank Carver and his research group for providing an exciting and enjoyable environment in which to interact and accomplish research. I thank Massimo Sivilotti, John Wawrzynek, John Tanner, and Lars Nielsen for helping to initiate me into the lab and to introduce me to scientific research. I am grateful for the chance that I have had to interact with Massimo, Michael Emerling, Misha Mahowald, Mary Ann Maher, Xavier Arreguit, John LeMoncheck, Lloyd Watts, Dick Lyon, and John Harris, all of whom have helped to shape my research and to enrich my understanding of science and engineering.

To my friends, I thank you for keeping me sane through the high and low periods. To Steve Burns, I thank you for being a true friend and a great roommate. To Michael and Massimo, I thank you for being there for me; we shared some great times. To Ruth, Nan, Andy, and the others who made my years in grad school enjoyable, I thank you all.

To my parents and my sister Suzie, I thank you for your love and encouragement throughout my life. To my children, Emily and Erik, I thank you for your understanding when I was unable to give you all the attention that you needed, and for your continuous confidence in me. To Valerie, words cannot describe the gratitude that I have for you; through your constant love and support, you carried me through the bad times and shared the good.

Abstract

This thesis presents a design framework and circuit implementations for integrating sensory and motor processing onto very large-scale integrated (VLSI) chips. The designs consist of analog circuits that are composed of bipolar and subthreshold MOS transistors. The primary emphasis in this work is the transformation from the spatially-encoded representation found in sensory images to a scalar representation that is useful for controlling motor systems.

The thesis begins with a discussion of the aggregation of sensory signals and the resulting extraction of high-level features from sensory images. An integrated circuit that computes the centroid of a visual image is presented. A theoretical analysis of the function of this circuit in stimulus localization and a detailed error analysis are also presented. Next, the control of motors using pulse trains is discussed. Pulse-generating circuits for use in bidirectional motor control and the implementation of traditional control schemes are presented. A method for analyzing the operation of these controllers is also discussed. Finally, a framework for the combination of sensory aggregation and pulse-encoded outputs is presented. The need for signal normalization and circuits to perform this task are discussed. Two complete sensorimotor feedback systems are presented.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The realization of new integrated circuit technologies gives the circuit designer the opportunities to explore novel, previously unavailable computational paradigms. However, innovative technologies also require the development of new computational primitives, implementation strategies, and analytical tools to solve traditional problems. Mead [Mea89] has presented a very large-scale integrated (VLSI) circuit design paradigm for implementing analog circuits using MOS transistors operated in their subthreshold regime. This paradigm allows for the implementation of highly collective architectures that are constructed from primitive elements, which are characterized by their low power, small physical size, and complex functionalities.

One of the primary advantages of this design paradigm is that sensing and processing (both local and global) can be integrated on a single VLSI chip. This advantage is especially important because the inherent dimensionality in sensory images requires communicating the high bandwidth signals from the sensing surface to the external environment. The traditional approach to handling this type of high-bandwidth data is to scan the sensory images serially and then to process the resulting data stream. By allowing for the combination of sensing and processing on a single chip, significant data processing can be accomplished on the chip, where the communication bandwidth is potentially much higher than that available at the electrical interface to the chip [SMM87].

One reason for incorporating processing onto the sensing surface is to reduce the data bandwidth that must be transmitted off the chip, and, in particular, to extract a class of features that can be encoded using a fixed number of wires (independent of the array size)

[Tan86] [DM88]. By reducing the number of continuous-time output signals, the need for time sharing the output wires of the chip is eliminated. Processing found in the nervous systems of biological organisms suggests similar computational paradigms that use highly parallel computational structures to aggregate sensory information causing the reduction in signal bandwidth and the extraction of useful features [RP76] [PR76].

By using this analog VLSI design paradigm, the designer can combine sensing, computation, and motor control on a single chip. Global computation facilitates the conversion of one- or two-dimensional spatially encoded sensory images into motor signals that can be used in feedback applications [DeW91]. In the past, there has been significant effort (in fields such as control theory [Wil60]) to build systems that combine sensing and motor control. The designer can take advantage of these traditional approaches by studying the effects of implementing such systems with analog VLSI circuits.

In this thesis, we present a novel conceptual framework and accompanying circuitry for accomplishing the task of converting sensory images to motor signals. The primary contribution of this work is in the design and analysis of circuits that are integrated on a single VLSI chip and convert spatially encoded sensory data into scalar signals that can be used to control motors. We discuss the advantages and limitation of the technology on the implementation of these circuits.

The first part of this thesis (Chapter 2) presents circuitry to aggregate sensory images, such that the bandwidth is reduced to that which can be communicated on a fixed number of wires. We present a circuit that uses this aggregation network to compute the centroid of a visual image. We develop a theoretical basis for analyzing this circuit when it is used to localize a bright constant illumination area (the stimulus) on a darker background, and present data to compare to this theory. We also describe, in detail, the sources of error for this circuit, and quantify the error from each source to demonstrate the limitations of the circuitry.

The second part of this thesis (Chapters 3, 4, and 5) discusses the use of pulses to drive motors, and presents VLSI circuits to generate these pulses. Chapter 3 is an overview of the operation of a DC motor when driven by pulses. We present analysis of pulse-driven systems in the presence of friction, and discuss the reasons that pulses are preferable to a DC voltage for controlling such systems. Chapter 4 is a description of circuits for generating

pulses. We present a modified version of the aggregation network that converts an array on inputs to a dual-rail encoded pulse output, which is sufficient for controlling the motion of a DC motor bidirectionally. Chapter 5 presents a proportional-derivative servo implemented within the pulse-output aggregation network. We present a method of analysis which takes into account the nonlinear elements in the system and which can be applied to controllers implemented within the aggregation network. We compare this analysis with data acquired from a feedback system for the control of motor position.

The third part of the thesis (Chapter 6) presents a circuit framework that combines sensory aggregation with pulse generation to implement complete sensorimotor feedback systems. We discuss the need for signal normalization in this framework, and present circuits for linear and winner-take-all normalization. We present an implementation of the framework that, when incorporated into a mechanical system, causes the system to track a bright visual stimulus (spot of light). We also discuss the advantages of winner-take-all normalization in this system. Finally, we present two electromechanical systems into which this circuit has been incorporated.

Chapter 2

Signal Aggregation and the Centroid Detector

In this chapter, we present circuitry that converts a one-dimensional array of input signals into an output representation that encodes a weighted combination of these inputs. We discuss how this circuit can be used to convert a place-encoded input into a pair of output signals. We then describe an integrated circuit that combines this aggregation circuitry with visual input and electrical feedback to compute the centroid of a visual image. Finally, we present data from a chip that performs this centroid computation, and discuss the ability of the chip to localize a bright stimulus in the visual field.

2.1 The Aggregation Network

We shall describe a circuit that aggregates an array of input currents and differential voltages to produce a pair of dual-rail output currents that encodes a weighted combination of the inputs. This aggregation network consists of an array of differential-pair circuits. We shall first discuss the operation of the differential pair; we shall follow with a description of the aggregation network.

2.1.1 The Differential-Pair Circuit

Dating from the days of the vacuum tube, the differential pair [Vit85] is one of the most common circuits used in analog VLSI systems. This circuit consists of two similar transistors

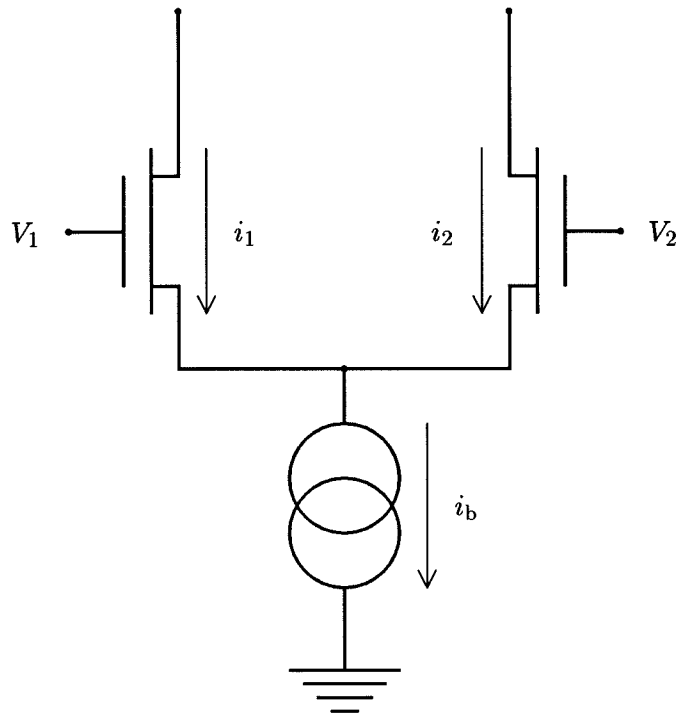

Figure 2.1: The differential-pair circuit. This two-transistor circuit differentially divides a bias current i_b between the two output currents i_1 and i_2 as a function of the differential input voltage $V_1 - V_2$.

with their sources connected. Figure 2.1 is a schematic of an n-channel MOS version of this circuit. The inputs to the circuit are a bias current i_b and a pair of voltages V_1 and V_2. The outputs are two currents (i_1 and i_2). Although the circuit can be designed with either MOS or bipolar transistors, for purposes of analysis, we shall describe it as designed with MOS devices. (We shall later discuss the advantages of using bipolar devices.)

The basic operation of this circuit occurs as follows. The bias current is divided between the two transistors to form the two output currents. The amount of the total current passing through each transistor depends on the relationship between the two input voltages. We know from Kirchoff's current law that the sum of the two output currents must be equal to the input current, or

$$i_b = i_1 + i_2$$

If we ignore drain conductance and assume that the transistors are saturated (see [Mea89] for

derivations of transistor equations), we obtain the following equations for the two transistors:

$$\begin{aligned} i_1 &= i_0 e^{\frac{\kappa V_1 - V}{U_t}} \\ i_2 &= i_0 e^{\frac{\kappa V_2 - V}{U_t}} \end{aligned}$$

where $U_t = \frac{kT}{q}$ is the thermal voltage.

By combining the three preceding equations, we can derive the following equation:

$$i_0 e^{\frac{-V}{U_t}} = \frac{i_b}{e^{\frac{\kappa V_1}{U_t}} + e^{\frac{\kappa V_2}{U_t}}}$$

If we use this equation to substitute for $i_0 e^{\frac{-V}{U_t}}$ in the previous two equations, we can derive the following for i_1 and i_2 in terms of the bias current and the differential input voltage $\kappa(V_1 - V_2)$:

$$i_1 = i_b \frac{1}{1 + e^{\frac{\kappa(V_2 - V_1)}{U_t}}} \tag{2.1}$$

$$i_2 = i_b \frac{1}{1 + e^{\frac{\kappa(V_1 - V_2)}{U_t}}} \tag{2.2}$$

Equations 2.1 and 2.2 describe a complementary pair of sigmoidal functions, as shown in Figure 2.2. In theory, when the differential input voltage is zero ($V_1 = V_2$), the two currents are equal. However, due to mismatches in the two transistors, this equality does not generally hold, and we are left with a nonideality in the circuit. We can describe this nonideality with an offset voltage, which is the differential input voltage necessary to produce equal output currents.

As the magnitude of the differential voltage increases, the magnitude of the difference between the two currents also increases, with one current asymptotically approaching zero and the other current asymptotically approaching i_b. Notice that mismatches in the transistors do not affect the values of the asymptotes. These values are ensured by the direct current path from input to output.

Quite often, the value with which we are concerned in our computations is the difference between the two output currents. This differential input current is the hyperbolic tangent

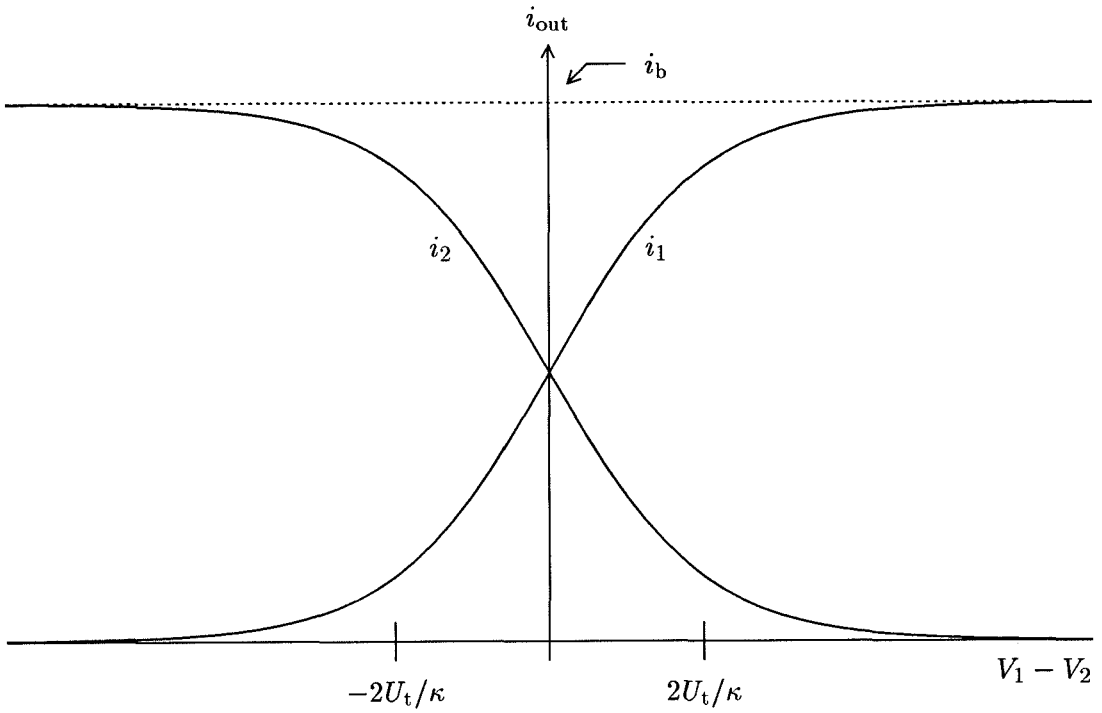

Figure 2.2: Differential-pair output currents. The sum of the output currents from the differential pair are equal to the bias current. When the differential input voltage is zero ($V_1 = V_2$), the two currents are equal. As the magnitude of the differential voltage increases, the magnitude of the difference between the two currents also increases, with one current asymptotically approaching zero and the other current asymptotically approaching i_{b}.

of the differential input voltage [Mea89]:

$$i_1 - i_2 = \tanh\left(\frac{\kappa(V_1 - V_2)}{2U_t}\right)$$

The use of the differential current $i_1 - i_2$ as a signal itself often simplifies the computations in larger networks (as we see later in this chapter). Obviously, if we are to use this current as a signal, the subtraction must actually be performed at some point in the system.

2.1.2 Signal Aggregation

We can use the differential-pair circuit as a basis for a circuit that aggregates a number of input signals. We implement this aggregation using an array of $N + 1$ differential pair elements configured as shown in Figure 2.3. Each element has a current input and a differential voltage input. The current outputs from the differential pairs are aggregated onto a pair of wires so that the aggregated currents retain the differential nature of outputs from the individual elements. The aggregated currents (I_1 and I_2) are the sums of the $N + 1$ input currents multiplied by their respective sigmoidally compressed differential input voltages ($\Delta V = \kappa(V_1 - V_2)$). The equations for these two currents are

$$I_1 = \sum_n \frac{i_n}{1 + e^{\frac{-\Delta V_n}{U_t}}} \tag{2.3}$$

$$I_2 = \sum_n \frac{i_n}{1 + e^{\frac{\Delta V_n}{U_t}}} \tag{2.4}$$

Kirchoff's current law mandates that the sum of the output currents ($I_1 + I_2$) must be equal to the sum of the input currents ($\sum i_n$). This restriction guarantees that the magnitudes of the signals are retained in this circuit. Since the difference of the output signals ($I_1 - I_2$) is the same as the sum of the individual differential output currents, the following equation describes the differential output current:

$$I_1 - I_2 = \sum_n i_n \tanh\left(\frac{\Delta V_n}{2U_t}\right)$$

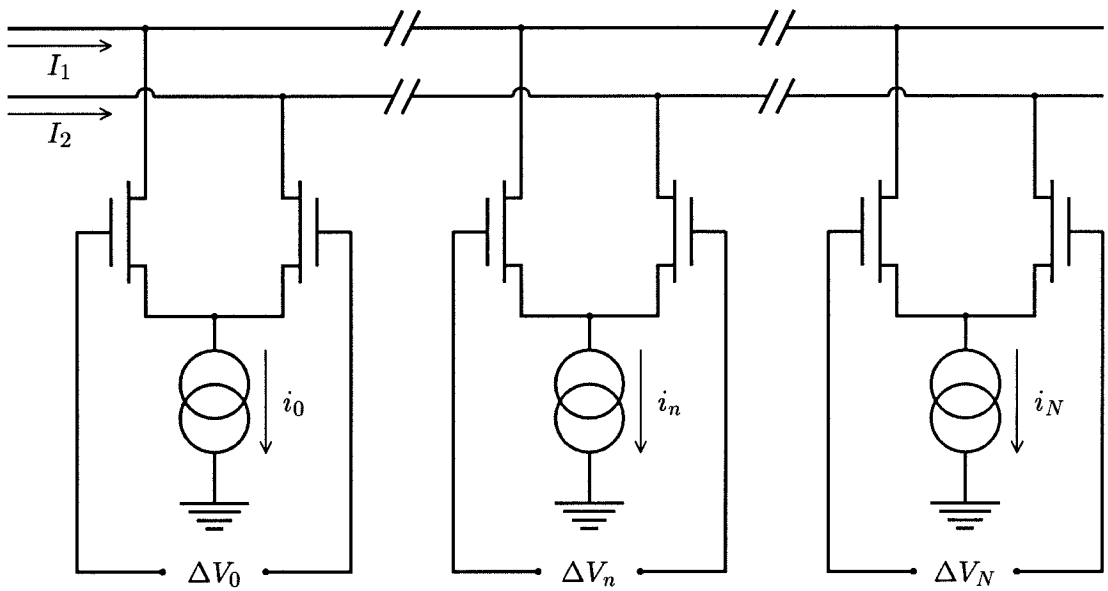

Figure 2.3: The aggregation network of differential-pair circuits. Each element takes as input a differential voltage ΔV_n and a current i_n. The output currents from the differential pairs are summed onto a pair of wires. The resulting aggregated currents are I_1 and I_2.

2.2 The Centroid Circuitry

In this section, we present a circuit that, combined with the aggregation network, computes a function similar to the centroid of a spatially-encoded array of currents. In a true centroid, this function is the mean of the positions of the elements weighted by their respective input currents. If our elements had linear transfer functions (as opposed to the hyperbolic tangent transfer function) the computation would be a true centroid [DM88]. In this circuit, however, the position is compressed by the hyperbolic tangent function. Although the function computed by the circuit is not a true (mathematical) centroid, we shall use this term to describe the circuit.

2.2.1 Position Encoding

In the previous section, the input voltages to the aggregation network were independent of one another. These voltages, however, can be used to encode the position of each element in the array. This encoding can be performed by a chain of $N+1$ series resistors used as a voltage divider, as shown in Figure 2.4. By applying voltages V_0 and V_N to the ends of the resistive line, we can create a linear voltage gradient along the array, where the voltage at the nth node is given by

$$V_n = \left(\frac{V_N - V_0}{N}\right) n + V_0 \tag{2.5}$$

This position-encoding scheme can be incorporated into the aggregation network as shown in Figure 2.5. The position of the nth element is encoded by the differential voltage $V_n - V_{\text{ref}}$, where V_n is set by the voltage divider and V_{ref} is a constant. V_{ref} sets the origin of the array (the position in the array at which the input current is equally distributed between the two output currents).

2.2.2 Subtraction Circuit

As stated earlier, we must eventually perform the subtraction of the two aggregated output currents to take advantage of the differential nature of the outputs from the aggregation network. The subtraction can be performed locally (at each element) or globally on the aggregated signal. A current mirror circuit [Mea89] can be used to perform this subtraction on the chip (either locally or globally). We can use this circuit to subtract two currents of

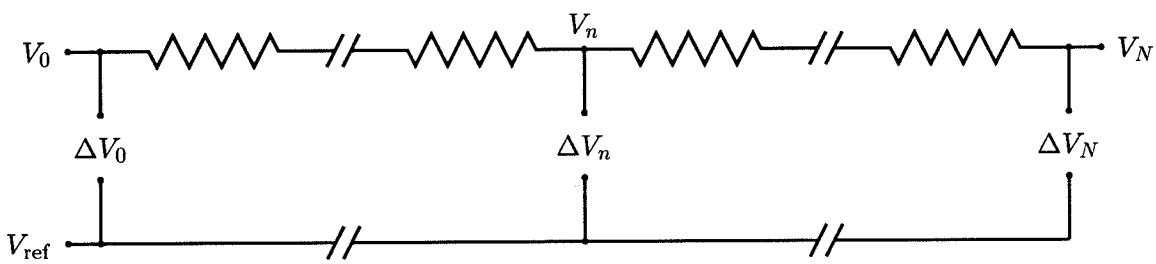

Figure 2.4: The resistive voltage divider. By applying a voltage to each end of a resistive chain, we can use this circuit to encode linearly the position of the elements in the array. The voltage at the nth element is $V_n = \left(\frac{V_N - V_0}{N}\right) n + V_0$.

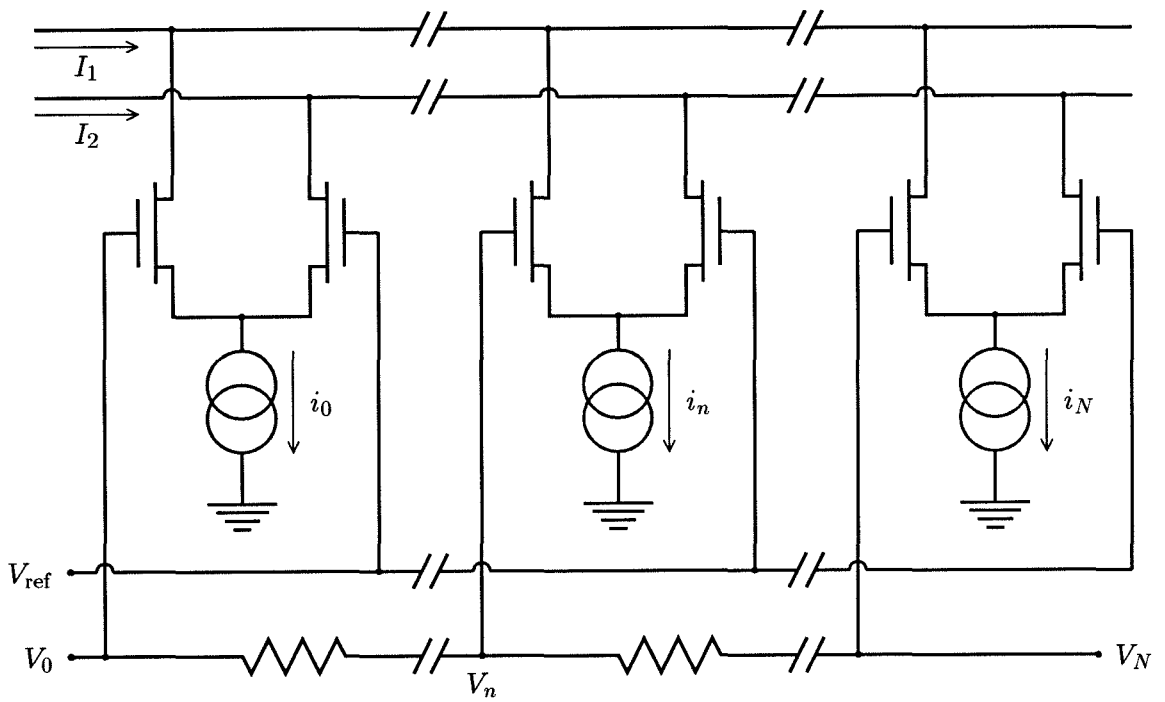

Figure 2.5: The aggregation network with position encoding. The position of the nth element is encoded by the differential voltage $V_n - V_{\text{ref}}$, where V_n is set by the voltage divider and V_{ref} is a constant.

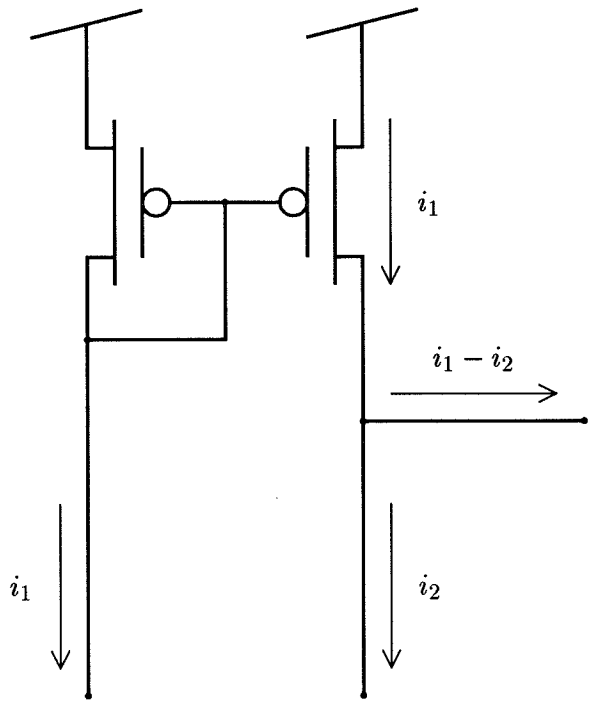

Figure 2.6: The current mirror used as a subtractor. The current mirror can be used to subtract two currents (i_1 and i_2) with the same sign. i_1 is mirrored and the resulting current is added to the output current. i_2 is subtracted directly from the output current with the result $i_{\text{out}} = i_1 - i_2$.

the same sign, such as those produced by a differential pair. Figure 2.6 shows how a current mirror can be used to perform the subtraction.

Due to nonidealities in the transistors, the current mirror does not produce a perfect replica of the original current, and we must take this fact into account when using it in our circuit. Although we can subtract the currents at each individual node and then sum these bidirectional currents onto a single wire, this approach would produce an offset at each element and thus a random error in the apparent position of each element [DM88].

A more desirable alternative is to sum the i_1 and i_2 currents separately (as in Figure 2.3, and use a single current mirror to subtract these aggregated currents. This scheme introduces only a single offset in the system, changing the apparent position of all the elements by the same constant value. A single current mirror, however, must be large enough to handle the sum of the currents that are generated by all of the input current sources.

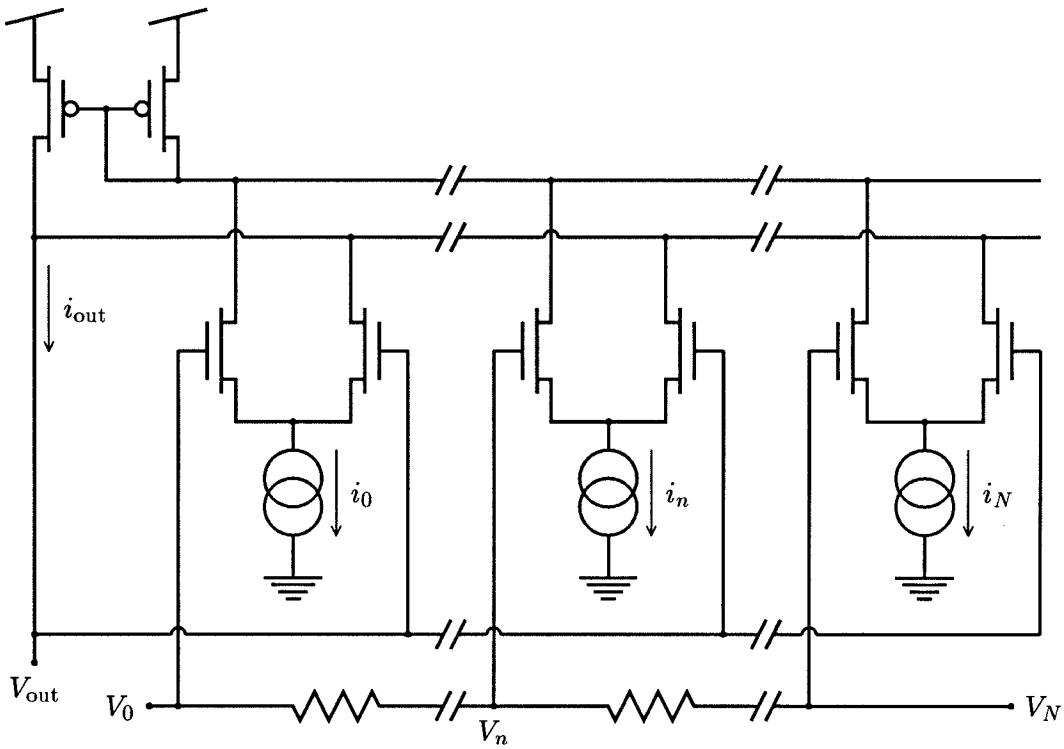

Figure 2.7: The centroid detection circuitry. This circuit consists of an aggregation network, a resistive divider for position encoding, and a current mirror. The current mirror computes the difference between the two aggregated currents and feeds this current back onto the voltage reference wire. The current is integrated on this wire until the differential current is 0, at which point the voltage on the wire (V_{out}) is equal to the centroid of the array referenced to the position voltages.

2.2.3 Electrical Feedback

The completed centroid circuit is shown in Figure 2.7. This figure shows the aggregation network with the addition of the position encoding circuitry and the feedback using a single current mirror. The current mirror computes the difference between the two aggregated currents and feeds this current onto the voltage reference wire. The current is integrated on this high-gain node until the differential current is 0, at which point the voltage on the wire (V_{out}) approximates the centroid of the array referenced to the position voltages.

The sum of the differential currents from all of the elements must be equal to 0. The differential current from each element is the hyperbolic tangent of the differential input

voltage weighted by the input current. Thus, the following equation describes the operation of the closed-loop system in the limit of infinite gain (i.e., ignoring the finite drain resistance of the transistors feeding the voltage reference wire):

$$0 = \sum_n i_n \tanh\left(\frac{V_n - V_{\text{out}}}{2U_t}\right) \tag{2.6}$$

Even with this simplifying assumption, we have not discovered a closed-form solution for V_{out} for Equation 2.6. After the discussion of the circuit implementation, we shall discuss methods for analyzing the functionality of this circuit under specific operating conditions.

We can set the voltage gradient along the resistive voltage divider by changing V_0 and V_N. By Equation 2.6, the output voltage (V_{out}) is a weighted combination of the amplifier input voltages, and is constrained to lie within the input voltage range. Thus, the maximum possible differential voltage at the input of any amplifier is $|V_N - V_0|$.

If $|V_N - V_0| \leq 2U_t$, all the amplifiers will be operating in their "linear" regimes, and the network will approximately compute the mean of the input voltages of the amplifiers weighted by their corresponding input currents. Since the input voltages encode the positions of the elements, performing this calculation is equivalent to computing the centroid of the image represented by the currents [DM88]. In this regime, the network approximates the function of a similarly configured network of linear resistors. This type of network has been shown to perform the centroid computation [DM88].

If $|V_N - V_0| \gg 2U_t$, the slope of the sigmoidal transfer function with respect to distance along the array will be very large. The differential pairs (except for those with input voltages very near the output voltage) then can be approximated as threshold elements, and the network will approximate the weighted-median calculation [DM88].

If $|V_N - V_0|$ falls between these two regimes, the calculation will be a combination of the two computation modes. In particular, the section of the array with input voltages near the output voltage (near the center of intensity) will compute a weighted mean, whereas the parts distant from the center of intensity will compute a weighted median.

2.3 Circuit Implementation

We shall present the circuit details necessary to implement a centroid detector. The current inputs to this circuit are generated by an array of photoreceptors. We shall describe the photoreceptors used in this system, and shall show how these elements, combined with the previous circuitry, can be used to implement both one- and two-dimensional centroid detectors for visual images.

2.3.1 The Photoreceptor

We can implement the photoreceptors in our system using either photodiodes or phototransistors. Photodiodes are implemented using the junctions formed between the substrate and the active regions of opposite doping. Phototransistors are implemented using the vertical parasitic bipolar transistor existing in the standard CMOS fabrication process [Mea85].

In a phototransistor or photodiode of this type, at a given wavelength of light,

$$i_{\text{photo}} \propto I \tag{2.7}$$

where i_{photo} is the current through the device (the emitter current of the phototransistor) and I is the light intensity applied to the device.

For receptors of similar sizes, we were able to obtain a much higher coverage factor (ratio of light collecting area to total area) with photodiodes than we were with phototransistors. For this reason, we chose to use photodiodes in our design. A close packing of receptors increases the coverage factor of the array, but allows minority carriers that are generated in the substrate near one photodiode to diffuse to neighboring diodes, thus increasing their photocurrents. This effect, known as **blooming**, rarely affects diodes more than two pixels away from the site of photocurrent generation (given the pixel sizes and light wavelengths that we normally use). Blooming effectively increases the apparent spatial extent of the stimulus to the array.

2.3.2 A One-Dimensional Visual Centroid Detector

Figure 2.8 shows a one-dimensional centroid detector with photodiode inputs. The photodiodes are placed in a regularly spaced pixel array. The resistive divider is implemented

using a resistive wire (either well or polysilicon) that runs the length of the array. Voltages V_0 and V_N are applied to the ends of the wire, and equally spaced taps are placed along the wire to encode the position of each pixel.

The differential pairs in the aggregation network are implemented using bipolar transistors. We replaced the MOS transistors with bipolar transistors because bipolar differential pairs have much smaller input offset voltages than do their MOS counterparts. The smaller offset voltages allow the circuit to have much better resolution and thus more accurate operation. One negative feature of bipolar transistors is that they exhibit a nonzero current generated at the base terminal (unlike MOS transistors, which have an infinite impedance at the gate terminal). This factor does not affect the operation of the circuit as long as the resistance of the voltage divider is small enough that the base currents have an insignificant effect on the position voltages (V_n).

2.3.3 A Two-Dimensional Array

The one-dimensional centroid circuit can be expanded to two dimensions. Because the two dimensions of the calculation are independent, the value of the intensity at any point along each axis is taken to be the sum of the currents from the receptors in a row (or column) perpendicular to that axis. The two-dimensional array can be implemented using dual-output photoreceptors and current summing along each row and column as shown in Figure 2.9. To produce a photoreceptor with two outputs, we use a three-transistor current mirror to produce two copies of the photocurrent. We then use the row and column currents as the input (bias) currents to two centroid circuits (one on each axis of the photoreceptor array), which compute the centroid of the image in each dimension.

This configuration is problematic, however; the layout area of a photoreceptor increases by more than a factor of two when the current mirror is added, reducing the resolution and the coverage factor of the array. For this reason, a second scheme of photoreceptor placement using single-output receptors can be used to tile the plane, as shown in Figure 2.10. In this array configuration, each photoreceptor contributes its current to only one axis. The receptors are alternated spatially so that the currents from adjacent receptors are added to opposing axes. When this scheme is used, only one-half of the receptors contribute current to each axis, but the effective density of receptors is increased due to the unit size reduction

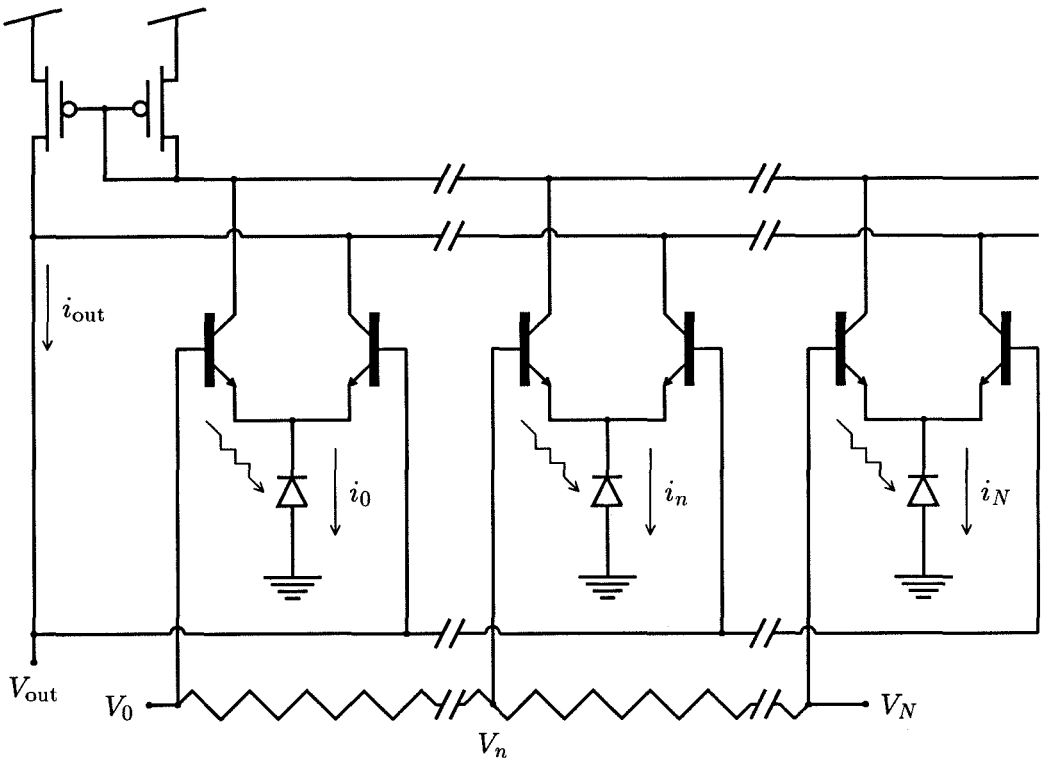

Figure 2.8: A one-dimensional centroid detector. This circuitry is used in the implementation of the centroid detector. The input image is generated by an array of photodiodes. The differential pairs are implemented using bipolar transistors because bipolar differential pairs generally have much smaller input offset voltages than do their MOS counterparts. The resistive divider is implemented using a polysilicon wire with voltages V_0 and V_N applied to the ends of the wire.

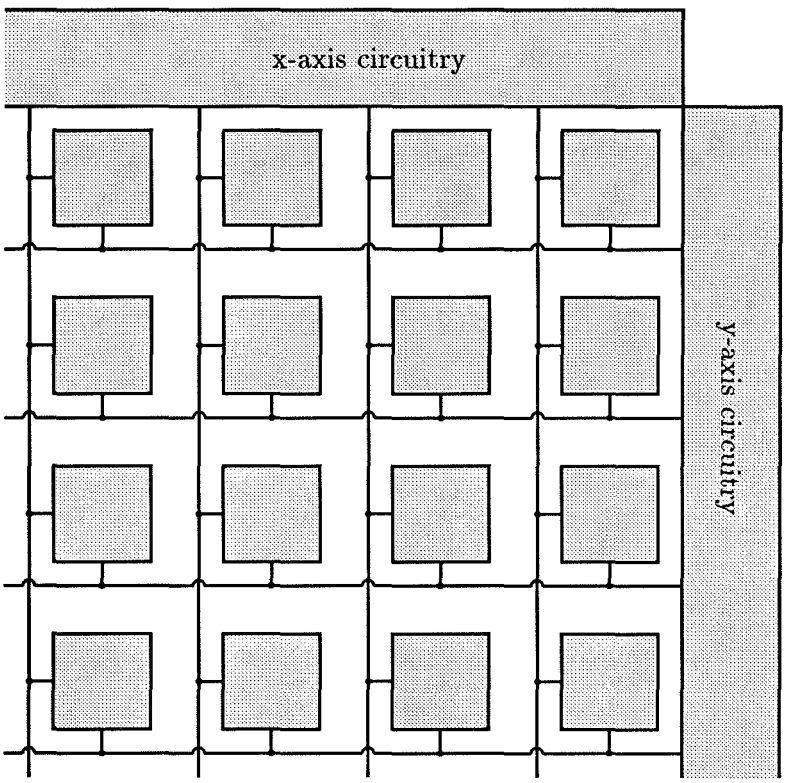

Figure 2.9: Two-dimensional configuration of dual-output photoreceptors. The photocurrents from the receptors are replicated and summed onto wires running parallel to each axis. These current sums are used as bias currents for the amplifier arrays along the edges of the receptor array. This edge circuitry performs the weighted mean/median computations.

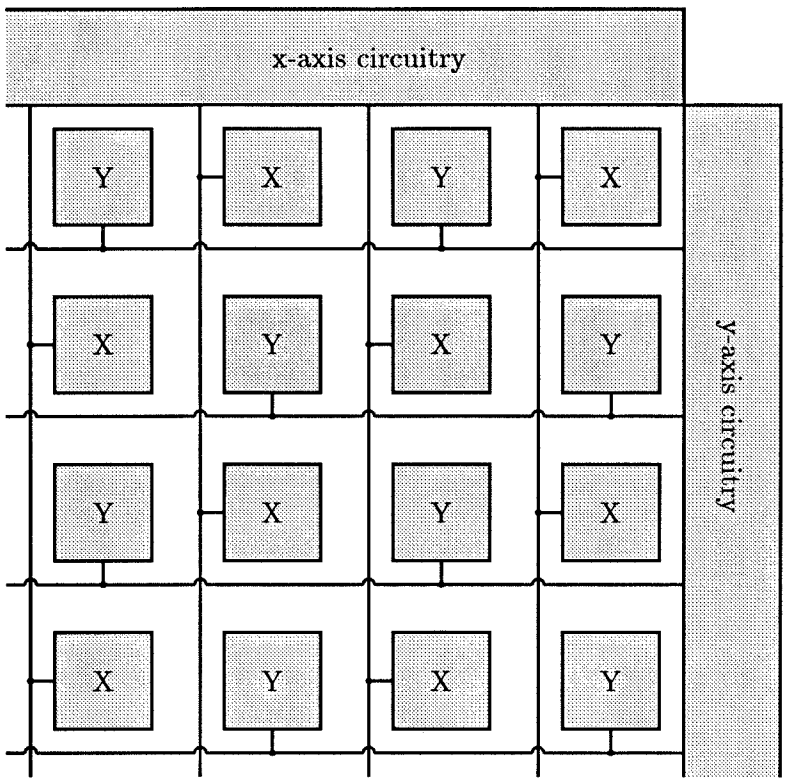

Figure 2.10: Two-dimensional configuration of single-output photoreceptors. In this configuration, each receptor consists of a single phototransistor and thus has only one output. The receptors are alternated spatially so that the currents from adjacent receptors are added to opposing axes.

made possible by the removal of the current mirror.

2.3.4 Experimental Apparatus

We studied the centroid circuit used as a sensor for localizing a bright visual stimulus (a spot of light) on a dark background. We fabricated a two-dimensional version of the circuit shown in Figure 2.8 using the alternating receptor scheme shown in Figure 2.10. The receptors were photodiodes in the substrate. We fabricated the chip with a 160 × 160-pixel array in 2 μ BiCMOS. We implemented the resistive dividers using polysilicon wires running the length of each of the two edges of the chip. The resistance of these lines was approximately 7 $k\Omega$.

To measure the response of the circuit, we placed the chip on a two-dimensional precision motion table (accurate to 0.2 μm) under a microscope. The stimulus was created by

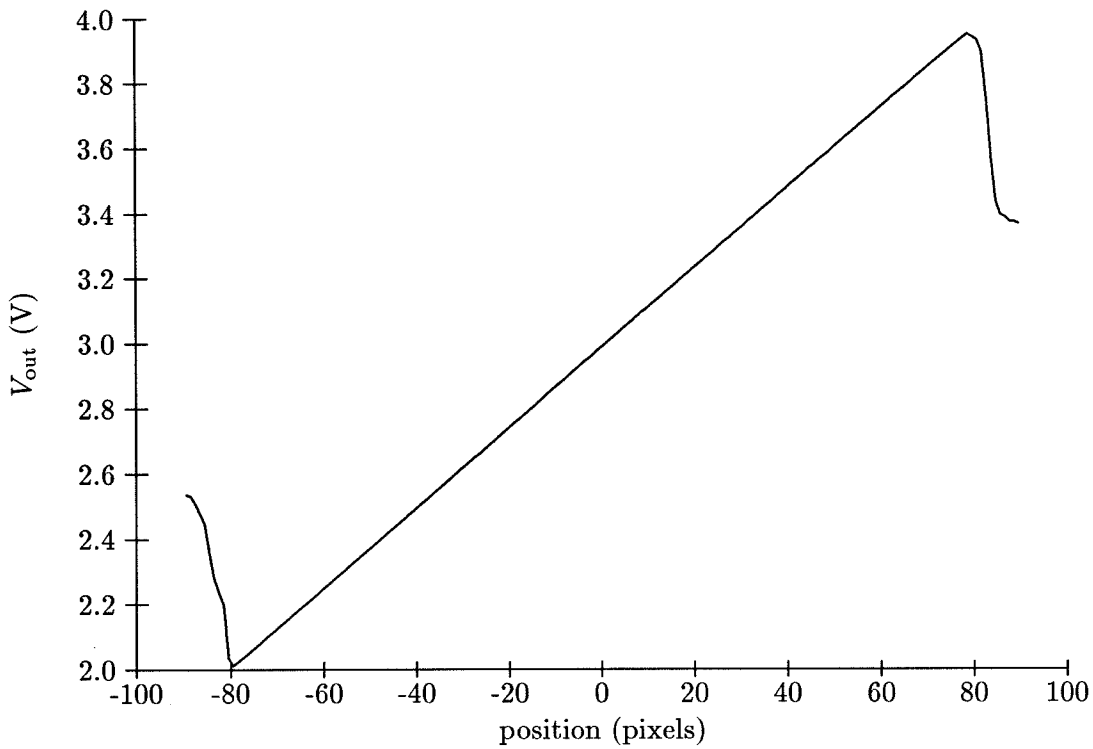

Figure 2.11: Output voltage versus stimulus position. This figure shows the output voltage as a function of the position of the image of the stimulus on the chip. The curve represents a sweep of the stimulus along one axis of the chip. As the image of the stimulus moves across the receptor surface, the output voltage increases linearly. As this stimulus passes off the edge of the surface, the output voltage returns to the uniform-intensity output value.

projection of the image of an incandescent lamp through the microscope onto the chip. The background was created with an incandescent lamp placed to the side of the microscope. We moved the stimulus across the surface of the chip by moving the motion table.

2.3.5 Stimulus Position to Output Voltage Relationship

We investigated the output response of the chip by moving the stimulus parallel to one axis of the chip (Figure 2.11). We set the voltage divider inputs (the voltage at the ends of the polysilicon wire) to 2.0 and 4.0 V. As the stimulus passed across the surface of the chip, the output voltage changed as a function of the position. At the ends of the sweep (as the image moved off the edge of the chip), the output voltage leveled off and then began to move rapidly toward the value obtained for a uniformly illuminated background.

2.4 Stimulus Localization

We are interested in using the moment chip in the localization of a bright stimulus. As shown in Figure 2.11, the chip produces an approximately linear relationship between stimulus position and output voltage. The range of the output voltage is dependent on the input voltage range, the stimulus and background intensities, and the stimulus width (in relationship to the width of the array). The output voltage range effectively sets the slope of the linear relationship between position and output voltage. We shall derive theory to predict this slope, and shall compare our theory with data taken from the chip.

2.4.1 The Continuous Approximation

To analyze the operation of the circuit, we can separate the current (i_{out}) on the output wire into multiple currents generated by multiple areas in which the illumination may be assumed to be constant. When the system is in steady state, Kirchoff's current law at the output node dictates that the sum of all of the separated currents must be equal to 0. Thus, we can analyze each of these constant illumination areas separately, and add their outputs to calculate the total system operation. Since the voltages on the resistive voltage divider are linear with respect to position, we can define the position of a point in voltage space to be the voltage on the resistive voltage divider.

We define the following variables for one region of constant illumination. The region falls on pixels (elements) n_1 to n_2, giving a width (in pixels) of $n_2 - n_1$. The position voltages (set by the voltage divider) at elements n_1 and n_2 are V_1 and V_2, respectively. The width of the region in voltage space, defined as the difference between the voltages on the voltage divider at the endpoints of the region, is $V_2 - V_1$. The total photocurrent generated by the illumination from this source is i_{in}; thus, the current from each pixel is $i_{\text{in}}/(n_2 - n_1)$. By applying these definitions, we derive the following equation for the output current from this region:

$$\begin{aligned} i_{\text{out}} &= \sum_{n=n_1}^{n_2} \frac{i_{\text{in}}}{n_2 - n_1} \tanh\left(\frac{V_n - V_{\text{out}}}{2U_{\text{t}}}\right) \\ &= \frac{i_{\text{in}}}{V_2 - V_1} \sum_{n=n_1}^{n_2} \frac{V_2 - V_1}{n_2 - n_1} \tanh\left(\frac{V_n - V_{\text{out}}}{2U_{\text{t}}}\right) \end{aligned}$$

When the voltage difference $((V_N - V_0)/N)$ between adjacent elements on the resistive voltage divider is smaller than $2U_t$, we can approximate the the piecewise linear function described by the preceding summation as a continuous hyperbolic tangent. In this case, we make the following continous integral approximation to the summation:

$$i_{out} = \frac{i_{in}}{V_2 - V_1} \int_{V_1}^{V_2} \tanh\left(\frac{V - V_{out}}{2U_t}\right) dV \tag{2.8}$$

We shall use this continuous approximation in our analysis of the operation of the circuit. The maximum voltage that we shall place across the resistive divider is 3.2 V. Since $N = 160$, the maximum voltage difference between adjacent elements is thus 20 mV, which is less than $2U_t$.

2.4.2 The Solution in the Continuous Limit

We can use Equation 2.8 to analyze the function of the moment circuit when presented with a constant illumination stimulus and a constant illumination background. We restrict the light coming into the circuit such that the current generated by the background is constant over the entire array. We define i_B to be the total background current, and define i_S to be the total stimulus current. Thus, the current in the region of the stimulus is a combination of both stimulus and background current. We define the voltage at the center of the voltage divider (the center of the background) to be 0, and the voltages at the endpoints to be $-V_B/2$ and $V_B/2$, giving a total background voltage range for the background of V_B. We define the voltage at the center of the stimulus (with reference to the voltage divider) to be V_{in} and the voltage range of the stimulus to be V_S. These definitions are illustrated in Figure 2.12. We also define α to be the ratio of the total stimulus current to total background current, and β to be the contrast ratio (the ratio of the stimulus current per distance to the background current per distance), as described by the following equations:

$$\alpha \equiv \frac{i_S}{i_B} \qquad\qquad \beta \equiv \frac{i_S/V_S}{i_B/V_B}$$

We shall analyze separately the aggregated currents generated by the stimulus and the background. Since the sum of all of the output currents must be equal to 0, we combine these equations by adding the values of the two aggregated currents and setting the sum

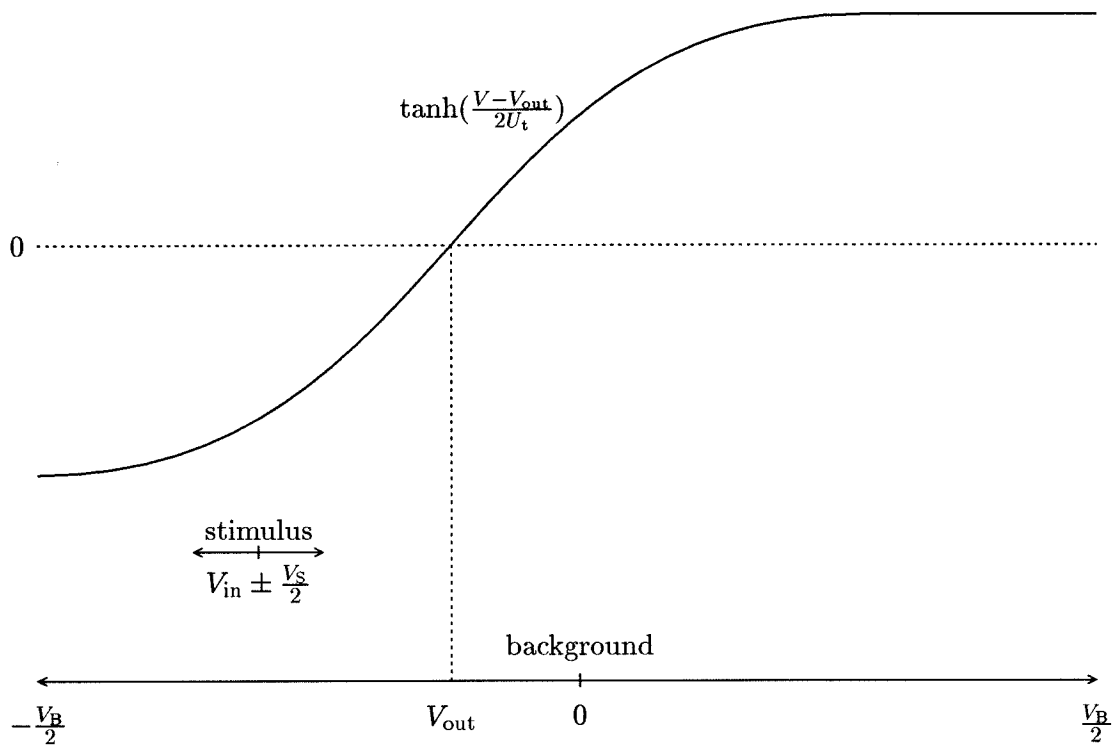

Figure 2.12: Illustration of stimulus localization parameters. The image is divided into stimulus and background, both of which have constant illumination. The voltage at the center of the background (the midpoint voltage on the resistive voltage divider) is defined as 0. The width of the background in voltage space (the difference between the endpoint voltages) is V_B. V_{in} is defined as the center of the stimulus, which has width of V_S. The output voltage V_{out} defines the position of the center of the aggregated transfer function, which is a hyperbolic tangent.

equal to 0. This separation is possible by linear superposition because, for a voltage range $[V_1, V_2]$ covered by both the stimulus and the background, the value of the output current is

$$\begin{aligned} i_{\text{out}} &= \frac{i_{\text{s}}}{V_2 - V_1} \int_{V_1}^{V_2} \tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV + \frac{i_{\text{B}}}{V_2 - V_1} \int_{V_1}^{V_2} \tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV \\ &= \frac{i_{\text{s}} + i_{\text{B}}}{V_2 - V_1} \int_{V_1}^{V_2} \tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV \end{aligned}$$

The equation for the stimulus (from Equation 2.8) is

$$\begin{aligned} i_{\text{out}_{\text{S}}} &= \frac{i_{\text{S}}}{V_{\text{S}}} \int_{V_{\text{in}} - V_{\text{S}}/2}^{V_{\text{in}} + V_{\text{S}}/2} \tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV \\ &= \frac{i_{\text{S}}}{V_{\text{S}}} \ln\left(\frac{e^{(V_{\text{in}} + V_{\text{S}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(-V_{\text{in}} - V_{\text{S}}/2 + V_{\text{out}})/2U_{\text{t}}}}{e^{(V_{\text{in}} - V_{\text{S}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(-V_{\text{in}} + V_{\text{S}}/2 + V_{\text{out}})/2U_{\text{t}}}}\right) \end{aligned}$$

The equation for the background (from Equation 2.8) is

$$\begin{aligned} i_{\text{out}_{\text{B}}} &= \frac{i_{\text{B}}}{V_{\text{B}}} \int_{-V_{\text{B}}/2}^{V_{\text{B}}/2} \tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV \\ &= \frac{i_{\text{B}}}{V_{\text{B}}} \ln\left(\frac{e^{(-V_{\text{B}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(V_{\text{B}}/2 + V_{\text{out}})/2U_{\text{t}}}}{e^{(V_{\text{B}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(-V_{\text{B}}/2 + V_{\text{out}})/2U_{\text{t}}}}\right) \end{aligned}$$

Since $i_{\text{out}_{\text{B}}} + i_{\text{out}_{\text{S}}} = 0$ at equilibrium, we set $i_{\text{out}_{\text{S}}} = -i_{\text{out}_{\text{B}}}$. By moving the constants inside of the logarithms (as exponentials) and exponentiating both sides of the equation, we derive the following equation for the system.

$$\left(\frac{e^{(V_{\text{in}} + V_{\text{S}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(-V_{\text{in}} - V_{\text{S}}/2 + V_{\text{out}})/2U_{\text{t}}}}{e^{(V_{\text{in}} - V_{\text{S}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(-V_{\text{in}} + V_{\text{S}}/2 + V_{\text{out}})/2U_{\text{t}}}}\right)^{\beta} = \left(\frac{e^{(-V_{\text{B}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(V_{\text{B}}/2 + V_{\text{out}})/2U_{\text{t}}}}{e^{(V_{\text{B}}/2 - V_{\text{out}})/2U_{\text{t}}} + e^{(-V_{\text{B}}/2 + V_{\text{out}})/2U_{\text{t}}}}\right) \tag{2.9}$$

We have not discovered a closed-form solution to this equation for V_{out}. We can, however, make simplifications to our system to derive approximate answers for the slope of the V_{out} versus V_{in} curves for specific cases.

2.4.3 Small Differential-Voltage Approximation

When the voltage (V_{B}) across the voltage divider is less than $2U_{\text{t}}$, all of the elements in the array are working within the central, unsaturated region of the hyperbolic tangent. In this

operating regime, we make the following linear approximation to the hyperbolic tangent:

$$\tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) \approx \frac{V - V_{\text{out}}}{2U_{\text{t}}}$$

We insert this approximation into Equation 2.8 to derive the following equations for the stimulus and background currents:

$$\begin{aligned} i_{\text{out}_{\text{S}}} &= \frac{i_{\text{S}}}{V_{\text{S}}} \int_{V_{\text{in}} - V_{\text{S}}/2}^{V_{\text{in}} + V_{\text{S}}/2} \left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV \\ &= \frac{i_{\text{S}}}{2U_{\text{t}}}(V_{\text{in}} - V_{\text{out}}) \end{aligned}$$

$$\begin{aligned} i_{\text{out}_{\text{B}}} &= \frac{i_{\text{B}}}{V_{\text{B}}} \int_{-V_{\text{B}}/2}^{V_{\text{B}}/2} \left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) dV \\ &= \frac{i_{\text{B}}}{2U_{\text{t}}}(-V_{\text{out}}) \end{aligned}$$

If we set the sum of these two aggregated currents equal to 0, we can derive the following equation relating the slope of the resulting line to the total stimulus and background currents:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{i_{\text{S}}}{i_{\text{S}} + i_{\text{B}}} = \frac{\alpha}{\alpha + 1} \tag{2.10}$$

This equation describes the asymptotic limit approached as the voltage across the resistive voltage divider is decreased toward 0. This equation is equivalent to the solution for the analysis of a discrete linear resistor array [DM88].

2.4.4 Large Differential-Voltage Approximation

When the voltage (V_{B}) across the voltage divider is much larger that $2U_{\text{t}}$, the differential voltages across most of the elements in the array are very large and the hyperbolic tangent is saturated. In this regime of operation, we make the following piecewise-linear approximation to the hyperbolic tangent:

$$\tanh\left(\frac{V - V_{\text{out}}}{2U_{\text{t}}}\right) \approx \begin{cases} -1 & V < V_{\text{out}} \\ 0 & V = V_{\text{out}} \\ 1 & V > V_{\text{out}} \end{cases}$$

Since the output voltage must lie within the background voltage range, we can combine this approximation with Equation 2.8 to derive the following equation for the background current:

$$\begin{aligned} i_{\text{out}_\text{B}} &= \frac{i_\text{B}}{V_\text{B}} \left(\int_{-V_\text{B}/2}^{V_\text{out}} (-1) dV + \int_{V_\text{out}}^{V_\text{B}/2} (1) dV \right) \\ &= \frac{i_\text{B}}{V_\text{B}} (-V_\text{out}) \end{aligned}$$

To analyze the current generated by the stimulus, we assume that the complete stimulus falls within the array, and that the output voltage lies within the stimulus voltage range ($V_\text{in} - \frac{V_\text{S}}{2} \leq V_\text{out} \leq V_\text{in} + \frac{V_\text{S}}{2}$). Given these assumptions, we derive the following equation describing the current generated by the stimulus:

$$\begin{aligned} i_{\text{out}_\text{S}} &= \frac{i_\text{S}}{V_\text{S}} \left(\int_{V_\text{in}-V_\text{S}/2}^{V_\text{out}} (-1) dV + \int_{V_\text{out}}^{V_\text{in}+V_\text{S}/2} (1) dV \right) \\ &= \frac{i_\text{S}}{V_\text{S}} (V_\text{in} - V_\text{out}) \end{aligned}$$

We can set the sum of these two currents to 0 and derive the following equation for the slope of the linear relationship between the input voltage and the output voltage:

$$\frac{V_\text{out}}{V_\text{in}} = \frac{i_\text{S}/V_\text{S}}{i_\text{S}/V_\text{S} + i_\text{B}/V_\text{B}} = \frac{\beta}{\beta + 1} \tag{2.11}$$

This equation describes the asymptotic limit approached as the voltage across the resistive voltage divider is increased. This equation is equivalent to the solution for the analysis of a discrete array of threshold elements (with the same functionality that we described in this section) [DM88].

Using this equation, we can verify our initial assumption that the output voltage falls within the range of the stimulus voltage. For $V_\text{in} \geq 0$, the following inequality must be true for this assumption to hold:

$$V_\text{out} = \frac{\beta}{\beta + 1} V_\text{in} \geq V_\text{in} - \frac{V_\text{S}}{2}$$

Since the difference between the input voltage and the output voltage is largest near the edges of the array, we check the assumption for the input voltage at the edge of the array

($V_{\text{in}} = V_{\text{B}}/2$). By making this replacement and rearranging the previous equation, we can derive the following relationship between the stimulus and background widths:

$$V_{\text{S}} \geq (1 - \alpha)V_{\text{B}}$$

Thus, if the stimulus intensity is at least as large as the background intensity ($\alpha \geq 1$), the output voltage will always fall within the stimulus voltage range.

2.4.5 Intermediate Approximation

The approximations in Sections 2.4.3 and 2.4.4 define asymptotic values for the slope of the input–output voltage curve for small and large differential voltages across the voltage divider. We would like to be able to approximate this slope for intermediate differential voltages.

In the intermediate region, we assume that the stimulus is small with respect to the hyperbolic tangent, thus we can treat it as a point stimulus. This assumption is valid for narrow stimuli or for wide hyperbolic tangents (small differential voltages across the voltage divider) in which the function can be approximated as linear over a narrow region. The output current (the difference between the two aggregated differential-pair output currents) from a point stimulus is

$$i_{\text{out}_{\text{S}}} = i_{\text{S}} \tanh\left(\frac{V_{\text{in}} - V_{\text{out}}}{2U_{\text{t}}}\right)$$

To compute the output current from the background, we assume that the magnitude of the output voltage is small enough that the central (unsaturated) region of the hyperbolic tangent does not extend to the ends of the array. Given this assumption, the currents from the unsaturated region sum to 0 (because the function is symmetic about the origin). Thus, we are only concerned with the current generated by the saturated region, and we can approximate the hyperbolic tangent as a threshold function. The output current (from Section 2.4.4) is

$$i_{out_{\text{B}}} = \frac{i_{\text{B}}}{V_{\text{B}}}(-V_{out})$$

By setting the sum of these two currents equal to 0, we derive the following equation:

$$i_S \tanh\left(\frac{V_{in} - V_{out}}{2U_t}\right) - i_B \frac{V_{out}}{V_B} = 0$$

We rearrange this equation to derive the following equation for V_{in}:

$$V_{in} = V_{out} + 2U_t \tanh^{-1}\left(\frac{V_{out}}{\alpha V_B}\right)$$

Since we want to determine the slope of the V_{out} versus V_{in} curve, we can take the derivative of this equation with respect to V_{out} to arrive at the following equation:

$$\frac{dV_{in}}{dV_{out}} = 1 + \frac{2U_t}{\alpha V_B} \frac{1}{1 + \left(\frac{V_{out}}{\alpha V_B}\right)}$$

From this equation, we see that the derivative when $V_{out} \approx 0$ (near the center of the array) can be approximated by the following equation:

$$\frac{dV_{in}}{dV_{out}} \approx 1 + \frac{2U_t}{\alpha V_B}$$

Finally, we invert this equation to get the following equation for the approximation to the slope of the V_{out} versus V_{in} curve:

$$\frac{dV_{out}}{dV_{in}} \approx \frac{V_B}{V_B + 2U_t/\alpha}$$

This equation is a reasonable approximation to the true slope when the voltage across the voltage divider is large enough that the unsaturated regions of the hyperbolic tangent do not reach the ends of the array.

2.4.6 Experimental Data

We measured the slope of the output voltage versus stimulus position curves as a function of the differential input voltage (the voltage across the resistive voltage divider). We took data for differential input voltages in the range of 10 mV to 3.2 V.

We held the stimulus and background intensities constant (for all of the curves) such that the total stimulus current (i_S) was 1.24 nA and the total background current (i_B) was

1.24 nA. The width of the stimulus was 3.0 pixels. From these values we computed $\alpha = 1.0$ and $\beta = 0.019$. From the analysis in Sections 2.4.3 and 2.4.4, we can compute values for the asymptotes for small and large values of the differential input voltage V_B. The asymptote for small currents (from Equation 2.10) is

$$\frac{\alpha}{\alpha + 1} = 0.50$$

The asymptote for large currents (from Equation 2.11) is

$$\frac{\beta}{\beta + 1} = 0.98$$

Figure 2.13 shows the data relating differential input voltage to normalized slope for this experiment. We computed the normalized slope by dividing the slope of output voltage versus stimulus position curves by the differential input voltage, and then dividing by the slope of a curve taken when the background current was equal to 0. (This normalizes the data to compensate for any discrepancies in the scaling of the position or the endpoint voltages of the voltage divider.)

Along with the data we plotted the horizontal asymptotes at 0.50 and 0.98. The data for small voltages fit the lower asymptote quite well. The data for large voltages approached the upper asymptote. We also show the theoretical curve from the analysis in Section 2.4.5. As expected the theory fits the data very poorly for small values of V_B where the hyperbolic tangent is not saturated. For large values of V_B the curves are similar. The theoretical values are larger than the experimental values because of the point stimulus assumption. The upper asymptote for a curve taken with a point stimulus has a value of 1, and thus the theoretical curve is slightly larger than the data curve, which approaches an upper asymptote of 0.98.

2.5 Error Analysis

When the centroid circuitry is used to localize a stimulus, the ideal response would be a perfectly linear relationship between the position of the stimulus and the output voltage. However, imperfections in the devices lead to errors that show themselves as non-linearities

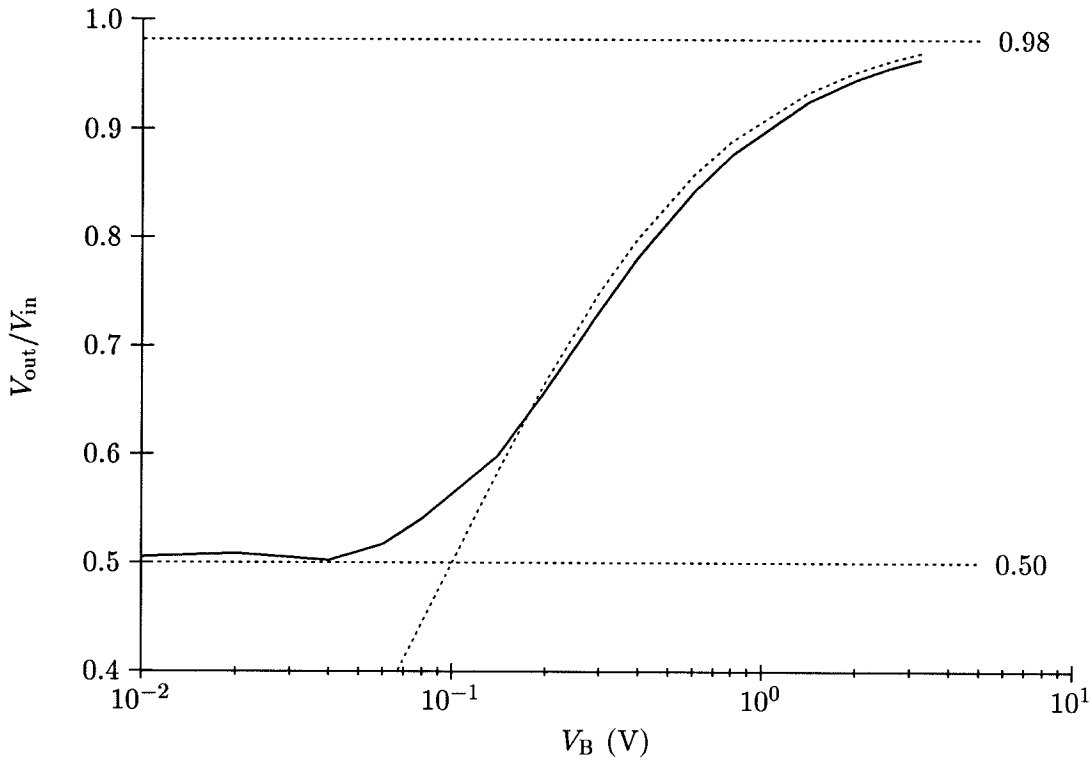

Figure 2.13: Effects of varying the voltage across the resistive divider (V_B). We took data for a number of differential input voltages by moving a stimulus across the chip and measuring the output voltage. The solid line represents the data taken for V_B in the range of 10 mV to 3.2 V. The horizontal lines represent the asymptotes for small and large values of V_B. The dotted curve was generated from the theory in Section 2.4.5. Since V_{out}/V_{in} approaches asymptotes as V_B approaches 0 and ∞, we plotted the horizontal axis on a logarithmic scale.

in the transfer functions. There are four primary device nonidealities and resulting effects:

- Variations among the photodiodes (or phototransistors) cause the individual diodes to output different photocurrents for the same applied light intensity.
- The finite gain of the feedback causes an output offset voltage.
- Mismatches between the differential pair transistors (due to variations in κ and i_0) cause fixed offset voltages at each node.
- Nonuniformities in the resistivity of the resistive divider lead to variations in the voltage-encoded position.

One metric for studying the effects of offsets and mismatches and their effects on localizing a stimulus is the amount of inaccuracy in the apparent position of the stimulus with respect to the true position of the stimulus. Since the ideal transfer function is linear, we can measure error as the deviation from the best-fit line to the position–voltage curve. This line is fixed for a given stimulus-to-background ratio and differential input voltage (across the resistive divider). We shall compute all offsets in pixels for the chip being tested, and shall use the worst-case analysis whenever there are multiple options. When measuring offsets, we shall delete the data from the outer 20 pixels at each end of the array, to account for effects (such as decreased photocurrents) that decrease the precision of the circuitry in that region.

2.5.1 The Distribution of Photocurrents

We can study the photocurrents generated on the chip by measuring the input current to the circuit while moving a fixed-intensity, single-pixel light source over one row of the chip perpendicular to the axis of measurement (to ensure that all other circuitry is the same). We acquire the value of the photocurrent from a transistor whose gate and source are connected to the common gate and source of the current mirrors. The output current from this transistor is a scaled version of the current through the current mirror. The current through the current mirror is equal to one-half of the photocurrent, because the feedback causes the two aggregated currents to be equal in steady state. Thus, this output current is a scaled version of the photocurrent generated by the chip.

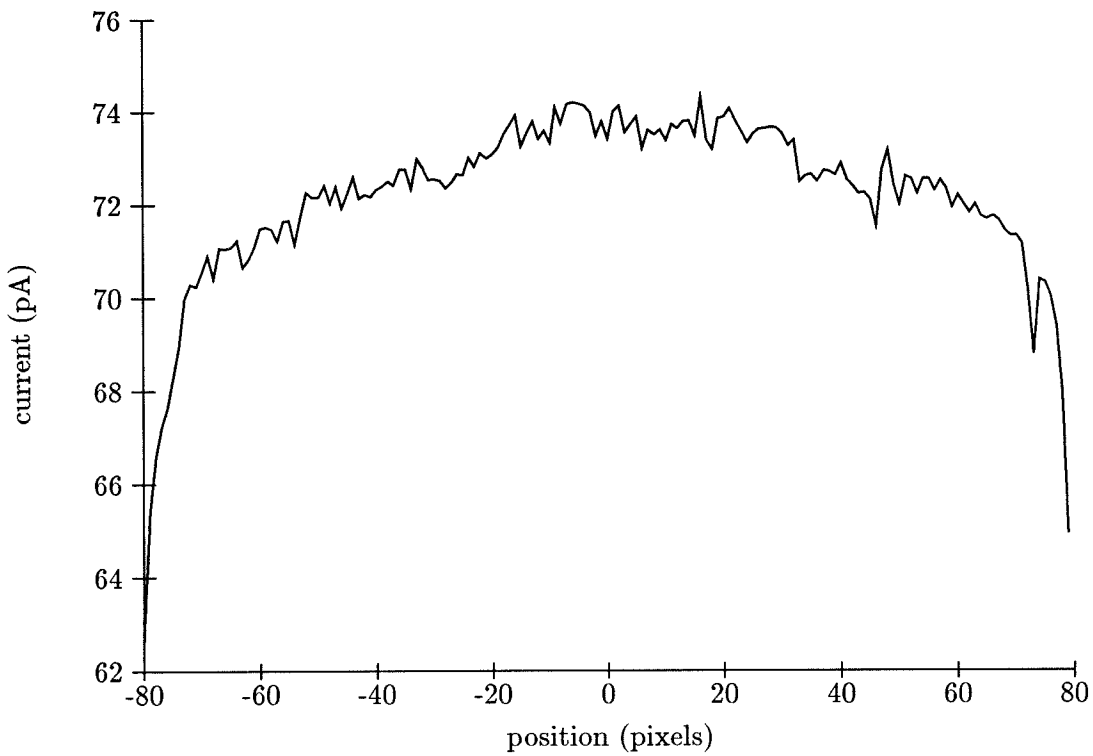

Figure 2.14: Photocurrent versus position of stimulus on chip. We measure the photocurrent generated by the circuit while moving a fixed-intensity, single-pixel light source over one row of the chip perpendicular to the axis of measurement. We acquire the value of the current using a transistor that replicates the current flowing through the current mirror. In steady state, this current is equal to one-half of the total photocurrent.

Figure 2.14 shows a plot of photocurrent data from one row of photodiodes. The curve shows an obvious increase in current toward the center of the chip. This variation in current probably is due to two effects that cause the stimulus not to appear as a point light source. Minority carrier diffusion [Sze81] causes blooming (the spread of photocurrent to neighboring pixels). This effect has an extent of less than $50\,\mu m$, and thus affects only nearby pixels. The second effect is the scattering of the light as that light passes through the optics in the microscope. This effect has the potential for a much wider spread, and thus can explain the general current variations over large numbers of pixels.

We can quantify the effects of this intensity spread by comparing the total photocurrents when the stimulus is at the edges of the array to those when the stimulus is in the center of the array. The current at each end of the array includes the current generated by the stimulus plus one-half of the distributed current, whereas the current at the center of the array should include all of the available current. Thus, if we add the currents at the two edges and subtract the center current from this sum, the resulting current should be equal to the current generated by the pixel on which the stimulus is placed. For the data shown in Figure 2.14, this equation is $63pA + 65pA - 74pA = 54pA$. This value indicates that approximately 73 percent of the current generated in the array comes from the column at the stimulus. This distribution of current causes a statistical averaging of errors on the chip, thus affecting the magnitude of the measured errors. In the worst case, this averaging would cause the measured error to be approximately 73 percent of the real error. We shall consider this fact when we analyze the errors from the circuitry.

2.5.2 Photodiode Mismatches

Figure 2.14 showed the distribution of photocurrents as a single-pixel stimulus was swept across the photodiode array. We computed the relative error for the central 120 pixels (75 precent) of the array using the average current as the expected value. Figure 2.15 is a plot of the percent relative error. Even in this region, there still is a tendency for the currents near the center to be larger than those near the edges. Even with this effect, however, the maximum error in this region is approximately 2 percent. The maximum local error (difference between two adjacent currents) is approximately 1 percent, so including the variation due to position increases the error by no more than a factor of two. Factoring in

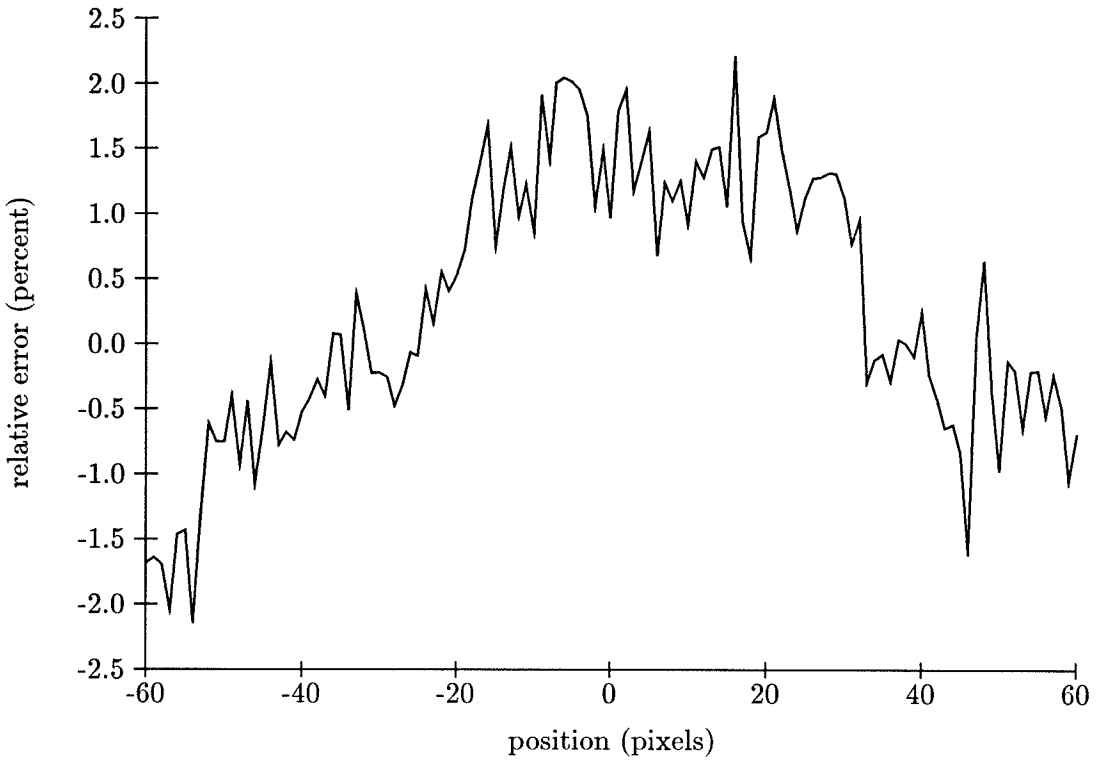

Figure 2.15: Relative error in the photocurrents. Using the data from Figure 2.14, we have generated a plot of the error relative to the average current value for the pixels in the center of the array. The variation is less than 2 percent.

the worst-case averaging due to the distributed photocurrents, we obtain a relative error of approximately 3 percent.

To determine the effect of these current variations on the operation of the system, we can convert the current variations to variations in the position (in pixels). The amount of positional error generated by these offsets is dependent on both the voltage across the resistive divider and the ratio of stimulus to background intensity. The effect of the voltage across the resistive divider is due to the functional variation (weighted mean or median) produced by variation of this voltage. Since the variation of a particular photocurrent has a much larger effect on the weighted-mean computation (and, from Equation 2.11, practically no effect on the weighted-median computation as long as the stimulus is larger than the background), we shall analyze the effects of photocurrent variation for the case of the weighted mean.

We can analyze the error by considering the case where there is a stimulus with a total

current i_S at some position x, a background with a total current i_B centered at the origin (position = 0), and a stimulus to background intensity ratio α. For the weighted mean, the output position would then be

$$x_{out} = \frac{i_S \cdot x + i_B \cdot 0}{i_S + i_B} = \frac{\alpha \cdot x}{1 + \alpha}$$

A relative error of ϵ in the intensity of the stimulus would give the following equation for the position:

$$x_{error} = \frac{(1+\epsilon)\alpha \cdot x}{1 + (1+\epsilon)\alpha}$$

We can compute the relative error in position due to the offset by taking the difference between these two values and dividing by the position of the stimulus, giving

$$\begin{aligned} \frac{x_{error} - x_{out}}{x} &= \frac{(1+\epsilon)\alpha}{1 + (1+\epsilon)\alpha} - \frac{\alpha}{1+\alpha} \\ &= \frac{\epsilon}{(1+\epsilon)\alpha + \frac{1}{\alpha} + (2+\epsilon)} \end{aligned}$$

From this equation, we can see that the largest positional error for a given error in photocurrent is produced when the stimulus and background intensities are equal ($\alpha = 1$). When the background intensity is very small with respect to the stimulus intensity, the variation in photocurrent will have practically no effect on the output voltage from the circuit, since this voltage is approximately equal to the input voltage at the center of the stimulus.

We can now derive the worst-case positional error due to the photocurrent variation. We shall use the value for ϵ of 3 percent that we computed from the data, and shall set the intensity ratio (α) equal to 1. Plugging these values into the preceding equation gives a relative positional error of approximately 0.7 percent. Since the maximum distance that a stimulus can be from the center of the array is 60 pixels (ignoring the outer 20 pixels), the maximum error is approximately 0.4 pixels. Since we used the worst-case analysis at every point in our computation, this value is much larger than it would be in normal operation, especially when either the contrast ratio or the voltage across the resistive divider is large.

2.5.3 Finite-Gain Offset

We can measure the offset due to the finite gain of the feedback by setting the endpoints of the resistive line to the same voltage and presenting to the chip a constant background intensity. We sweep the input voltage (the voltage on the resistive wire) and measure the output voltage. The other sources of error are statistical and should cancel each other out, leaving only the finite-gain offset.

In this circuit configuration, the corresponding nodes of the individual elements are all connected and the current mirror feeds back the differential pair. Thus, the complete system can be analyzed as if it were a single transconductance amplifier configured as a voltage follower. There is an additional amplifier configured as a voltage follower at the output of the chip whose input offset voltage is also included in this analysis. This output amplifier is designed so that it has a very high gain, and thus we shall not consider this gain in the analysis. We shall consider the finite gain of the circuit and the offset voltage of the output amplifier to be a single offset.

The gain A of a transconductance amplifier is set by the ratio of the transconductance to the output conductance of the amplifier [Mea89]. The relationship between the input and the output voltages in the voltage follower configuration is described by the following equation:

$$\frac{dV_{\text{out}}}{dV_{\text{in}}} = \frac{A}{1+A} \tag{2.12}$$

Thus the finite gain causes the follower to have a gain less than unity.

Figure 2.16 shows a plot of the output voltage (V_{out}) as a function of the input voltage to the ends of the resistive wire (V_{in}). This plot shows the voltage operating range of the chip that results from restrictions set by the differential pairs, current mirror, and output amplifier. The central voltage region (1.0 to 4.0 V) appears quite linear. Thus, we shall use this range of voltages for our error analysis and restrict the circuitry to operating voltages within this range.

We measure the average amplifier gain by performing a linear regression over the central region of the curve in Figure 2.16. The slope of this line is equal to the average value of $A/(1+A)$. We measured a best-fit slope of 0.985. This value gives an average gain of 66.

We measure the error due to the finite gain and the feedback amplifier by taking the

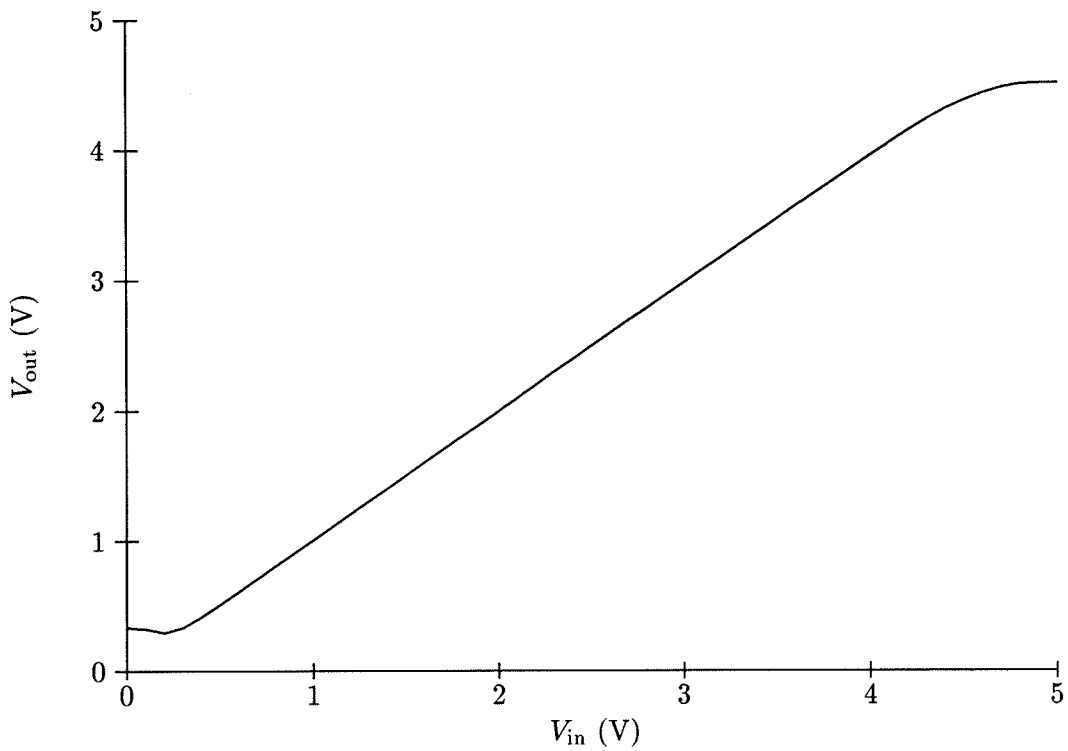

Figure 2.16: V_{out} versus V_{in} given uniform illumination. We set a uniform background illumination on the chip and swept the input voltage from 0 to 5 V while measuring the output voltage. The curve is approximately linear except at the two ends at which points the circuitry limits the output voltage range.

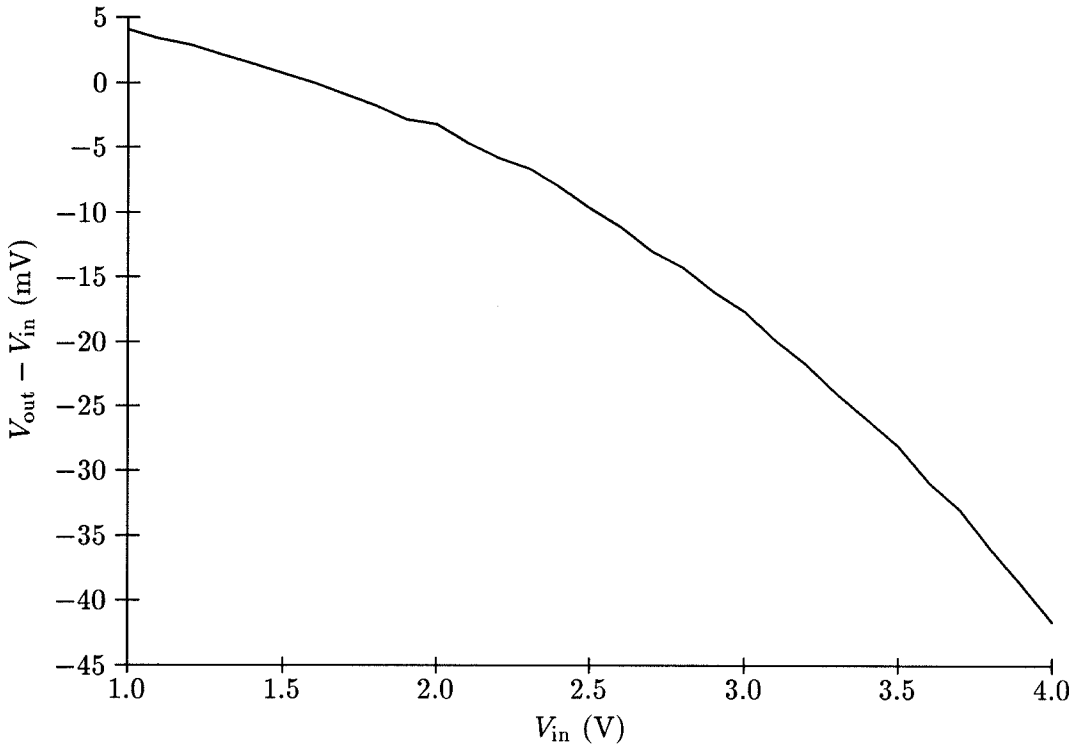

Figure 2.17: Absolute error due to the finite feedback gain. We computed the absolute error by subtracting the input voltage from the output voltage in Figure 2.16 for the input voltage range of 1.0 to 4.0 V. We plotted this error as a function of the input voltage.

difference between the output and input voltages in Figure 2.16. Figure 2.17 shows a plot of this absolute error over the input voltage range of 1.0 to 4.0 V. The absolute error changes monotonically with the input voltage. If the gain of the amplifier was constant, this error curve would be linear, and the error would have no effect on the total error of the system since we are comparing our data to a best-fit line. However, there is some curvature which denotes a variation in the gain over the input voltage range. We computed the local gains by taking local differences in output voltage divided by the input voltage difference for the same points. A smoothed version of the resulting curve showed a gain of approximately 125 at the low end of the voltage range to a gain of approximately 45 at the high end.

To see how the finite-gain error affects the circuitry in operation, we can examine the case in which the resistive divider endpoint voltages are 1.0 and 4.0 V. The deviation of the curve in Figure 2.17 from a best-fit line is in the range $\pm 6\,mV$. If we divide this value by the total range of the input voltage (3.0 V) we get a relative error range of ± 0.2 percent. If we take into account the averaging due to the distribution of photocurrent, this error is less than ± 0.5 pixels over a 160 pixel range.

We can calculate the error for any input voltage range by dividing the absolute error in that range by the voltage range and multiplying by the number of pixels (160). We can center our input ranges around 2.5 V, and set the error (relative to a best-fit line at that point) equal to zero. Since the absolute error curve is monotonic, the maximum error will then occur at one of the endpoints of the curve. We can thus divide the deviation from the best-fit line by the range at a given voltage ($2 \cdot |V_{\text{in}} - 2.5\,V|$), multiply by the number of pixels, and account for photocurrent distribution. Figure 2.18 shows the results of this computation. The maximum error of less than 1 pixel occurs at the point 4.0 V when the voltage range is 3.0 V. Note that this error range is the same as the error calculated previously (± 0.5 pixels) for the 3.0 V range.

2.5.4 Differential-Pair Offsets

We can measure the differential pair offsets by sweeping a single-pixel stimulus along the array while measuring the output voltage. The other errors can be isolated properly to make this measurement as follows. We isolate the photodiode variations by making the background intensity significantly smaller than the stimulus intensity. This restriction causes

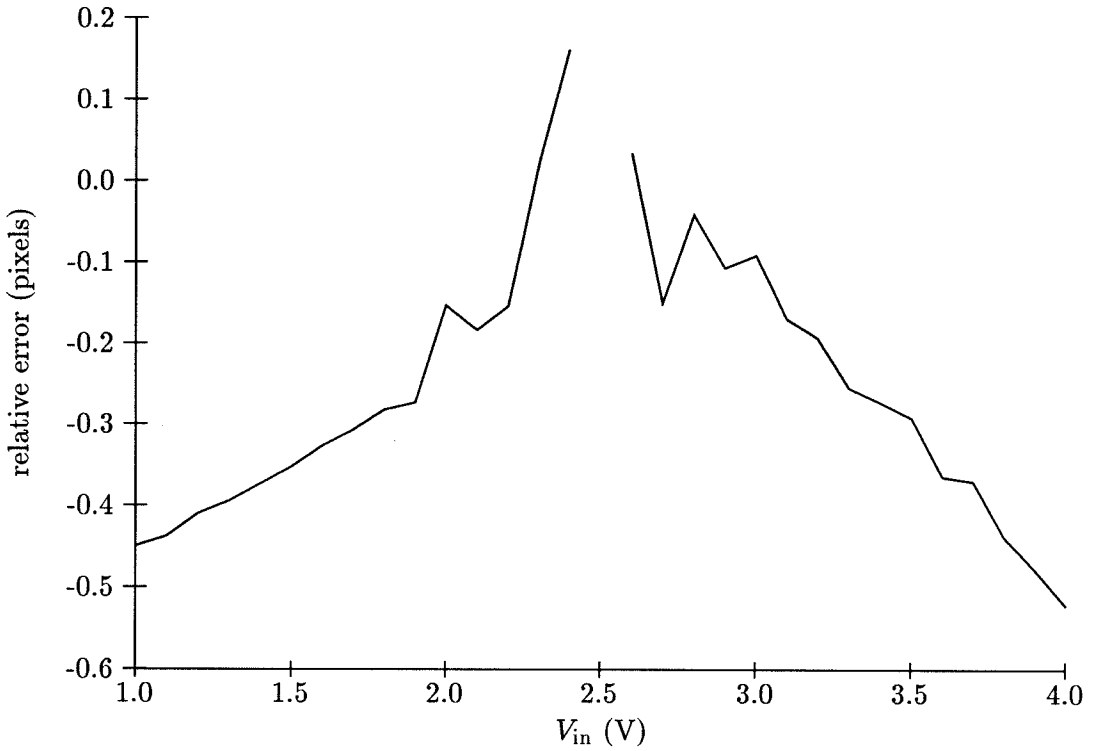

Figure 2.18: Error (in pixels) due to finite feedback gain. We computed this error from the data in Figure 2.17 by computing the error values from a best fit line (that intersects the curve at an input voltage of 2.5 V), dividing these values by twice the difference between the input voltage for the point and 2.5 V, and multiplying by the total number of pixels. The resulting curve represents a measure of the maximum error (in pixels) generated by the finite gain for differential input voltages centered about 2.5 V.

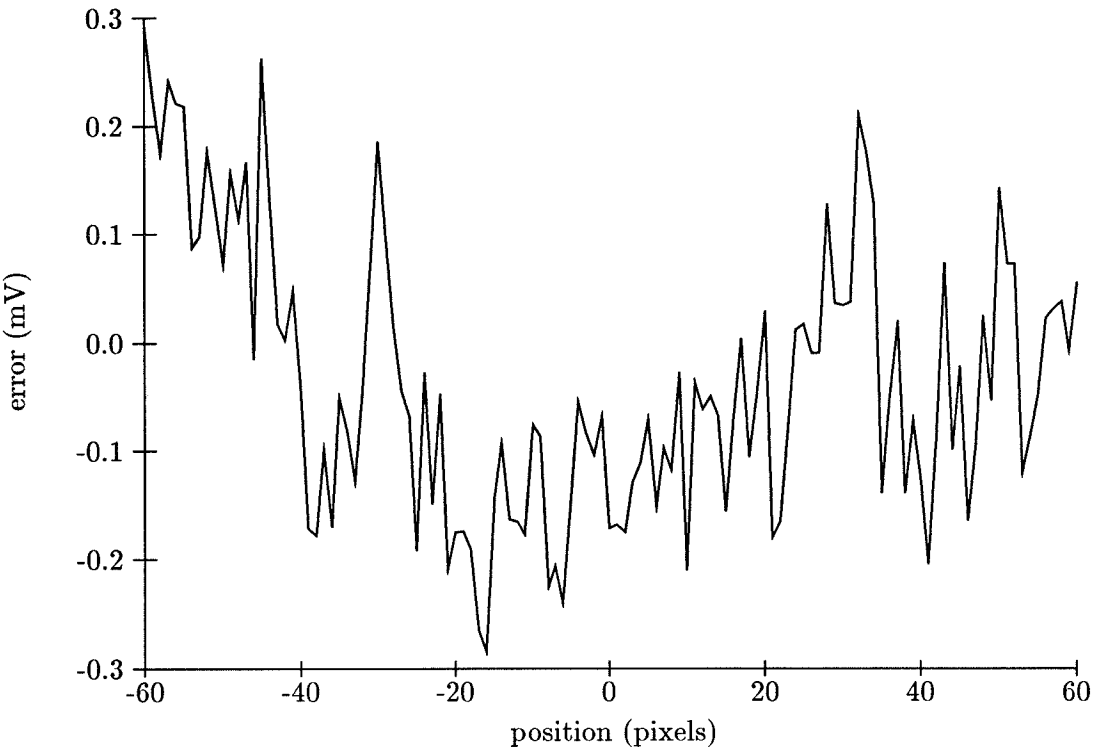

Figure 2.19: Differential-pair offset voltages. We are able to measure the offset voltages of the differential pairs by setting the ends of the resistive divider to the same voltage and moving a one-pixel stimulus across the chip. The voltages displayed are the difference between the output voltage and the input voltage (to the resistive divider). The common offset voltage ($\approx 17.3\,\mathrm{mV}$) in all of these data is due to the finite gain error. The variation around this point is due to the differential-pair offset voltages. The range of offset voltage among the differential pairs is less than $\pm 0.4\,\mathrm{mV}$.

the output voltage to be equal to the input voltage to the differential pair (modulo the offset) because only the one differential pair is active. We eliminate the resistive divider variation by setting the two ends to the same voltage. Since there is no voltage drop along the wire, the voltage along the resistive divider will be constant (and will be equal to the endpoint voltages) independent of resistivity variations. Finally, we can remove the constant offset voltage due to the finite feedback gain by subtracting the average output voltage from the data.

Figure 2.19 shows a plot of the difference between the output and input voltages (for an input voltage of $3.00\,V$) versus the position of the stimulus over a 100-pixel range. The

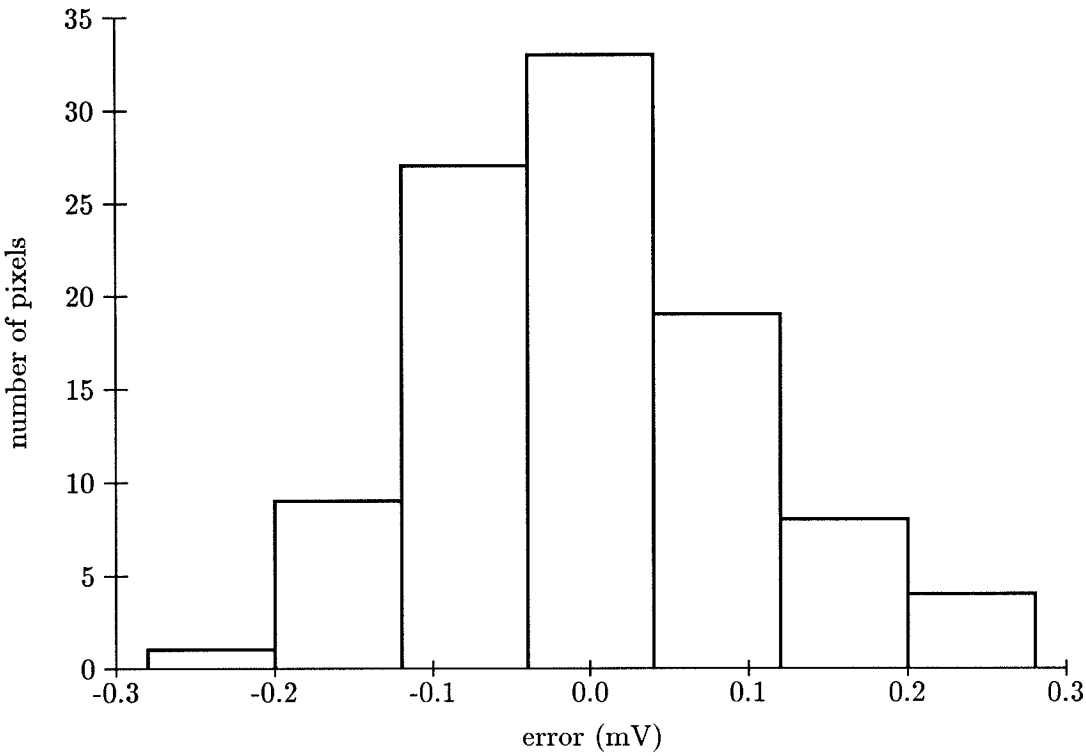

Figure 2.20: Histogram of differential-pair offset voltages. This figure is a histogram of the offset voltages seen in Figure 2.19. The average voltage is subtracted from each voltage to account for effects from the finite gain.

offsets appear to be random and to be fairly independent of position in the array. Figure 2.20 shows a histogram of the distribution of the offset voltages. The variation in the offset is less than $\pm 0.4\,\mathrm{mV}$. When the possible effect of the light-source distribution is taken into account, the range of offset voltages is less than $\pm 0.5\,\mathrm{mV}$.

We can analyze the effect of these offset voltages on the relative error of the centroid circuit by dividing the offset voltage by the voltage across the resistive divider, and multiplying by the total number of pixels. The result of this calculation is the error in pixels due to the differential-pair offset voltages. Since these offset voltages are constant (independent of the voltage range on the resistive divider), there is an inverse relationship between their effect on the relative offset in the system and the range of voltages across the resistive divider. To derive a number for this relative offset, we must first assume a value for the voltage across the resistive divider. For example, for a voltage of $1.0\,V$, the relative error from this source is less than 0.1 pixels.

2.5.5 Resistive-Divider Nonuniformities

We can study the nonuniformity of the resistive divider by placing a large differential voltage (1.0 to 4.0 V) across the resistive divider and measuring the output voltage from the chip as a single-pixel stimulus is swept across the chip. Since the error is due to the voltage drops along the resistive line, the absolute error will scale with input-voltage range, and the relative error will be independent of input-voltage range.

In Figure 2.21, we show data for the relative error of the output voltage from a best-fit line, in which the error is divided by the range of the input voltages. We also show a plot of the relative finite gain offset (with respect to a best-fit line) described previously with the horizontal axis scaled to the position of the appropriate input voltages. Most of the deviation in the data is due to the finite gain. We eliminate the effect of the finite gain by subtracting the finite-gain data from the error data. From this difference, we computed an error for the resistive divider of less than ± 0.1 pixels. This error takes into account both the differential pair offset voltages, which are relatively small at this differential voltage input range, and the resistive divider nonuniformities.

2.5.6 Overall Error

We can now analyze the overall error generated on the chip. The maximum total error is the sum of the errors from all sources. The sum of the numbers derived in the analyses in the previous sections is less than one pixel of error. Since there are 120 pixels in the region that we analyzed, the circuit has a resolution of approximately 7 bits. Since these were all worst-case estimates and the actual error can be made to be much smaller, we can expect to achieve resolutions much better than the worst-case prediction.

In this error analysis, we computed our error from a best-fit line. In a controlled situation (where the stimulus and background intensities are constant), we can compute (or measure) the slope of this best-fit line, and can use this information to calibrate the output voltage versus position curve. In this situation, the error analysis presented here is valid. If we do not know the intensity levels of the environment (or if they change with time), we cannot calibrate our system by using the best-fit line. In this situation, we must consider the effects of lighting changes on the error. In this situation, the system generates the least error when the voltage across the resistive divider is large, and the system is operating

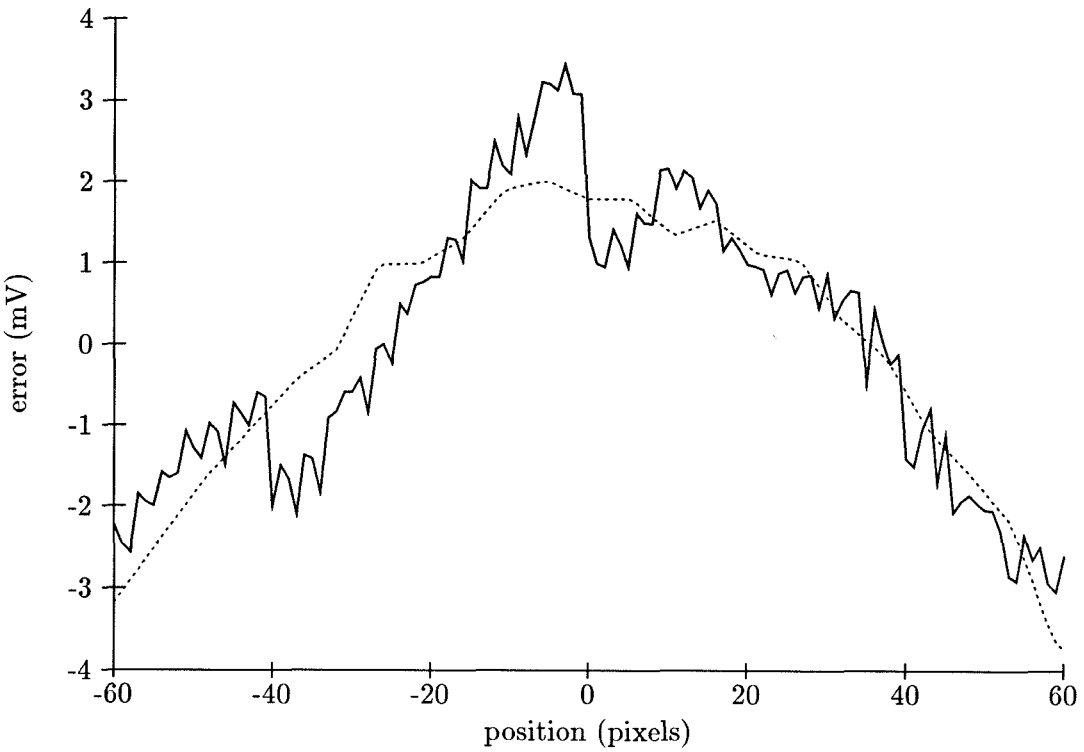

Figure 2.21: Resistive-divider offset voltages. The solid line is a plot of the relative error of the output voltage from a best-fit line, in which the error is divided by the range of the input voltages ($3.0\,V$). The dotted line is a plot of the relative finite-gain error (with respect to a best-fit line) described previously with the horizontal axis scaled to the position of the appropriate input voltages. We can see that most of the deviation shown in the solid line can be explained by the finite gain error. If we take the difference between the two curves (scaling by the number of pixels and accounting for the photocurrent distribution), we obtain an error of less than ± 0.1 pixels.

near the weighted mean functionality. If we assume a slope of 1 for the output voltage versus input position voltage line (i.e., assume $V_{out} = V_{in}$), the relative error (derived from Equation 2.11 for the weighted median) is described by the following equation:

$$\frac{V_{in} - V_{out}}{V_{in}} = \frac{1}{\beta + 1}$$

For example, if the total stimulus and background currents are equal and the stimulus width is 5 percent of the background width, the relative error is less than 5 percent.

Chapter 3

DC Motors and Pulse Control

In typical use, DC motors are controlled by analog currents or voltages. However, a DC motor may have other types of signals as inputs; in particular, it may be controlled with a binary pulse train. This type of input signal has many advantages over the typical analog control signals. These advantages include ease of amplification, resistance to noise, and ability to overcome friction for small motor velocities. Ease of amplification is important when we build small, robust systems using VLSI circuits as the controllers. In these cases, we need to amplify our signals using a minimal amount of circuitry external to the chip.

One example of the use of pulses to drive mechanical systems can be found in biological systems. The nervous systems of higher animals use pulses as a means of communicating information throughout the animal's body. It is generally accepted that this pulse representation of the information improves reliability in the transfer of information. However, nature also makes a virtue out of using pulses when controlling muscles. Muscles are complicated mechanical actuators with both damping and internal friction [WCH85] [EL90]. Thus, muscles pose a complictated control problem for which pulses were chosen as an evolutionary solution.

In this chapter, we formulate a general model for a DC motor, and present characterization data for the motors that we use in our experiments. We explain why the use of pulses is an effective method for driving motors (and other DC actuators). We discuss the effects of friction on DC systems, and explain the advantages of pulse control in these conditions.

3.1 The DC Servo Motor

Before we can analyze the effectiveness of pulses at driving DC motors, we must first develop a sufficient model of the electromechanical properties of these devices. We present a standard linear model that takes into account the basic electrical properties of the motor [Edw86]. Simplifying assumptions will be made and noted. An analysis similar to that presented here can be applied to other linear and rotational DC actuators.

3.1.1 Motor Operation

The DC servo motor is a device that converts a DC current into a torque applied to a shaft. Each motor consists of a fixed part called the *stator*, and a rotating part called the *rotor*. One of these two parts of the motor consists of a set of permanent magnets of alternating polarity; the other consists of a set of current-carrying coils. As the shaft of the motor rotates, the direction of the current through each coil is switched, depending on the position of that coil with respect to the adjacent magnets.

The two basic types of DC servo motor differ in whether the coils are part of the stator or rotor, and by what method the current switching is accomplished. In the traditional DC motor, the coils are part of the rotor and the magnets are part of the stator. The current is distributed to the rotating coils by a set of brushes on the stator. The brushes contact a set of commutators on the rotor that are connected to the coils. Each brush is connected to one of the two motor terminals, depending on the position of the brush with respect to the magnets. Thus, the motion of the rotor causes the direction of the current through each coil to change, depending on the position of the coil with respect to the magnets.

In the other type of motor, called a *brushless DC motor*, the coils are part of the stator and the magnets are part of the rotor. The mechanical switching is replaced by electrical switching; thus, the need for brushes and commutators is removed.

Aside from the control of switching, the operation of both types of motor is essentially equivalent. A magnetic field of flux density β is generated by the permanent magnets, and passes through the coils, which lie normal to the field. A voltage across the terminals of the motor causes an electric current i to flow through the coils. This current flow normal to the magnetic field causes a torque to be generated by the motor and applied to the

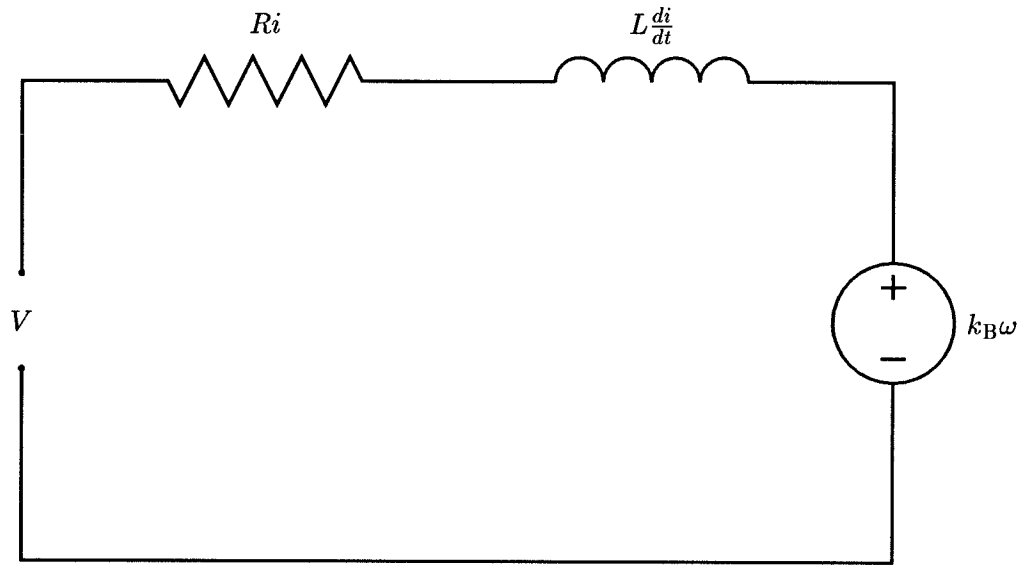

Figure 3.1: An electrical model of a DC motor. The motor consists of three primary elements: a resistor, an inductor, and a voltage source. The resistor and inductor represent the resistance and the inductance of the coil of the motor. The voltage source represents the back EMF of the motor; thus, the voltage is proportional to the angular velocity of the motor.

rotor. Rotation of the rotor causes the coils to move with respect to the magnetic field. This motion generates an electric potential across the coils. The potential is called the *back EMF* of the motor.

3.1.2 A DC Motor Model

Figure 3.1 shows a model of the electrical system for the motor. This circuit consists of three primary elements: a resistor, an inductor, and a voltage source. The resistor and inductor represent the resistance R and the inductance L of the coils of the motor. The voltage source represents the back EMF induced in the motor by the movement of the coils through the electric field. This voltage is proportional to the angular velocity ω of the

motor. By Kirchoff's voltage law, the equation for this electric circuit is

$$V = L\frac{di}{dt} + Ri + k_B\omega \tag{3.1}$$

where $k_B\omega$ is the back EMF generated by the motor.

If we neglect friction and loads, the mechanical system for the motor, from Newton's second law of motion, is described by the equation:

$$J\frac{d\omega}{dt} = k_T i \tag{3.2}$$

where J is the moment of inertia of the rotor and $k_T i$ is the torque generated by the current flowing through the coils normal to the magnetic field.

The constants k_B (the back-EMF constant) and k_T (the torque sensitivity) are both in units of V·sec (or kg·m·A^{-1}·sec^{-2}). By studying the effects that lead to the generation of torque from current flow and the generation of back EMF by magnetic-field motion [GB42], we can see that not only are the units for the two constants the same, but also the theoretical values of the constants are equal. Given coils with a radius of r, N total number of turns for all coils active at any given time, and a rotor with an effective radius (distance from center of rotor to center of coils) of d, we can derive the relationship

$$k_B = k_T = 2\pi r \cdot N \cdot \beta \cdot d$$

This constant is called the *electromechanical-coupling constant* of the system.

We can combine Equations 3.2 and 3.1 by solving for i in Equation 3.2 and substituting this solution into Equation 3.1:

$$V = \frac{LJ}{k_T}\frac{d^2\omega}{dt^2} + \frac{RJ}{k_T}\frac{d\omega}{dt} + k_B\omega$$

From this equation, we can derive the Laplace transfer function for the motor from input voltage to output angular velocity:

$$\frac{\omega}{V} = \frac{1/k_B}{\frac{LJ}{k_B k_T}s^2 + \frac{RJ}{k_B k_T}s + 1} \tag{3.3}$$

Equation 3.3 describes a second-order system in which one time constant is a function of the electrical system and the other time constant is a function of the mechanical system. The electrical time constant τ_e is a result of the resistance and the inductance of the coils, and is defined by the equation:

$$\tau_e = \frac{L}{R} \tag{3.4}$$

If we assume that the mechanical time constant τ_m is much larger than the electrical time constant, we can approximate the mechanical time constant by the equation:

$$\tau_m = \frac{R \cdot J}{k_B \cdot k_T} \tag{3.5}$$

Given these time constants, we can approximate the transfer function in Equation 3.3 by the following equation:

$$\frac{\omega}{V} = \frac{1/k_B}{(\tau_m s + 1)(\tau_e s + 1)}$$

Since $\tau_e \ll \tau_m$, the motor transfer equation can be even further simplified (for small values of s) to the following first-order equation:

$$\frac{\omega}{V} = \frac{1/k_B}{(\tau_m s + 1)} \tag{3.6}$$

In our experiments, we shall use this simplified model to characterize the type of motor that we use in most electromechanical systems. The motor is a Swiss-made combination motor and tachometer measuring approximately 10 cm in length and 3 cm in diameter.

3.2 Pulse Control of First-Order Linear Systems

In Section 3.1, we presented transfer functions relating the angular velocity of a DC motor to the input voltage. Generally, these equations are used to describe the behavior of the system given smoothly varying input voltages. We are, however, interested in controlling motors using a switched input or binary pulse train. This technique for driving DC electromechanical systems is traditionally referred to as *pulse-width modulation* [AH91]. We shall describe the behavior of a generic first-order linear system when driven by a pulse train. We shall apply this analysis to both the torque and velocity for a DC motor.

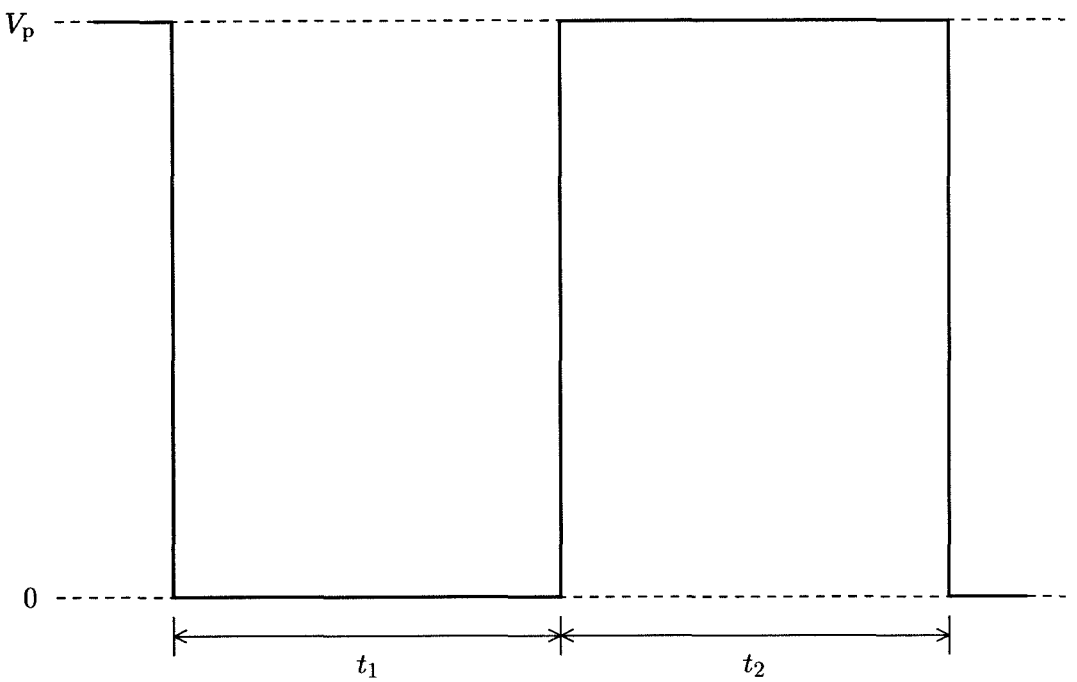

Figure 3.2: The anatomy of a single period of a pulse train. We define the voltage when the pulse is low (off) to be equal to ground or 0, and the voltage when the pulse is high (on) to be equal to V_p. The time intervals t_1 and t_2 are the periods during which the pulse is low and high, respectively.

3.2.1 The Constant Pulse Train

We shall discuss the system operation for the case of a constant signal, which, for a pulse train, means that the characteristics of the signal (high and low voltages, frequency, and pulse width) remain constant. Given this restriction, we need to consider only one pulse period in our analysis of a particular set of pulse-train parameters.

Figure 3.2 is a diagram of a single pulse with the time and voltage values defined. The voltage when the pulse is low (off) is equal to 0 (ground), and the voltage when the pulse is high (on) is equal to V_p. The time interval t_2 is the period during which the pulse is high (the pulse width), and the time interval t_1 is the period during which the pulse is low (the inverse pulse width). Using these values, we can define the following figures of merit for the

pulse train:

$$\begin{aligned} \text{period} \quad & t_p = t_1 + t_2 \\ \text{frequency} \quad & f = \frac{1}{t_1 + t_2} \\ \text{duty cycle} \quad & \alpha = \frac{t_2}{t_1 + t_2} \end{aligned}$$

3.2.2 Output Response

We shall describe the behavior of a generic first-order linear system when driven by a pulse train (as shown in Figure 3.2). We define a first-order linear system with input voltage $V(t)$, output signal $Y(t)$, and transfer function

$$\frac{Y}{V} = \frac{k}{\tau s + 1} \tag{3.7}$$

The response of this system, when driven by a constant pulse train, is shown in Figure 3.3. We shall extend the analysis of this generic system to the analysis of first-order systems for torque and angular velocity. This extension is accomplished by replacing Y and k by the appropriate output variable and proportionality constant for the system being analyzed.

Because the input pulse train (square wave) is a piecewise constant function, we can analyze the response of the system by separating each pulse period into two sections: the interval when the pulse is low ($\Delta t = (1 - \alpha)t_p$) and the interval when the pulse is high ($\Delta t = \alpha t_p$). To analyze the response of the system, we define the time t to be equal to 0 at the beginning of each interval. Given this definition, the equations that describe the response of this system as the pulse switches low and high, respectively, are:

$$Y_{lo}(t) = Y_{max} e^{-t/\tau} \tag{3.8}$$

$$Y_{hi}(t) = kV_p + (Y_{min} - kV_p)e^{-t/\tau} \tag{3.9}$$

where Y_{max} and Y_{min} are the output values at the beginning of the low and high intervals, respectively.

If we set $t = 0$ at the beginning of each interval, the solutions at the endpoints of the

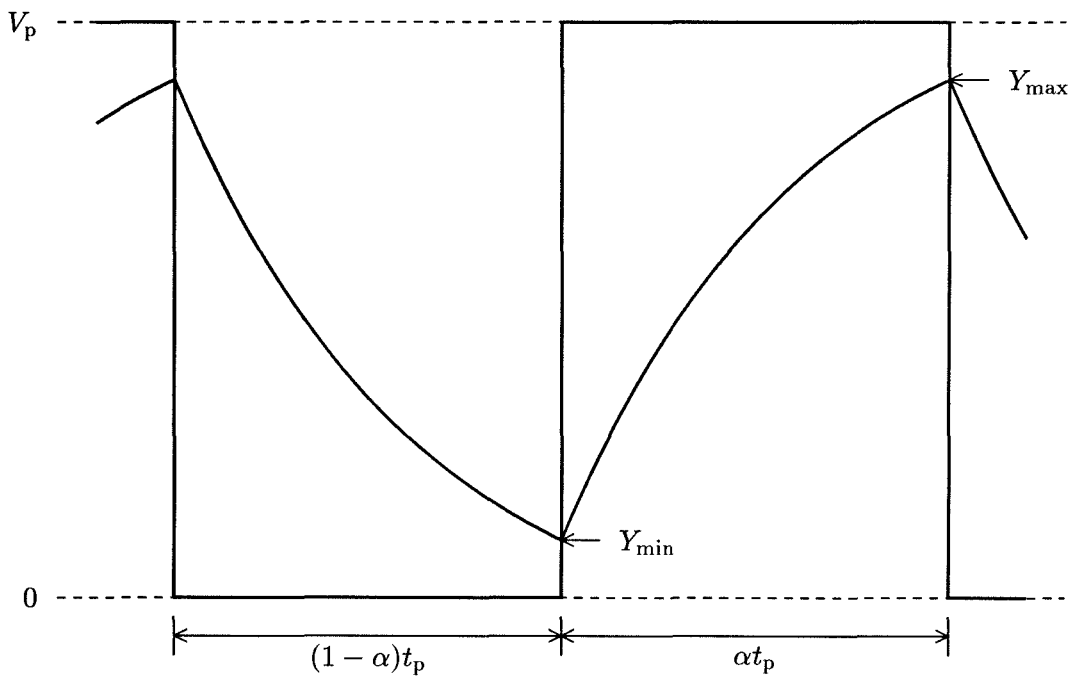

Figure 3.3: The response of a first-order linear system driven by a pulse train. The input and output waverforms are shown for a system with proportionality constant $k = 1$. The input is a square wave with amplitude V_p, pulse period t_p, and duty cycle α. The output waveform decays exponentially, approaching asymptotic values of 0 and kV_p. The time constant for the exponential decay is τ. Y_{min} and Y_{max} are defined to be the minimum and maximum output values.

two intervals give the following two equations in two unknowns ($Y_{\min}$ and $Y_{\max}$):

$$\begin{aligned} Y_{\min} &= Y_{\max} e^{-(1-\alpha)t_p/\tau} \\ Y_{\max} &= kV_p + (Y_{\min} - kV_p)e^{-\alpha t_p/\tau} \end{aligned}$$

Solving for these two equations, we derive the following values for $Y_{\min}$ and $Y_{\max}$:

$$Y_{\min} = kV_p \frac{e^{-(1-\alpha)t_p/\tau} - e^{-t_p/\tau}}{1 - e^{-t_p/\tau}} \tag{3.10}$$

$$Y_{\max} = kV_p \frac{1 - e^{-\alpha t_p/\tau}}{1 - e^{-t_p/\tau}} \tag{3.11}$$

We can replace $Y_{\min}$ and $Y_{\max}$ in Equations 3.8 and 3.9 with these values to derive the following equations for the output response of the system during the low and high intervals:

$$Y_{lo}(t) = kV_p \frac{1 - e^{-\alpha t_p/\tau}}{1 - e^{-t_p/\tau}} e^{-t/\tau} \tag{3.12}$$

$$Y_{hi}(t) = kV_p \left(1 - \frac{1 - e^{-(1-\alpha)t_p/\tau}}{1 - e^{-t_p/\tau}} e^{-t/\tau}\right) \tag{3.13}$$

3.2.3 Effective Input Voltage

When the system described by Equation 3.7 is driven by a DC input voltage the output generated by the system is described by the equation:

$$Y = k \cdot V$$

and the magnitude of the control signal is the input voltage V.

When the system is driven by a constant pulse train, the magnitude of the effective control signal should also be constant. We can derive this value by considering the response of the system during one pulse period. The average output during one pulse period is the integral of the output Y (Equation 3.12 for time t_1 plus Equation 3.13 for time t_2) divided by the time period of the integration (t_p). The average output is described by the following equation:

$$Y_{av} = \frac{1}{t_p}\left(\int_0^{(1-\alpha)t_p/\tau} Y_{lo}(t)dt + \int_0^{\alpha t_p/\tau} Y_{hi}(t)dt\right)$$

The result of this integration produces the following equation for the average output gen-

erated by the system:

$$Y_{av} = k \cdot (\alpha V_p) \tag{3.14}$$

The effective control voltage in this equation is αV_p (the product of the duty cycle and the amplitude of the pulse train). This value is the average voltage of the pulse train and is independent of the frequency of the input pulse train.

3.2.4 Minimum and Maximum Output Values

In some applications, it is important to know the minimum and maximum values of the output of a pulse-driven system. The values Y_{min} and Y_{max} for a pulse-driven first-order system were given in Equations 3.10 and 3.11. We can analyze these two values with respect to the pulse frequency by considering the values to be functions of t_p. Figure 3.4 shows the values of Y_{min} and Y_{max} (divided by kV_p) versus normalized pulse period (t_p/τ).

For very low frequencies ($t_p \gg \tau$), the values of Y_{min} and Y_{max} move toward asymptotic values. The limits of these values (from Equations 3.10 and 3.11) as the frequency decreases toward 0 are:

$$\begin{aligned} \lim_{t_p \to \infty} Y_{min} &= 0 \\ \lim_{t_p \to \infty} Y_{max} &= kV_p \end{aligned}$$

The output waveform in this regime closely approximates the input square wave (scaled by the constant k).

As the frequency increases, the output waveform becomes averaged by the integrating functionality of the system, and for very high frequencies ($t_p \ll \tau$) the output will approach a constant value. The limits of Y_{min} and Y_{max} as the frequency increases toward infinity are:

$$\begin{aligned} \lim_{t_p \to 0} Y_{min} &= k(\alpha V_p) \\ \lim_{t_p \to 0} Y_{max} &= k(\alpha V_p) \end{aligned}$$

The magnitude of the ripple at the output of the system is another important parameter. This magnitude is defined as the difference between the maximum and minimum output

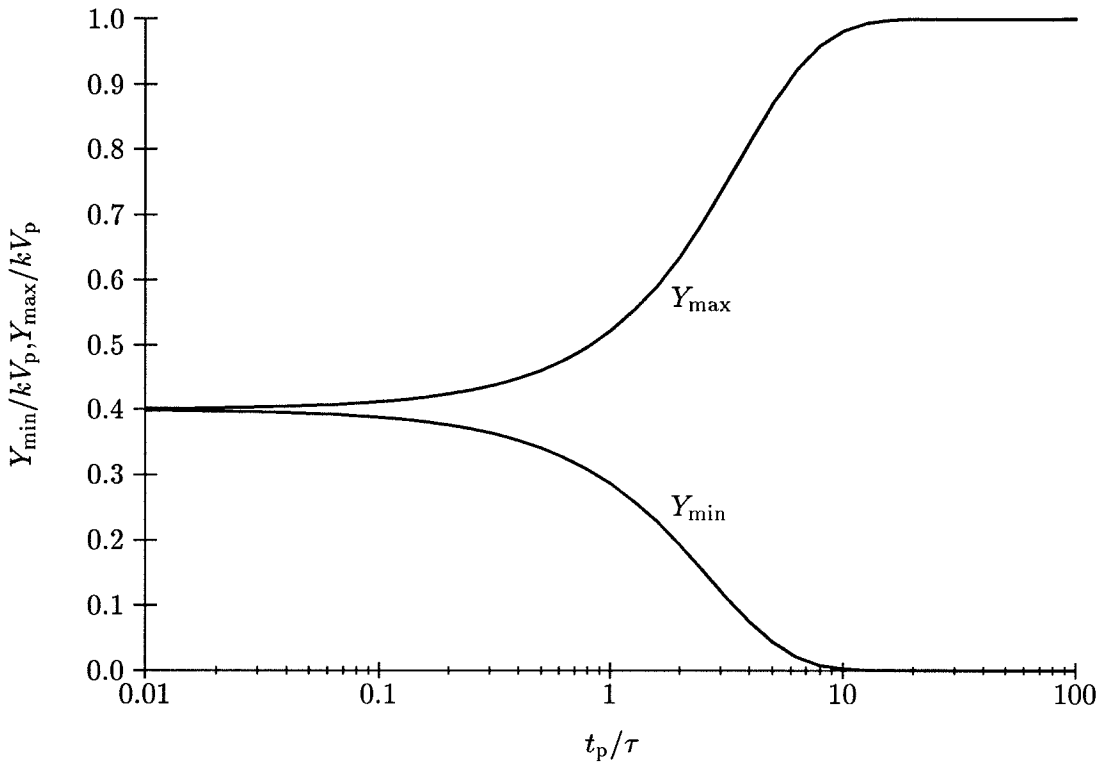

Figure 3.4: Output extremes for a pulse-driven first-order system. The two curves represent $Y_{\min}$ and $Y_{\max}$ as a function of t_p for $\alpha = 0.40$. The horizontal axis is normalized by the time constant τ. The vertical axis is normalized by kV_p. For small values of t_p/τ, $Y_{\min}$ and $Y_{\max}$ asymptotically approach $\alpha(kV_p)$. For large values of t_p/τ, $Y_{\min}$ and $Y_{\max}$ asymptotically approach 0 and kV_p, respectively.

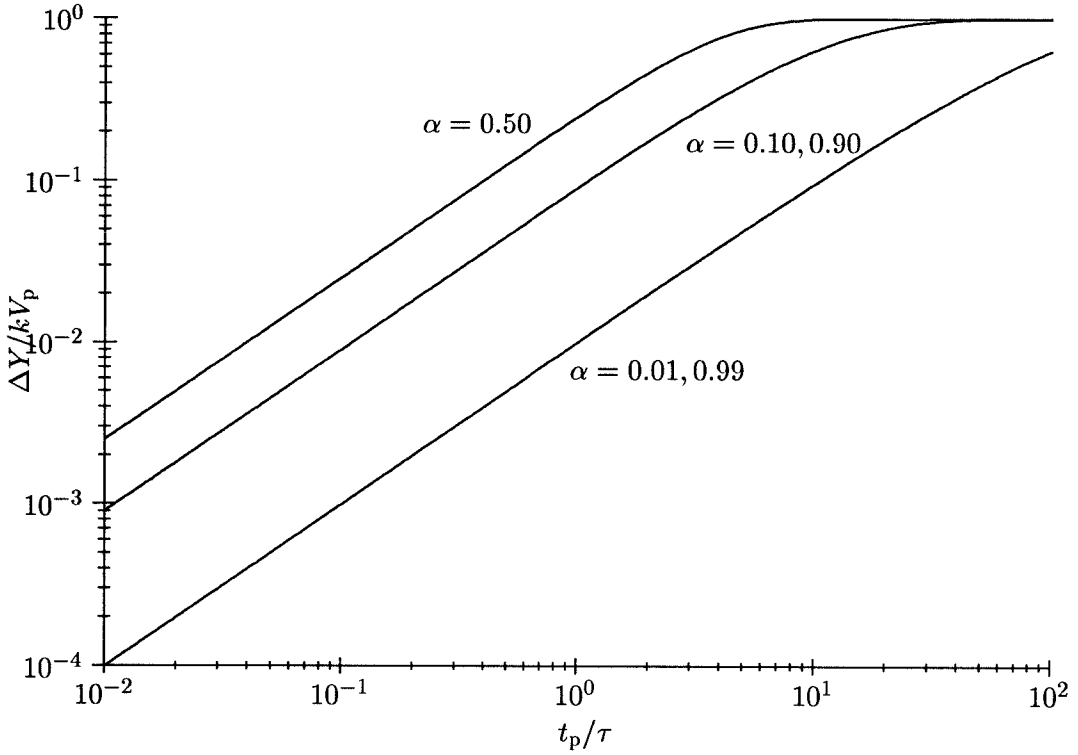

Figure 3.5: Magnitude of ripple in a pulse-driven first-order system. The magnitude of the ripple (the difference between the maximum and the minimum values of the output of the first-order system) is shown for a number of values of α. For high frequency inputs ($t_p \ll \tau$), this magnitude is approximately equal to 0, and the output is almost constant. For low freqency inputs ($t_p \gg \tau$), this magnitude is approximately equal to kV_p, and the output approximates a square wave.

values, and is given by the equation:

$$\Delta Y = Y_{max} - Y_{min} = kV_p \frac{(1 - e^{-\alpha t_p/\tau})(1 - e^{-(1-\alpha)t_p/\tau})}{1 - e^{-t_p/\tau}} \tag{3.15}$$

From the above limits, we see that this function approaches kV_p for low frequency inputs, and approaches 0 for high frequency inputs. This function is symmetric about $\alpha = 0.5$. In Figure 3.5, we show ΔY versus t_p/τ for different values of α.

3.3 Torque and Pulses

The stall torque of a motor is defined as the torque generated by the motor in the absence of motion. Since the angular velocity is 0, we can reduce Equation 3.1 to the following equation:

$$V = Ri + L\frac{di}{dt}$$

Since the torque T produced by the motor is proportional to the current i (with proportionality constant k_T), the stalled motor has a first-order relationship between the input voltage and the stall torque defined by the electrical time constant alone. The stall torque thus has a transfer function:

$$\frac{T}{V} = \frac{k_T/R}{\tau_e s + 1} \tag{3.16}$$

Since this is a first-order system, we can analyze the effects of driving the system with pulses using the theory presented in Section 3.2.

3.3.1 Stall Torque versus Input Voltage

Equation 3.14 showed that the effective voltage generated by a pulse train is equal to the product of the duty cycle and the amplitude of the pulse train. From our analysis of pulse-driven first-order systems, we see that the average stall torque produced by a pulse-driven motor is:

$$T_{av} = \frac{k_T}{R}(\alpha V_p)$$

We can measure the stall torque produced by the motor using an experimental apparatus consisting of a motor connected via a spindle and cable to a load cell (a device that linearly converts force to voltage). This apparatus converts the torque produced by the motor to a voltage. We calibrate the load cell, and use the calibration curve (and the radius of the spindle) to determine the torque produced by the motor.

Figure 3.6 shows stall torque versus input voltage data for a DC voltage input and three pulsed inputs (of different frequencies). These data all fit a linear model well, with an error from a best-fit line of less than 1 percent. The slope of the best-fit line to these curves is $4.91 \cdot 10^{-3}\,\mathrm{N{\cdot}m{\cdot}V^{-1}}$. The slope should be equal to $\frac{k_T}{R}$, and we computed the value of k_T

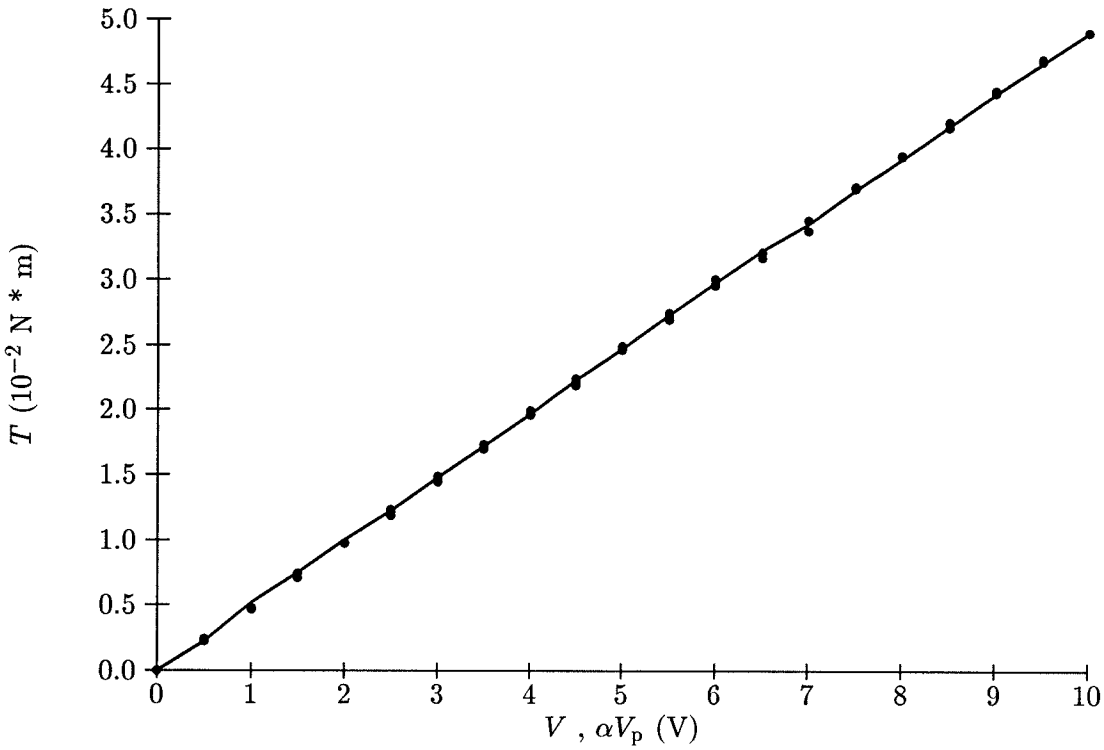

Figure 3.6: Stall-torque data for DC and pulsed inputs. The solid line represents data taken when the inputs were DC voltages in the range 0 to 10 V. The points represent data taken when the inputs were pulse trains with an amplitude of 10 V and with a duty cycle in the range between 0 and 100 percent. For the pulsed data, the horizontal axis represents the amplitude multiplied by the duty cycle. Three sets of data with pulse frequencies of 30 Hz, 300 Hz, and 3 kHz are shown superimposed.

(from the slope and a previously measured value of R) to be:

$$k_T = 3.07 \cdot 10^{-2}\,\mathrm{V} \cdot \mathrm{sec}$$

3.3.2 Effects of Pulses on Torque

We can study the effects of pulses on the stall torque by fixing the rotor in place, applying an input pulse train, and measuring the resulting torque using the load cell. The inclusion of the load cell, however, may introduce another time constant that would affect the operation of the system. To alleviate this problem, we can measure the current through the motor (since the torque is proportional to the current). We add an external resistance R_{ext} in series with the motor, and apply the input voltage across both resistor and motor. We measure the current through the motor by measuring the voltage drop V_{ext} across the resistor with an oscilloscope. The resulting torque is proportional to this voltage, and is described by the equation:

$$T_{ext} = \frac{k_T}{R_{ext}} V_{ext}$$

The addition of the external resistor, however, causes the torque produced to be less than it would be with the resistor removed (given the same input voltage in both cases). For the input pulse train presented here, the torque produced by the motor with the external resistor removed would be:

$$T = \frac{k_T}{R_{ext}} \frac{R + R_{ext}}{R} V_{ext} \tag{3.17}$$

Figure 3.7 shows the temporal relationship between stall torque and input voltage for a pulse-driven motor. The response appears first-order with the waveform exponentially approaching asymptotic values. The torque in this figure is scaled by the constant T_p, which is the stall torque produced by the motor when the input voltage is V_p. T_p is described by the following equation:

$$T_p = \frac{k_T}{R} V_p$$

3.3.3 Measurement of the Electrical Time Constant

We measure the electrical time constant τ_e of the motor using the experimental apparatus described in the previous section. The addition of the external resistor causes the test

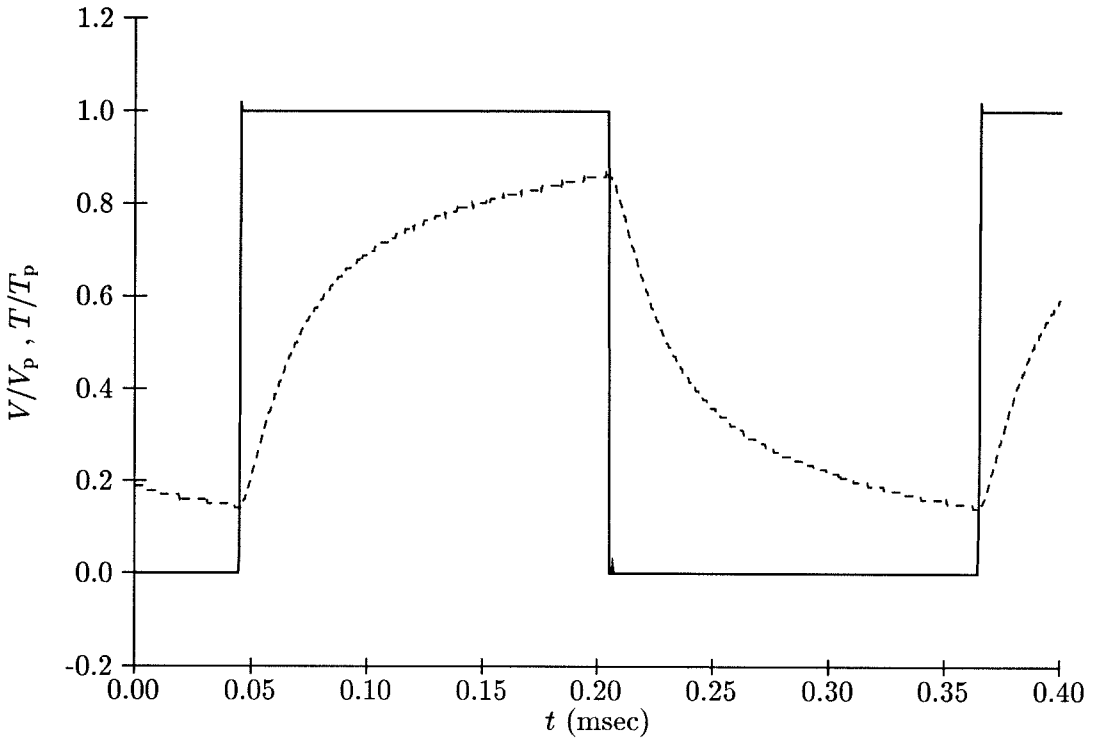

Figure 3.7: Temporal stall torque characteristics from a pulse-driven motor. The temporal waveforms for normalized input voltage and normalized output torque are shown. The input is normalized by dividing by the pulse amplitude ($V_p = 10\,V$). The output is normalized by dividing by the stall torque (T_p) that is generated at a DC voltage input of V_p.

circuit to have a time constant different from that of the motor. The time constant τ_{ext} of this circuit is

$$\tau_{ext} = \frac{L}{R + R_{ext}}$$

We measure the time constant τ_{ext} by applying an input voltage step of magnitude V_p to the motor and measuring the time for the output to reach $(1-e^{-1})T_p$, where T_p is the stall torque that is generated at a DC voltage input of V_p. By measuring the time constant at other points, we found that the response of the electrical system is not an inverse exponential as expected. This response implies that the electrical system is not a first-order linear system. In our analyses, however, we shall approximate the transfer function as in Equation 3.16.

The measured value of the external resistance was $R_{ext} = 5.04\Omega$. We measured the value of τ_{ext} to be $1.1 \cdot 10^{-4}$ sec. We measured the following motor resistance, and computed the following motor inductance and eletrical time constant:

$$\begin{aligned} R &= 6.25\,\Omega \\ L &= 1.2 \cdot 10^{-3}\,\mathrm{H} \\ \tau_e &= 2.0 \cdot 10^{-4}\,\mathrm{sec} \end{aligned}$$

3.4 Angular Velocity and Pulses

A motor experiencing no frictional load spins freely and an average torque of 0 is produced in steady state. Equation 3.6 gave the following approximate transfer function for a DC motor with no frictional load:

$$\frac{\omega}{V} = \frac{1/k_B}{(\tau_m s + 1)}$$

This approximation is valid only if $\tau_e \ll \tau_m$. We shall analyze the relationship between input voltage and angular velocity using this first-order transfer function and the theory presented in Section 3.2.

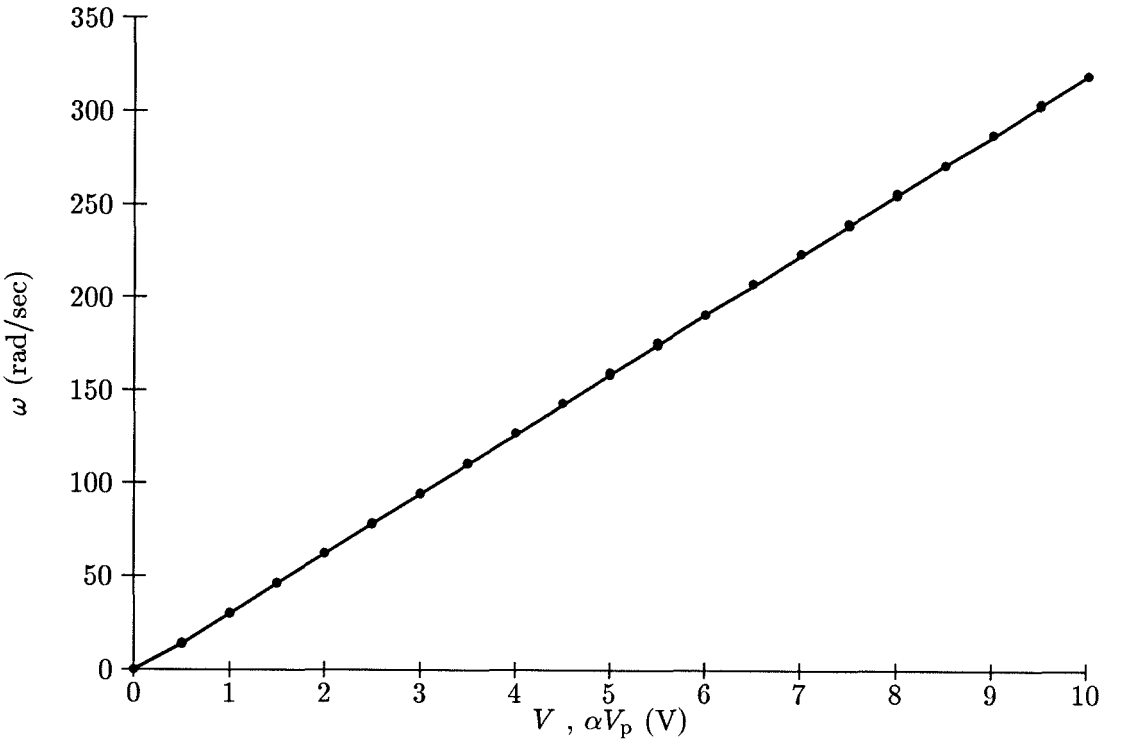

Figure 3.8: No-load speed data for DC and pulsed inputs. The solid line represents data taken when the inputs were DC voltages in the range 0 to 10 V. The points represent data taken when the inputs were pulse trains with an amplitude of 10 V and with a duty cycle in the range between 0 and 100 percent. We took three sets of data at 30 Hz, 300 Hz, and 3 kHz. The error from a best-fit line for all of these data is less than 1 percent.

3.4.1 Angular Velocity versus Input Voltage

By using Equation 3.14, we derive the following equation describing the average angular velocity of a pulse-driven motor:

$$\omega_{\mathrm{av}} = \frac{1}{k_{\mathrm{B}}}(\alpha V_{\mathrm{p}}) \tag{3.18}$$

To measure the average speed (angular velocity) of the motor, we attached an optical encoder to the shaft of the motor, and measured its output frequency, which is proportional to the speed of the motor. The addition of the optical encoder adds an inertial load to the system. This load increases the moment of inertia J and the time constant τ_{m} of the mechanical system. The added inertia does not affect the DC operation (or the average angular velocity) of the motor.

Figure 3.8 shows angular velocity versus input voltage data for a DC voltage input and three pulsed inputs (of different frequencies). These data all fit the linear model in Equation 3.18 well, with an error from a best-fit line of less than 1 percent. The slope of the best-fit line to these curves was $32.1\,\mathrm{V}^{-1}\cdot\mathrm{sec}^{-1}$. The slope is equal to $\frac{1}{k_\mathrm{B}}$, and we computed the value of k_B to be

$$k_\mathrm{B} = 3.12 \cdot 10^{-2}\,\mathrm{V}\cdot\mathrm{sec}$$

This value differs from the measured value of k_T by less than 2 percent. Also, using these two values and Equation 3.5, we calculated the moment of inertia J to be

$$J = 2.0 \cdot 10^{-6}\,\mathrm{kg}\cdot\mathrm{m}^2$$

We removed the rotor from the motor and measured its radius and effective mass to be approximately 14 g and 1.14 cm, respectively. These values give an approximate moment of inertia of $J \approx 1.8 \cdot 10^{-6}\,\mathrm{kg}\cdot\mathrm{m}^2$.

3.4.2 Effects of Pulses on Angular Velocity

We can study the effects of pulses on the angular velocity of the motor by measuring the voltage generated by the tachometer attached to the motor. This voltage is proportional to the angular velocity of the motor. By using the tachometer, we are able to remove all external frictional and inertial loads from the motor. We can make temporal measurements to determine whether the first-order transfer function approximation (for ω/V) is valid.

Figure 3.9 shows the temporal relationship between angular velocity (speed) and input voltage for a pulse-driven motor. The response appears first-order with the waveform exponentially approaching asymptotic values. The angular velocity in this figure is scaled by the constant ω_p, which is the angular velocity produced by the motor (with no load) when the input voltage is V_p. ω_p is described by the following equation:

$$\omega_\mathrm{p} = \frac{1}{k_\mathrm{B}} V_\mathrm{p}$$

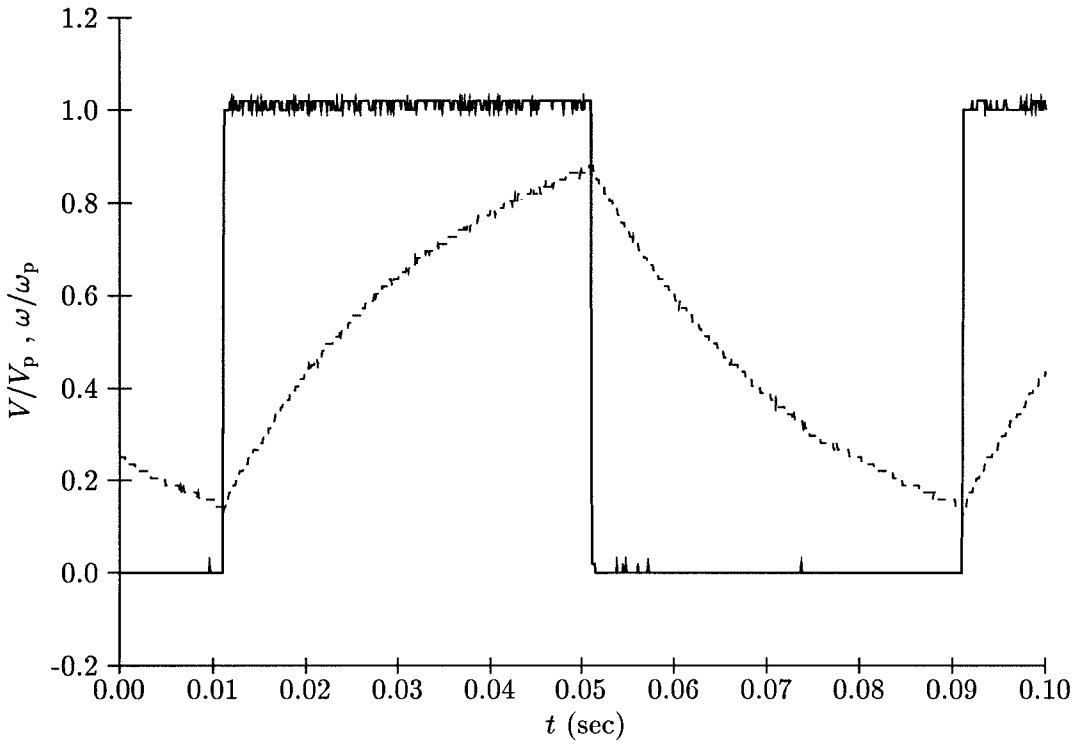

Figure 3.9: Temporal speed characteristics from a pulse-driven motor. The temporal waveforms for normalized input voltage and normalized output speed are shown. The input is normalized by dividing by the pulse amplitude ($V_p = 10\,V$). The output is normalized by dividing by the angular velocity (ω_p) that is generated at a DC voltage input of V_p.

3.4.3 Measurement of the Mechanical Time Constant

We measure the mechanical time constant τ_m of the motor by measuring the temporal response of the angular velocity of the motor (via the tachometer) when a voltage step of magnitude V_p is applied at the input. The time constant is defined as the time that it takes for the output to reach $(1 - e^{-1})\omega_p$. We obtained a mechanical time constant value of

$$\tau_m = 1.3 \cdot 10^{-2} \sec$$

Comparing this value to the value for the electrical time constant, we can see that τ_m is indeed much larger than τ_e (by approximately two orders of magnitude). Thus, our first-order approximation for the motor (in Equation 3.6) is valid.

3.5 Torque and Velocity Characteristics

In Section 3.3, we described the effect of pulses on the torque of a motor with infinite load ($\omega = 0$). In that case, the torque is only dependent on the electrical system (and electrical time constant). When the motor is rotating, however, the torque is dependent on the response of both the electrical and mechanical systems. We shall describe the effects of pulses on the torque and angular velocity characteristics of a motor with no load.

When there is no load on the motor, the first-order transfer function in Equation 3.6 describes the angular velocity of the system. We derive the following solution to this equation (for the low and high periods of the input pulse train) from Equations 3.12 and 3.13:

$$\omega_{lo}(t) = \frac{V_p}{k_B}\left(\frac{1 - e^{-\alpha t_p/\tau_m}}{1 - e^{-t_p/\tau_m}} e^{-t/\tau_m}\right) \tag{3.19}$$

$$\omega_{hi}(t) = \frac{V_p}{k_B}\left(1 - \frac{1 - e^{-(1-\alpha)t_p/\tau_m}}{1 - e^{-t_p/\tau_m}} e^{-t/\tau_m}\right) \tag{3.20}$$

When there is no load on the motor, the response of the electrical system is described by Equation 3.1. We can combine this equation with the first-order transfer function for the mechanical system (Equation 3.6) to derive the following torque–voltage transfer function for a motor with no load:

$$\frac{T}{V} = \frac{k_T}{R} \cdot \frac{\tau_m s}{\tau_m s + 1} \cdot \frac{1}{\tau_e s + 1}$$

We can derive the following approximate solution of this transfer function for a system presented with an input pulse train with a period large enough that the electrical system nearly reaches its asymptotic limit:

$$T_{\mathrm{lo}}(t) = \frac{k_{\mathrm{T}} V_{\mathrm{p}}}{R}\left[-\left(\frac{1-e^{-\alpha t_{\mathrm{p}}/\tau_{\mathrm{m}}}}{1-e^{-t_{\mathrm{p}}/\tau_{\mathrm{m}}}}\right)e^{-t/\tau_{\mathrm{m}}} + e^{-t/\tau_{\mathrm{e}}}\right] \tag{3.21}$$

$$T_{\mathrm{hi}}(t) = \frac{k_{\mathrm{T}} V_{\mathrm{p}}}{R}\left[\left(\frac{1-e^{-(1-\alpha)t_{\mathrm{p}}/\tau_{\mathrm{m}}}}{1-e^{-t_{\mathrm{p}}/\tau_{\mathrm{m}}}}\right)e^{-t/\tau_{\mathrm{m}}} - e^{-t/\tau_{\mathrm{e}}}\right] \tag{3.22}$$

where T_{lo} is the torque generated by the motor when the pulse is low, and T_{hi} is the torque generated by the motor when the pulse is high.

For pulse trains with periods short enough that the electrical system does not reach its asymptotic limit, we consider the angular velocity constant and equal to $\alpha V_{\mathrm{p}}/k_{\mathrm{B}}$, and the following equations describe the torque during the low and high parts of the pulse train:

$$T_{\mathrm{lo}}(t) = \frac{k_{\mathrm{T}} V_{\mathrm{p}}}{R}\left(\frac{1-e^{-\alpha t_{\mathrm{p}}/\tau_{\mathrm{e}}}}{1-e^{-t_{\mathrm{p}}/\tau_{\mathrm{e}}}}e^{-t/\tau_{\mathrm{e}}} - \alpha\right) \tag{3.23}$$

$$T_{\mathrm{hi}}(t) = \frac{k_{\mathrm{T}} V_{\mathrm{p}}}{R}\left((1-\alpha) - \frac{1-e^{-(1-\alpha)t_{\mathrm{p}}/\tau_{\mathrm{e}}}}{1-e^{-t_{\mathrm{p}}/\tau_{\mathrm{e}}}}e^{-t/\tau_{\mathrm{e}}}\right) \tag{3.24}$$

3.5.1 Torque and Velocity Data

In this section, we shall compare theoretical predictions for angular velocity in Equations 3.19 and 3.20 and for torque in Equations 3.21 and 3.22. We shall show experimental data and theoretical curves for the input voltage, angular velocity, and torque when the system is presented with pulse trains in differing frequency regimes. The data are normalized by dividing by V, T, and ω by V_{p}, T_{p}, and w_{p}, respectively. We measure the torque by measuring the voltage drop across an external resistor as described in Section 3.3. We measure the angular velocity by measuring the tachometer voltage. The system with the added external resistor has time constants of $t_{\mathrm{e}} = 1.1 \cdot 10^{-4}$ sec and $t_{\mathrm{m}} = 2.3 \cdot 10^{-2}$ sec.

When the frequency of the input pulse train is small ($f < 1/\tau_{\mathrm{m}}$), the mechanical system dominates the operation of the system and the torque produced by the motor. For very low frequencies, the angular velocity approximates a square wave (with exponentially rounded leading edges). At each input pulse edge, the motor produces a broad torque impulse, which starts or stops the motion of the motor. The magnitude of the torque decreases as

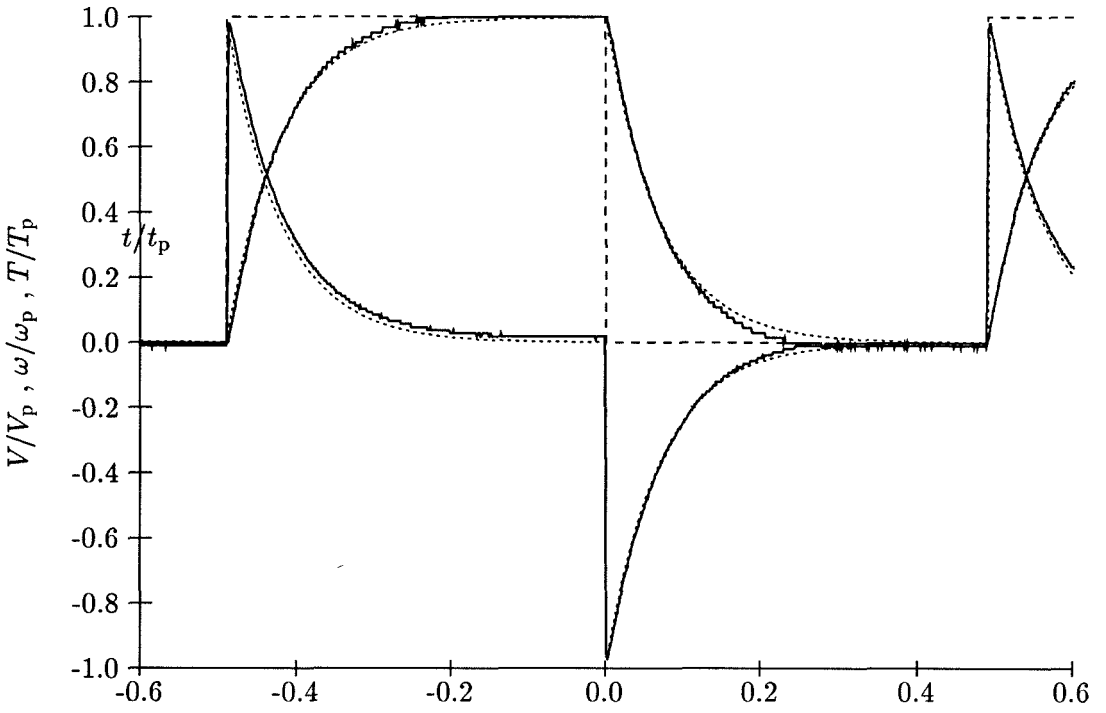

Figure 3.10: System response to a low-frequency pulse train. The input frequency is 3.125 Hz. The dashed curve represents the input pulse train. The solid curves represent the experimental torque and speed data, and the dotted curves represent the theoretical torque and speed. The mechanical time constant dominates the operation in this regime. Thus, the angular velocity approximates the input pulse train and the torque is a broad impulse.

the angular velocity approaches its asymptotic value. Figure 3.10 shows the response of the system to a pulse train with a frequency in this range.

For higher frequency pulses, the angular velocity response becomes smoothed by the low-pass effect of the mechanical system. The torque response remains highly oscillatory. The effects of the electrical time constant become apparent, and the leading edge of each torque impulse becomes rounded as the frequency increases. The trailing edge of the torque ripple becomes flatter as the deviation in the angular velocity response decreases. Figure 3.11 shows the response of the system to a pulse train with a frequency in this range.

As the frequency is increased even farther, the smoothing (low-pass) effect of the mechanical system causes the angular velocity to be practically constant. The torque response in now completely dominated by the effects of the electrical time constant, showing a first-

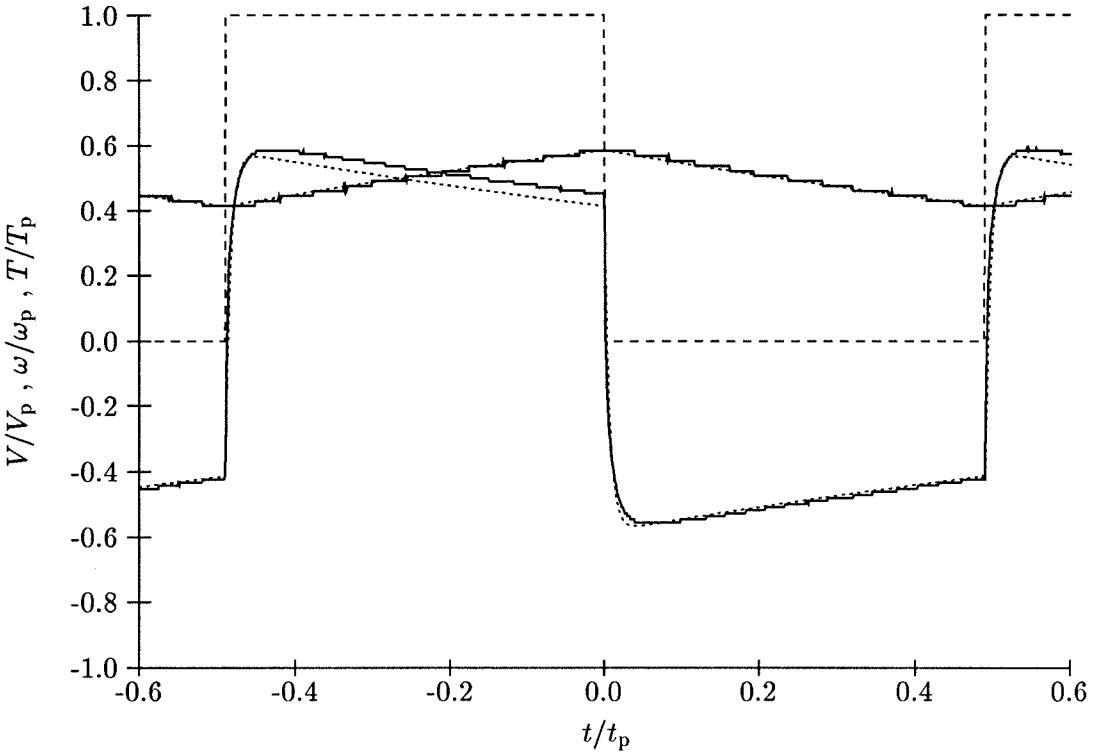

Figure 3.11: System response to intermediate-frequency pulse trains. The input frequency is 62.5 Hz. The dashed curve represents the input pulse train. The solid curves represent the experimental torque and speed data, and the dotted curves represent the theoretical torque and speed. The angular velocity is more smoothed than in Figure 3.10, and the effect of the electrical time constant on the (leading edge of) the torque pulses is apparent.

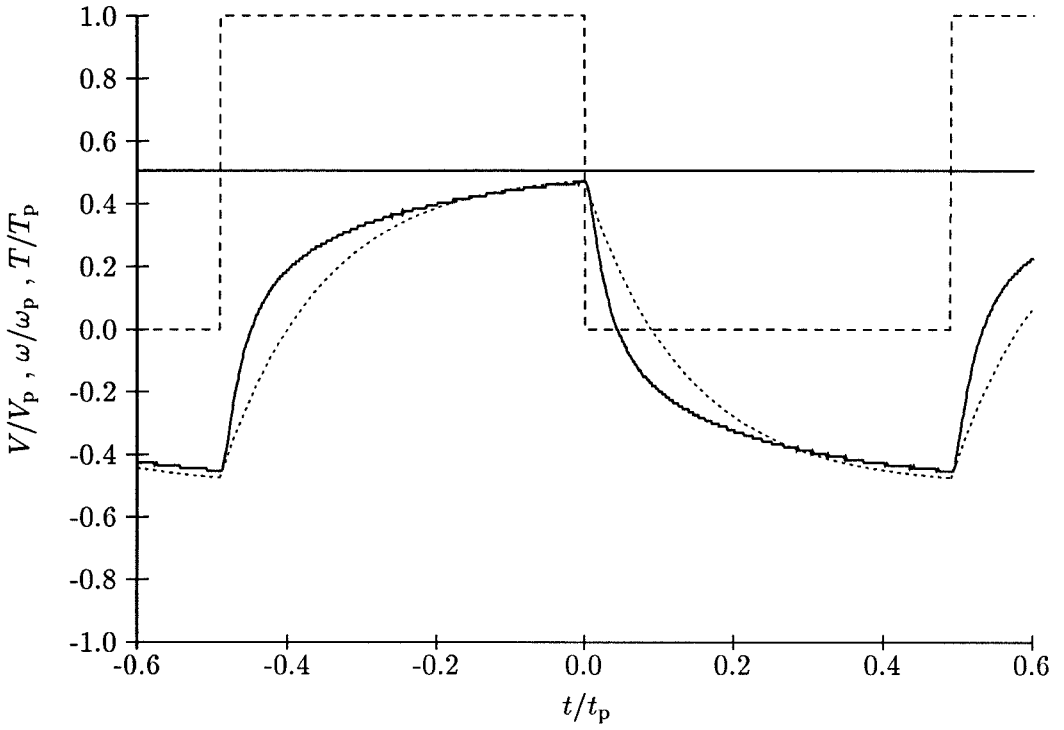

Figure 3.12: System response to high-frequency pulse trains. The input frequency is 1250 Hz. The dashed curve represents the input pulse train. The solid curves represent the experimental torque and speed data, and the dotted curves represent the theoretical torque and speed. The angular velocity remains almost constant, and the torque response is dominated by the electrical time constant.

order response each time the pulse train switches. In this regime of operation the torque can be described by Equations 3.23 and 3.24. Figure 3.12 shows the response of the system to a pulse train with a frequency in this range. There is an obvious discrepency between the theory and data in this figure, due to the nonlinear electrical system behavior described in Section 3.3.

In the limit of very high frequencies ($t_p \ll \tau_e$), the low-pass effect of the electrical system causes the torque to become nearly constant, and the response of the pulse-driven system approximates the response of the system to a DC input voltage.

3.5.2 The Relationship Between the Time Constants

In the preceding description of the operation of the motor in differing frequency regimes, we showed that, since the mechanical time constant is much greater than the electrical time constant, the angular velocity smooths at much lower frequencies than the torque.

We show that, in response to low-frequency input pulse trains, the motor demonstrates a significant amount of ripple in the angular velocity. In some applications, it may be necessary to limit the magnitude of this ripple. By applying Equation 3.15, we can derive the following equation for the magnitude of angular velocity variation:

$$\Delta\omega = \frac{V_{\mathrm{p}}}{k_{\mathrm{B}}} \frac{(1 - e^{-\alpha t_{\mathrm{p}}/\tau})(1 - e^{-(1-\alpha)t_{\mathrm{p}}/\tau})}{1 - e^{-t_{\mathrm{p}}/\tau}} \tag{3.25}$$

When we consider the effects of kinetic and static friction in Section 3.6, we shall see that the benefits of pulses in allowing the motor to overcome friction are lost when the maximum torque falls below the level necessary to break free of the static friction in the system. From Equation 3.11, we can derive the following equation for the maximum stall torque produced by the motor:

$$T_{\max} = \frac{k_{\mathrm{T}} V_{\mathrm{p}}}{R} \frac{1 - e^{-\alpha t_{\mathrm{p}}/\tau_{\mathrm{e}}}}{1 - e^{-t_{\mathrm{p}}/\tau_{\mathrm{e}}}} \tag{3.26}$$

Thus, we would like to generate pulse trains in frequency regime that allows for both low amplitude ripple in the angular velocity and large values of the maximum torque. This concern is especially apparent when driving the system with a pulse train with a small duty cycle, since the average stall torque is small in this case. Figure 3.13 shows the magnitude of the angular velocity ripple and the maximum stall torque as functions of the pulse period t_{p} for a system in which the time constants are separated by two orders of magnitude. From this figure, we can make a good argument for operating the system in the frequency range between the two time constants (where the angular velocity is constant, but the torque remains oscillatory).

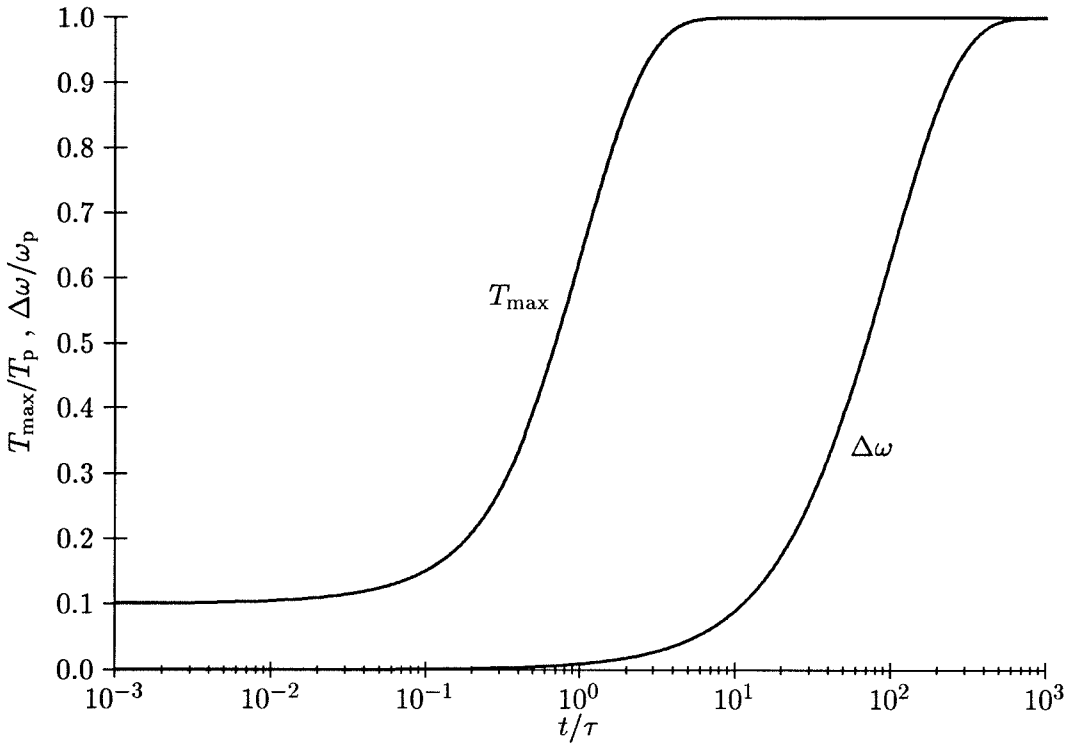

Figure 3.13: $\Delta\omega$ and T_{max} as a function of t_p. The curves are for a pulse train with a duty cycle of 0.1 and time constants separated by two orders of magnitude. The horizontal axis is normalized by the τ, where $\tau = 10\tau_e = 0.1\tau_m$. The values on the vertical axis are normalized by ω_p and T_p. The region between the two time constants (and especially nearer to τ_m) shows the best combination of small $\Delta\omega$ and large T_{max}.

3.6 The Effects of Friction

Up to this point, we have assumed a frictionless system. In most real-world problems, however, friction must be considered. The existence of friction-limited systems is one of the primary reasons for choosing a pulse-control paradigm [AH91]. In this section, we shall consider the effects of both kinetic (sliding) and static friction on the operation of systems driven by analog voltages and systems driven by pulses.

3.6.1 Kinetic Friction

By definition, kinetic friction on a moving system acts as a force or torque on the system that opposes motion and has constant magnitude (depending on only the sign of the velocity of the system) [HR81]. We can add kinetic friction to our motor model by modifying Equation 3.2 to produce the following equation:

$$J\frac{d\omega}{dt} = k_{\mathrm{T}}i - \mathrm{sgn}(\omega)T_{\mathrm{f}}$$

where T_{f} is a constant (representing the magnitude of the torque caused by kinetic friction), and $\mathrm{sgn}(\omega)$ is -1, 1, and 0 for negative, positive, and 0 angular velocity, respectively. We can combine this equation with Equation 3.1 to derive the transfer function for the system with kinetic friction included. Since kinetic friction is a discontinuous function of angular velocity, however, we must place restrictions on the system to determine a closed-form equation. For example, if we restrict the angular velocity to be positive, the transfer function is

$$\frac{\omega}{V - V_{\mathrm{f}}} = \frac{1/k_{\mathrm{B}}}{\tau_{\mathrm{m}}s + 1} \tag{3.27}$$

where we define V_{k} to be a constant voltage offset due to friction, defined by the equation:

$$V_{\mathrm{f}} \equiv \frac{R}{k_{\mathrm{T}}}T_{\mathrm{f}}$$

The DC version of Equation 3.27 is

$$\omega = \frac{1}{k_{\mathrm{B}}}(V - V_{\mathrm{f}}) \tag{3.28}$$

This equation is valid only when ω is positive ($V > V_f$). When V is positive and less than V_f, ω is equal to zero because the torque produced by the motor is smaller than the frictional torque.

3.6.2 Static Friction

Static friction is the friction on a system at rest, and is generally greater than kinetic friction. As described previously, when our system (with kinetic friction included) is at rest ($\omega = 0$), we must apply a voltage greater than V_f so that the torque generated by the motor is larger than the frictional torque. Given that static friction is greater than kinetic friction, however, the generated torque must actually exceed the torque due to static friction (T_s) for the motor to break free of the static friction. Once the system breaks free of the static friction, the frictional torque immediately decreases (to the level of the kinetic friction). This effect causes a discontinuity in the motor speed when the system is at rest and the magnitude of the applied voltage is increased gradually.

We define the voltage necessary to break free of the static friction to be

$$V_s \equiv \frac{R}{k_T} T_s$$

Starting with a system at rest, if the applied voltage is less than V_s, the motor speed is equal to 0; if the applied voltage is greater than V_s, the motor speed is described by Equation 3.28. Static friction thus limits the minimum speed that can be generated by the application of a DC voltage to a system at rest. This minimum speed is $(V_s - V_f)/k_B$.

3.6.3 Low-Speed Control of Frictional Systems

The minimum-speed restriction (due to static friction) greatly limits the reliable operation of DC-driven systems that have significant friction and are required to operate at low speeds. A DC voltage is unable to control a system at low speed because, if the system begins at rest, the control voltage must be large enough for the motor to generate a torque capable of overcoming the static friction. For systems with appreciable static friction, the resulting discontinuity (in the voltage–speed curve) can be very large.

One solution to the static-friction problem is to guarantee that the system is in motion,

and then to decrease the speed to the correct value. This approach might be feasible in some applications, such as when the system is driven unidirectionally and the motor is always rotating. This solution, however, does not guarantee operation of the system at low speeds. Even if the system is in motion, low-speed control can be difficult with a DC voltage due to variations in the kinetic friction. Since the system is moving very slowly, a slight increase in friction might be sufficient to cause the moving system to stop, at which point the generated torque will be unable to break free of the static friction. Thus, unless there is very little variation in kinetic friction, a DC analog voltage will be ineffective at low-speed control.

Another problem with DC-driven frictional system is that angular velocity is not an increasing function of the input voltage, but rather, the angular velocity is equal to 0 for voltages near 0. Also, when the system starts at rest, there is a discontinuity in the curve due to static friction. This discontinuity can result in hysteresis in such systems.

3.7 Pulses and Friction

We shall analyze the effects of friction on a pulse-driven system for two regimes of operation. The first regime is characterized by the fact that the motor is continuously moving. In this regime, the angular velocity may have ripple, but cannot decrease to the point where $\omega = 0$. The second regime is characterized by the fact that the motor stops during the low period of the input pulse train. In this regime, each pulse generates a discrete motion of the motor. In both regimes, we assume that the pulse width is large enough that we can ignore the effects of the electrical time constant on the operation of the system.

To test the effects of friction, we shall take data from our system with the addition of an apparatus that generates friction. Two aluminum plates are mounted to the shaft of the motor. An asbestos plate is sandwiched between the rotating plates, and is fixed so that it cannot move. The rotating plates apply a force on the fixed plate. This force can be varied to change the frictional torque generated by the apparatus. We measure the angular velocity with the optical encoder, which is also attached to the shaft of the motor.

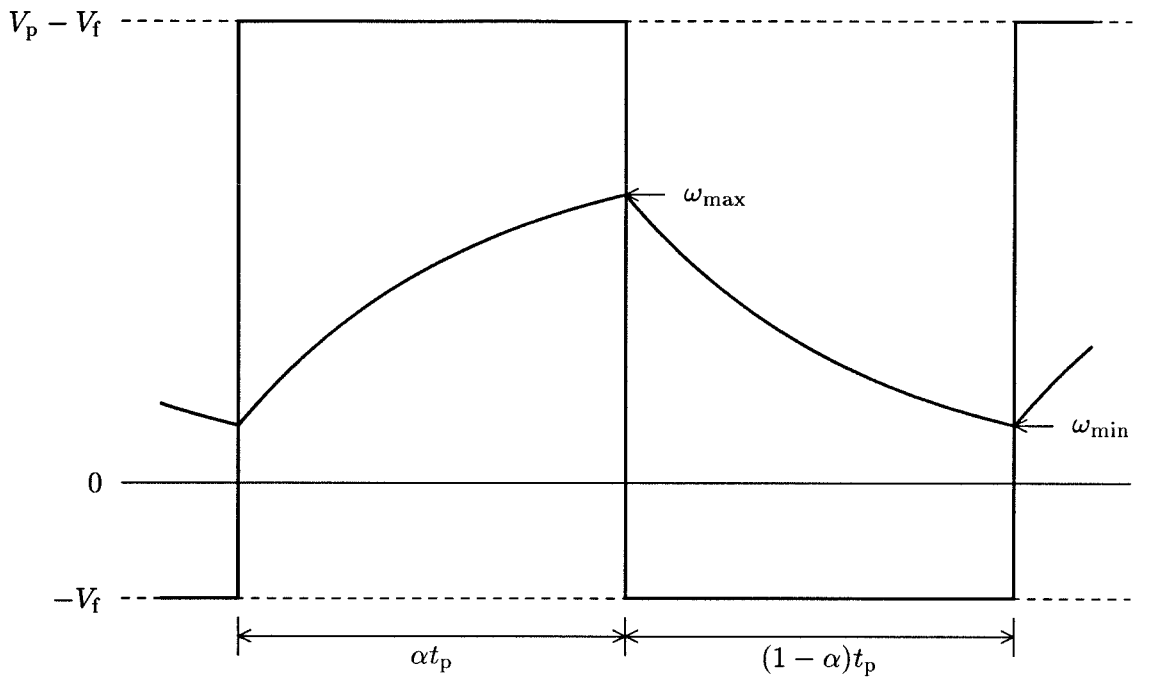

Figure 3.14: Continuous motion in a pulse-driven frictional system. Friction affects the system by decreasing the effective input voltage by a constant amount (V_f), thus shifting the horizontal axis of the plot of input voltage by V_f. This shift effectively decreases the angular velocity at any given point by a constant amount (ω_f).

3.7.1 Continuous Motion

From the transfer function in Equation 3.27, we see that, in a system with kinetic friction, the effective input voltage is equal to the applied voltage minus a voltage (V_f) representing the effect of friction. In a pulse-driven system, the friction effectively translates the input pulse train vertically by $-V_f$, resulting in an effective pulse train with low and high voltages of $-V_f$ and $V_p - V_f$, respectively.

This translation in input voltage causes a translation in angular velocity by $-\omega_f$, where ω_f is defined by the following equation:

$$\omega_f \equiv \frac{V_f}{k_B}$$

Figure 3.14 shows the effect of the friction on the effective pulse train and on the resulting

angular velocity when the velocity is always greater than 0. For this angular velocity to remain positive, the minimum angular velocity generated by the frictionless system must be greater than ω_f, and from Equation 3.10:

$$\omega_f < \frac{V_p}{k_B} \frac{e^{-(1-\alpha)t_p/\tau_m} - e^{-t_p/\tau_m}}{1 - e^{-t_p/\tau_m}}$$

When this inequality is valid, the average angular velocity is described by the equation:

$$\omega_{av} = \frac{\alpha V_p - V_f}{k_B} \tag{3.29}$$

Since the effective DC input voltage V for a pulse train is equal to αV_p, the response of the system driven by pulses is equivalent to the response of the system driven by a DC voltage (as defined in Equation 3.28) when the angular velocity remains positive.

Figure 3.15 shows data for a frictional system driven by both DC and pulsed inputs. We generated two curves from the DC-driven system. One of these curves was taken with the system initially at rest (the voltage was increased from 0 V), and shows the effect of the static friction. The other DC curve was taken with the system initially in motion (the voltage was decreased toward 0 V), and shows no static friction effect. For input voltage less than V_f, the angular velocity for the DC-driven system is equal to 0. The angular velocity for the pulse-driven system is positive in this region.

The pulse data are approximately equal to the DC data for voltages greater than a particular input voltage. This voltage represents the interface between the continuous-motion and discrete-motion regimes of operation ($t_f = (1 - \alpha)t_p$). We confirmed this fact by measuring the angular velocity with an oscilloscope. The point at which the DC and pulse curves meet is indeed the point at which the minimum angular velocity in the pulse-driven system was 0. Below this voltage, the motion of the pulse-driven system is no longer described by Equation 3.29. The motion in this regime will be discussed in the following section.

For large input voltages, the curves from the DC- and pulse-driven systems are quite well matched as predicted by Equations 3.28 and 3.29. The curves are not linear in the region, however, because the friction is not constant over the entire range of motion. We can measure the friction at a given point by calculating the difference between the angular

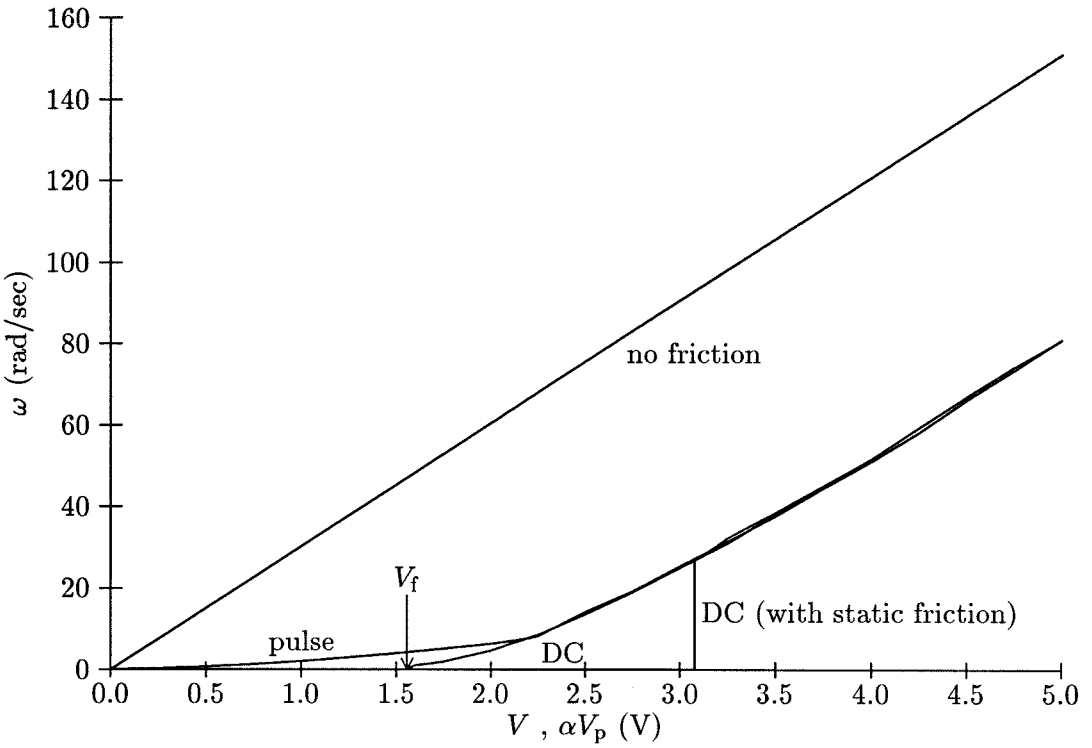

Figure 3.15: Angular velocity versus input voltage for a frictional system. One curve represents data taken when the system begins at rest and is driven by a DC voltage. This curve has a discontinuity at the point when the input voltage is large enough to break free of the static friction ($V > V_s$). A second curve shows the effects of a DC input voltage on a previously moving system. This curve is continuous and decreases smoothly to the horizontal axis at $V = V_f$. The third curve represents data taken from a pulse-driven system. This curve matches the DC curves for high voltages. At lower voltages this curve decreases more slowly, such that the angular velocity is positive for positive input voltages. The final curve shows the response of the system with no friction.

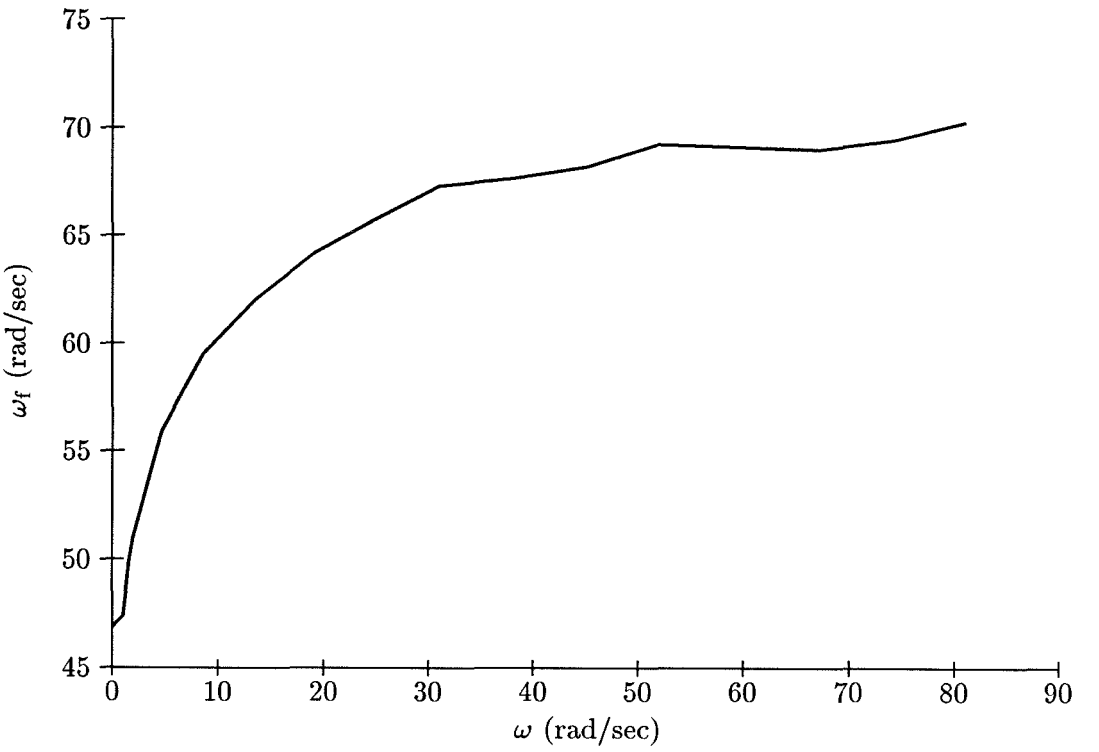

Figure 3.16: The variation in the friction generated by the system. The curve represents the angular velocity offset ω_f due to the friction generated by the mechanical system versus the angular velocity ω generated by the motor in the frictional system. ω_f is proportional to the frictional torque T_f; thus, if the friction was constant (independent of angular velocity), this curve would be a horizontal line.

velocity in the frictionless system and the angular velocity in this frictional system. The resulting difference is equal to ω_f. Figure 3.16 shows a plot of this value versus the angular velocity in the frictional system.

3.7.2 Discrete Motion

If the angular velocity in a pulsed system decreases to 0 during the period that the input is low, the system no longer reacts in the manner described in the preceding section. When the motor stops rotating, the frictional torque is no longer generated and the system remains at rest until the pulse train again goes high. Thus, each pulse produces a discrete motion. Figure 3.17 shows the response of the system in this regime.

At the beginning of each pulse, the system is at rest ($\omega = 0$). During the period in which

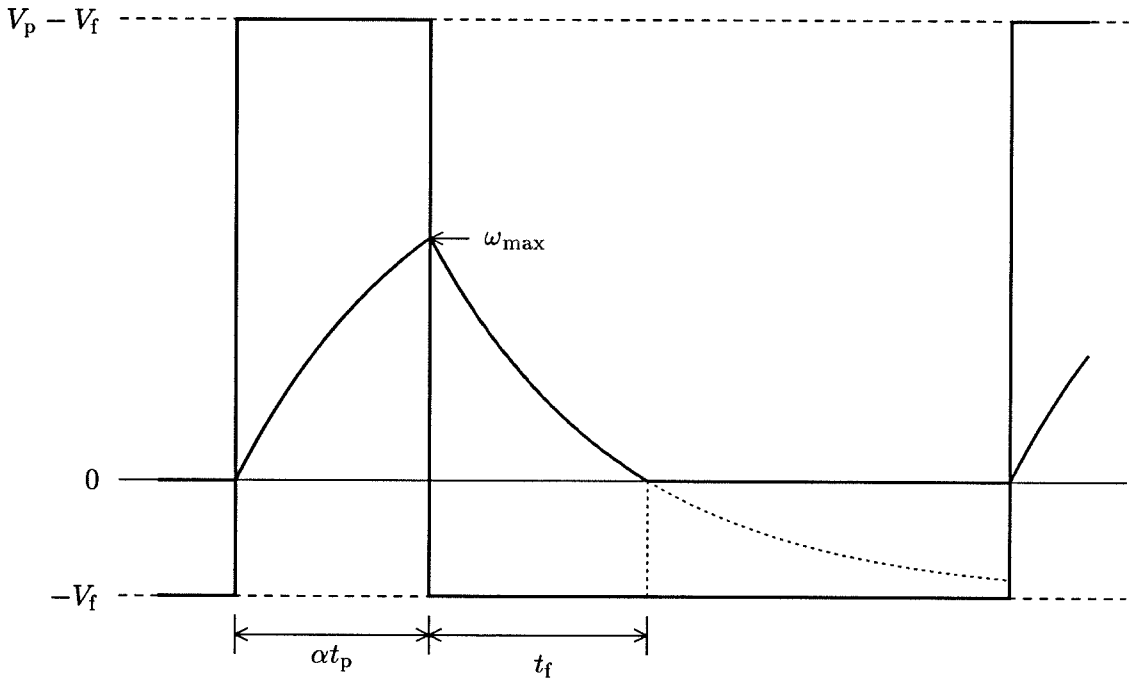

Figure 3.17: Discrete motion in a pulse-driven frictional system. During the low period of each pulse, the angular velocity of the motor decreases to 0. Thus, the average velocity is no longer the same as it is in a DC-driven system. We define t_f to be the time (into the low portion of the pulse train) at which the motor stops.

the pulse train is high, the angular velocity of the system increases toward an asymptotic value of $\omega_p - \omega_f$. For a constant value of ω_f, the behavior is described by the following equation:

$$\omega_{hi}(t) = (\omega_p - \omega_f)(1 - e^{-t/\tau_m})$$

The maximum angular velocity ω_{max} occurs at the point when the pulse train goes low, and is described by the following equation:

$$\omega_{max} = (\omega_p - \omega_f)(1 - e^{-\alpha t_p/\tau_m})$$

During the period in which the pulse train is low, the angular velocity decreases toward an asymptotic value of $-\omega_f$, and is described by the following equation:

$$\omega_{lo}(t) = -\omega_f + (\omega_{max} + \omega_f)e^{-t/\tau_m}$$

We define t_f to be the time (into this interval) when the angular velocity is equal to 0. By setting $\omega = 0$ and $t = t_f$ in the previous equation, we can derive the following equation for t_f:

$$t_f = \tau_m \ln\left(\frac{\omega_{max} + \omega_f}{\omega_f}\right)$$

We can compute the average angular velocity in this regime of operation by integrating the velocity over one complete period of the pulse train, and dividing by the pulse period t_p. This integral is the sum of the integrals of the angular velocity during both the low and the high periods of the pulse train, and is described by the following equation:

$$\omega_{av} = \frac{1}{t_p}\left(\int_0^{\alpha t_p} \omega_{hi}(t)dt + \int_0^{t_f} \omega_{lo}(t)dt\right)$$

By integrating and rearranging this equation, we derive the following equation for the average angular velocity in the regime in which each pulse generates a discrete motion:

$$\omega_{av} = \alpha(\omega_p - \omega_f) - \frac{t_f}{t_p}\omega_f \tag{3.30}$$

Since t_f is an increasing function of the pulse width αt_p, the average angular velocity (and thus the displacement generated by each pulse) is also an increasing function of αt_p. At

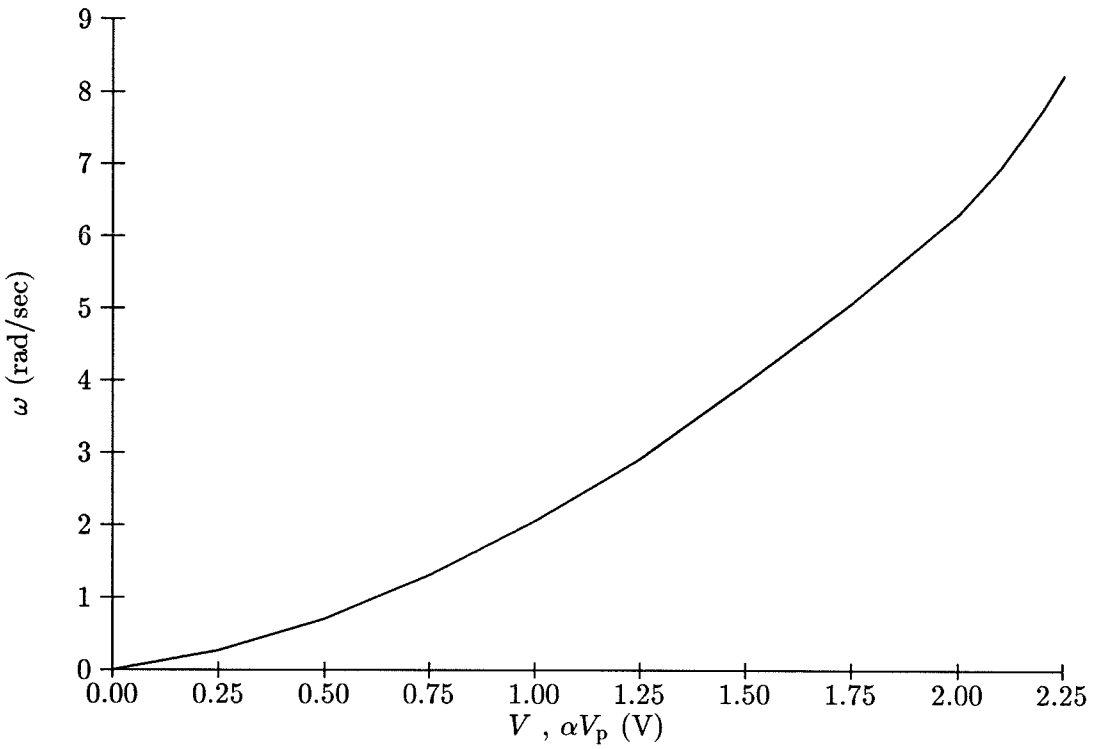

Figure 3.18: Discrete motion in a frictional system. The curve is the subset of the pulse data in Figure 3.15 for which the system is operating in the discrete motion regime. The angular velocity is plotted versus the effective input voltage αV_{p}.

the point at which the system crosses from the regime of continuous motion to the regime of discrete motion, $t_{\mathrm{f}} = (1 - \alpha)t_{\mathrm{p}}$, and this equation is equivalent to Equation 3.29.

Figure 3.18 shows data for the frictional system driven in this regime. The data in this figure are taken from the region in Figure 3.15 where the curve for the pulse-driven system deviates from the curve for the DC-driven system. In this region, the duty-cycle of the pulse train is small and the pulse-driven system is operating in its discrete motion regime. To generate the data curve, the frequency of the input pulse train was held constant and the pulse width (and thus the duty cycle) was varied. This curve shows that as the pulse width decreases the average angular velocity decreases faster than linear; thus, the displacement generated by each pulse decreases with decreasing pulse width.

3.7.3 The Electrical Time Constant and Maximum Torque

In our previous analysis, we considered pulse trains with pulse widths large enough that the electrical time constant could be ignored (the maximum torque was reached during each pulse). If the pulse width is small (in the range of the electrical time constant), however, we can no longer ignore the electrical effects in the system. If the pulses become too narrow, the exponential decay due to the electrical system causes the torque generated by the motor for an individual pulse not to reach the level needed to overcome the static friction. We need to consider only the torque generated when the angular velocity is 0 (during the time that the input pulse is low). The maximum value of the torque ($T_{\max}$) in this situation is described by the following equation:

$$T_{\max} = \frac{k_T V_p}{R}\left(1 - e^{-\frac{t_2}{\tau_e}}\right)$$

If the value of $T_{\max}$ exceeds T_s, the pulse will cause the system to rotate. To guarantee this motion, we must guarantee a minimum pulse width, such that

$$t_2 > \tau_e \ln\left(1 - \frac{RT_s}{k_T V_p}\right)$$

If this minimum pulse-width criterion is met, there is no theoretical limit on the minimum average speed that the motor can generate.

We used the friction-generating apparatus described previously, and measured values of angular velocity for a DC-driven system starting at rest and in motion, and for a pulse-driven system. When the system started at rest, the minimum angular velocity generated using a DC voltage to overcome the static friction was approximately 7 rad/sec. When the system started in motion and the input voltage was decreased slowly, we obtained a minimum angular velocity of approximately 1 rad/sec, due to the variation in the friction between the two surfaces (aluminum against asbestos). When the system was driven with pulses, we achieved a minimum angular velocity of approximately $1 \cdot 10^{-3}$ rad/sec (about 2 hours per revolution). We could, however, conceivably decrease this speed further.

Chapter 4

Pulse-Generating Circuits

In Chapter 3, we explained the effectiveness of pulses in the control of the motion of mechanical systems. In this chapter, we present circuits that generate pulses using the subthreshold analog VLSI design paradigm. We describe the dual-rail pulse encoding, and discuss its use in controlling mechanical systems bidirectionally. Finally, we present a circuit that aggregates a number of inputs and produces a dual-rail pulse encoding at its output.

4.1 The Neuron Circuit

In this section, we present a VLSI circuit that converts analog signals (currents and voltages) into temporally modulated pulse trains. We describe the operation and functionality of this circuit, and explain the reasons that is well suited for driving mechanical systems. We also present two variations of the original circuit.

4.1.1 The Original Neuron Circuit

Mead [Mea89] describes a circuit that converts an analog current into a pulse train. This *neuron circuit* takes its name from the biological cells that produce pulses in the nervous systems of animals. The schematic of this circuit is shown in Figure 4.1. The circuit consists of a current mirror, an amplifier (two inverters), two capacitors, and two series transistors. The input current is integrated by the capacitance of the input node, causing the input voltage to increase. When this voltage reaches the switching voltage of the amplifier, the output voltage switches high, causing the reset current to be enabled. This

current discharges the input node until the output switches low. A feedback capacitor causes hysteresis by capacitively coupling the output node to the input node. If the input and reset currents are held constant, this circuit produces a constant pulse train at the output.

We can analyze the quantitative operation of the circuit given a constant input current. We define t_1 and t_2 to be the periods of time that the output is low and high, respectively. The magnitude of the charge placed on the input node by the capacitive feedback each time the output switches is $C_2 V_{\mathrm{dd}}$. The input and reset currents must counteract this charge before the output will switch again. During the period when the output is low, the reset pathway is disabled and the current at the input node is equal to i_{in}. The integrated charge during this period is $t_1 i_{\mathrm{in}}$ and must be equal to $C_2 V_{\mathrm{dd}}$. When the output is high, the reset pathway is enabled, and the current is $i_{\mathrm{in}} - i_{\mathrm{r}}$ (for $i_{\mathrm{r}} > i_{\mathrm{in}}$). The integrated charge during this period is $t_2(i_{\mathrm{in}} - i_{\mathrm{r}})$ and must be equal to $-C_2 V_{\mathrm{dd}}$.

In our discussion of driving motors with pulses in Section 3.2, we described two attributes of pulse trains that are important for the control of motors: duty cycle and minimum pulse width. We shall now analyze the behavior of the neuron circuit with respect to these attributes.

Since the duty cycle is the important temporal measure of a pulse train for determining the drive that a pulse train provides to a motor, we must show that the neuron circuit performs an acceptable transformation between the input current and the output duty cycle. By setting the sum of the charges at the input node during a complete pulse period equal to 0 (because of charge conservation), we can derive the following equation for the pulse duty cycle:

$$\frac{t_2}{t_1 + t_2} = \frac{i_{\mathrm{in}}}{i_{\mathrm{r}}} \tag{4.1}$$

This equation shows that the duty cycle of the pulse train is proportional to the input current, with a proportionality constant of $\frac{1}{i_{\mathrm{r}}}$. Thus, the transfer function of the neuron circuit (from i_{in} to output duty cycle) is constant, and the form of the input signal is retained at the output. This linearity is valid only when $i_{\mathrm{r}} > i_{\mathrm{in}}$; the circuit saturates to a duty cycle of 1 if this condition does not hold. Figure 4.2 shows data for the duty cycle versus the input current for the neuron circuit.

In Section 3.6, we also discussed why it is important that, for a pulse train to control a

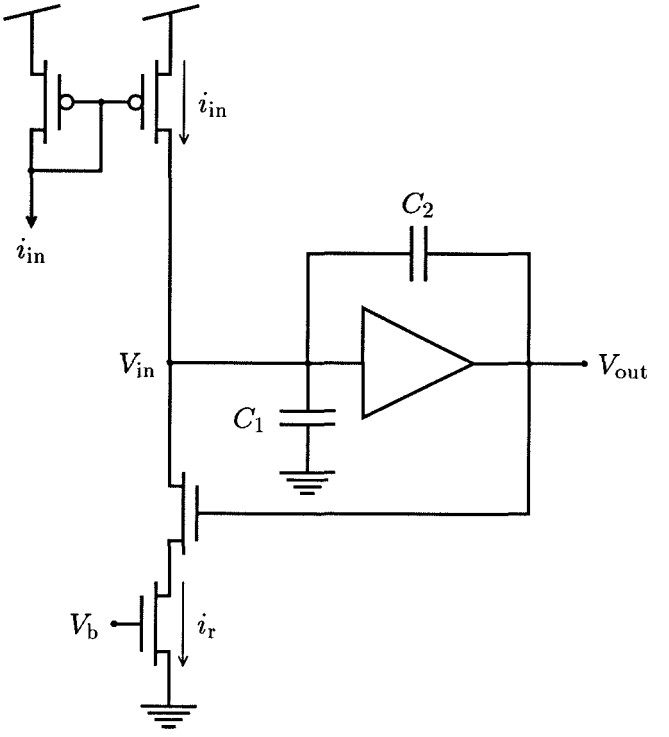

Figure 4.1: The neuron circuit. A current mirror produces a current i_{in}, which is integrated by the input capacitance ($C = C_1 + C_2$), causing V_{in} to increase. When V_{in} reaches the switching threshold of the amplifier ($V_{sw} \approx \frac{V_{dd}}{2}$), the output voltage switches high. This switching causes charge to be dumped on the input node due to capacitive feedback from capacitor C_2, increasing V_{in} by $\Delta V = \frac{C_2}{C} V_{dd}$. The high output also causes the two-transistor reset pathway to be activated, sinking reset current i_r off the input node (assuming $i_r > i_{in}$) and decreasing V_{in}. When V_{in} reaches V_{sw}, the output switches low, capacitively decreasing V_{in} (by ΔV), and deactivating the reset pathway. At this point, the circuit is reset and the process begins again.

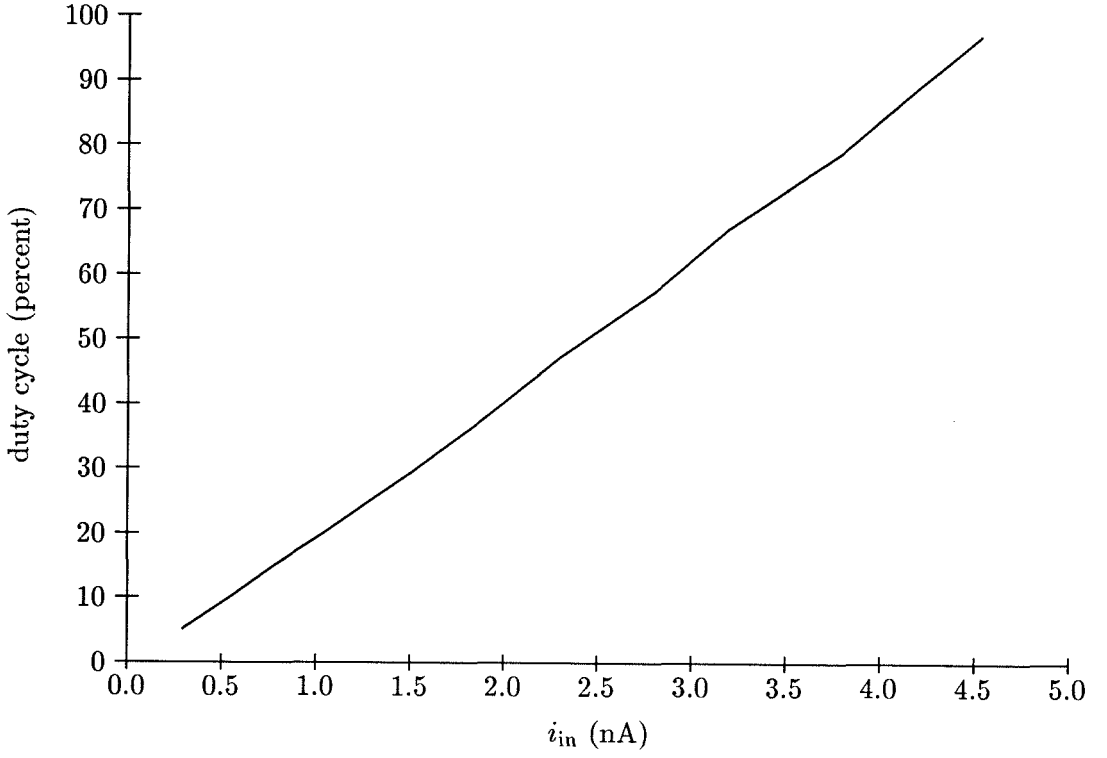

Figure 4.2: Duty cycle versus input current for the neuron circuit. The duty cycle should be linear with respect to the input current, with a slope of $\frac{1}{i_r}$.

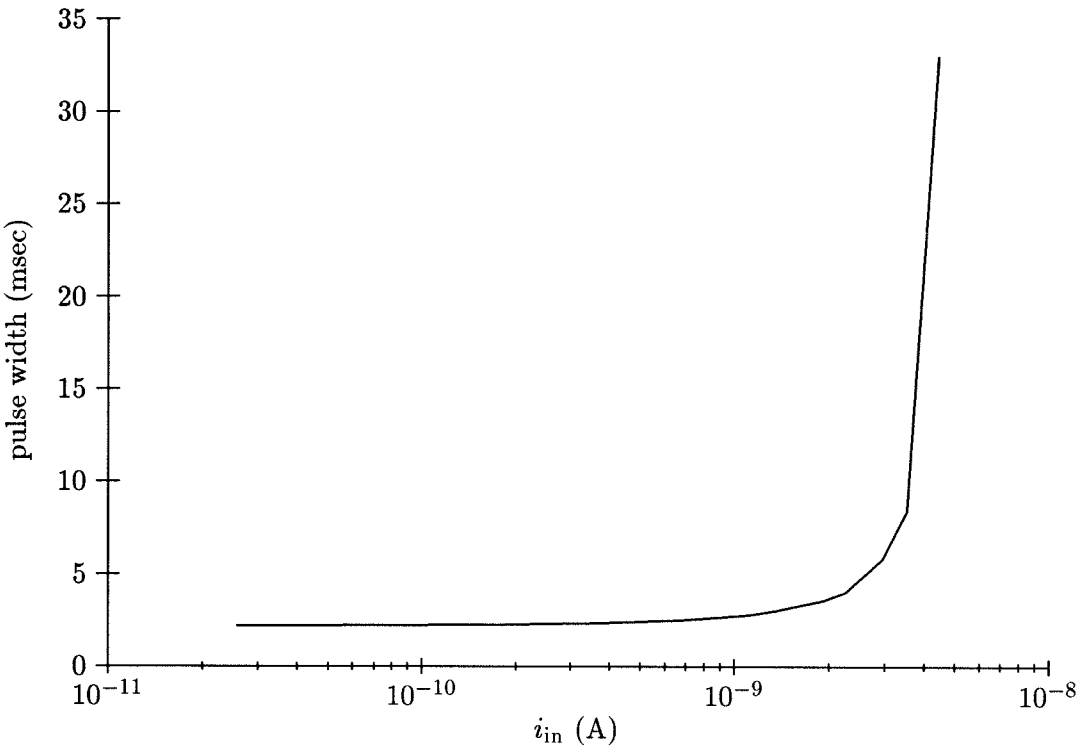

Figure 4.3: Pulse width versus input current for the neuron circuit. The pulse width is a function of the input current i_{in}. The pulse width asymptotically approaches C_2V_{dd}/i_r as i_{in} approaches 0.

DC motor system effectively, the pulse width not become smaller than some constant value. For the neuron circuit, the pulse width is described by

$$t_2 = \frac{C_2V_{dd}}{i_r - i_{in}} > \frac{C_2V_{dd}}{i_r} \tag{4.2}$$

We can see that the pulse width is bounded below by a constant that is independent of the input current. Thus, we are able to set the minimum pulse width for the circuit that is consistent with the restriction for a driving DC motor. Figure 4.3 shows the pulse width versus the input current for the neuron circuit.

4.1.2 Variations on the Neuron Circuit

In our preceding analysis of the neuron circuit, i_{in} is considered the only variable input signal (we assume that i_r is a constant). It may be advantageous, however, to use both

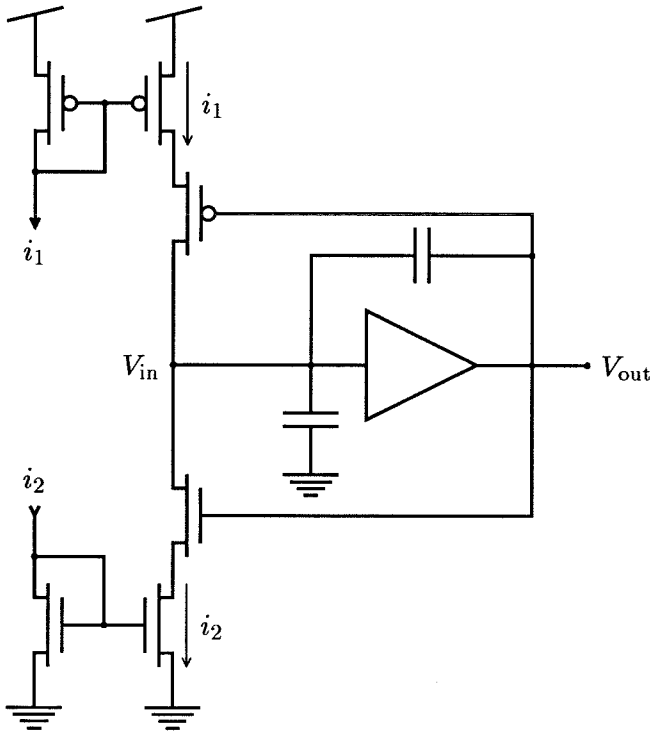

Figure 4.4: A neuron circuit with symmetric inputs. In this circuit, the positive input current i_1 is allowed to flow only when the output is low; the negative input current i_2 is allowed to flow only when the output is high. This behavior gives the circuit symmetric outputs, such that, when i_1 and i_2 are switched, the output pulse train is the inverse of the pulse train with the currents in their original configuration.

the input and reset currents as input signals. For example, the duty cycle of the circuit in Figure 4.1 is the ratio of these two currents; thus, the circuit can be used to compute a division.

The neuron circuit is asymmetric with respect to its two input signals, because the input current flows continuously, whereas the reset current flows only when the output is high. In situations where two complementary input signals are generated, a neuron circuit in which the inputs are symmetric would be useful. Figure 4.4 shows a circuit with an output that is symmetric with respect to its two input currents (i_1 and i_2). In this circuit, the positive input current i_1 is allowed to flow only when the output is low; the negative input current i_2 is allowed to flow only when the output is high. Thus, the inverse pulse period t_1 and the pulse period t_2 are dependent on only one input current each (i_1 and i_2, respectively).

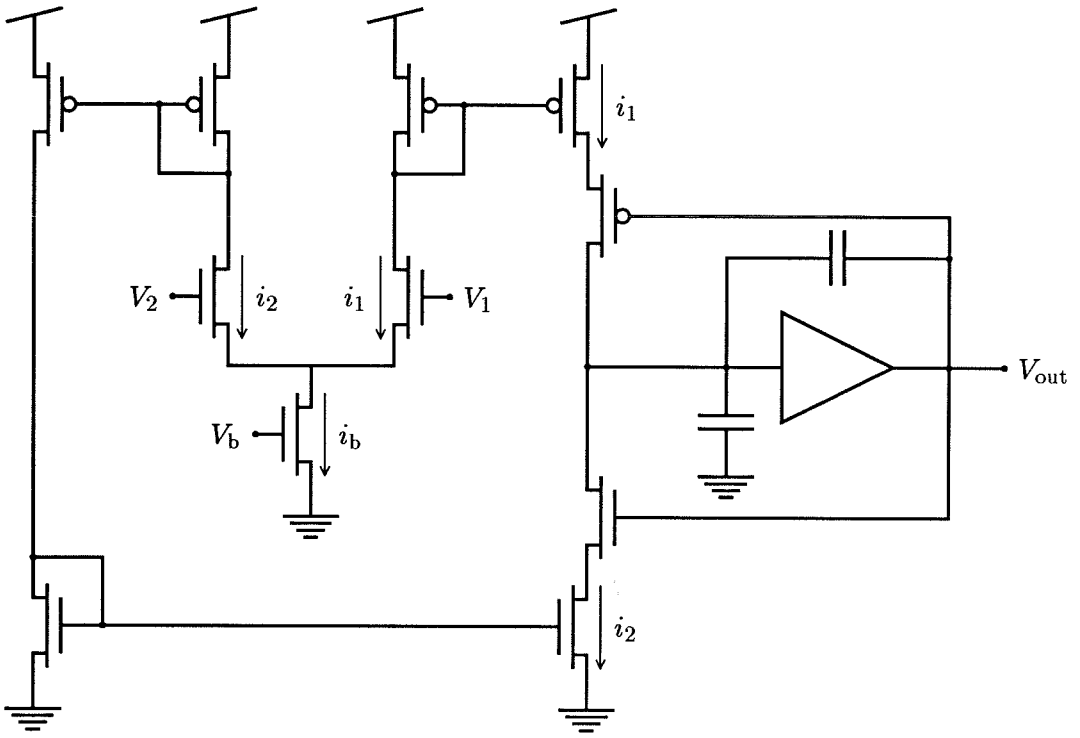

Figure 4.5: A neuron circuit with differential-pair input circuitry. The output currents from a differential pair are mirrored, and are used as input to a symmetric neuron circuit. The differential input voltage ($V_1 - V_2$) sets the duty cycle of the output, and the bias current (i_b) sets the time scale, such that, for a given duty cycle, the output frequency is proportional to the bias current.

We can see the symmetric nature of the circuit if we switch i_1 and i_2. In this configuration, the output pulse train is the inverse of the pulse train with the currents in their original configuration.

The differential pair is an example of a circuit that can be used to generate complementary currents. By using current mirrors at the output of the differential pair, we generate a pair of complementary currents that are the inputs of the symmetric neuron circuit, as shown in Figure 4.5. The input signals to the differential neuron circuit are the bias current and the differential voltage of the differential pair. This circuit has the advantage over the original neuron circuit that the duty cycle and the time scale of the pulse train are set by independent input signals.

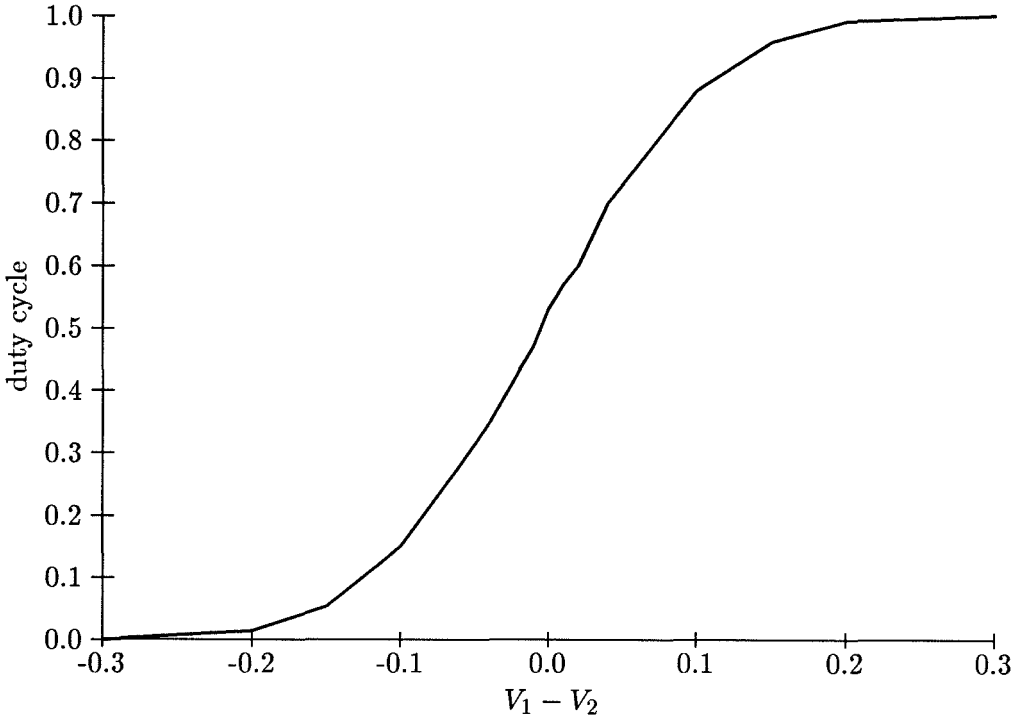

Figure 4.6: Duty-cycle data from the differential neuron circuit. The duty cycle of the differential neuron circuit is a sigmoidal function of the differential input voltage saturating at duty cycles of 0 and 1.

The duty cycle for the differential neuron circuit is defined by the following equation:

$$\frac{t_2}{t_1 + t_2} = \frac{1}{1 + e^{\kappa(V_2 - V_1)}}$$

From this equation, we can see that the duty cycle of the circuit is a sigmoidal function of the differential input voltage and is independent of the input current. Since $i_1 + i_2$ is constant, there is no factor multiplying the sigmoid. Thus, the asymptotic values to which the duty cycle saturates are fixed at 0 and 1. Figure 4.6 shows data of duty cycle versus differential voltage taken from a differential neuron circuit.

The output frequency for the differential neuron circuit is described by the following equation:

$$\frac{1}{t_1 + t_2} = i_b \frac{1}{C_2 V_{dd}} \frac{1}{1 + e^{\kappa(V_2 - V_1)}} \frac{1}{1 + e^{\kappa(V_1 - V_2)}}$$

For a given duty cycle (differential input voltage), the frequency of the output pulse train is proportional to the bias current. Figure 4.7 shows frequency versus bias current data for this circuit.

4.2 Bidirectional Motion Control

To control many systems, we must be able to drive the system bidirectionally. To control the motion of a one-dimensional motor system bidirectionally, we must use a signed control signal. For example, bidirectional control using DC voltages is generally accomplished using a control signal that can be either positive or negative, depending on the direction of the intended motion. The use of pulses as control signals, however, presents a problem, due to the inherently unsigned (unidirectional) character of a pulse train. It is possible, however, to achieve bidirectional control by using the motor system to subtract two unsigned signals to generate the equivalent of signed signal. This subtraction is useful if the input to the motor system is a dual-rail configuration in which two unsigned signals are generated such that the difference between these signals is the intended control signal for the system.

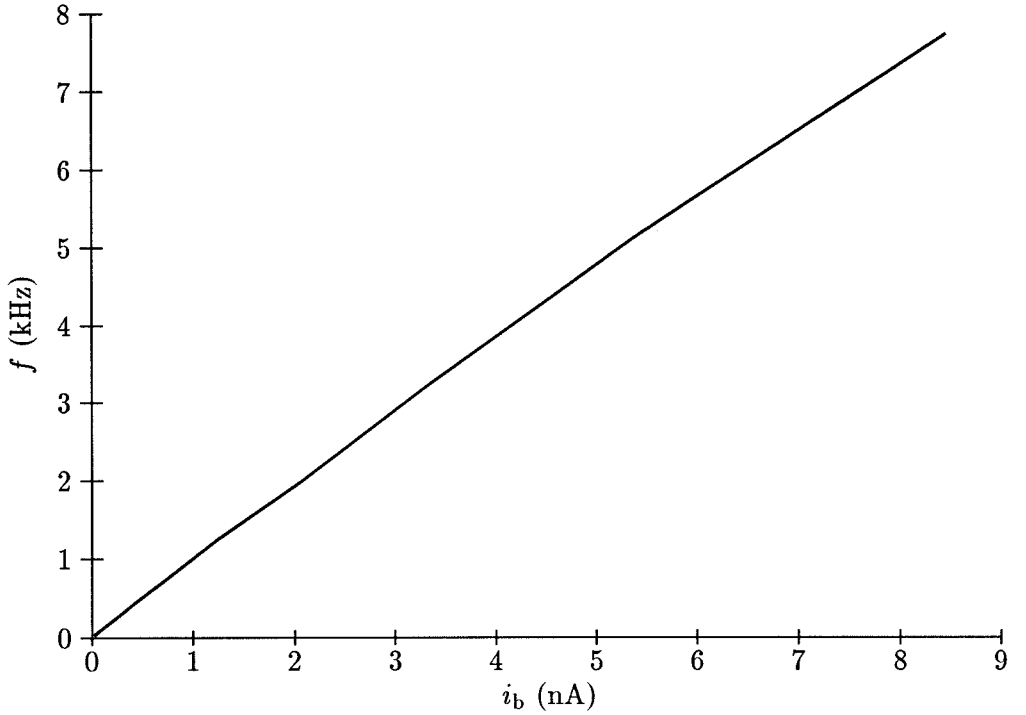

Figure 4.7: Frequency data from the differential neuron circuit. For a given duty cycle, the frequency of the differential neuron circuit is a linear function of the bias current, as shown here.

P_1	P_2	Fraction
low	low	$(1-\alpha_1)(1-\alpha_2)$
high	low	$\alpha_1(1-\alpha_2)$
low	high	$(1-\alpha_1)\alpha_2$
high	high	$\alpha_1\alpha_2$

Table 4.1: Overlapping states of two pulse trains.

4.2.1 The Dual-Rail Pulse Encoding

A dual-rail pulse encoding consists of a pair of pulse trains in which the difference between the duty cycles encodes the control signal. We can configure a DC motor system such that the two pulse trains independently drive the system in opposite directions. When both of the pulse trains are active, the input signals affect the system as follows. If one of the pulse inputs is high and the other is low, the motor system will be driven in the direction in which the high pulse train drives the system. However, if the inputs are either both high or both low, the effective control signal will be 0.

To analyze the control quantitatively, we can define two pulse trains P_1 and P_2 with duty cycles α_1 and α_2, respectively. The combination of the two pulse trains can occur in one of four possible states: both low, P_1 high (P_2 low), P_2 high (P_1 low), or both high. The fraction of the time (in terms of α_1 and α_2) that the pulse trains are, on average, in each of these states is shown in Table 4.1. Each fraction is the product of the fraction of the time that each pulse train is in that particular state (A is low, for example, $1-\alpha_1$ of the time). These fractions are true on average, or over a combined period of the two pulse trains (the least common multiple of the individual periods).

If we define A to be the signal that drives the system in the positive direction and B to be the signal that drives the system in the negative direction, then the control signal is proportional to the difference between the fraction of the time that A is high (and B is low) and the fraction of the time that B is high (and A is low). From Table 4.1, this difference is $\alpha_1(1-\alpha_2) - \alpha_2(1-\alpha_1) = \alpha_1 - \alpha_2$, or the difference between the two duty cycles.

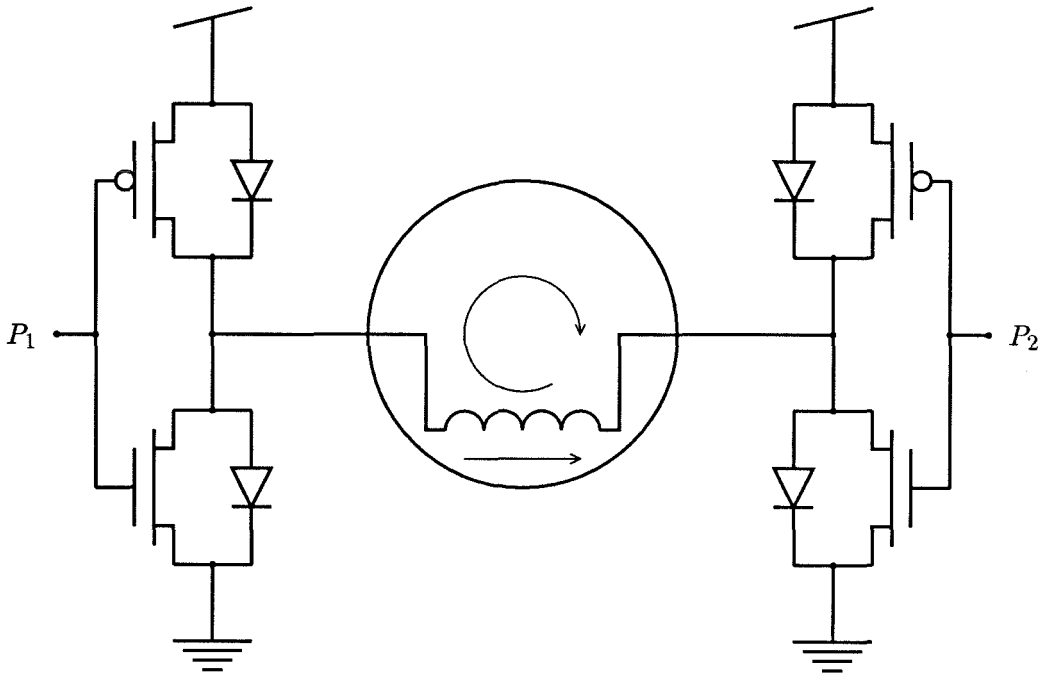

Figure 4.8: H-bridge configuration for bidirectional control. Two transistors, configured as an inverter, are used to drive each terminal of the motor. When the input pulse train is high, the signal to the motor will be low; when the input pulse train is low, the signal to the motor will be high. Thus, the effect of the control signal is inverted. Each transistor has an accompanying diode that prevents the voltage across the transistor from inverting. In other words, the diodes restrict the terminal voltages to be within the power-supply voltage rails.

4.2.2 Bidirectional Driving of a Single Motor

We can control the motion of a single motor using a dual-rail pulse encoding by driving each terminal of the motor with one of the two pulse trains. If the pulse inputs are either both high or both low, the voltage across the motor is 0 V; if only one pulse input is high, the voltage across the motor is either positive or negative, depending on which pulse input is high. Since the drive to the motor is proportional to the voltage across the terminals, this configuration subtracts the two pulse trains, and the effective control signal is proportional to the difference of the two duty cycles.

Since we are driving the motor using low-power circuitry, we must use circuitry ex-

ternal to the chip to amplify the pulse trains before we input them to the motor. The standard configuration for driving a motor bidirectionally using pulses is called an *H-bridge* (Figure 4.8). Each terminal of the motor is driven by an inverter that amplifies the input signals (and inverts the sign of the control signal). Diodes are used to counteract inductive effects from the motor by restricting the terminal voltages to be within the power-supply voltage rails.

4.2.3 Antagonistic Control

It is often advantageous to use multiple actuators to effect motion in a system with only a single degree of freedom. Biological systems, for example, perform bidirectional control through the use of antagonistic actuators (muscles) that pull in opposite directions on the same joint. Because muscles are unidirectional actuators (their only active motion is to contract), biological systems must use at least two muscles to control a single joint.

Antagonistic control has advantages even when we use actuators (such as motors) that have the potential for active bidirectional motion. The primary advantage of antagonistic motor control is that the system can be kept under tension even when the differential control signal is 0 (and the system is at rest). In a biological motor system, the motor neurons driving the antagonistic muscles have a background firing rate that causes the muscles to be under tension at all times [She79]. This tension keeps the system firm so that it can respond more smoothly to varying loads. The tension also keeps the systems from experiencing backlash when the direction of motion is changed.

Figure 4.9 shows a diagram of a linear antagonistic mechanical system. In this system, two motors are used to control the motion of the system. Each motor generates a force on the load such that the two forces are applied in opposite directions. The total force on the system is the difference between these two forces. Since each force is a linear function of the duty cycle of the related pulse train, the total force is proportional to the difference between the two duty cycles.

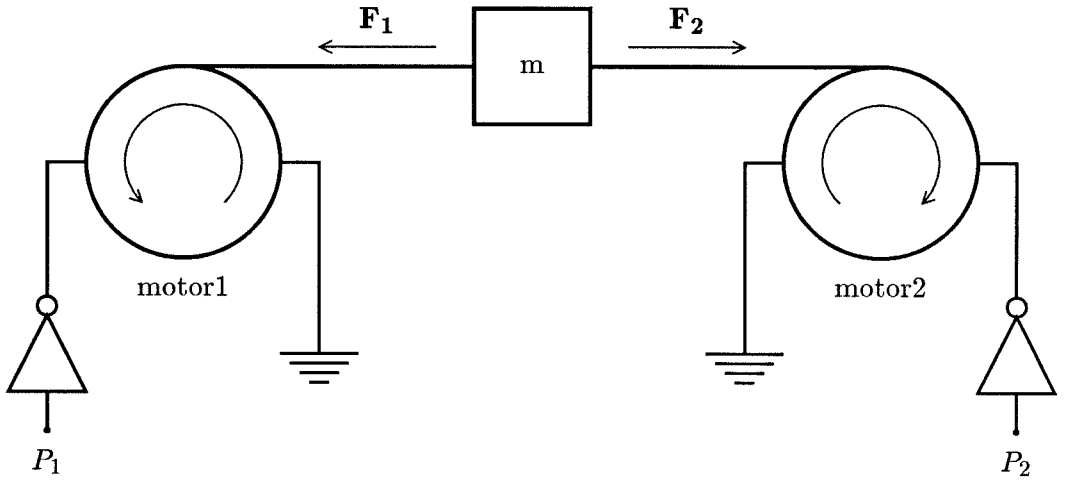

Figure 4.9: Antagonistic configuration for bidirectional control. A pair of motors is driven by two input pulse trains P_1 and P_2. The pulse trains are amplified using two half H-bridge circuits. Each motor generates a torque proportional to the duty cycle of the corresponding pulse train. These torques are converted to forces F_1 and F_2 that effect bidirectional motion on a mass m. The total force (in the direction of F_1) is $F_1 - F_2$. Since F_1 and F_2 are linear functions of the duty cycles P_1 and P_2, the combined force is proportional to the differential duty cycle $P_1 - P_2$.

4.3 Dual-Rail Pulse Generation

We shall now describe circuitry to generate dual-rail pulse encodings for bidirectional motor control. We have chosen an input scheme that uses the differential pair as the input device. The sigmoidal functionality of the neuron circuit gives the output of the circuit a smoothly saturating behavior, rather than the hard saturation that occurs when the system is driven linearly and the duty cycle reaches 0 or 100 percent. The two output currents from the differential pair also have the characteristic that, at 0 differential output current, neither of the currents is 0. This fact makes the neuron circuit an ideal input device to generate antagonistic motor signals because, when converted to pulses, the currents will generate an inherent background firing rate in the pulse outputs.

4.3.1 A Single-Input Circuit

The differential neuron (Figure 4.5) is an example of a circuit that can be used to generate a smoothly saturating dual-rail pulse encoding. The duty cycles of output and the inverse of the output of this circuit constitute a pair of sigmoidal functions. If we use these two signals as inputs to the two terminals of a bidirectional motor system, the system will subtract the two signals. The resulting control signal will be proportional to the hyperbolic tangent of the differential input voltage, and the system will have a smoothly saturating, bidirectional behavior.

The use of a pulse train and its inverse (such as that used in the differential neuron circuit) for bidirectional control is limited because we have no control over the maximum magnitude of the differential duty cycle. For this control paradigm to work correctly and be symmetric, the duty cycle must saturate at 0 and 100 percent. The differential neuron circuit has two input signals that control the duty cycle and the time scale. If we want to have control over the maximum differential duty cycle and thus the magnitude of the control signal, we must have a circuit that has a third degree of freedom to control the magnitude of the output.

In Figure 4.10, we show a circuit that uses two neuron circuits to generate pulse trains whose duty cycles can be controlled such that they sum to any value between 0 and 100 percent. This circuit has three degrees of freedom that arise from three control signals (the

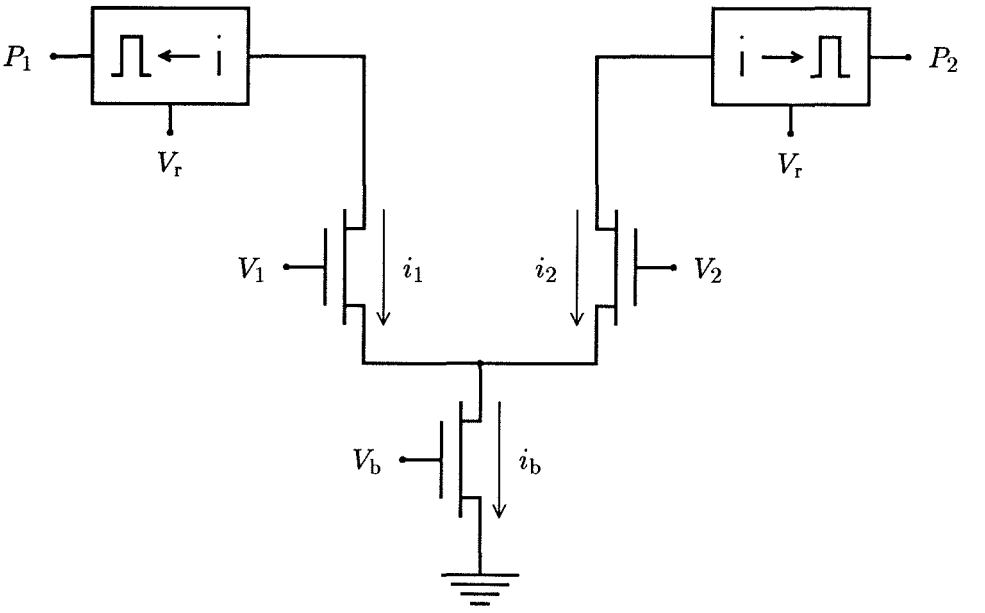

Figure 4.10: Dual-rail pulse generation using a differential pair. The circuit consists of a differential pair (with bias transistor) and two neuron circuits. The circuit has three input signals (i_b, i_r, and $\Delta V = V_1 - V_2$), and thus, three degrees of freedom. The outputs from the two neuron circuits are symmetric sigmoidal functions of ΔV, with the maximum magnitude of the duty cycle defined by the ratio i_b/i_r. The time scale of both outputs is defined by i_b and i_r.

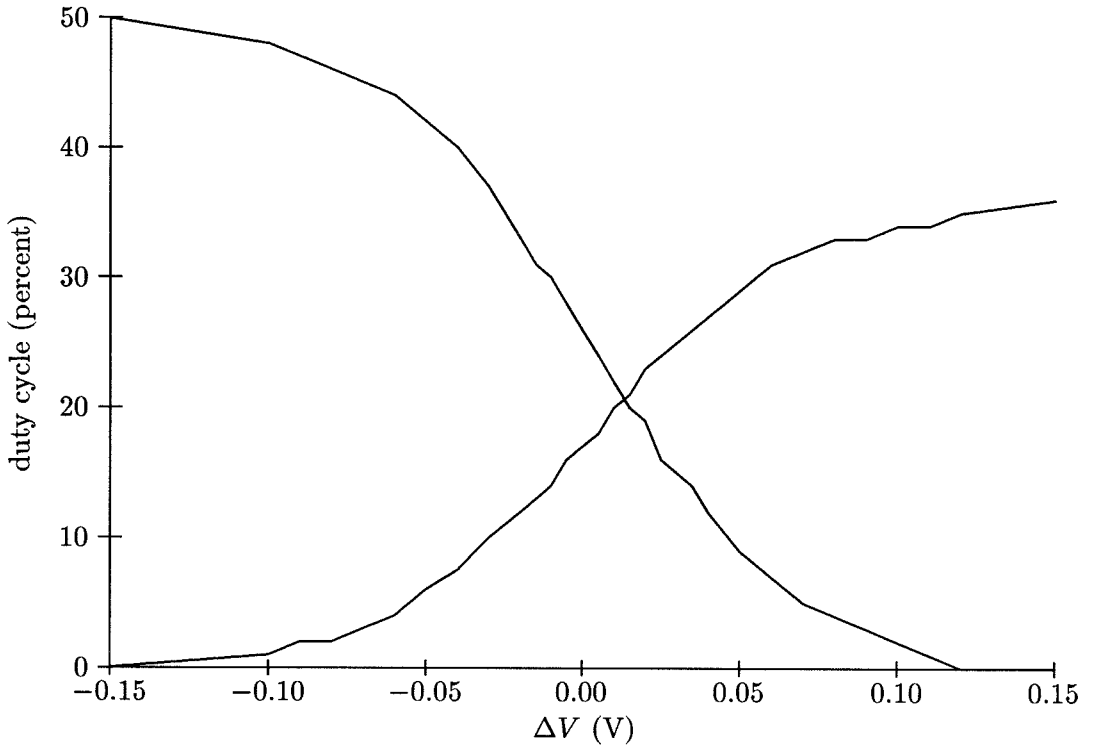

Figure 4.11: Duty cycle data from the dual-rail pulse-generating circuit. The outputs from the circuit are two pulse trains with symmetric sigmoidal duty cycles. The positive asymptote is defined by the ratio i_b/i_r. There is an obvious difference between the positive asymptotes for the two neuron circuits due to mismatches in the current mirrors and reset transistors.

differential voltage, the bias current, and the neuron reset current). As in the differential neuron circuit, the differential voltage controls the sigmoidal response of the circuit. The time scale of the circuit is controlled by the common mode of the two input currents. In this circuit, however, we can also control the maximum magnitude of the duty cycle (the position of the asymptotes) by varying the differential mode (the ratio) of the two currents. The duty cycles of the outputs are inverse sigmoidal functions with maximum magnitude equal to the ratio of the bias current to the input current. In Figure 4.11, we show duty-cycle data taken from this dual-output circuit. The general form of these data is similar to that of antagonistic motoneuron data taken from biological systems [Bal86].

4.3.2 Aggregation and Pulse Generation

If we are to develop sensorimotor systems using the analog VLSI design paradigm, we must develop circuitry that converts a sensory representation into a motor representation. This circuitry should allow for the aggregation of any number of sensory input signals, and should produce a dual-rail pulse encoding at the output. In Chapter 2, we presented an aggregation network of differential-pair circuits (Figure 2.3) that takes an array of input signals (currents and differential voltages) and produces a pair of output currents. Using the methods for generating pulses presented in this chapter, we can convert the aggregation network into a circuit that converts an array of inputs into a dual-rail pulse encoding at its output. This pulse-output aggregation network is shown in Figure 4.12.

We shall use the pulse-output aggregation circuit as a basis for our development of sensorimotor systems in the remaining chapters. We shall show how this circuit can be used to aggregate sensory input information and to control motor systems, resulting in complete sensorimotor systems consisting of analog VLSI circuitry, actuators, and feedback.

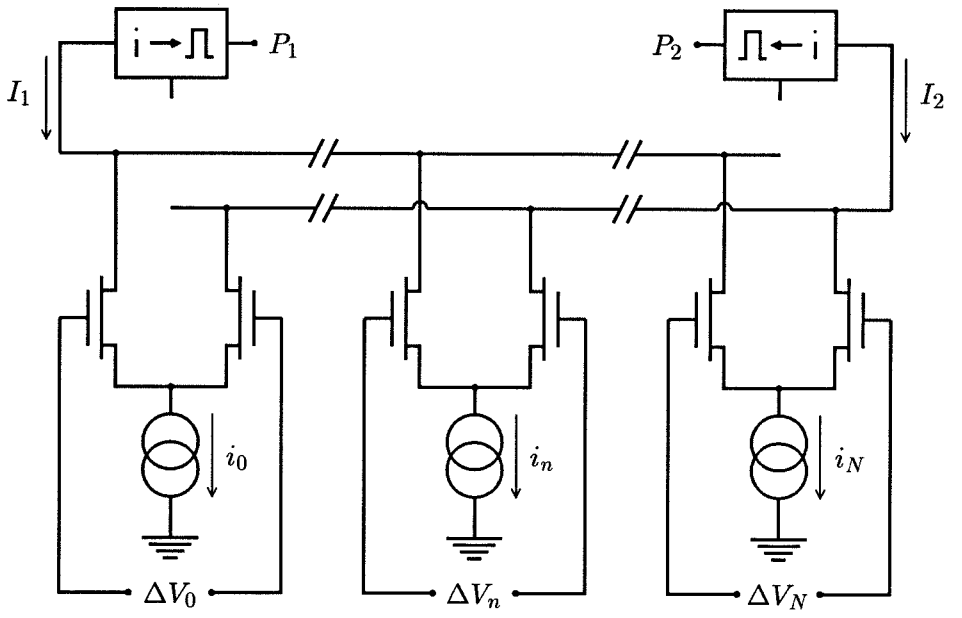

Figure 4.12: Dual-rail pulse generation using the aggregation network. The inputs to each differential-pair element are a differential voltage ΔV_n and a bias current i_n. The output currents are aggregated onto a pair of wires, and these aggregated currents are used as input to two neuron circuits. The complete circuit produces a dual-rail pulse encoding in which the differential duty cycle is set by the aggregated input, and the maximum duty cycle is equal to the ratio $(\sum i_n)/i_r$.

Chapter 5

Circuits for Feedback Control

Technology has strongly influenced the way that engineers think about control systems. Implementations using relays, operational amplifiers, and microprocessors have individual requirements that affect their uses. The microcomputer revolution, for example, has had a major influence on control-system implementations [ÅW84]. As a consequence, control theory has developed in new directions, and more advanced control schemes, such as that of adaptive control, have become practical [ÅW89].

We have developed controllers impemented using the analog VLSI circuits presented in Chapter 4 [DNMÅ90] [DNMÅ91]. The use of analog VLSI circuits in the implementation of control systems also affects the type of control that can be accomplished. By using this paradigm, we can make controllers with a large number of control elements in a small area. These systems require silicon surface area and power dissipation that is many orders of magnitude smaller than a conventional system with A/D, CPU, memory, and D/A.

In this chapter, we describe an analog VLSI circuit framework into which traditional controllers can be embedded. We discuss the operation of a proportional-derivative controller implemented within this framework. We also develop a method of analyzing controllers containing certain nonlinear elements.

5.1 The Analog VLSI Control Framework

We have chosen a weighted combination of sigmoidally compressed differential inputs as the generic form for our controller. The control signal u generated by the controller is described

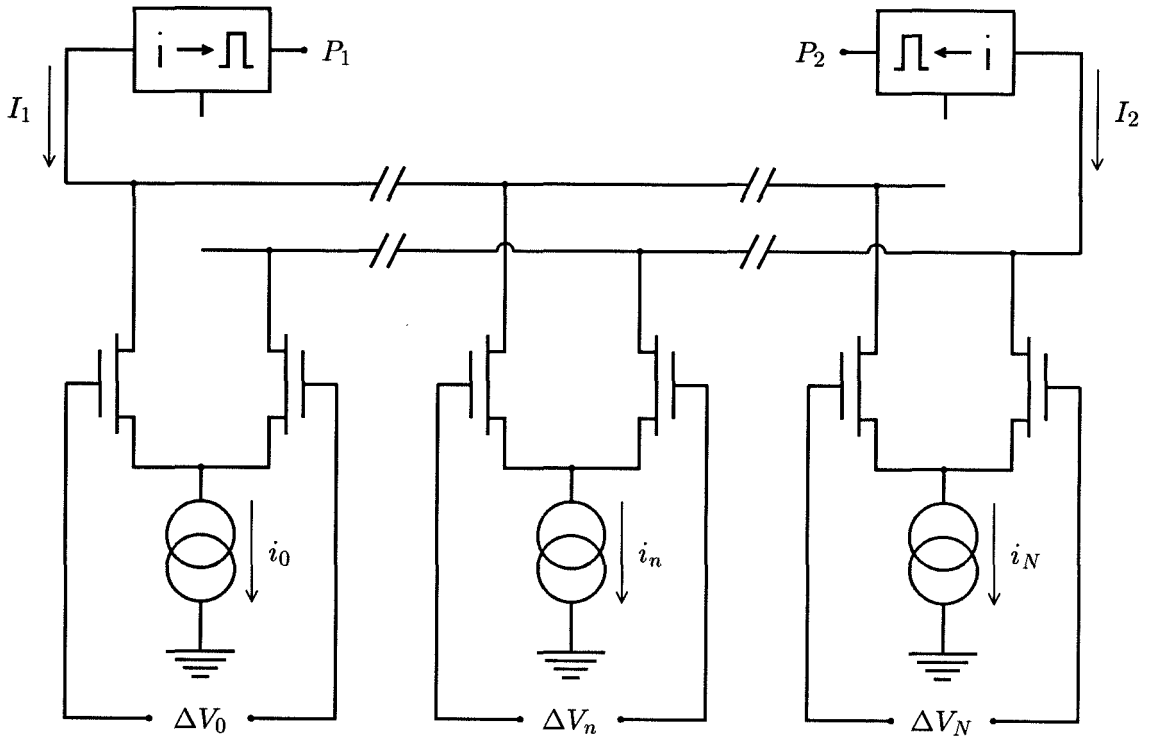

Figure 5.1: Circuit implementation of the control framework. The framework is a multiple-input circuit with a dual-rail pulse output. The input elements are differential pairs, and the output elements are neuron circuits. The input signal $r_n - x_n$ to the nth element in encoded in the differential-pair input voltage ΔV_n. The gain of the nth element is the ratio of the bias current i_n to the neuron reset current i_r.

by the following equation:

$$u = \sum_n k_n \tanh(r_n - x_n) \tag{5.1}$$

where k_n is the proportionality constant and $r_n - x_n$ is the differential input for the nth stage. This system allows for the implementation of control elements that consist of functional elements multiplied by constants, and have input signals that can be turned into a differential voltages. This system also allows for simple replication of the individual functional elements.

In Figure 5.1, we show the circuit implementation for the control design frame. We implement the design frame using the dual-rail, pulse-output aggregation network described in Chapter 4. The differential input to the nth element $(r_n - x_n)$ is encoded as the differential

input voltage (ΔV_n) of the differential pair. The multiplicative constant k_n is set by the ratio of the bias current at the nth element to the reset current of the neuron circuits. The control signal u of the circuit is encoded in the differential duty cycle of the output pulse trains, and is described by the equation

$$u = \sum_n \frac{i_n}{i_\mathrm{r}} \tanh\left(\frac{\Delta V_n}{2U_\mathrm{t}/\kappa}\right) \tag{5.2}$$

which is equivalent in form to Equation 5.1.

In the small-signal regime ($\Delta V_n < 2U_\mathrm{t}/\kappa$), we approximate Equation 5.2 by the following equation:

$$u = \sum_n \frac{i_n}{i_\mathrm{r}} \frac{\Delta V_n}{2U_\mathrm{t}/\kappa}$$

Controllers embedded in this framework and operated in the small-signal regime can be analyzed using traditional linear control theory.

In the large-signal regime ($\Delta V_n > 2U_\mathrm{t}/\kappa$), we approximate Equation 5.2 by the following equation:

$$u = \sum_n \frac{i_n}{i_\mathrm{r}} \mathrm{sgn}(\Delta V_n)$$

Controllers embedded in this framework and operated in the large-signal regime act as relay controllers [Tsy84]. We shall develop theory to approximate the operation in this regime.

5.2 A Proportional–Derivative Controller

We have studied the effects of feedback on the control the angular displacement (position) of a motor. We measured the angle of the motor using a potentiometer connected to the shaft of the motor. Since the angular position of a rotating system is the integral of the angular velocity, the transfer function relating the position θ of a motor to its input voltage V can be derived from Equation 3.6. This positional transfer function is

$$\frac{\theta}{V} = \frac{1/k_\mathrm{B}}{s(\tau_\mathrm{m} s + 1)} \tag{5.3}$$

The intertial load due to the potentiometer and the connecting structure increase the time constant τ_m of the system.

We have developed a controller for this system using the traditional proportional–derivative (P–D) feedback control scheme [FPEN86]. We have analyzed the system operating in the small-signal regime using the standard method for analyzing linear P–D controllers. We have developed a method for analyzing the system operating in the large-signal regime. This method approximates the operation of the controller in this regime and is extendable to other controllers that have piecewise constant inputs. This method is similar to methods used for analyzing switching power supplies [MĆ76].

5.2.1 Proportional–Derivative Feedback

When the control signal in a feedback controller is linearly proportional to the error signal, the resulting control scheme is called *proportional feedback* [FPEN86]. Given a system with voltage input V_{in} and voltage output V_{out}, the form of this feedback is

$$V = K_{\text{P}}(V_{\text{in}} - V_{\text{out}})$$

where V is the control voltage generated by the controller and K_{P} is the *proportional gain.*

When the control signal in a feedback controller is linearly proportional to the derivative of the error signal, the resulting control scheme is called *derivative* or *rate feedback* [FPEN86]. We have chosen a feedback scheme where the derivative is computed only on the output (and not on the input) of the system. This variant on rate feedback causes the control signal to be independent of changes in the input voltage. The control voltage V generated by the controller is a function of the output voltage V_{out} as described by the following equation:

$$V = T_{\text{D}} \frac{dV_{\text{out}}}{dt} = T_{\text{D}} s V_{\text{out}}$$

where T_{D} is a constant called the *derivative time.*

Figure 5.2 shows the proportional-derivative feedback control system. The transfer function of this system is described by the following equation:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{1}{\frac{\tau_{\text{m}}}{k_0 K_{\text{P}}} s^2 + (\frac{T_{\text{D}}}{K_{\text{P}}} + \frac{1}{k_0 K_{\text{P}}})s + 1} \tag{5.4}$$

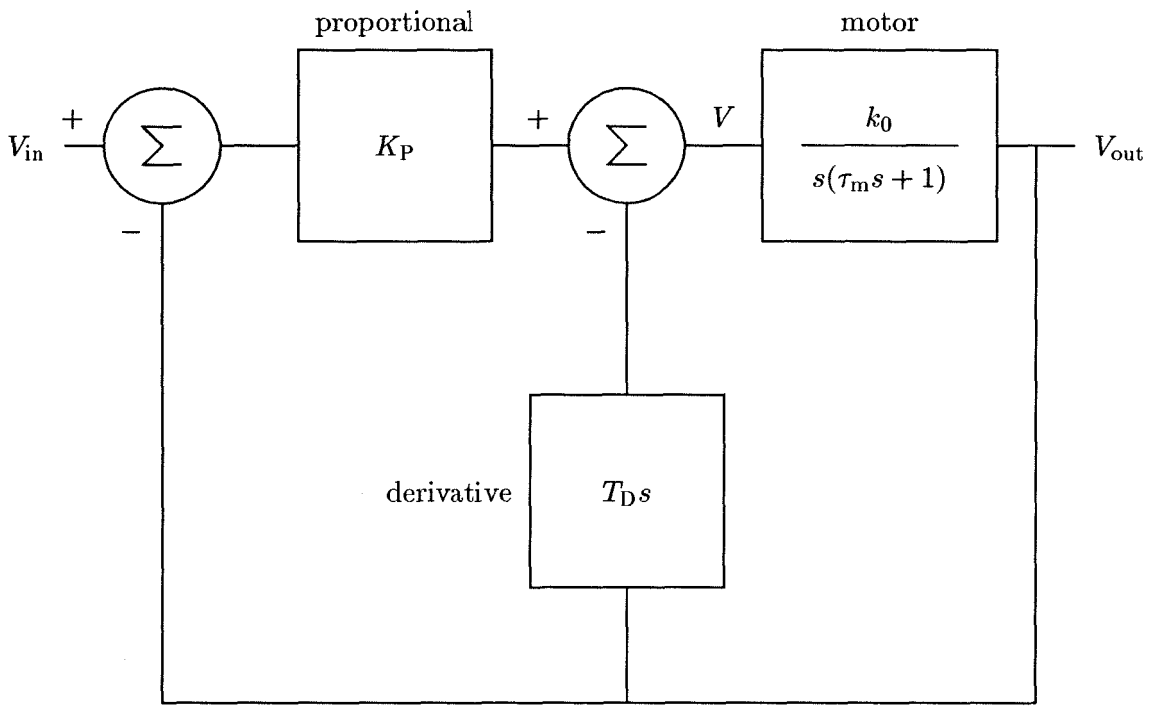

Figure 5.2: The proportional–derivative feedback system. The system consist of a motor, a proportional element, and a derivative element. The voltage V_{in} sets the set-point position of the system. The voltage V_{out} is produced by a potentiometer attached to the shaft of the motor, and encodes the angular position of the shaft. The proportional element multiplies the error signal ($V_{\text{in}} - V_{\text{out}}$) by the proportional gain (K_{P}). The derivative element multiplies the derivative of the V_{out} by the derivative time (T_{D}).

A standard second-order form for transfer equations [Wil60] is

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{1}{\frac{1}{\omega_{\text{n}}^2}s^2 + \frac{2\zeta}{\omega_{\text{n}}}s + 1} \tag{5.5}$$

where ω_{n} is the *undamped natural frequency* and ζ is the *damping ratio.* At $\zeta = 1$, the poles coincide on the negative real axis and the system is critically damped. When ζ is less than 1, the poles of this equation are complex and the system reponse is underdamped.

The natural frequency and damping ratio for the system described by Equation 5.4 are defined by the following equations:

$$\begin{aligned} \omega_{\text{n}} &= \sqrt{\frac{k_0 K_{\text{P}}}{\tau_{\text{m}}}} \\ \zeta &= \frac{1 + k_0 T_{\text{D}}}{\sqrt{4 k_0 K_{\text{P}} \tau_{\text{m}}}} \end{aligned}$$

In general, the proportional gain is set to a large value to decrease the DC offset of the system (by making the closed-loop gain near 1), and the derivative time is set such that the system is damped sufficiently for the appropriate application [FPEN86].

There are three parameters of interest that are often analyzed for damped oscillations in control-system applications [FPEN86]. The first of these is the *overshoot* or *peak magnitude*, M_{p}. This parameter is defined as the magnitude of the first peak in the oscillation of the system when presented with an input step of magnitude equal to 1. The second parameter is the *rise time*, t_{r}. This parameter is defined as the time that it takes for the the angular position to go from 10 to 90 percent of the difference between the initial position and the asymptotic position (set-point). The third parameter is the *settling time* t_{s}. This parameter is defined as the time required for the oscillations to decrease to a small value (we use 1 percent) so that the output is almost in steady state. The values for these parmeters in a linear system (from [FPEN86]) are

$$\begin{aligned} M_{\text{p}} &= e^{-\pi\zeta/\sqrt{1-\zeta^2}} \\ t_{\text{r}} &\approx \frac{1.8}{\omega_n} \\ t_{\text{s}} &= \frac{4.6}{\zeta\omega_n} \end{aligned}$$

5.2.2 Circuit Implementation

We implement the proportional–derivative controller in the control framework as shown in Figure 5.3. A single differential pair acts as the proportional section of the controller. The input is the difference between the input voltage and the output voltage. The proportional gain is set by the ratio of the bias current i_{P} of this circuit divided by the reset current i_{R} of the neuron circuits. The average control voltage generated strictly by the proportional circuit is described by the following equation:

$$V_{\mathrm{av}} = V_{\mathrm{p}} \frac{i_{\mathrm{P}}}{i_{\mathrm{R}}} \tanh\left(\frac{V_{\mathrm{in}} - V_{\mathrm{out}}}{2U_{\mathrm{t}}/\kappa}\right)$$

where V_{p} is the amplitude of the pulse train (as described in Chapter 3). In the small signal regime, this equation can be approximated by the following equation:

$$V_{\mathrm{av}} \approx \frac{i_{\mathrm{P}}}{i_{\mathrm{R}}} \frac{V_{\mathrm{p}}}{2U_{\mathrm{t}}/\kappa}(V_{\mathrm{in}} - V_{\mathrm{out}})$$

This equation describes a proportional controller with

$$K_{\mathrm{P}} = \frac{i_{\mathrm{P}} V_{\mathrm{p}}}{i_{\mathrm{R}} 2U_{\mathrm{t}}/\kappa}$$

We implement the derivative section of the controller with a differential pair and a follower–integrator circuit [Mea89]. The follower–integrator circuit is a low-pass filter with a time constant (τ_{D}) set by the bias current of its transconductance amplifier. The input voltage to the follower–integrator circuit is the output voltage of the system (V_{out}). The differential pair subtracts V_{out} from the low-passed version of V_{out} (the output of the follower–integrator) to form an approximation to the derivative. The derivative time is set by the time constant of the follower–integrator multiplied by the ratio of the bias current i_{D} of this circuit divided by the reset current i_{R} of the neuron circuits. Assuming the follower–integrator circuit is linear, the average control voltage generated strictly by the derivative circuit is described by the following equation:

$$V_{\mathrm{av}} = V_{\mathrm{p}} \frac{i_{\mathrm{D}}}{i_{\mathrm{R}}} \tanh\left(\frac{\tau_{\mathrm{D}} s}{\tau_{\mathrm{D}} s + 1} \frac{V_{\mathrm{out}}}{2U_{\mathrm{t}}/\kappa}\right)$$

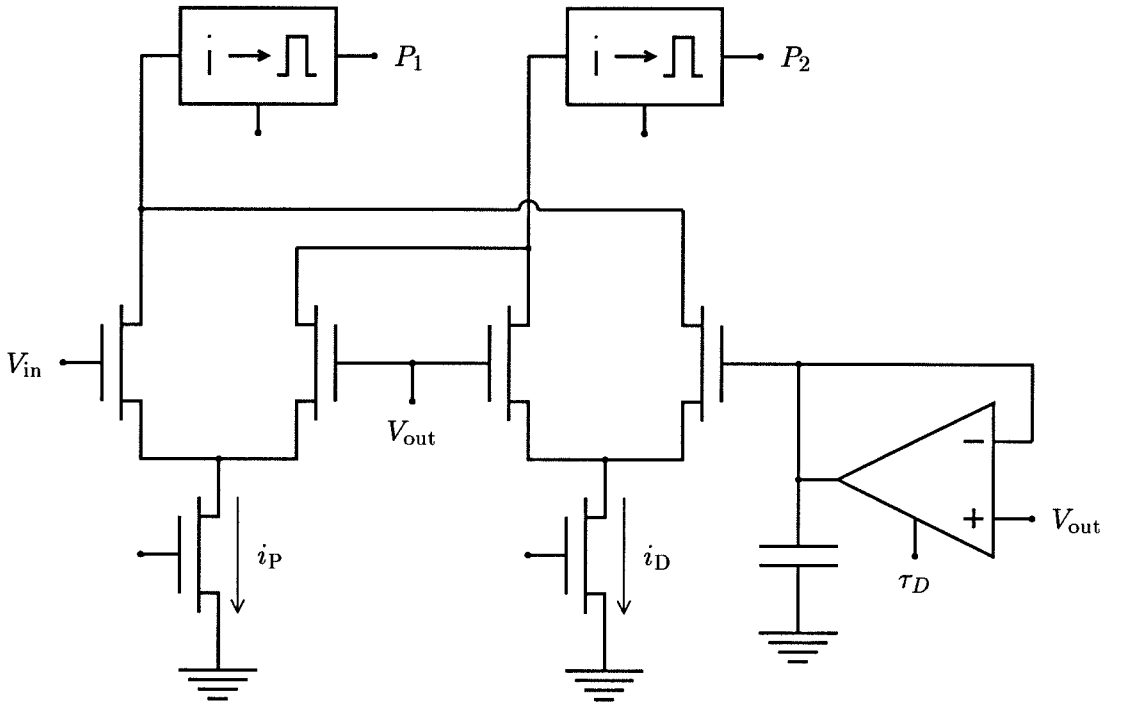

Figure 5.3: Proportional–derivative controller circuitry. The proportional section of the controller consists of a single differential pair, which has input $V_{\text{in}} - V_{\text{out}}$ and gain (bias) current i_{P}. The derivative section consists of a differential pair and a follower–integrator circuit [Mea85] with time constant τ_{D}. The differential pair subtracts V_{out} from the temporally delayed version of V_{out} that is produced by the follower–integrator. This difference encodes the derivative of V_{out}. The gain of the element is set by the current i_{D}.

If τ_D is set such that $\tau_D s$ is much less than 1, the preceding equation can be approximated by the following equation:

$$V_{av} \approx V_p \frac{i_D}{i_R} \tanh\left(\tau_D s \frac{V_{out}}{2U_t/\kappa}\right)$$

In the small-signal regime, the preceding equation can be approximated by the following equation:

$$V_{av} \approx \frac{i_D}{i_R} \frac{V_p}{2U_t/\kappa} \tau_D s V_{out}$$

The preceding equation describes a derivative controller with

$$T_D = \frac{i_D V_p}{i_R 2U_t/\kappa} \tau_D$$

5.3 Large-Signal Analysis

Differential-pair offset voltages in the P–D controller make it difficult to obtain reliable operation in the small-signal regime, because the magnitudes of the input signals are in the same range as the offset voltages. The linear approximation for the transfer function of the differential pairs in the control framework is not valid in the large-signal regime, however. Thus, we must analyze the operation of the circuit using a method different from that used in the analysis of the linear system.

As stated in Section 5.1, we approximate the differential output current of a differential pair operated in the large-signal regime as

$$i_{out} = i_b \mathrm{sgn}(\Delta V_{in})$$

The implementation of this function in our controller causes the outputs of the controller to be thresholded functions of the inputs. Controllers that exhibit this type of thresholded behavior are referred to as *relay controllers*. Researchers have analyzed the behavior of such systems using many different methods [Tsy84]. We shall present a method for analyzing the step-response of these systems using the fact that the input to the system (the output from the controller) is a piecewise constant function whose value changes at each point at which one of the control elements switches.

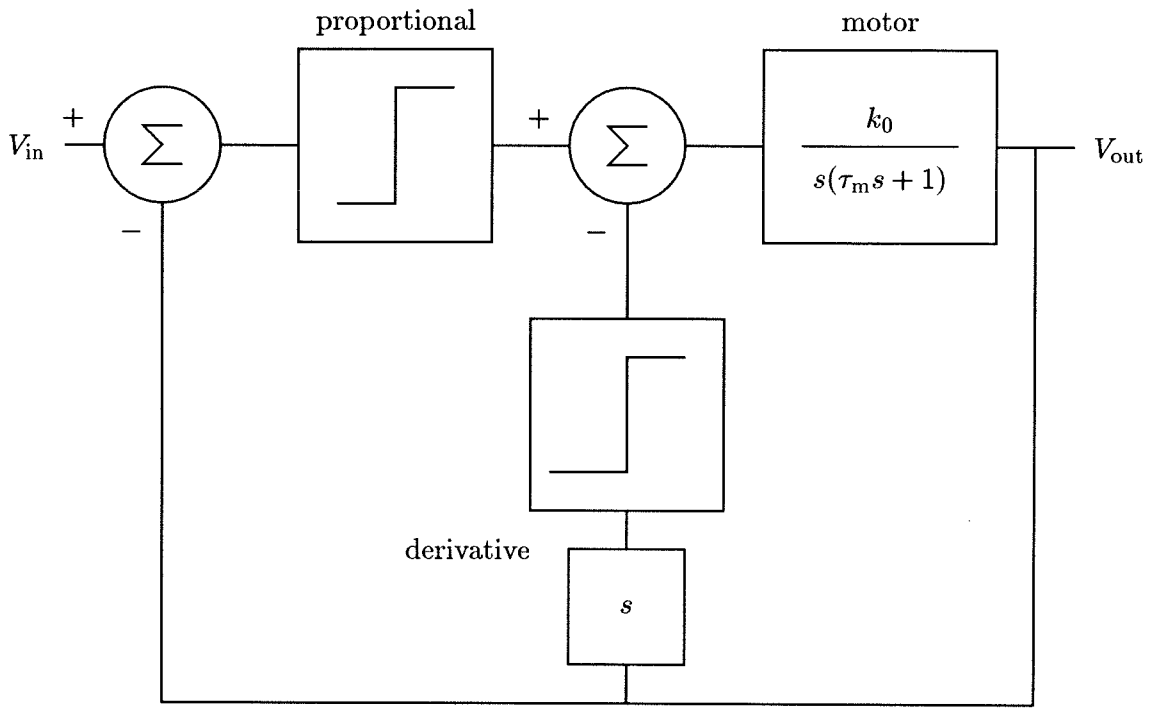

Figure 5.4: The large-signal proportional–derivative control system. In the limit of a large input voltage range, the system in Figure 5.2 can be approximated by this configuration. The proportional derivative controllers can be approximated as relays with a constant magnitude (as described by Equations 5.6 and 5.7).

For our analysis, we shall consider the P–D servo implementation presented previously. Figure 5.4 shows the block diagram of this controller when the threshold approximation is used in the analysis of the differential pairs (in the large-signal regime). In this system, the proportional and derivative elements produce outputs that are thresholded functions of the position and the velocity of the system, respectively.

The equation for the average output voltage from the proportional section of the controller in this system is described by the following equation:

$$V_{\text{av}} = V_{\text{p}} \frac{i_{\text{P}}}{i_{\text{R}}} \text{sgn}(V_{\text{in}} - V_{\text{out}}) \tag{5.6}$$

The equation for the average output voltage from the derivative section of the controller

(with $\tau_D s \ll 1$) in this system is described by the following equation:

$$V_{av} = V_p \frac{i_D}{i_R} \text{sgn}(sV_{out}) \tag{5.7}$$

Since the effects of these two sections of the controller are summed by the control circuitry (on the current aggregation wires), the average output voltage of the controller is the sum of the average voltage in these two equations, and is described by the following equation:

$$V_{av} = V_p \frac{i_P}{i_R} \text{sgn}(V_{in} - V_{out}) + V_p \frac{i_D}{i_R} \text{sgn}(sV_{out}) \tag{5.8}$$

In Chapter 3, we discussed the effect of a constant pulse train used as a motor-control signal. We showed that the average effect of a pulse train on the angular velocity of the motor is equivalent to the effect of a DC voltage V. In our analysis, we shall assume that the period of the pulse train is much smaller than the mechanical time constant of the system. Thus, we can approximate V_{av} with a DC voltage V.

5.3.1 Approximations to System Operation

Since the input to the motor is a piecewise constant function, we can analyze the operation of the system by analyzing separately the response during each constant section of the input (similar to the analysis of pulse-driven motors in Chapter 3). When the motor is presented with a constant control voltage, the angular velocity decays exponentially toward a constant value (as described in Chapter 3). The equation for this decay is

$$\omega(t) = \frac{V}{k_B} + \left(\omega_0 - \frac{V}{k_B}\right) e^{-t/\tau_m}$$

where V is the constant control voltage and ω_0 is the initial angular velocity. If we assume that the time during which the input to the system is at any particular constant voltage is short with respect to τ_m, we can approximate the effect of the voltage on the system as generating a constant torque. Making this assumption is equivalent to approximating the exponential by the first two terms of its Taylor series. Given this approximation, the

angular velocity of the system can be described by the following equation:

$$\omega(t) = \omega_0 + \left(\frac{V}{k_B} - \omega_0\right)\frac{t}{\tau_m}$$

The error between this approximate value and the actual value described in the previous equation is an increasing function of t. The value of the relative error when $t = \tau_m$ is approximately 37 percent.

We shall also assume that ω_0 is also small (with respect to V/k_B). Given this assumption, we can simplify the equation for ω to the following equation:

$$\omega(t) = \omega_0 + \frac{V}{k_B\tau_m}t = \omega_0 + at \tag{5.9}$$

where a is a constant angular acceleration generated by input voltage V. This equation describes a system in which the angular velocity is the integral of a constant torque. In other words, by making these simplifying assumptions, we have approximated the pole of the mechanical system (at $1/\tau_m$) as being equal to 0.

Since the angular position θ is the integral of the angular velocity, the following equation describes the angular position as a function of time:

$$\theta(t) = \theta_0 + \omega_0 t + \frac{a}{2}t^2 \tag{5.10}$$

where θ_0 is the position of the system at time $t = 0$.

5.3.2 Angular Acceleration Generated by the P–D Controller

From Equations 5.6 and 5.9, we can derive the following equation for the magnitude of the angular acceleration a_θ generated by the proportional controller:

$$a_\theta = \frac{V_p i_P}{i_R k_B \tau_m}$$

The sign of this term is dependent on the the position of the system. Since the controller operates with negative feedback, the effect of a_θ is positive when $V_{in} > V_{out}$ and is negative when $V_{in} < V_{out}$.

From Equations 5.7 and 5.9, we can derive the following equation for the magnitude of the angular acceleration a_ω due to the derivative controller:

$$a_\omega = \frac{V_\mathrm{p} i_\mathrm{D}}{i_\mathrm{R} k_\mathrm{B} \tau_\mathrm{m}} + \frac{T_\mathrm{f}}{J}$$

The term T_f/J (the angular acceleration due to kinetic friction) is added to this equation because the effect of friction is identical to the effect of the derivative feedback in the relay controller. In both cases, the sign of the effect is dependent on the the sign of the angular velocity of the system, and the magnitude of the effect is constant. Since both effects act to slow the system (negative feedback), the effect of a_ω is positive when $\omega < 0$ and is negative when $\omega > 0$.

5.3.3 Piecewise Analysis of the P–D Controller

Figure 5.5 shows the theoretical response of a the P–D control system to a large step in the input position voltage (using the theory presented in the following sections). The initial angular position of the motor is defined as $-\theta_0$ and the set-point is defined as 0, giving a step magnitude of θ_0. The initial angular velocity is 0. We separate the analysis of the system in to discrete time intervals, defined as the intervals in which the voltage produced by the controller (and, given our previous assumptions, the angular acceleration) is constant. Since the two terms in the equation for angular acceleration are dependent on the sign of the angular velocity and the sign of the angular position, the voltage generated by the controller (and, thus, the angular acceleration) changes at each point where there is a change in the sign of either the angular velocity or the angular position.

We define the following parameters for the mth discrete time interval: Δt_m is the length of time from the beginning to the end of the time interval, ω_m is the angular velocity at the end of the time interval, and θ_m is the angular position at the end of the time interval. From Equation 5.8, we see that the angular acceleration during each time interval is the sum $\pm a_\theta \pm a_\omega$, with the sign of the two terms dependent on the sign of the angular position and the angular velocity. From these definitions, we derive the values for the first four time intervals $m = 1$ to $m = 4$ (the cycle repeats every four intervals) as shown in Table 5.1.

We solve for the values of Δt_m, ω_m (m odd), and θ_m (m even) for each interval. We set

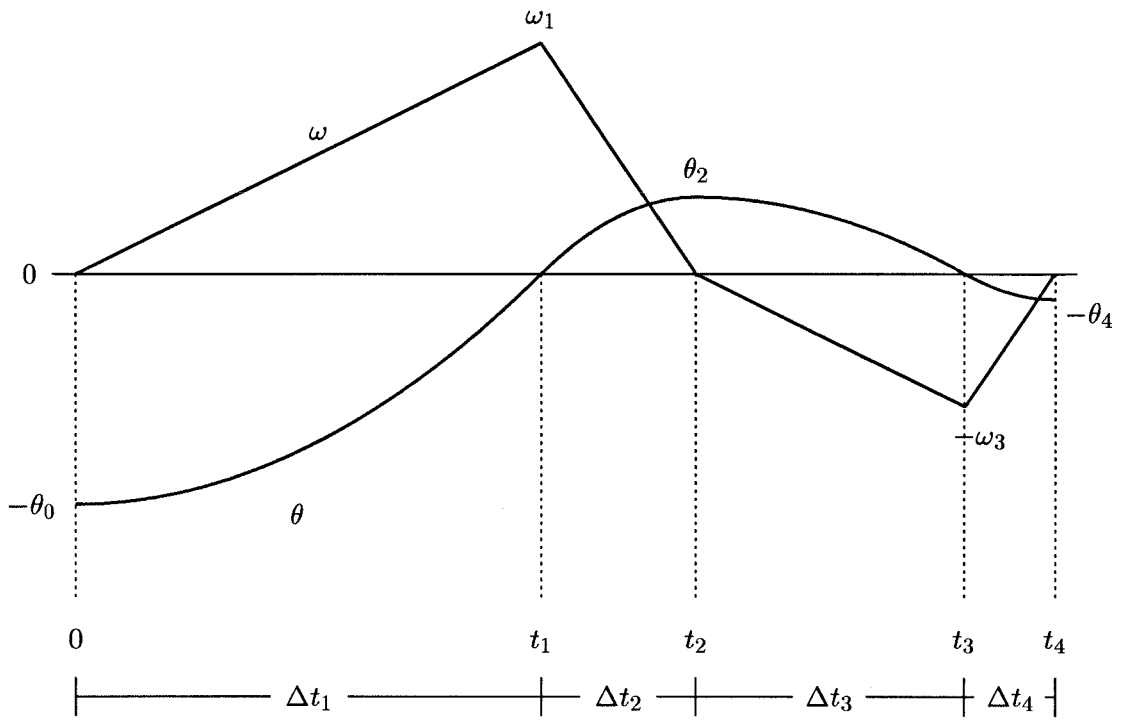

Figure 5.5: Definition of parameters in large-signal operation. The response of the system to a large step in angular position is shown for $\theta_0 = 1$, $a_\theta = 1$, and $a_\omega = 0.5$. In the large-signal analysis, time is divided into discrete intervals whose endpoints are defined as the points when either the angular velocity or the angular position is equal to 0. The length of the mth interval is Δt_m, and the time at the end of the interval is t_m. The extremes (peaks) of the angular velocity and angular position curves are ω_m and θ_m, respectively. The angular velocity peaks occur when m is odd, and the angular position peaks occur when m is even.

m	$\omega(t)$	ω_m	$\theta(t)$	θ_m
1	$(a_\theta - a_\omega)t$	ω_1	$(a_\theta - a_\omega)\frac{t^2}{2} - \theta_0$	0
2	$(-a_\theta - a_\omega)t + \omega_1$	0	$(-a_\theta - a_\omega)\frac{t^2}{2} + \omega_1 t$	θ_2
3	$(-a_\theta + a_\omega)t$	$-\omega_3$	$(-a_\theta + a_\omega)\frac{t^2}{2} + \theta_2$	0
4	$(a_\theta + a_\omega)t - \omega_3$	0	$(a_\theta + a_\omega)\frac{t^2}{2} - \omega_3 t$	θ_4

Table 5.1: ω and θ in the large-signal analysis.

$t = \Delta t_m$ in the equations for ω and θ for the interval. For each interval, this formulation gives two equations in two unknowns. The unknowns are Δt_m and either ω_m or θ_m, since either ω_m or θ_m is equal to 0 at the end of each interval. We define the following two parameters to simplify the solutions of the equations:

$$\varphi \equiv \sqrt{\frac{a_\theta - a_\omega}{a_\theta + a_\omega}}$$

$$t_1 \equiv \sqrt{\frac{2\theta_0}{a_\theta - a_\omega}}$$

From these equations, we can see that a_θ must be greater than a_ω for φ and t_1 to be real, and for the system to operate as predicted. This restriction corresponds with the assumption that the angular acceleration during the first interval ($m = 1$) in the oscillation is positive.

Using the preceding analysis, we derive the following solutions for Δt_m, ω_m, and θ_m:

$$\begin{aligned}
&m \text{ odd} && \Delta t_m = t_1 \varphi^{\frac{m-1}{2}} \\
& && \omega_m = t_1 (a_\theta - a_\omega) \varphi^{\frac{m-1}{2}} \\
&m \text{ even} && \Delta t_m = t_1 \varphi^{\frac{m+2}{2}} \\
& && \theta_m = \theta_0 \varphi^m
\end{aligned}$$

From these equations, we can see that t_1 is a measure of the frequency of the system, and is thus conceptually similar to the parameter w_n in linear systems. Also, φ is a measure of the damping of the system, and is thus conceptually similar to ζ. The natural frequency is defined as the frequency when the system is undamped. When this system is undamped ($a_\omega = 0$), the frequency of the oscillations is $\frac{1}{4t_1}$, which, unlike in the linear system, depends on the initial position θ_0. The damping of the system can be seen in the preceding equation for θ_m. For $\varphi = 1$, the magnitudes of all the peaks in the oscillation are the same, and the system is undamped. As φ is decreased toward 0, the damping in the system increases.

5.3.4 The Envelope of the Oscillation

We shall derive an expression for the magnitude of the peak height in the position curve in terms of the elapsed time. The continous-time version of this function describes the envelope

that bounds the oscillations in the position of the motor. The time interval between two adjacent peaks in the position curve is the sum of the lengths of the two subintervals that make up this interval. This sum is described by the following equation:

$$\begin{aligned}\Delta t_{m-1} + \Delta t_m &= t_1 \varphi^{\frac{m-2}{2}} + t_1 \varphi^{\frac{m+2}{2}} \\ &= t_1(1+\varphi^2)\varphi^{\frac{m-2}{2}}\end{aligned}$$

To determine t_m (the elapsed time from t_0 to the end of interval m) for m even, we compute the following summation:

$$\begin{aligned}t_m &= \sum_{i=1}^{m/2} t_1(1+\varphi^2)\varphi^{2i-1} \\ &= t_1(1+\varphi^2)\sum_{i=0}^{m/2-1} \varphi^{2i} \\ &= t_1(1+\varphi^2)\frac{\varphi^{m/2}-1}{\varphi-1}\end{aligned}$$

We can rearrange the preceding equation to derive the following equation for φ^m in terms of t_m:

$$\varphi^m = \left(\frac{\varphi-1}{t_1(1+\varphi^2)}t_m + 1\right)^2$$

We can combine the preceding equation with the equation for the magnitude of the mth peak (θ_m) to derive the following equation for θ_m in terms of t_m for m even:

$$\theta_m = \theta_0\left(\frac{\varphi-1}{t_1(1+\varphi^2)}t_m + 1\right)^2$$

Thus, the envelope θ_{env} that encloses the oscillations is described by the following equation:

$$\theta_{\text{env}}(t) = \theta_0\left(\frac{\varphi-1}{t_1(1+\varphi^2)}t + 1\right)^2 \tag{5.11}$$

5.3.5 Overshoot

The overshoot M_p in the system is the magnitude of the the first peak in the position curve. We derive the following equation for this magnitude:

$$M_p = \theta_0 \varphi^2 = \theta_0 \left(\frac{a_\theta - a_\omega}{a_\theta + a_\omega} \right) \tag{5.12}$$

Figure 5.6 shows the relationship between the overshoot and the values of a_θ and a_ω. We can see that, as a_ω approaches a_θ, the overshoot becomes smaller. At the point where $a_\omega = a_\theta$, the oscillations become critically damped (no overshoot). In the first interval, however, the angular acceleration is equal to $a_\theta - a_\omega$; it is equal to 0 when $a_\omega = a_\theta$. This restriction means that a_ω must always be less than a_θ for the system to operate. Thus, although we can set the parameters such that the overshoot is smaller than any positive value, we cannot, in the large-signal limit, cause the system to be critically damped.

5.3.6 Rise Time

We use a standard definition for rise time [FPEN86], which defines the rise time as the time required for the angular position to change from 10 to 90 percent of the difference between the initial position to the asymptotic position (set-point). In this situation, these values are $-\theta_0$ and 0, respectively. Thus, we must measure the time required for the position to change from $\theta = -0.9\theta_0$ to $\theta = -0.1\theta_0$. The equation for this position in this interval is

$$(a_\theta - a_\omega)\frac{t^2}{2} - \theta_0$$

By using this equation, we compute the time at the 10- and 90-percent points, and subtract these two values to derive the following equation for rise time:

$$\begin{aligned} t_r &= (\sqrt{0.9} - \sqrt{0.1})t_1 \\ &\approx 0.63\sqrt{\frac{2\theta_0}{a_\theta - a_\omega}} \end{aligned}$$

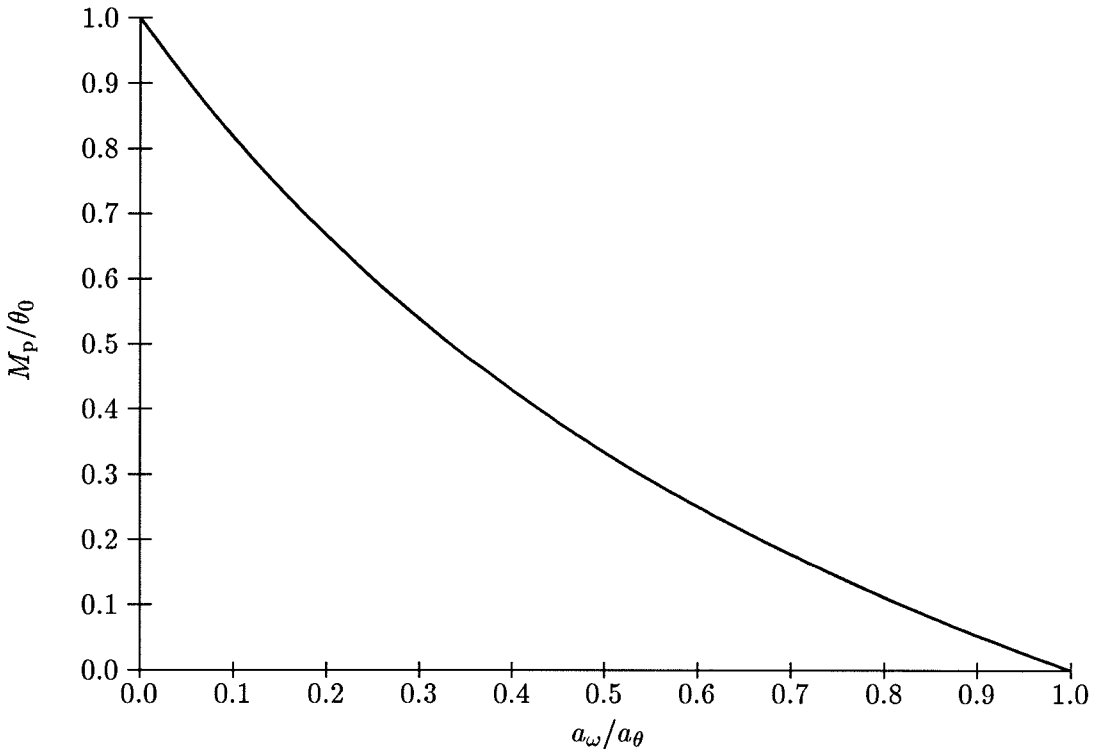

Figure 5.6: The magnitude of the overshoot in the large-signal approximation. The magnitude of the overshoot M_p is plotted versus the ratio of the acceleration terms (a_ω and a_θ). The overshoot is scaled by the magnitude of the initial angular position. The ratio of the acceleration terms is restricted to be between 0 and 1.

5.3.7 Settling Time

The settling time (t_s) is the time required for the transient response of the system to decay to the point at which the maximum error from the fixed point at any point after that is less than some given value [FPEN86]. We shall use 1 percent as a reasonable measure of this error.

We can solve for the settling time by using Equation 5.11, which describes the envelope of the oscillation. When this envelope decays to the $0.01\theta_0$, we are guaranteed that the error after this point will always be less than 1 percent. This point is defined by the following equation:

$$\theta_0 \left(\frac{\varphi - 1}{t_1(1+\varphi^2)} t_s + 1 \right)^2 = 0.01\theta_0$$

We rearrange this equation to derive the following equation for the settling time:

$$t_s = \frac{0.9t_1(1+\varphi^2)}{1-\varphi}$$

Figure 5.7 shows the settling time of this system versus the ratio of the accelerations (a_ω/a_θ). The settling time is scaled by the factor $\sqrt{2\theta_0/a_\theta}$ (t_1 for $a_\omega = 0$), which is a measure of the period of the oscillations. The figure shows that, for constant values of θ_0 and a_θ, the settling time is minimized when $a_\omega \approx 0.56a_\theta$.

5.3.8 Experimental Data

We built an experimental apparatus to test the operation of our proportional–derivative controller. We directly connected a potentiometer to the shaft of a motor. We placed a differential voltage across the fixed terminals of the potentiometer so that the voltage at the sliding terminal encoded the angular position of the shaft. We tested the system by generating a voltage step at the input. The voltage step was made large so that the circuit would operate in the large-signal regime.

Figure 5.8 shows theoretical and experimental angular-velocity and angular-position curves that represent the response of the system to this voltage step. We fit the experimental data by measuring t_1 and θ_2, and by calculating a_ω and a_θ from these values. We generated the theoretical curves using these values of a_ω and a_θ. The angular position curve fits

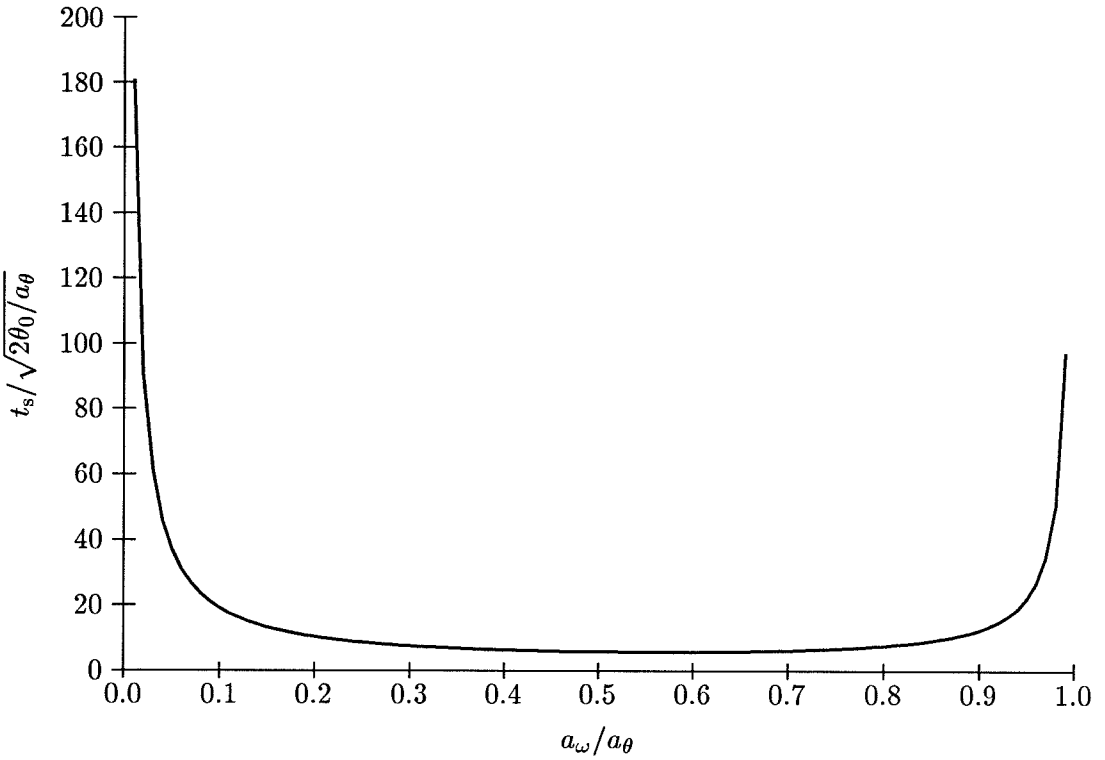

Figure 5.7: Settling time for the large-signal approximation. The horizontal axis is given in terms of the ratio of the accelerations a_ω and a_θ. The vertical axis (settling time) is scaled by the term $\sqrt{2\theta_0/a_\theta}$, which is a measure of the period of the oscillations.

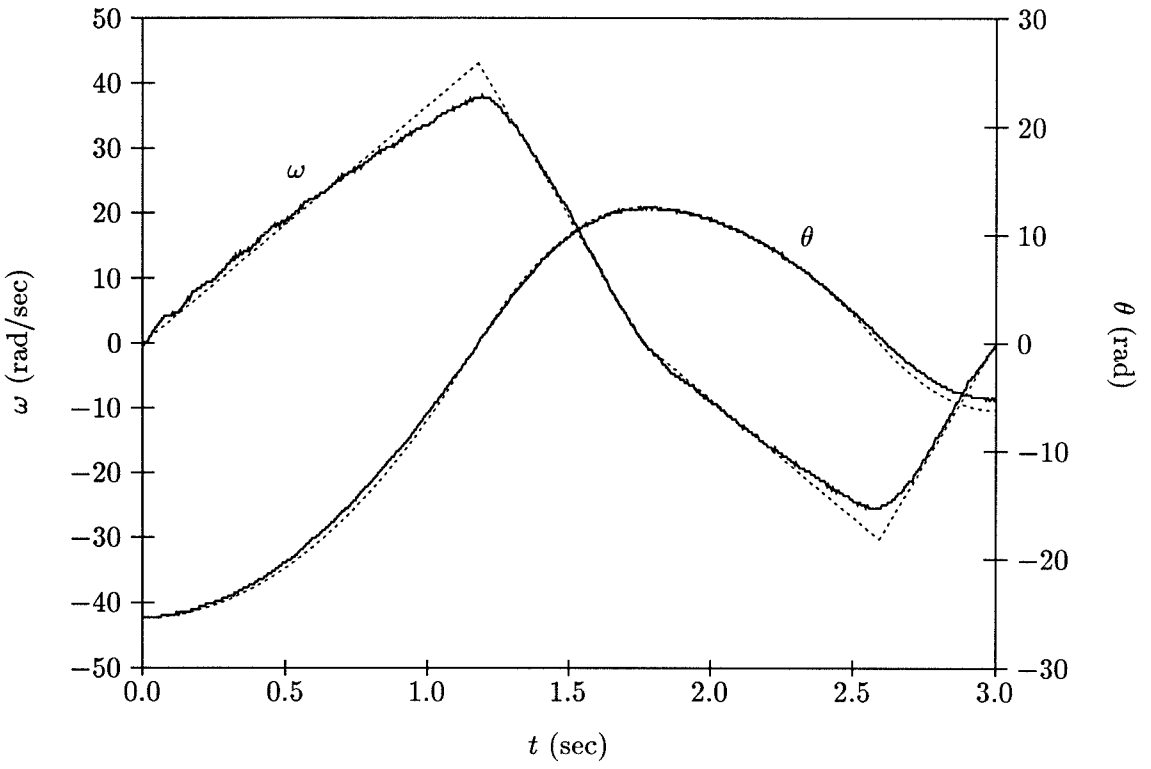

Figure 5.8: Data and theory for the proportional–derivative feedback system. The curves represent the angular velocity and angular position of the system when presented with a large step in angular position. The solid lines represent the experimental data, and the dotted lines represent the theoretical approximation. The values of ω and θ are scaled by their maximum values, which occur at ω_1 and θ_0, respectively.

the theory quite well. The slopes of the sections of the angular velocity curve fit the piecewise-linear theory fairly well, although the effects of the the linear approximation to the exponential and the unsaturated region of the hyperbolic tangent can be seen.

We also measured the overshoot as a function of the initial position (θ_0). From Equation 5.12, we see that the overshoot should be proportional to the initial position. Figure 5.9 shows the expermental data for this measurement. The data fit a line (that passes through the origin) fairly well. The ratio of the overshoot to the initial position decreases slightly over the range of input. This decrease can be explained by the increased inaccuracy in the linear approximation to the exponential for larger values of t_1, which increases as θ_0 increases.

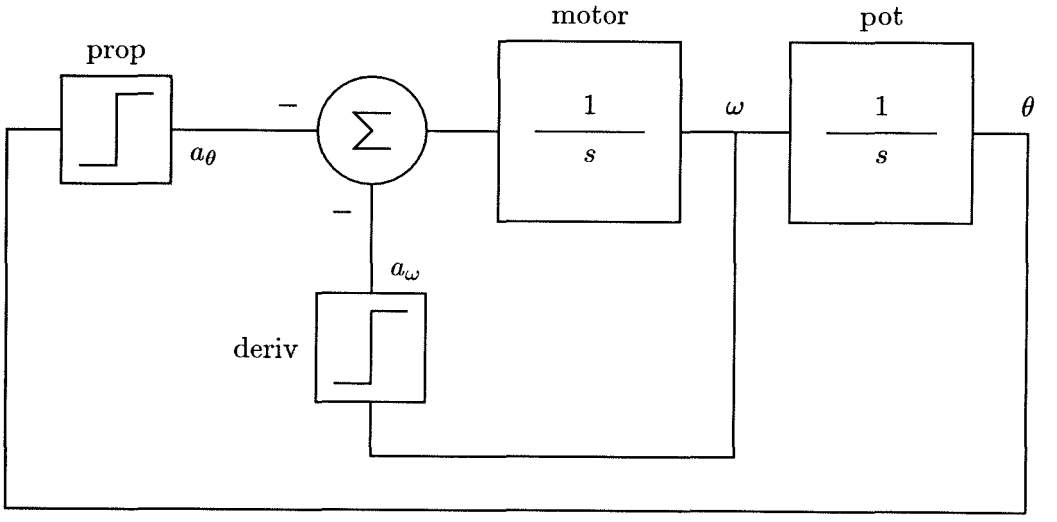

Figure 5.9: Experimental data for overshoot versus initial position. The data points represent the overshoot for a number of different initial positions. The best-fit line passing through the origin is also shown.

Chapter 6

Stimulus Tracking

In this chapter, we present circuitry that converts a spatially encoded sensory image into a dual-rail pulse-encoded output. These output pulses can, in turn, be used to control bidirectionally the motion of a one-dimensional motor system. The basic circuitry is presented as a framework that facilitates the tracking of a stimulus such that an electromechanical system can move to cause the stimulus to be in the center of the sensory field. We describe an implementation of this framework that uses visual intensity as its input signal. We incorporated this circuitry into mechanical systems that track a bright visual stimulus (a bright spot of light) on a darker background.

This type of circuitry has analogs in the sensory systems of many animals. The oculomotor system in mammals is one example. The smooth pursuit system in these animals reacts to moving stimuli by following or tracking the stimulus. Also, the superior colliculus region of the mammalian brain contains circuitry (similar in function to the circuits presented here) that controls saccadic eye motion (fast point-to-point motions) [Spa83]. The circuit concepts presented in this chapter possibly could be used to implement engineering analogs of these biological systems.

6.1 The Tracking Framework

The basic task that we consider in this chapter is the tracking of a sensory stimulus by a motor system. A block diagram of the circuitry used to accomplish this task is shown in Figure 6.1. The circuitry consists of four functional layers: the input layer, the normalization

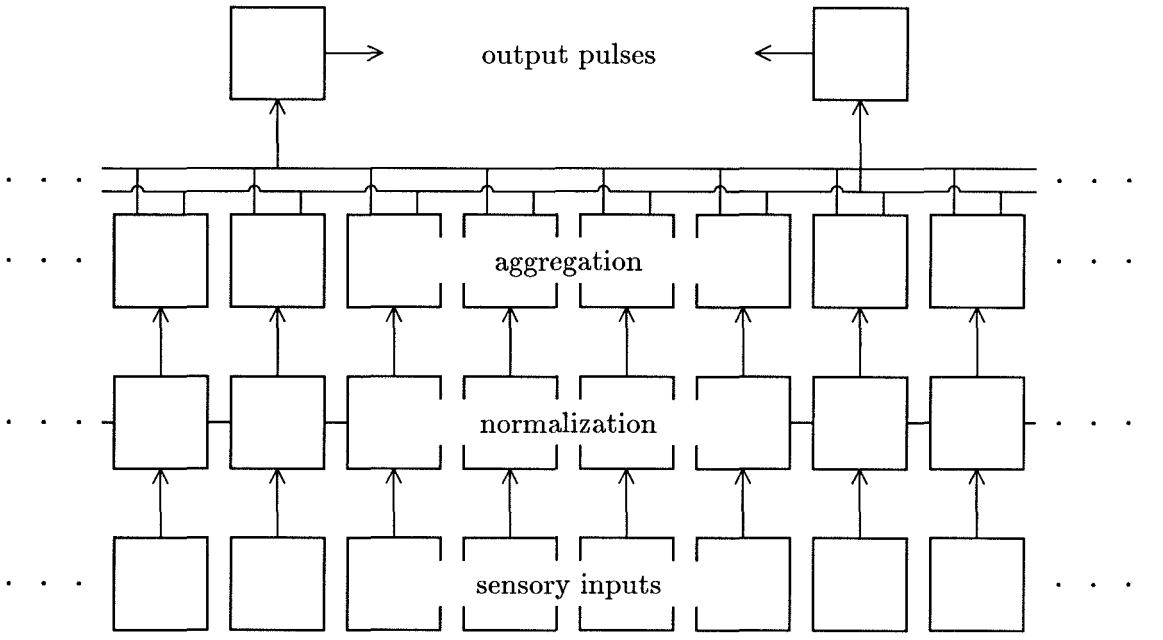

Figure 6.1: A block diagram of the tracking framework. The circuitry is divided into four layers: sensory input, normalization, aggregation, and motor output. Each layer processes its input and passes its output to the subsequent layer.

layer, the aggregation layer, and the output layer. An image is presented to the input layer, which generates a one-dimensional array of currents. The normalization circuitry scales these input currents so that variations in global signal magnitude do not affect the subsequent processing. The circuitry in the aggregation layer computes a weighted combination of the input signals, resulting in a pair of complementary output currents, thus reducing the one-dimensional sensory representation into a scalar motor representation. Finally, the output circuitry converts these two currents into pulse trains that can be used to control the bidirectional motion of a one-dimensional motor system.

The aggregation and output layers compute the transformation from the one-dimensional sensory representation to the scalar motor representation. The circuitry accomplishes this task by encoding the position of the individual elements in the array, and by using this position information to redirect each input current to a pair of output wires. We implement this circuitry with a modified version of the pulse-output aggregation network described in

Section 4.3.2. We augment this circuit by adding position encoding in the form of a resistive voltage divider, as described in Section 2.2.1. The combined aggregation and output circuitry is shown in Figure 6.2. This circuit computes an error signal in the form of a dual-rail pulse encoding.

6.2 Sensory Input

The input layer generates a one-dimensional, spatially encoded sensory image. There is a wide range of possible sensory-input types and circuit functionality at this level. The circuitry for this layer consists of input devices combined with local and global processing. The only restriction placed on this layer is that the resulting output consist of a spatially encoded array of currents. The source of the input signals may be derived from sensors that respond to many different types of input, including visual, tactile, auditory, sonar, and magnetic inputs. This layer can include a variety of types of processing circuitry combined with the sensory imaging.

Given inherently spatially encoded inputs (such as visual images), the function of this circuitry can range from merely passing the input signals to the next layer to complex spatiotemporal processing. Circuitry that computes spatial derivatives (such as that found in the silicon retina [MM88]) can be used to preprocess the image locally, giving the chip the behavior of tracking an edge or discontinuity in the image. Optical-flow circuitry [Tan86] in the input layer could be used in a system that tracks moving objects. Input-layer circuitry that computes the focus of expansion of a visual image could be used to guide a vehicle through a tunnel, or to position a mechanical robot head such that it points in the direction of motion of the robot's body.

Images that are not initially encoded spatially (such as auditory signals) require processing to generate a positionally encoded output. One example of circuitry that performs this conversion is circuitry to compute the binaural time disparity from a pair of auditory inputs [LM89]. This circuitry generates an array of output signals. These signals encode the spatial positions of auditory stimuli. By using these signals as input to the tracking circuitry, we can design a system that tracks a moving sound source.

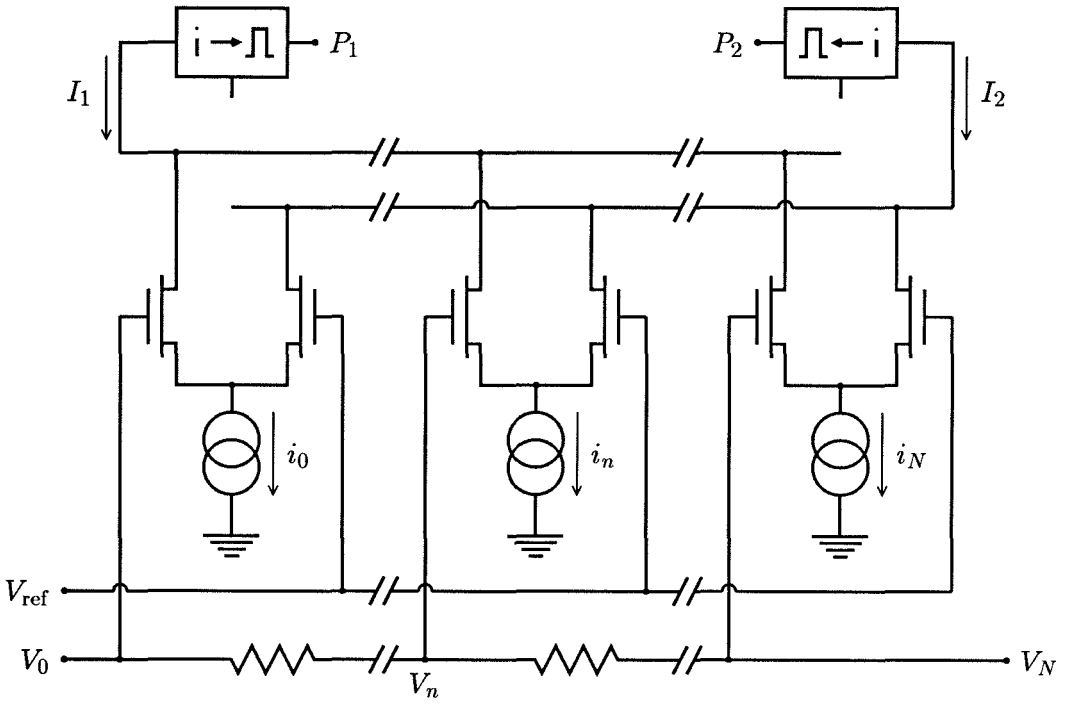

Figure 6.2: Pulse-output aggregation network with position encoding. This circuit combines aggregation, position encoding, and dual-rail pulse generation. The voltage V_n encodes the position of element n with respect to V_{ref}, which effectively sets the origin of the array. The current i_n is the input signal for the nth element. The input currents are divided (according to their positions) and are aggregated onto two wires by the differential-pair circuits. These aggregated currents are the inputs to the neuron circuits. The currents near the ends of the array are directed primarily to one neuron circuit or to the other, whereas those in the middle are divided evenly between the two wires. The differential duty cycle of the pulse trains encodes an error signal between the first moment of the input currents and the origin of the array.

6.3 Signal Normalization

The input to the aggregation layer and the signals produced by the input layer are both spatially encoded arrays of currents. Thus, we could conceivably implement the tracking framework by connecting the input layer directly to the aggregation layer. This solution does not work correctly for many types of input signals, however. Sensory input (such as light intensity) can often vary over many orders of magnitude. Unless the input-layer circuitry compensates for this variation, the currents produced by the input layer can also vary greatly. Since the aggregation network merely routes its input currents to the neuron circuits, the input variation would have a direct effect on the magnitude of the drive to the motor system. Without intervening circuitry to modify the current levels, global changes in the input signals would cause unacceptable changes in velocity of the mechanical system.

To counteract this global input variation, we insert normalization circuitry between the input layer and the aggregation layer. This normalization layer must guarantee that, independent of the variation of its input signals, its total output current will remain constant. Thus, the normalization (gain-control) circuit must scale the currents generated by the input layer such that the sum of the currents input to the aggregation and output layers is fixed and is independent of the input signals. By defining normalization to restrict only the sum of the output currents, but not the relative scaling of different input currents, we allow a variety of possible functionalities at this layer.

Figure 6.3 shows a representation of the generic normalization circuit. Since the inputs and outputs to this circuitry are each arrays of currents, the circuitry can be implemented by an array of circuit elements with each element having one input current and one output current. The individual circuit elements are all connected by a single wire that performs global communication for the circuit by distributing a voltage to and summing a current from each element. The circuitry in each element is restricted such that the output current (i_{out_n}) is equal to the current that that element passes to the global wire. Thus, by Kirchoff's current law on the global wire, the sum of the output currents must be equal to the bias current (i_{b}) that is sunk from the global wire:

$$i_{\text{b}} = \sum_n i_{\text{out}_n} \tag{6.1}$$

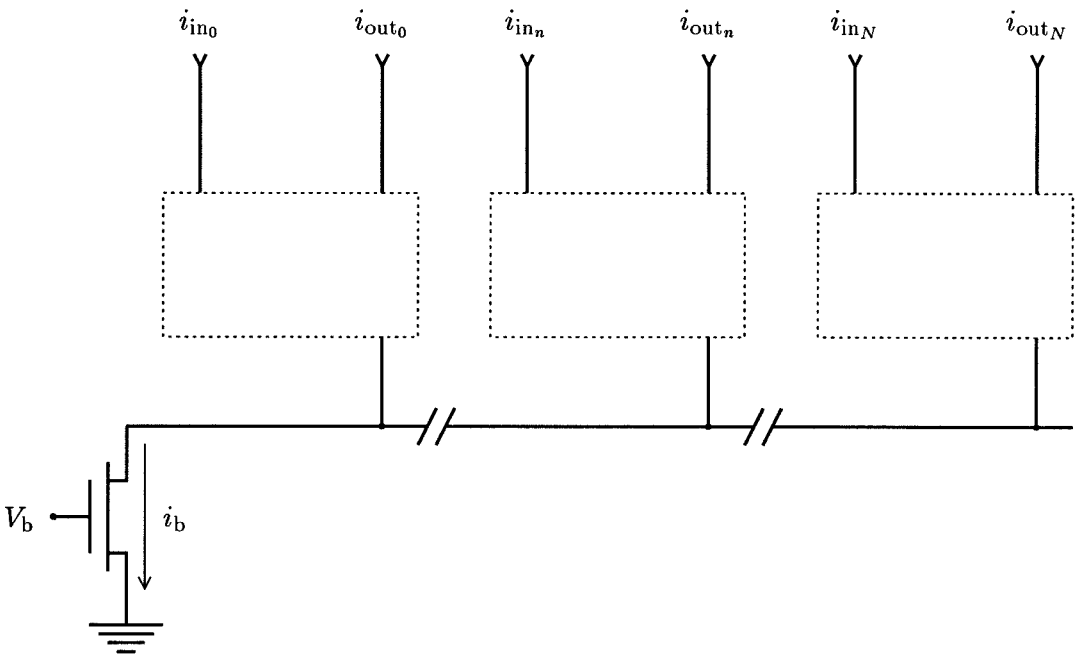

Figure 6.3: The generic normalization circuit. A global wire sums a current from each element and distributes a voltage V to all of the elements. Each element consists of circuitry (in the dotted boxes) that has the restriction that the output current i_{out_n} must be equal to the current that the element sources onto the global wire. This restriction guarantees, by Kirchoff's current law, that the sum of the output currents is equal to the constant bias current i_{b}.

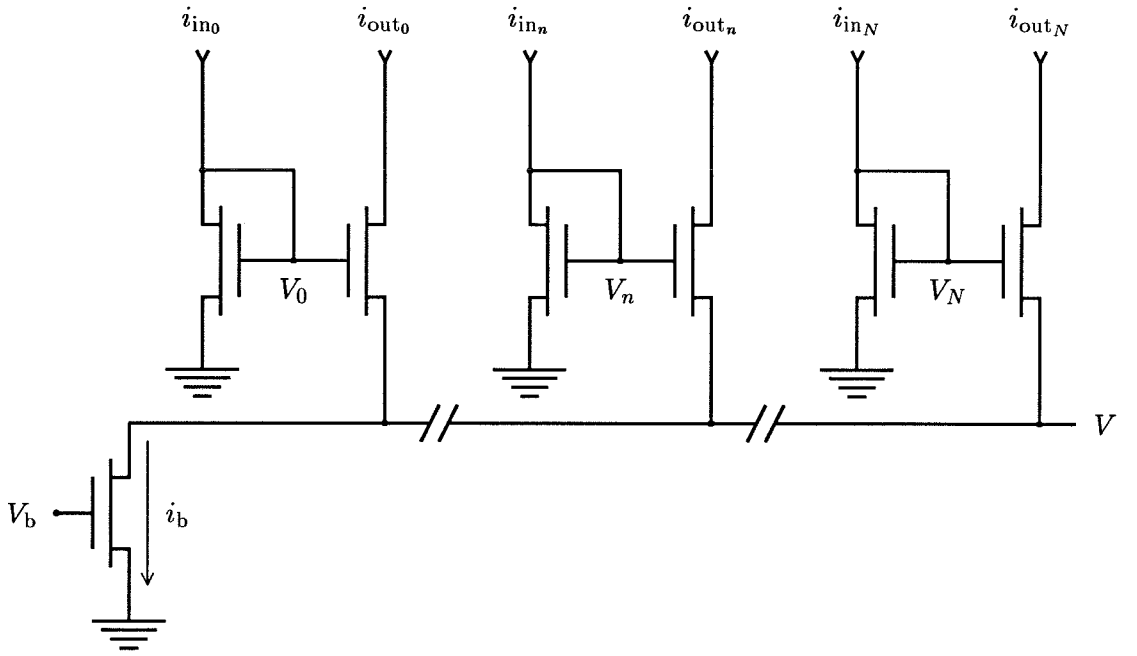

Figure 6.4: A linear normalization circuit. Each element takes an input current (i_{in_n}) and produces an output current (i_{out_n}) that is proportional to its corresponding input, with the restriction that the sum of all output currents is equal to i_{b}.

The sum of the output currents is set directly by the bias current. Thus, the sum of the output currents will be constant independent of the input currents. In the following sections, we shall present two examples of normalization circuits that use this basic design and perform different functions.

6.3.1 A Linear Normalization Circuit

The circuit in Figure 6.4 [Gil84] is a one example of a normalizing circuit. Each circuit element consists of two transistors configured as a current mirror modified such that the source of the output transistor is connected to the global wire. The output current for each element flows into the drain of the output transistor, guaranteeing that the output current is equal to the current flowing from the element onto the global wire.

If we assume that the two transistors in the nth element are saturated, the equations

for the nth input current (i_{in_n}) and the nth output current (i_{out_n}) are

$$\begin{aligned} i_{\text{in}_n} &= i_0 e^{\kappa V_n} \\ i_{\text{out}_n} &= i_0 e^{\kappa V_n - V} \end{aligned}$$

We can combine and rearrange these two equations to derive the following relationship:

$$i_{\text{out}_n} = e^{-V} i_{\text{in}_n}$$

If we take the sum over all the inputs for both sides of this equation and set the sum of the outputs equal to i_{b} (by Equation 6.1), we can derive the following equation for e^{-V}:

$$e^{-V} = \frac{i_{\text{b}}}{\sum_n i_{\text{in}_n}}$$

We can combine the two previous equations and derive the following equation for i_{out_n}:

$$i_{\text{out}_n} = i_{\text{b}} \cdot \frac{i_{\text{in}_n}}{\sum_n i_{\text{in}_n}} \tag{6.2}$$

The circuit in Figure 6.4 performs *linear normalization*, or normalization in which all input currents are scaled by the same constant. Since all the currents are scaled equally and the sum of the output currents is constant, the scaling constant must be the ratio of the total output current to the total input current. Given a particular image, this circuitry will eliminate the effects of changing global signal levels (such as overall illumination), while retaining at the output the relative relationships (ratios) between the input signals. We present data for a two-element version of this circuit in Figure 6.5. This figure shows that, as the two input currents vary, the output currents respond by varying linearly.

6.3.2 A Winner-Take-All Normalization Circuit

Figure 6.6 shows a variant of the winner-take-all circuit [LRMM89] that obeys the normalization criterion. This circuit has a highly nonlinear functionality, and thus differs functionally from the linear normalization circuit.

Each circuit element consists of two transistors, and, as in the linear normalization

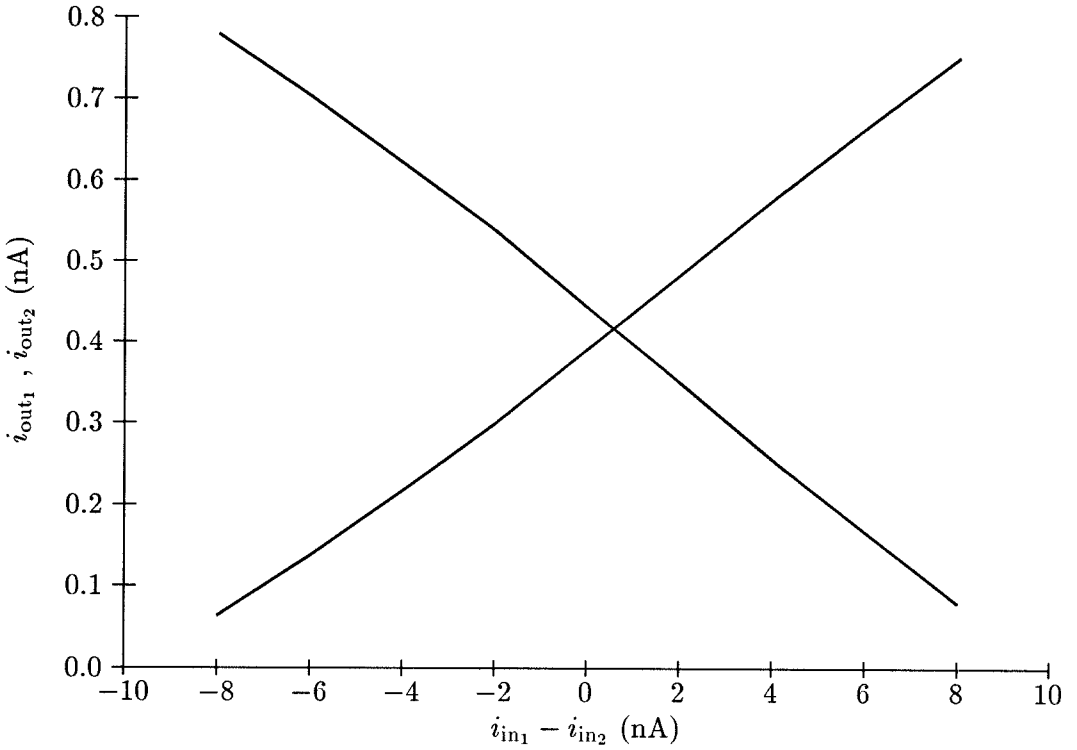

Figure 6.5: Data from a two-element linear normalization circuit. The two input currents (i_{in_1} and i_{in_2}) were swept such that their sum remained constant at 10 nA. The bias current (i_b) was set at 1 nA. The two curves represent the output currents i_{out_1} and i_{out_2} versus the difference ($i_{in_1} - i_{in_2}$) between the input currents. The data demonstrate the linear relationship between the output currents and the input currents.

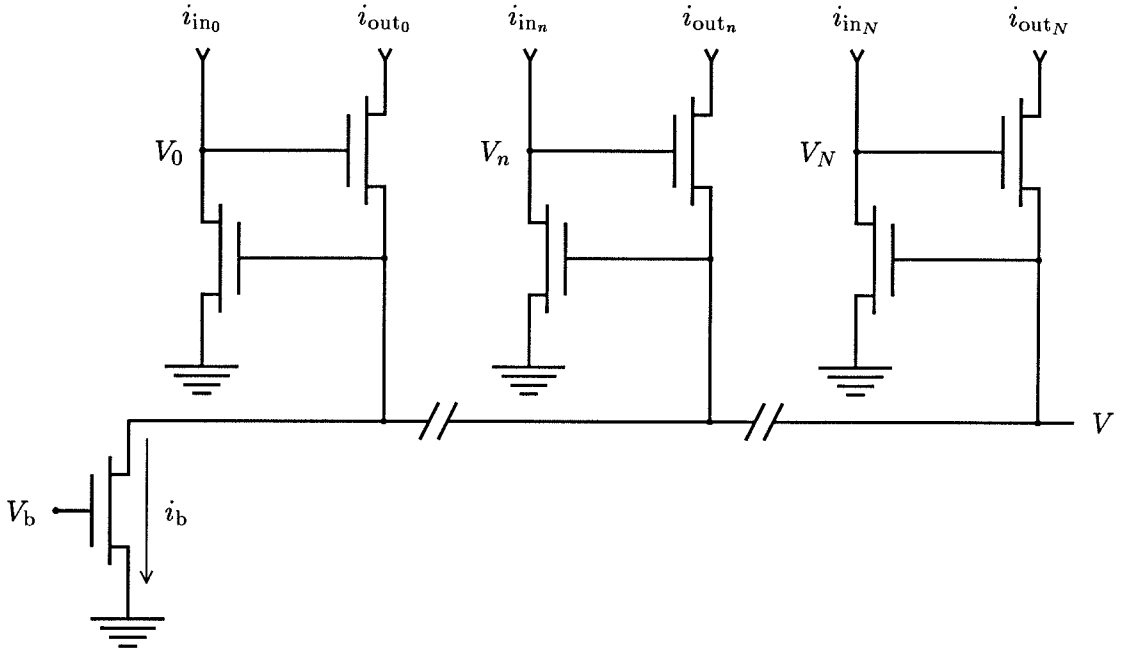

Figure 6.6: A winner-take-all normalization circuit. Since all the input transistors share a common gate voltage V, if all of the input currents are equal, then V_n will be the same for all the nodes and the output current from all the nodes will be equal. If the input current at node n is increased, V will increase to allow i_{in_n} to be sunk. The increase in V will cause the local voltages at the other nodes to decrease to force the input transistors at these nodes out of saturation (V_ds small) so that these transistors properly sink their input currents. This voltage decrease effectively forces the gate voltages of the output transistors in these nodes toward ground, causing the output currents of these nodes to decrease greatly. For the bias current i_b to be sourced, V_n must increase, which makes $i_{\text{out}_n} \approx i_\text{b}$.

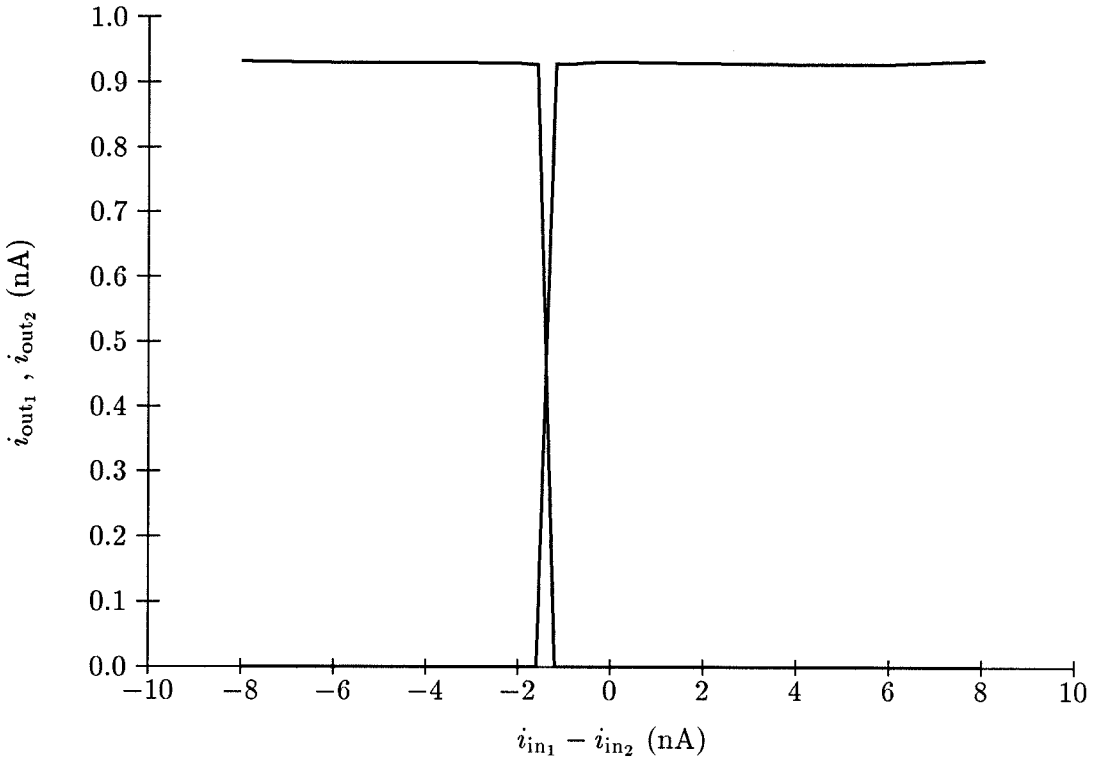

Figure 6.7: Data from a two-element winner-take-all normalization circuit. The two input currents (i_{in_1} and i_{in_2}) were swept such that their sum remained constant at 10 nA. The bias current (i_b) was set at 1 nA. The two curves represent the output currents i_{out_1} and i_{out_2} versus the difference ($i_{in_1} - i_{in_2}$) between the input currents. The data demonstrate the highly nonlinear relationship between the output currents and the input currents.

circuit, passes its output current directly onto the global wire. The transistors are connected such that the output currents have a highly nonlinear response as a function of the input currents. If one element has an input current higher than that of the others, this element will produce an output current approximately equal to the bias current, and all the other output currents will be approximately equal to 0. If more than one element shares equal input currents that are greater than the input currents of the other elements, the output currents at the elements with maximum input current will be equal.

In Figure 6.7, we present data from a two-element version of this circuit. This figure shows the highly nonlinear response at the outputs of the circuit to variation in the input currents. Because of this nonlinearity, multiple output currents rarely occur in a multiple-element circuit except when the winner (the element with the largest input current) is

changing. Thus, the circuit functions as a winner-take-all in which the winner takes all the output current. When used in the tracking framework, the winner-take-all circuit has the distinct advantage over the linear normalization circuit in that it produces only one output, and thus has a response that is more resistant to background sensory levels. The linear normalization circuit averages over multiple pixels, however, and thus allows for increased resolution to make the localization of a stimulus more accurate.

6.4 Framework Implementation and Functionality

We implemented the tracking framework using the circuits presented in the preceding sections. We generated a sensory image using a one-dimensional array of phototransistors (with no additional circuitry). These devices were bipolar phototransistors incorporated directly onto the silicon surface with the processing circuitry of the subsequent layers. A lens is used to focus a visual image onto this array. The output from this layer is an array of photocurrents, each of which is proportional to the intensity of the light at the respective photosensor. The normalization layer was a winner-take-all normalization circuit. Although this circuit reduces the effective resolution of the system, it is much more independent of background illumination than is the linear normalization circuit. We implemented the aggregation and output layers with the circuitry shown in Figure 6.2.

We can analyze the functionality of the complete system under the intended operating conditions. The chip localizes the largest intensity in the visual field, and generates a pair of output signals that encodes this location in a representation that is suitable for driving motors.

Figure 6.8 shows the duty cycles of the output signals versus the position of the stimulus on the array. We present a stimulus to the chip in the form of a bright source of light. As we move a stimulus across the array of phototransistors, the duty cycles of the output pulse trains change. As a result of the operation of the differential pairs in the aggregation network, the resulting duty cycles are approximately sigmoidal. The width of the sigmoid is controlled by the differential voltage across the resistive divider. The maximum duty cycle (the upper asymptote) is the ratio of the bias current in the aggregation network to the reset current of the neuron circuit.

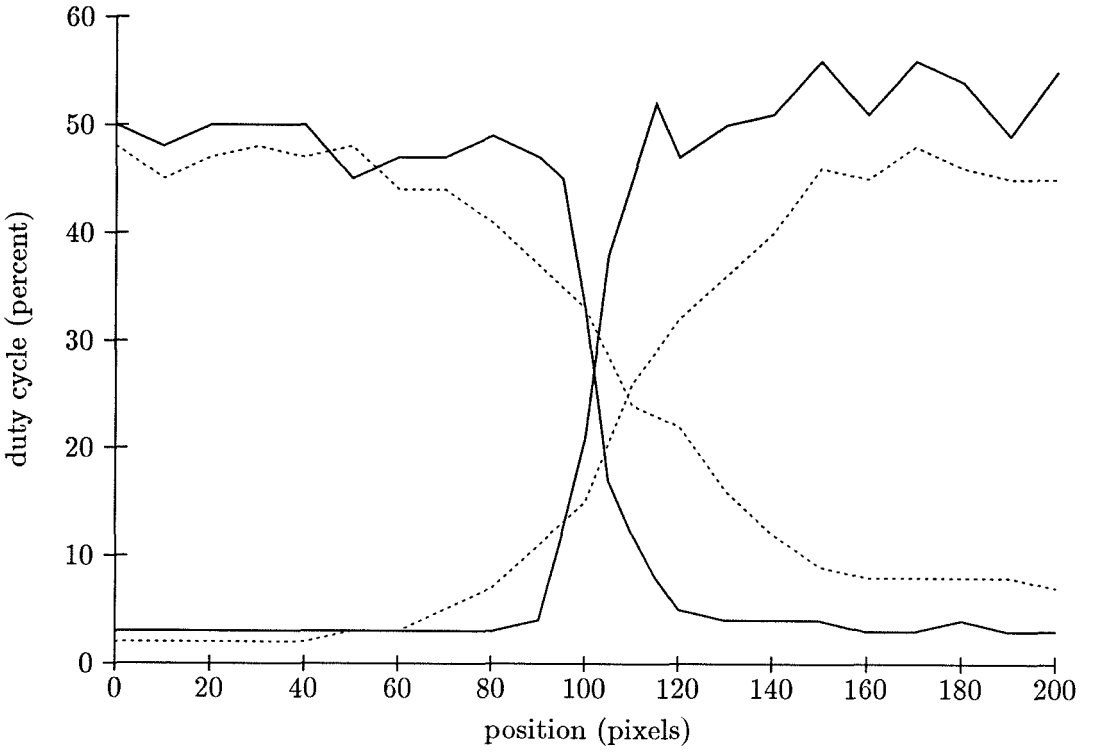

Figure 6.8: Duty cycle versus position for the tracking imager. The curves represent the duty cycles of the two output pulse trains with respect to the position of a bright stimulus on the photosensors. The solid lines represent data taken with the voltage across the resistive divider equal to 0.5 V. The dotted lines represent data taken with the voltage across the resistive divider equal to 2.0 V. The curves are approximately sigmoidal due to the functionality of the differential pairs, with the width being controlled by the voltage across the resistive divider. The maximum duty cycle (the upper asymptote) is the ratio of the bias current in the aggregation network to the reset current of the neuron circuit.

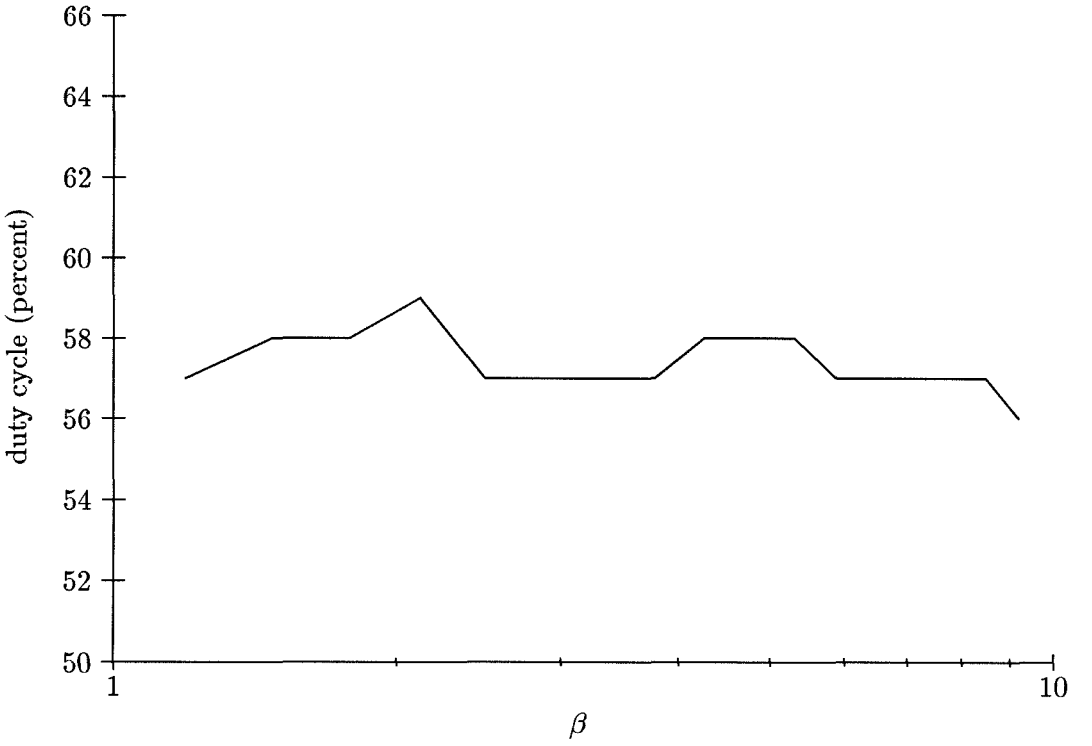

Figure 6.9: Output independence from stimulus intensity. The data represent the maximum duty cycle of the output pulse-train versus the contrast ratio (β). The intensity of the stimulus was varied while the background intensity was held constant. Because of the winner-take-all normalization, the duty cycle is independent of the stimulus and background intensities when the maximum intensity at an individual pixel is generated by the stimulus ($\beta > 1$). In this experiment, the duty cycle varied by less than ± 2 percent.

Figure 6.9 shows the effects of varying the stimulus intensity while holding the background intensity constant. The winner-take-all normalization circuit generates a single current output at the element with the largest input current, independent of the intensity of the stimulus and background. Thus, the pulse outputs from the system are independent of ratio of the background intensity to the stimulus intensity, as long as the background intensity at all pixels is less than the maximum stimulus intensity.

6.5 Feedback System Examples

We have implemented two complete sensorimotor tracking systems by combining the chip described in the preceding section with mechanical systems that move the chip such that, when the chip is presented with a visual stimulus, the origin of the array is moved toward the stimulus.

6.5.1 The Planar Eye Model

One mechanical system that we have implemented is a model of a planar eye and two ocular muscles. Figure 6.10 shows a diagram of this mechanical system. We model the eye with a turntable that rotates at its center. The chip and lens are placed on the turntable, such that rotation of the turntable changes the field of view of the chip. We model each ocular muscle with a motor and cable that apply a torque to the eye. The two motors are placed in an antagonistic configuration such that they produce torque on the eye in opposite directions. Each pulse output generated by the chip controls one of the motors.

If the chip is stimulated with a bright spot of light that is not at the center of the photoreceptor array, the motors will produce a differential torque that rotates the turntable to direct the center of the chip toward the stimulus. Because of the background firing rates of the neuron circuits, the motors will produce nonzero (equal and opposite) torques when the stimulus is centered, keeping a constant torque on the system.

6.5.2 The Tracking Vehicle

Figure 6.11 shows a two-wheeled vehicle that we have designed and built. The chip and a lens are mounted on the vehicle such that the image of the area in front of the vehicle is

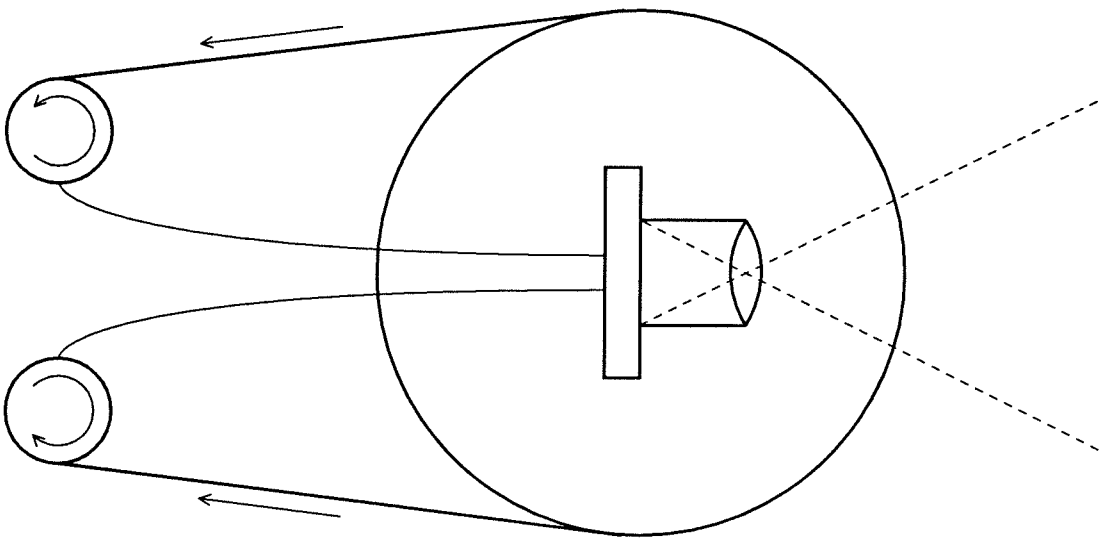

Figure 6.10: A mechanical planar eye model. The tracking chip is placed at the center of a turnatable that is free to rotate on its axis. A lens is positioned to focus the image of the region in front of the chip onto the sensory surface. The chip generates a two pulse trains, each of which is amplified by a half H-bridge circuit. The amplified pulse trains drive the two motors, which are configured in an antagonistic format such that they apply torque to the turntable in opposite directions. A visual stimulus is placed in the field of view of the chip, and the system rotates the turntable such that the stimulus is centered on the chip.

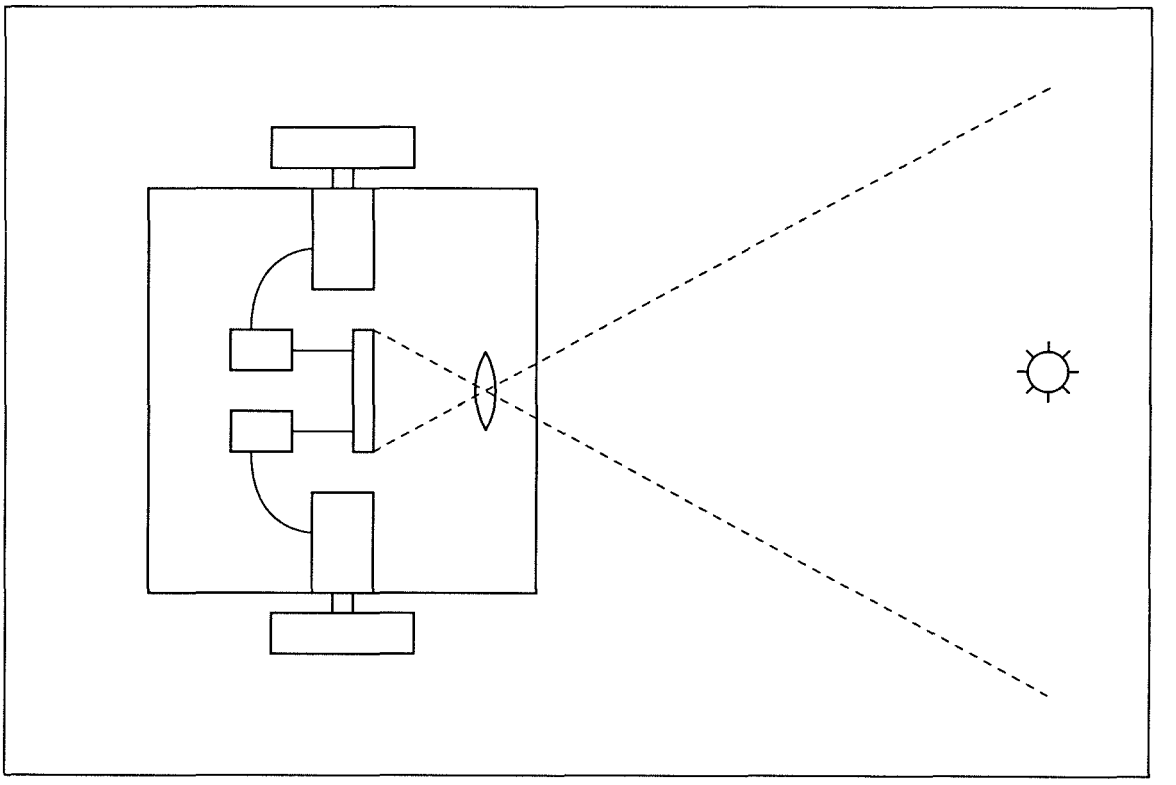

Figure 6.11: A vehicle that follows a stimulus. A lens focuses the image of the area in front of the vehicle onto the chip. Each output pulse train generated by the chip is amplified by a half H-bridge circuit, and is used to drive a motor that rotates one of the wheels. A stimulus positioned away from the center of the chip produces a differential drive that turns the vehicle toward the stimulus. The vehicle will move toward a bright stimulus at a speed set by the background firing rate (the common mode) of the pulse trains.

focused onto the chip. Each wheel is directly connected to a motor that is unidirectionally driven by one of the pulse trains generated by the chip. When a bright stimulus is within the field of view of the chip, the vehicle will move toward the stimulus at (a rate set by the background firing rate of the pulse trains). The vehicle will turn, when necessary, such that the motion is in the direction of the stimulus.

6.6 Summary

In this chapter, we presented a circuit framework that converts a sensory image in the form of a positionally encoded array of currents into a dual-rail pulse encoding for driving a motor

system bidirectionally. The framework circuitry, in conjunction with a motor system, tracks a stimulus that is defined by the sensory input to the system. The framework consists of sensory-input circuitry that provides the system with a positionally encoded image, circuitry to normalize the output from the sensory circuits, and aggregation and pulse-generating circuitry to convert the image into two pulse trains. We discussed methods for implementing these different circuit layers.

We discussed the design and implementation of a version of this circuitry that uses photoreceptors at its input and normalizes their output with the winner-take-all normalization circuit. We have used mechanical hardware to close the loop around the circuitry, and have implemented feedback systems that track a bright visual stimulus. The mechanical systems are driven by negative feedback from the chip, and move such that the stimulus remains centered on the photosensor array. We described how we tested the circuitry by incorporating it into a mechanical model of a planar eye (with associated eye muscles) that rotates to face a stimulus. We also explained how we incorporated the circuitry onto a vehicle that moves toward a stimulus.

References

[AH91] B. Armstrong-Hélouvry. *Control of Machines with Friction.* Kluwer Academic Publishers, Boston, MA, 1991.

[ÅW84] K.J. Åström and B. Wittenmark. *Computer Controlled Systems.* Prentice-Hall, Englewood Cliffs, NJ, 1984.

[ÅW89] K.J. Åström and B. Wittenmark. *Adaptive Control.* Addison-Wesley, Reading, MA, 1989.

[Bal86] D.A. Ballard. Corticle connections and parallel processing: Structure and function. *Behavioral and Brain Sciences*, 9:67–120, 1986.

[DeW91] S.P. DeWeerth. A VLSI-based system for tracking visual stimuli. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, pages 1336–1341, Los Alamitos, CA, 1991. IEEE Computer Society Press.

[DM88] S.P. DeWeerth and C.A. Mead. A VLSI model of adaptation in the vestibulo-ocular reflex. In *Advanced Research in VLSI*, pages 259–275, Cambridge, MA, 1988. MIT Press.

[DNMÅ90] S.P. DeWeerth, L. Nielsen, C.A. Mead, and K.J. Åström. A neuron-based servo for motion control. In *Proceedings of the 1990 IEEE International Conference on Robotics and Automation*, pages 1698–1703, Los Alamitos, CA, 1990. IEEE Computer Society Press.

[DNMÅ91] S.P. DeWeerth, L. Nielsen, C.A. Mead, and K.J. Åström. A simple neuron servo. *IEEE Transactions on Neural Networks*, 2:248–251, 1991.

[Edw86] J.D. Edwards. *Electrical Machines: An Introduction to Principles and Characteristics.* Macmillan, New York, 1986.

[EL90] K.A.P. Edman and F. Lou. Changes in force and stiffness induced by fatigue and intracellular acidification in frog muscle fiber. *J. Physiol.*, 424:133–149, 1990.

[FPEN86] G.F. Franklin, J.D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems.* Addison-Wesley, Reading, MA, 1986.

[GB42] M.F. Gardner and J.L. Barnes. *Transients in Linear Systems*, volume 1. Wiley, New York, 1942.

[Gil84] B. Gilbert. A 16-channel array normalizer. *IEEE Journal of Solid-State Circuits*, 19:956–963, 1984.

[HR81] D. Halliday and R. Resnick. *Fundamentals of Physics.* Wiley, New York, 1981.

[LM89] J. Lazzaro and C.A. Mead. A silicon model of auditory localization. *Neural Computation*, 1:47–57, 1989.

[LRMM89] J. Lazzaro, S. Ryckebusch, M.A. Mahowald, and C.A. Mead. Winner-take-all networks of $o(n)$ complexity. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems*, volume 1, pages 703–711. Morgan Kaufmann, San Mateo, CA, 1989.

[MĆ76] R.D. Middlebrook and S. Ćuk. A general unified approach to modelling switching-converter power stages. In *Proceedings of the IEEE Power Electronics Specialists Conference.* IEEE Press, 1976.

[Mea85] C.A. Mead. A sensitive electronic photoreceptor. In H. Fuchs, editor, *Proceedings of the 1985 Chapel Hill Conference on Very Large Scale Integration*, pages 463–471, Rockville, MD, 1985. Computer Science Press.

[Mea89] C.A. Mead. *Analog VLSI and Neural Systems.* Addison-Wesley, Reading, MA, 1989.

[MM88] C.A. Mead and M.A. Mahowald. A silicon model of early visual processing. *Neural Networks*, 1:91–97, 1988.

[PR76] T. Poggio and W. Reichardt. Visual control of orientation behavior in the fly: Part ii. towards the underlying neural interactions. *Quarterly Reviews of Biophysics*, 9:377–438, 1976.

[RP76] W. Reichardt and T. Poggio. Visual control of orientation behavior in the fly: Part i. quantitative analysis. *Quarterly Reviews of Biophysics*, 9:311–375, 1976.

[She79] G.M. Shepherd. *The Synaptic Organization of the Brain.* Oxford, New York, 1979.

[SMM87] M.A. Sivilotti, M.A. Mahowald, and C.A. Mead. Real-time visual computation using analog CMOS processing arrays. In *Proceedings of the 1987 Stanford Conference on Very Large Scale Integration*, pages 295–312, Cambridge, MA, 1987. MIT Press.

[Spa83] D. Sparks. The role of primate superior colliculus in sensorimotor integration. In M. Arbib and A. Hanson, editors, *Proceedings of the Conference on Vision, Brain, and Cooperative Computation*, Cambridge, MA, 1983. MIT Press.

[Sze81] S.M. Sze. *Physics of Semiconductor Devices.* Wiley, Englewood Cliffs, NJ, 1981.

[Tan86] J.E. Tanner. *Integrated Optical Motion Detection.* PhD thesis, Department of Computer Science, California Institute of Technology, Pasadena, CA, 1986. 5223:TR:86.

[Tsy84] Ya.Z. Tsypkin. *Relay Control Systems.* Cambridge University Press, Cambridge, U.K., 1984.

[Vit85] E.A. Vittoz. Micropower techniques. In Y. Tsividis and P. Antognetti, editors, *Design of MOS VLSI Circuits for Telecommunications*, pages 104–144. Prentice-Hall, Englewood Cliffs, NJ, 1985.

[WCH85] R.C. Woledge, N.A. Curtin, and E. Homser. Energetic aspects of muscle contraction. *Monographs of the Physiological Society*, 41, 1985.

[Wil60] C.H. Wilts. *Principles of Feedback Control.* Addison-Wesley, Reading, MA, 1960.